DeepMIP: Model intercomparison of early Eocene climatic optimum (EECO) large-scale climate features and comparison with proxy data

Daniel J. Lunt[1], Fran Bragg[1], Wing-Le Chan[2], David K. Hutchinson[3], Jean-Baptiste Ladant[4], Igor Niezgodzki[5,6], Sebastian Steinig[1], Zhongshi Zhang[7,8], Jiang Zhu[4], Ayako Abe-Ouchi[2], Agatha M. de Boer[3], Helen K. Coxall[3], Yannick Donnadieu[9], Gregor Knorr[5], Petra M. Langebroek[7], Gerrit Lohmann[5], Christopher J. Poulsen[4], Pierre Sepulchre[10], Jess Tierney[11], Paul J. Valdes[1], Tom Dunkley Jones[12], Christopher J. Hollis[13], Matthew Huber[14], and Bette L. Otto-Bliesner[15]

[1]School of Geographical Sciences, University of Bristol, UK
[2]University of Tokyo, Japan
[3]Department of Geological Sciences, Stockholm University, Sweden
[4]University of Michigan, USA
[5]AWI, Germany
[6]Institute of Geological Sciences Polish Academy of Sciences, Research Center in Kraków, Biogeosystem Modelling Group, Poland
[7]NORCE Norwegian Research Centre, Bjerknes Centre for Climate Research, Norway
[8]China University of Geoscience (Wuhan), China
[9]CNRS, France
[10]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, France
[11]University of Arizona, USA
[12]Birmingham University, UK
[13]GNS, New Zealand
[14]Purdue University, USA
[15]NCAR, USA

Correspondence: Dan Lunt (d.j.lunt@bristol.ac.uk)

Abstract. We present results from an ensemble of seven climate models, each of which has carried out simulations of the early Eocene climate optimum (EECO, ∼50 million years ago). These simulations have been carried out in the framework of DeepMIP (www.deepmip.org), and as such all models have been configured with identical paleogeographic and vegetation boundary conditions. The results indicate that these non-CO_2 boundary conditions contribute between 3 and 5°C to Eocene

5 warmth. Compared to results from previous studies, the DeepMIP simulations show reduced spread of global mean surface temperature response across the ensemble, for a given atmospheric CO_2 concentration. In a marked departure from the results from previous simulations, at least two of the DeepMIP models (CESM and GFDL) are consistent with proxy indicators of global mean temperature, and atmospheric CO_2, and meridional SST gradients. The best agreement with global SST proxies from these models occurs at CO_2 concentrations of around 2400 ppmv. At a more regional scale the models lack skill in

10 reproducing the proxy SSTs, in particular in the southwest Pacific, around New Zealand and south Australia, where the modelled anomalies are substantially less than indicated by the proxies. However, in these regions modelled continental surface

air temperature anomalies are consistent with surface air temperature proxies, implying an inconsistency between marine and terrestrial temperatures in either the proxies or models in this region. Our aim is that the documentation of the large scale features and model-data comparison presented herein will pave the way to further studies that explore aspects of the model simulations in more detail, for example the ocean circulation, hydrological cycle, and modes of variability; and encourage sensitivity studies to aspects such as paleogeography and aerosols.

1 Introduction

Paleoclimate model-data comparisons allow us to assess confidence in the results from model sensitivity studies that explore the mechanisms that drove past climate change, and allow us to assess confidence in future climate predictions that are carried out with equivalent models. Past warm climates, particularly those associated with high atmospheric CO_2 concentrations, are especially relevant because they share characteristics with possible future climates (Burke et al., 2018). In this context, there has been a community focus on the Pliocene ($\sim$3-5 million years ago; Haywood et al., 2013) and Eocene ($\sim$50 million years ago; Lunt et al., 2012), which provide natural examples of past worlds with high CO_2 concentrations of $\sim$300-400 ppmv and $\sim$900-2500 ppmv respectively. In this paper, we focus on the Eocene, presenting model results that have recently been carried out in the framework of the DeepMIP project (www.deepmip.org; Lunt et al., 2017; Hollis et al., 2019), and associated model-data comparisons. Given the similarity of Eocene CO_2 concentrations and climate to those that are attained under high growth/low mitigation future scenarios considered by the IPCC (Burke et al., 2018), the Eocene provides a test-bed for state-of-the-art climate model predictions of the future.

Eocene modelling and model-data comparisons have a long history (e.g. Barron, 1987; Sloan and Barron, 1992). More recently, Lunt et al. (2012) carried out a synthesis of a group of models that had all carried out Eocene simulations (Lunt et al., 2010b; Heinemann et al., 2009; Winguth et al., 2010; Huber and Caballero, 2011; Roberts et al., 2009), with a focus on surface temperatures. Subsequent work also explored the precipitation in the simulations (Carmichael et al., 2016) and the implications for ice sheet growth (Gasson et al., 2014). This was an "ensemble of opportunity" in that the model simulations were carried out independently, using a variety of paleogeographical and vegetation boundary conditions, and carried out under a range of different CO_2 concentrations. A proxy data synthesis was also produced as part of the Lunt et al. (2012) study, consisting of SSTs, and a previously compiled continental temperature dataset (Huber and Caballero, 2011). That model-data comparison showed that: (a) For a given CO_2 concentration, there was a wide spread in global mean temperature response across the models. For example, at CO_2 concentrations $\times$4 those of preindustrial, the range in modelled global mean continental near-surface air temperature was 5.8°C. (b) Given CO_2 concentrations of $\times$16, the CCSM3 model was able to reproduce the mean climate and meridional temperature gradient indicated by the proxies. (c) The HadCM3 model had very weak polar amplification compared with the other models. (d) The climate sensitivity across the models was fairly similar, but HadCM3

had a notable non-linearity in sensitivity, in contrast to CCSM3. (e) Interpreting middle and high latitude proxy SSTs as representing summer temperatures brought the modelled temperatures closer to those indicated by the proxies.

45   At that time, due to uncertainties in pre-ice core CO_2 proxies, it was not possible to rule out the high CO_2 concentrations needed by CCSM3 to match the proxies, although such high values were outside the range of many CO_2 compilations (Beerling and Royer, 2011). As such, the Intergovernmental Panel on Climate Change (IPCC) concluded that "While recent simulations of the EECO... exhibit a wide inter-model variability, there is generally good agreement between new simulations and data, particularly if seasonal biases in some of the marine SST proxies from high-latitude sites are considered" (Masson-Delmotte

50   et al., 2013). However, more recent work has indicated that early Eocene CO_2 concentration was very likely (95% confidence interval) in the range 900 ppmv to 2500 ppmv during the EECO, with a best estimate of 1400 ppmv (Anagnostou et al., 2016). As such, none of the models examined in Lunt et al. (2012) agreed well with the proxy reconstructions at realistic CO_2 concentrations.

Following on from that initial modelling work, two studies (Sagoo et al., 2013; Kiehl and Shields, 2013) showed that the

55   representation of clouds in models could be modified to give greater polar amplification and climate sensitivity, resulting in simulations that were more consistent with both temperature and CO_2 proxies of the Eocene. Kiehl and Shields (2013) decreased the cloud drop density and increased the cloud drop radius to represent the effect of reduced cloud-condensation nucleii in the Eocene compared with modern, and found that, at a CO_2 concentration of 1375 ppmv and CH_4 of 760 ppbv (their "pre-PETM" simulation) they obtained a good agreement with data. Sagoo et al. (2013) perturbed ten atmospheric and

60   oceanic variables in an ensemble, of which those associated with clouds were judged the most important, and found that two ensemble members were able to simulate temperatures in good agreement with proxies at a CO_2 concentration of 560 ppmv. Although both of these studies indicated that clouds could be the key to reconciling proxies and models, neither of the changes applied were physically based. Furthermore, more recent work has indicated that the response to modifying cloud albedo is very similar to that of increasing CO_2, at least in terms of meridional temperature gradient (Carlson and Caballero, 2017), such

65   that prescribing cloud changes can result in a system that is somewhat unconstrained. As such, the relevance of these studies for future prediction or to other paleo time periods remains unclear.

To facilitate an intermodel comparison, a standard set of boundary conditions and experimental design has been proposed for a coordinated set of model simulations of the early Eocene (Lunt et al., 2017). In addition, there has been a community effort to better characterise the uncertainties in proxy temperature and CO_2 estimates of the latest Paleocene, Paleocene–Eocene

70   Thermal Maximum (PETM) and EECO (Hollis et al., 2019). Furthermore, many current-generation models now include cloud parameterisations that are substantially advanced on their predecessors, in particular in their representation of cloud microphysics. In this paper, we present an ensemble of early Eocene simulations from a range of climate models, carried out in this framework, and compare them with the latest paleo data of the EECO. Some key scientific questions that we address are:

– What are the large-scale features of the DeepMIP Eocene simulations?

75   – What is the model spread in our predictions of the Eocene?

– How well do the models fit the proxy data?

- Has there been an improvement in model fit to data with the latest models?

2 DeepMIP model simulations

Here we briefly describe the standard experimental design, and for each model give a brief description of the model and any
departures from the standard experimental design.

2.1 Experimental design

The standard experimental design for the DeepMIP model simulations, and underlying motivation, is described in detail in
Lunt et al. (2017). In brief, the simulations consist of a preindustrial control, and a number of Eocene simulations at various
atmospheric CO_2 concentrations ($\times$3, $\times$6, and $\times$12, although in practice many groups chose different concentrations). The
paleogeography, vegetation, and river routing for the Eocene simulations are prescribed according to the reconstructions of
Herold et al. (2014) (see Figure 3a and 4 in Lunt et al. (2017)). The solar constant, orbital configuration, and non-CO_2 green-
house gas concentrations are set to preindustrial values. Soil properties are set to homogeneous global mean values derived
from the preindustrial simulation, and there are no continental ice sheets in the Eocene simulations. A suggested initial con-
dition for ocean temperature and salinity was given, but many groups diverged from this. The prescription of calculation of
atmospheric aerosols were left to each individual group's discretion.

2.2 Individual model simulations

An overview of the model simulations is presented in Table 1. Here we describe each model in turn, and the experimental
design of the simulations where this diverged from that described in Lunt et al. (2017).

2.2.1 CESM

CESM model description

The Community Earth System Model version 1.2 (CESM) is used, which consists of the Community Atmosphere Model 5.3
(CAM), the Community Land Model 4.0 (CLM), the Parallel Ocean Program 2 (POP), the Los Alamos sea ice model 4 (CICE),
the River Transport Model (RTM), and a coupler connecting them (Hurrell et al., 2013). In comparison to previous versions of
the CESM models that have been used for Eocene simulation, e.g. CCSM3 (Huber and Caballero, 2011; Winguth et al., 2010;
Kiehl and Shields, 2013) and CESM1(CAM4) (Cramwinckel et al., 2018), CESM1.2(CAM5) represents a nearly complete
overhaul of physical parameterizations in the atmosphere model, including new schemes for radiation, boundary layer, shallow
convection, cloud microphysics and macrophysics, and aerosols (Hurrell et al., 2013). The new two-moment microphysical
scheme predicts both the cloud water mixing ratio and particle number concentration. The new aerosol scheme predicts the
aerosol mass and number, and is coupled with the cloud microphysics, allowing the inclusion of aerosol indirect effects. The
new boundary layer and shallow convection schemes improve the simulation of shallow clouds in the marine boundary layer.
These new parameterizations in CAM5 produce a cloud simulation that agrees much better with satellite observations (Kay

Table 1. A summary table of the DeepMIP Eocene model simulations described and presented in this paper. In addition to the simulations listed, each model has an associated preindustrial control.

Model	CO_2	length [years]	TOA inbalance [Wm^{-2}]	CMIP generation	Reference
CESM	$\times 1, \times 3, \times 6, \times 9$	2000+	$-0.25, -0.37,$ $0.33, 0.64$	CMIP5	Zhu et al. (2019)
IPSL	$\times 1.5, \times 3$	4000	0.59, 0.77	CMIP5	*This paper*, Sepulchre et al. (2019)
NorESM	$\times 2, \times 4$	2100,2000	0.05,0.26	CMIP5-6	*This paper.*
GFDL	$\times 1, \times 2, \times 3, \times 4, \times 6$	6000	$0.05, -0.16, -0.18,$ $-0.18, -0.28$	CMIP3	*This paper*, Hutchinson et al. (2018)
HadCM3	$\times 1, \times 2, \times 3$	>4000 + 900,500,1100	$0.00, -0.13, -0.12$	CMIP3	*This paper.*
MIROC	$\times 3$	5000	0.98	CMIP3	*This paper.*
COSMOS	$\times 3$	1000	2.06	CMIP3	*This paper.*

et al., 2012) and a larger present-day equilibrium climate sensitivity ($\sim 4^\circ$C) than previous versions ($\sim 3^\circ$C) (Gettelman et al., 2012). CESM1.2(CAM5) reproduces key features in state and variability of past climates, including the mid-Piacenzian Warm period (Feng et al., 2019), the last glacial maximum (Zhu et al., 2017a), Heinrich events (Zhu et al., 2017b), and the last

110 millennium (Otto-Bliesner et al., 2015; Thibodeau et al., 2018). To make the model suitable for a paleoclimate simulation with a high CO_2 level, the model code has been slightly modified to incorporate an upgrade to the radiation code that corrects the missing diffusivity angle specifications for certain longwave bands. As a result of the code modification, CAM5 has been re-tuned with a different relative humidity threshold for low clouds (rhminl=0.8975, versus the default value of 0.8875). These code and parameter changes are not found to alter the present-day climate sensitivity in CESM (Zhu et al., 2019).

115 *CESM model simulations*

The CESM Eocene simulations are run at $\times 1$, $\times 3$, $\times 6$, and $\times 9$ CO_2 concentrations (Table 1). The atmosphere and land have a horizontal resolution of $1.9 \times 2.5^\circ$ (latitude $\times$ longitude) with 30 hybrid sigma-pressure levels in the atmosphere. The ocean and sea ice are on a nominal 1° displaced pole Greenland grid with 60 vertical levels in the ocean. CAM5 runs with a prognostic aerosol scheme with prescribed preindustrial natural emissions that have been redistributed according to the Eocene

120 paleogeography following the method in Heavens et al. (2012). The vegetation type from Herold et al. (2014) is prescribed in the land model with active carbon and nitrogen cycling. A modified marginal sea balancing scheme was applied for the Arctic Ocean, which removes any gain/deficit of freshwater over the Arctic Ocean and redistributes the mass evenly over the global ocean surface excluding the Arctic. This implementation conserves ocean salinity and is necessary to prevent the occurrence of negative salinity that results from high precipitation and river runoff under warm conditions. A similar balancing scheme has

125   been included for marginal seas in all the previously published CESM simulations (Smith et al., 2010). The ocean temperature
       and salinity were initialized from a previous PETM simulation using CCSM3 (Kiehl and Shields, 2013). Sea ice model was
       initialized from a sea ice-free condition. All simulations have been integrated for 2000 model years, with the exception of ×1
       which was run for 2600 model years.

2.2.2   IPSL

130   *IPSL model description*

       The IPSL simulations are performed with the IPSL-CM5A2 earth system model (Sepulchre et al., 2019). IPSL-CM5A2 is
       based on the CMIP5-generation previous IPSL earth system model IPSL-CM5A (Dufresne et al., 2013) but includes new
       revisions of each components, a re-tuning of global temperature, and technical improvements to increase computing efficiency.
       It consists of the LMDZ5 atmosphere model, the ORCHIDEE land surface and vegetation model and the NEMOv3.6 ocean

135   model, which includes the LIM2 sea ice model and the PISCES-v2 biogeochemical model. LMDZ5 and ORCHIDEE run at a
       horizontal resolution of $1.9 \times 2.5°$ (latitude × longitude) with 39 hybrid sigma-pressure levels in the atmosphere. NEMO runs
       on a tripolar grid at a nominal resolution of $2°$, enhanced up to $0.5°$ at the Equator, with 31 vertical levels in the ocean. The
       performances and evaluation of IPSL-CM5A2 on preindustrial and historical climates are fully described in Sepulchre et al.
       (2019). Sepulchre et al. (2019) also provides a description of the technical changes that were implemented in IPSL-CM5A2 to

140   carry out deep time paleoclimate simulations. In particular, the tripolar mesh grid on which NEMO runs has been modified to
       ensure that there are no singularity points within the ocean domain. Modern parameterizations of water outflows across specific
       straits, such as the Gibraltar or Red Sea straits, are also turned off.

       IPSL model simulations

       The IPSL simulations are run at ×1.5 and ×3 CO_2 concentrations. The bathymetry is obtained from the Herold et al. (2014)

145   dataset, with additional handmade corrections in some locations, for instance in the West African region, to maintain sufficiently
       large oceanic straits. Modern boundary conditions of NEMO include forcings of the dissipation associated with internal wave
       energy for the M2 and K1 tidal components (de Lavergne et al., 2019). The parameterization follows Simmons et al. (2004)
       with refinements in the modern Indonesian Through Flow (ITF) region according to Koch-Larrouy et al. (2007). To create
       an Early Eocene tidal dissipation forcing, we directly interpolate the Herold et al. (2014) M2 tidal field (obtained from the

150   tidal model simulations of Green and Huber (2013)) onto the NEMO grid using bilinear interpolation. In the absence of any
       estimation for the Early Eocene, we prescribe the K1 tidal field to 0. In addition, the parameterization ofKoch-Larrouy et al.
       (2007) is not used here because the ITF does not exist in the Early Eocene. The geothermal heating distribution is created
       from the 55 Ma global crustal age distribution of Müller et al. (2008), on which is applied the age-heatflow relationship of the
       Stein and Stein (1992) model: $q(t) = 510 \times t^{-1/2}$ if t $\leq$ 55 Ma and $q(t) = 48 + 96 \exp(-0.0278 \times t)$ if t $>$ 55Ma. In regions of

155   subducted seafloor where age information is not available, we prescribe the minimal heatflow value derived from known crustal
       age. The $1° \times 1°$ resulting field is then bilinearly interpolated onto the NEMO grid. It must be noted that the Stein and Stein
       (1992) parameterization becomes singular for young crustal ages, which yields unrealistically large heatflow values. Following
       Emile-Geay and Madec (2009), we set an upper limit of 400 mW m^{-2} for heatflow values after the interpolation procedure.

Salinity is initialized as globally constant to a value of 34.7 psu following Lunt et al. (2017). The initialization of the model
160 with the proposed DeepMIP temperature distribution (Lunt et al., 2017) led to severe instabilities of the model during the
spin-up phase. The initial temperature distribution has thus been modified to follow: T(°C) = (1000-z)/1000 * 25 cos(ϕ) +
10 if z ≤ 1000 m and T(°C) = 10 if z > 1000 m. With ϕ the latitude and z the depth of the ocean (meters below surface).
This new equation gives an initial globally constant temperature of 10°C below 1000 m and a zonally symmetric distribution
above, reaching surface values of 35°C at the equator and 10°C at the poles. This corresponds to a 5°C surface temperature
165 reduction compared to DeepMIP guidelines (Lunt et al., 2017). No sea ice is prescribed at the beginning of the simulations.
In IPSL-CM5A2, the NEMO ocean model is inherently composed of the PISCES biogeochemical model. Biogeochemical
cycles and marine biology are directly forced by dynamical variables of the physical ocean and may affect the ocean physics
via its influence on chlorophyll production, which modulates light penetration in the ocean. However, because this feedback
does not much affect the ocean state (Kageyama et al., 2013) and because the Early Eocene mean ocean colour is unknown,
170 we have prescribed a constant chlorophyll value of 0.05 g.Chl/L for the computation of light penetration in the ocean. As a
consequence, marine biogeochemical cycles and biology do not alter the dynamics of the ocean and as such biogeochemical
initial and boundary conditions have been kept to modern. The topographic field is created from the Herold et al. (2014)
topographic dataset; LMDZ includes a subgrid scale orographic drag parameterization that requires high-resolution surface
orography (Lott and Miller, 1997; Lott, 1999). We apply a similar procedure for the standard deviation of orography provided
175 by Herold et al. (2014). Aerosol distributions are left identical to pre-industrial values. The ×3 simulation is initialized from
rest and run for 4000 years. The ×1.5 simulation is branched from year 1500 of the ×3 simulation and run for 4000 years. At
the end of the two simulations, the ocean has reached a quasi-equilibrated state and the trend in the deep ocean temperatures
over the final 1000 years of both simulations is smaller than 0.05°C/century.

2.2.3   NorESM

180 *NorESM model description*
The NorESM simulations are carried out with the NorESM1-F version of the model, which is described in detail in Guo
et al. (2019). The NorESM version that contributes to CMIP5 is NorESM1-M. It has a ∼2° resolution atmosphere and land
configuration, and a nominal 1° ocean and sea ice configuration. In NorESM1-F, the same atmosphere–land grid is used as
NorESM1-M (CMIP5 version), whereas a tripolar grid is used for the ocean–sea ice components in NorESM1-F, instead of the
185 bipolar grid in NorESM1-M. The tripolar grid is also used in the CMIP6 version of NorESM (NorESM2). NorESM1-F runs
about 2.5 times faster than NorESM1-M.
NorESM model simulations
The NorESM simulations are carried out at ×2, and ×4 CO_2 concentrations (Table 1). The ocean temperatures were initialized
from the ×2 CO_2 Eocene simulations with the lower resolution NorESM-L model (Zhang et al., 2012). The ocean salinity was
190 initialized with constant values of 25.5 psu for the Arctic and 34.5 psu for other oceans. From the initial conditions, the ×2
CO_2 experiment in total run for 2100 years. The ×4 CO_2 was branched from the end of the 100th year of ×2 CO_2 experiment,
and run for 2000 years. The results from the last 100 years were used in the study. Note that the NORESM simulations were

carried out with the Baatsen et al. (2016) paleogeography (based on a paleomagnetic reference frame), not the Herold et al. (2014) paleogeography (based on a mantle reference frame), in contrast to the other simulations described in this paper.

195 **2.2.4 GFDL**

GFDL model description

These simulations use a modified version of GFDL CM2.1 (Delworth et al., 2006), similar to the late Eocene configuration in Hutchinson et al. (2018, 2019). The ocean component uses the modular ocean model (MOM) version 5.1.0, while the other components of the model are the same as in CM2.1; Atmosphere Model 2, Land Model 2 and the Sea Ice Simulator 1. The

200 ocean and sea ice components use a horizontal resolution of 1° latitude × 1.5° longitude. We use a tripolar grid as in Hutchinson et al. (2018), with a regular latitude-longitude grid south of 65° N, and a transition to a bipolar Arctic grid north of 65° N, with poles over North America and Eurasia. There is no refinement of the latitudinal grid spacing in the tropics. The ocean uses 50 vertical levels, with the same vertical spacing as CM2.1 The atmospheric horizontal grid resolution is 3° × 3.75°, with 24 vertical levels, as in CM2Mc (Galbraith et al., 2010). This configuration enables relatively high resolution ocean and coastlines,

205 with the advantage of a faster-running atmosphere. The topography (both land and ocean) uses the 55 Ma reconstruction of Herold et al. (2014), re-gridded to our ocean and atmosphere components. Manual adjustments are made to ensure that no isolated lakes or seas exist, and that any narrow ocean straits are at least 2 grid cells wide to ensure non-zero velocity fields. The minimum depth of ocean grid cells is 25 m; any shallow ocean grid cells are deepened to this minimum depth. In the atmosphere, the topography is smoothed using a 3-point mean filter to ensure a smoother interaction with the wind field. This was

210 introduced to remove numerical noise over the Antarctic continent, due to convergence of meridians on the topography grid. Vegetation types are based on Herold et al. (2014), adapted to the corresponding vegetation type in CM2.1. Aerosol forcing is also adapted from Herold et al. (2014) to our model, and is a fixed boundary condition. Ocean vertical mixing is identical to that in Hutchinson et al. (2018); i.e. a uniform bottom-roughness enhanced mixing with a background diffusivity of 1.0×10^{-5} $m^2 \, s^{-1}$.

215 *GFDL model simulations*

The model was initiated from idealised conditions, similar to those outlined in Lunt et al. (2017) with reduced initial temperatures: T(°C) = (5000-z)/5000 * 25 cos(ϕ) + 10 if z ≤ 5000 m and T(°C) = 10 if z > 5000 m; where ϕ is latitude, z is the depth of the ocean (positive downwards). The initial salinity was a constant of 34.7 psu. The above initial conditions were used for the ×1, ×2, ×3 and ×4 CO_2 experiments. These simulations were initially run for 1500 years, after which the ocean temper-

220 atures were adjusted in order to accelerate the approach to equilibrium. This adjustment consisted of calculating the average temperature trend for the last 100 years at each model level below 500m, taking a level-by-level global average of this trend, and applying a 1000-year extrapolation uniformly across the ocean at that level. This choice was based on the observation that all model levels below the mixed layer were consistently cooling at a slow rate, and the rate of temperature adjustment was consistent over a long time scale. After a further 500 years, a second step-wise using the same method was performed.

225 After the second adjustment, all simulations were continuously integrated with no further adjustments for a further 4000 years. Thus the simulations were run for a total of 6000 years. For the ×6 CO_2 experiment, the initial conditions described above led

to transient instabilities due to overheating the surface. So we instead initialised the ×6 experiment using a globally uniform temperature of 19.32 °C. This represents the same global average temperature as in the other experiments, hence the same total ocean heat content. For the ×6 CO_2, no step-wise adjustments were made; the model was run continuously for 6000 years. By

230 the end of the 6000 year period, all simulations exhibited steady SST for at least 2000 years, while at 4000 m depth the ocean was cooling at a rate of approximately 0.2 °C per 1000 years or less. The TOA radiative imbalance varied between −0.28 and 0.05 depending on the CO_2 level, however this radiative imbalance was not reflected in any surface temperature trends.

2.2.5 COSMOS

COSMOS model description

235 The atmosphere is represented by means of the ECHAM5 atmosphere general circulation model Roeckner et al. (2003). ECHAM5 is based on a spectral dynamical core and includes 19 vertical hybrid sigma-pressure levels. The series of spectral harmonics is curtailed via triangular truncation at wave number 31 (approx. 3.75×3.75). Ocean circulation and sea ice dynamics are computed by the MPIOM ocean general circulation model Marsland et al. (2003) that is employed at 40 unequally-spaced levels on a bipolar curvilinear model grid with formal resolution of 3.0×1.8 longitude by latitude. The cou-

240 pled model system ECHAM5/MPIOM is described by Jungclaus et al. (2006). A concise description of the application of the COSMOS for paleoclimate studies is given by Stepanek and Lohmann (2012). The COSMOS version used here has proven to be a suitable tool for the study the Earth's past climate, from Holocene (Wei and Lohmann (2012); Wei et al. (2012); Lohmann et al. (2013)) and previous interglacials (Pfeiffer and Lohmann (2016); Gierz et al. (2017)), glacial (Gong et al. (2013); Zhang et al. (2013); Zhang et al. (2014); Abelmann et al. (2015); Zhang et al. (2017)) to tectonic time scales (Knorr et al. (2011);

245 Knorr and Lohmann (2014); Walliser et al. (2016); Huang et al. (2017); Niezgodzki et al. (2017); Stärz et al. (2017); Walliser et al. (2017); Niezgodzki et al. (2019)). The standard model code of the Community Earth System Models (COSMOS) version COSMOS-landveg r2413 (2009) is available upon request from the Max Planck Institute for Meteorology in Hamburg (https://www.mpimet.mpg.de).

COSMOS model simulations

250 The COSMOS simulations are carried out at ×1, ×3, and ×4 CO_2 concentrations. The ocean temperatures in $3\times$ CO_2 concentration were initialized with uniformly horizontal and vertical temperatures of 10° Celcius. The initial ocean salinity was set to 34.7 psu. The simulations with $1\times$ and $4\times$ CO_2 concentrations were restarted from $3\times$ CO_2 after 1000 years and are currently running. Therefore they are not included herein. The analysis of the simulation with $3\times$ CO_2 is based on the average between the model years 900-1000.

255 ### 2.2.6 HadCM3

HadCM3 model description

The HadCM3 simulations are carried out with the HadCM3B-M2.1aN version of the model, as described in detail in Valdes et al. (2017). Equations are solved on a Cartesian grid with horizontal resolutions of $3.75° \times 2.5°$ in the atmosphere and $1.25° \times 1.25°$ in the ocean with 19 and 20 vertical levels, respectively. We make a few changes to the version described in

260 Valdes et al. (2017) to make it suitable for deep-time paleoclimate modelling: (a) We apply a salinity flux correction to the global ocean in order to conserve salinity. (b) We turn off the various modern-specific parameterisations in the ocean model, such as associated with Mediterranean and Hudson Bay outflow, and North Atlantic mixing. (c) We use a prognostic 1D ozone scheme instead of a fixed vertical profile of ozone. The standard configuration uses a prescribed ozone climatology which is a function of latitude, height, and month of the year that does not change with climate and can become numerically unstable at

265 high CO_2 levels. The prognostic ozone scheme uses the diagnosed model tropopause height to assign three distinct ozone concentrations for the troposphere, tropopause, and stratosphere ($2.0x10^{-8}$, $2.0x10^{-7}$ and $1.5x10^{-6}$ in mmr, respectively). This allows for a dynamic update of the 1D ozone field in response to the thermally driven vertical expansion of the troposphere. Absolute values for the three levels are chosen to minimise the effects on global mean and overall tropospheric temperature changes compared to the standard 2D climatology. Concentrations at the uppermost model level are fixed to the higher strato-

270 spheric value to constrain the lower bound of total stratospheric ozone. Significant differences to the standard configuration are limited to the stratospheric meridional temperature gradient and zonal winds and are related to the missing latitudinal variations in the 1D field. Although HadCM3 has been used previously to simulate the Pliocene (e.g. Lunt et al., 2008, 2010a), the presented simulations represent the first published application of HadCM3 to pre-Pliocene boundary conditions. However, the lower resolution HadCM3L model has been previously used to simulate a range of pre-Quaternary climates (e.g. Lunt et al.,

275 2016; Farnsworth et al., 2019)

HadCM3 model simulations

The HadCM3 simulations are carried out at ×1, ×2, and ×3 CO_2 concentrations. Several ocean gateways were artificially widened to allow unrestricted throughflow and maximum water depths in parts of the Arctic Ocean were reduced. The ocean temperatures were initialised from the final state of Eocene model simulations using HadCM3L. The HadCM3L simulations

280 were set up identically to the corresponding HadCM3 simulations, but with lower ocean resolution ($3.75° \times 2.5°$ as opposed to $1.25° \times 1.25°$). The HadCM3L simulations were initialised from a similar idealised temperature and salinity state as described in Lunt et al. (2017), but with a function that scales with $cos^2(lat)$ rather than $cos(lat)$ and overall reduced initial temperatures to ensure numerical stability in tropical regions. Ocean temperatures below 600 m were set to constant values of 4, 8 and 10 °C (at ×1, ×2, and ×3 CO_2 respectively) based on results from previous Ypresian simulations. The HadCM3 simulations were

285 branched off from the respective HadCM3L integrations after 4400 to 4900 years of spin up and run for a further 500 to 1100 years. The initial 50 years of all HadCM3 runs used the simplified vertical diffusion scheme from HadCM3L (Valdes et al., 2017) to reduce numerical problems caused by the changed horizontal ocean resolution. The remaining years of the runs use the standard HadCM3 diffusion scheme (Valdes et al., 2017).

2.2.7 MIROC

290 *MIROC model description*

The version of the Model for Interdisciplinary Research on Climate (MIROC) used here is MIROC4m, a mid-resolution model composed of atmosphere, land, river, sea ice and ocean components. Full documentation of the model can be found in K-1 model developers (2004) and a summary in Chan et al. (2011). The atmosphere has a horizontal resolution of T42

and 20 vertical sigma levels. Details of the land-surface model, Minimal Advanced Treatments of Surface Interaction and
Runoff (MATSIRO), can be found in Takata et al. (2003). The ocean component is basically version 3.4 of the CCSR Ocean
Component Model (COCO) - refer to Hasumi (2000). The horizontal resolution is set to 256×196, with higher resolution
in the tropics, and the vertical resolution is set to 44 levels, with the top 8 in sigma coordinates. Present day bathymetry is
derived from ETOPO5 data. For present day experiments, areas of water such as the Hudson Bay and the Mediterranean Sea
are represented as isolated basins. As such, ocean salinity and heat are artificially exchanged with the open ocean through a
2-way linear damping. This damping and all isolated basins and lakes are removed in the DeepMIP simulation.

MIROC model simulations

Out of the three standard DeepMIP simulations, MIROC is used with $\times 3$ CO_2 concentration only and run for 5000 model
years. Higher concentrations, even $\times 4$ CO_2, lead to rising temperatures until instability sets in. The atmosphere is initialised
from a previous experiment without ice sheets and with $\times 2$ CO_2 concentration. For the initial ocean state, salinity is set to
a constant value of 34.7 psu, as recommended in Lunt et al. (2017). However, the ocean temperatures are 15°C cooler than
those recommended, i.e. T(°C) = (5000-z)/5000 * 25 cos(ϕ) if z $\leq$ 5000 m and T(°C) = 0 if z > 5000 m. Previous MIROC
experiments similar to this $\times 3$ CO_2 DeepMIP simulation show that this initialisation should be much closer to the final climate
state.

3 Results

We discuss the results from the model simulations, focusing on two aspects: the large-scale features of the modelled temperature
response compared with preindustrial (Section 3.1) and a comparison with paleo proxy data (Section 3.2).

3.1 Documentation of large-scale features

Here we present global mean quantities, latitudinal gradients, and spatial patterns in the DeepMIP ensemble, with a focus
on temperature. The aim here is not to fully understand the whole ensemble from a mechanistic viewpoint, but to document
the large-scale features. Further papers emerging from DeepMIP will analyse the mechanisms and the individual ensemble
members in detail.

Figure 1a shows the global mean near-surface temperature in each DeepMIP simulation and associated preindustrial control,
plus some older Eocene simulations carried out with other boundary conditions (Lunt et al., 2012; Kiehl and Shields, 2013;
Sagoo et al., 2013). All the DeepMIP model simulations are substantially warmer for a given CO_2 concentration than the
previous CCSM3 and HadCM3L simulations; the DeepMIP models with more than one CO_2 concentration (except NorESM)
have a greater climate sensitivity (global mean surface air temperature change due to a doubling of atmospheric CO_2) than
CCSM3. There is a non-linearity (i.e. global mean temperature that increases with CO_2 differently than would be expected
from a purely logarithmic relationship) in the CESM model simulations (as previously noted by Zhu et al. (2019)), and also
in HadCM3B and (to a lesser extent) GFDL. In CESM the climate sensitivity, normalised to a CO_2 doubling, increases from
4.2°C at $\times 1$ to 4.8°C and 9.7°C at $\times 3$ and $\times 6$, respectively. In GFDL the climate sensitivity increases from 3.7°C at $\times 1$ to
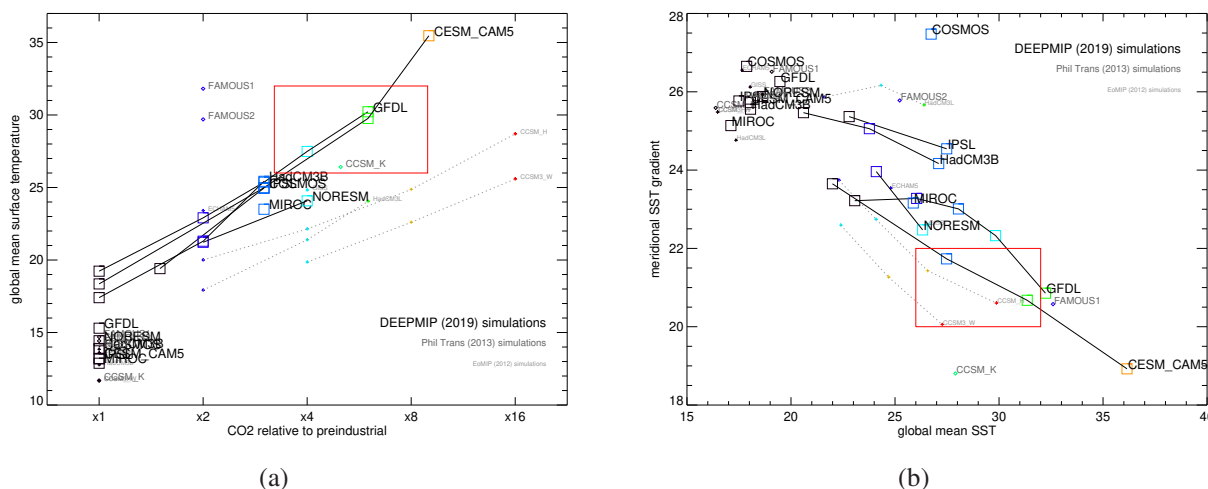
11

(a)

(b)

Figure 1. (a) Global annual mean near-surface (2m) air temperature in the DeepMIP simulations, as a function of atmospheric CO_2. Also shown are results from previous Eocene simulations (Lunt et al., 2012; Kiehl and Shields, 2013; Sagoo et al., 2013) as well as associated preindustrial control simulations. The models that have carried out simulations at more than one CO_2 concentration are joined by a straight line. The red square encompasses best estimates of the global mean temperature and CO_2 derived from proxies, from Zhu et al. (2019) and Anagnostou et al. (2016) respectively. (b) As (a), but for meridional temperature gradient as a function of global mean SST. Meridional SST gradient is defined here as the average SST equatorwards of $\pm30°$ minus the average SST polewards of $\pm60°$. The red square encompasses best estimates of the global mean temperature from Zhu et al. (2019) and the SST gradient from Cramwinckel et al. (2018).

5.1°C at ×3, but then decreases to 4.7°C at ×4. In HadCM3B the climate sensitivity increases from 3.9°C at ×1 to 7.1°C at ×2. In CESM, this non-linearity arises from an increase in strength of the positive shortwave cloud feedback as a function of temperature (Zhu et al., 2019), and is most apparent in the transition from ×6 to ×9. However, the TOA inbalance of +0.64 Wm^{-2} (Table 1) indicates that this ×9 simulation may not be fully spun-up, and could potentially become unstable if it was

330   integrated further towards full equilibrium. The DeepMIP simulations are fairly consistent in terms of global mean temperature across the ensemble; the scatter in the DeepMIP simulations is substantially less than in the previous Eocene simulations. In particular, at ×3 CO_2, the MIROC, IPSL, COSMOS, HadCM3B, and GFDL simulations are within 1.9°C, compared with 5.0°C at ×4 for the previous simulations. The reduced spread of the DeepMIP simulations compared with previous simulations may be related to the fact that all the DeepMIP model simulations have the same prescribed paleogeography, land-sea mask, and

335   vegetation. CESM, GFDL, and HadCM3B all carried out simulations at ×1 CO_2; comparison with the associated preindustrial controls indicates that the non-CO_2 component of global warmth (i.e. that due to changes paleogeography, vegetation, and aerosols, and removal of continental ice sheets) is 5.1°C, 3.9°C, and 3.1°C for CESM, GFDL, and HadCM3B respectively. This is for comparison with previous simulations using CCSM3 (Caballero and Huber, 2013) which indicated a non-CO_2 warming of ∼5°C. The latitudinal gradient in SST, defined as the average SST equatorwards of $\pm30°$ minus the average temperature

340   polewards of $\pm60°$, is shown in Figure 1b. All DeepMIP models that have carried out simulations at more than one CO_2 concentration show a decrease in meridional SST gradient as temperature increases. The ×1 CO_2 Eocene simulations indicate
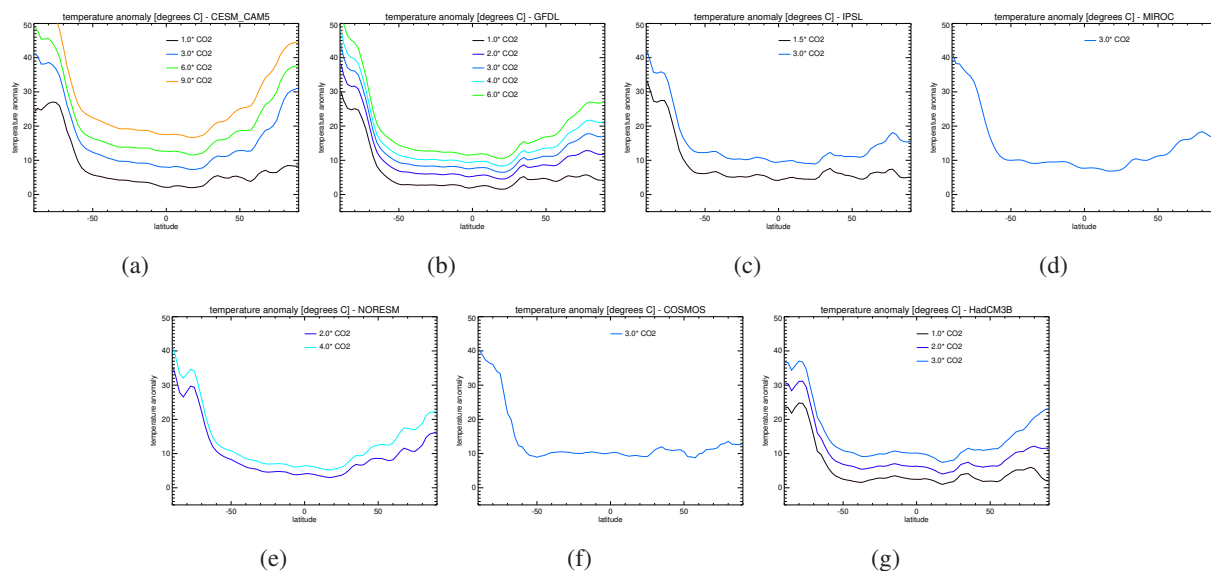
Figure 2. Zonal mean near-surface air temperatures in the DeepMIP simulations, as a function of latitude and prescribed atmospheric CO_2 concentration, expressed as anomalies relative to the equivalent preindustrial control.

that the non-CO_2 DeepMIP boundary conditions decrease the latitudinal gradient by 3.0°C for GFDL, 2.1°C for CESM, and 0.1 °C for HadCM3B. The GFDL model displays a non-linear response, with a more rapidly decreasing temperature gradient as a function of temperature at higher temperatures than at lower temperatures. In contrast to the global mean temperature, the

345 DeepMIP models show substantial spread in meridional temperature gradient across the ensemble; in particular, COSMOS has very weak polar amplification in the Eocene at ×3 CO_2, and HadCM3B abd IPSL also have relatively weak polar amplification, similar to previous Eocene simulations with HadCM3L (Lunt et al., 2010b). Part of the relatively weak response in COSMOS may be due to the fact that it has been integrated for the shortest time (see Table 1), and was initialised from a homogeneous ocean state (see Section 2.2.5).

350   The near-surface air temperature anomaly, relative to preindustrial, as a function of latitude is shown in Figure 2. Polar amplification is clear in both hemispheres for all models at $CO_2 > $×1, although it is weak in the Northern Hemisphere for COSMOS. There is greater amplification in the Southern Hemisphere than in the Northern, due to the replacement of the Antarctic ice sheets with vegetated land surface, with associated local warming due to the altitude and albedo change. There is a similar response across the models for a given CO_2 concentration. However, although the models have a similar response

355 in the Southern Hemisphere, the CESM model has greater polar amplification than other models in the Northern Hemisphere for a given CO_2 concentration (in particular at ×3 CO_2). The warming in the ×1 simulations is also similar between the GFDL, CESM, and HadCM3B models. In particular, they all exhibit warming around 30-40° North, which coincides with lower topography in the Tibetan plateau region in the Eocene relative to preindustrial. There is also consistent warming in the Northern Hemisphere Arctic which coincides with lower Eocene albedo, due to the absence of the Greenland ice sheet and due
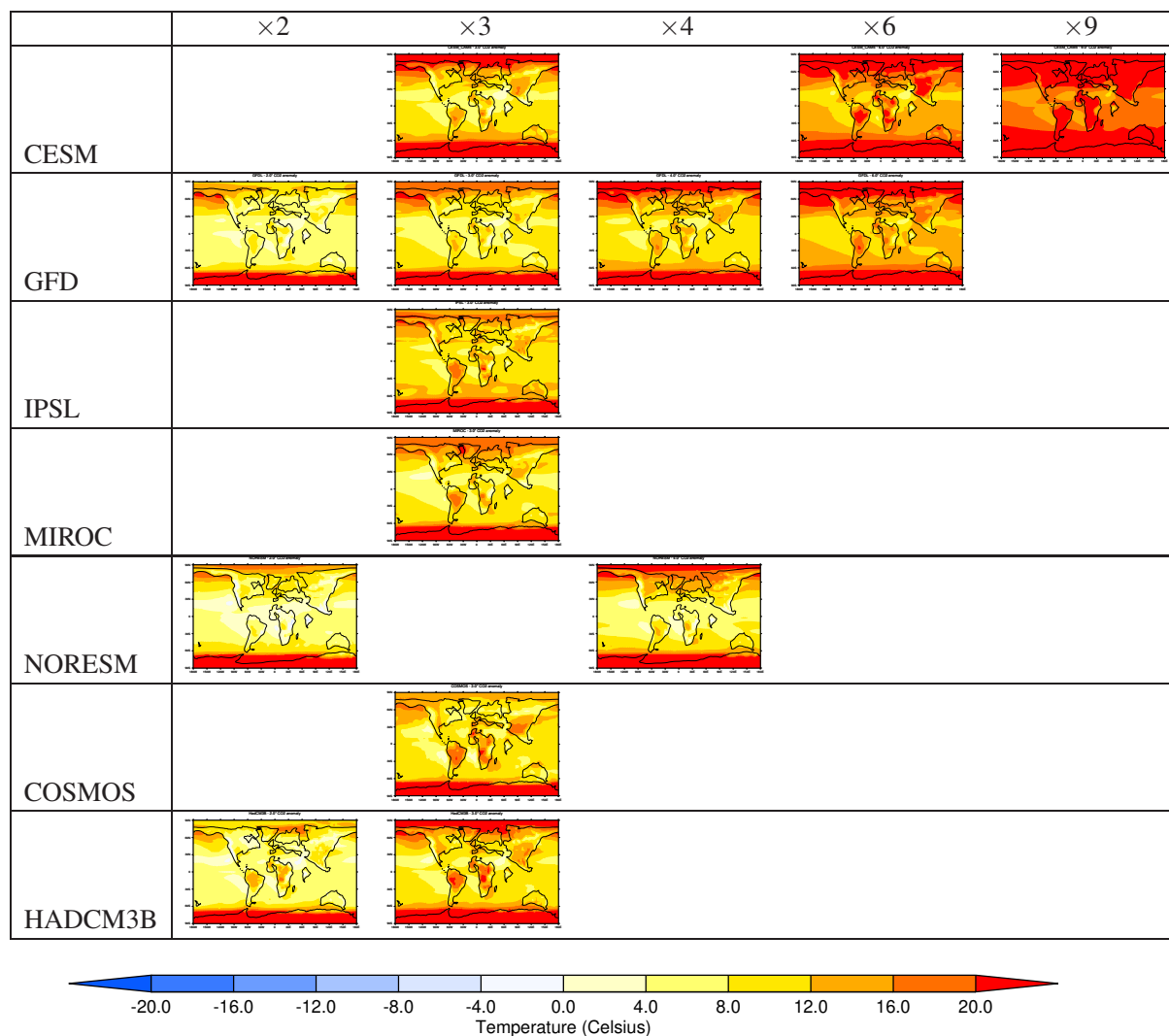
Figure 3. DeepMIP near-surface air temperature anomalies, relative to the zonal mean of the associated preindustrial simulation, ordered by CO_2 concentration and by model. Simulations with CO_2 equal or greater than $\times 2$ are shown. The variable plotted is $GAT_e^m - \overline{GAT_e^m}$ in the nomenclature of Lunt et al. (2012).

360   to boreal forest in place of tundra and bare soil in the preindustrial. The same underlying structure is seen in the higher CO_2 simulations (see for example in GFDL, Figure 2b).

The spatial pattern of surface air temperature response is shown in Figure 3. Because of the difference in continental positions between the preindustrial and Eocene, we show the difference between the Eocene and the zonal mean of the preindustrial, i.e. $GAT_e^m - \overline{GAT_e^m}$ in the nomenclature of Lunt et al. (2012). This shows some consistent responses across the ensemble. In

365   particular, in addition to the polar amplification, the response is characterised by greater warming over land than over ocean.

14

Many of those continental regions where the warming is more muted (such as the Rockies, tropical east Africa, India, and the mid-latitudes of East Asia) are associated with regions of high topography in the Eocene. There is also substantial warming in the North Pacific in all simulations. This is likely associated with deep water formation in this region driving poleward heat transport in the Pacific, but the ocean circulation in these simulations will be explored in a subsequent study.

3.2 Model-data comparison

Here we present a comparison of the models with proxy data of Eocene temperature. The most comprehensive temperature dataset available is that of Hollis et al. (2019), who synthesised temperature records of the latest Paleocene, PETM, and EECO. Here we focus on the EECO dataset, because that one is the largest in terms of number of data points, and is also the time period that our paleogeography is designed to most closely represent. For the purposes of this model-data comparison, we also focus on the proxy SSTs because these are better constrained in time than the proxy terrestrial temperatures, but we do include some comparison with terrestrial surface air temperature (SAT) data. We exclude those SST estimates that were identified by Hollis et al. (2019) as being "frosty" and therefore likely affected by diagenesis. We also make use of the global mean near-surface air temperature estimates from Zhu et al. (2019), which themselves are based on the Hollis et al. (2019) temperatures, and the SST gradient estimates from Cramwinckel et al. (2018). Furthermore, we use the CO_2 range for the EECO presented by Anagnostou et al. (2016).

The red boxes in Figures 1(a,b) show those regions of CO_2-temperature space that can be considered consistent with the published estimates for EECO global mean surface air temperature (26-32°C; Zhu et al., 2019), meridional SST gradient (20-22°C for the EECO; Cramwinckel et al., 2018), and CO_2 (900-2500ppmv; Anagnostou et al., 2016). Note that global mean SSTs and global mean near-surface air temperatures are not the same climatological variable, but have similar values for the Eocene simulations, and we use the same observational constraint for both. None of the pre-DeepMIP simulations are consistent with these three constraints. CCSM_H and CCSM3_W are consistent in terms of global mean and temperature gradient, but only at an inconsistently high CO_2 concentration (×16). Conversely, FAMOUS1 is close to consistency for the temperature constraints but at an inconsistently low CO_2 concentration (×2). CCSM_K comes close to consistency across all three constraints, especially considering the large range of reconstructed meridional temperature gradients from proxies through the Paleogene, which encompass values as low as 18°C in the earliest Eocene (Cramwinckel et al., 2018). However, as discussed in Section 1, CCSM_K includes somewhat arbitrary modifications to cloud parameters that are designed to enable the model to better fit the Eocene observations, and as such, in contrast to the DeepMIP simulations, cannot be considered entirely independent from the temperature data. Of the DeepMIP simulations, GFDL and CESM at ×4 and ×6 are the only simulations that are consistent with all three constraints. COSMOS has very weak polar amplification (even weaker than in the preindustrial), and unless there are strong non-linearities in the response of the model at higher CO_2 concentrations, this model appears to be inconsistent with the proxies. IPSL and HadCM3B also have weak polar amplification of SSTs. It is worth noting that despite this, HadCM3B has greater polar amplification than the pre-DeepMIP HadCM3L simulation of Lunt et al. (2010b); this may be due to the higher resolution in the ocean in HadCM3B, which allows increased flow through Atlantic gateways into the Arctic. It is possible that some of the other DeepMIP models (NORESM and MIROC) could have been consistent with

400   the the proxies if they had carried out simulations at higher CO_2 concentrations. However, as explained in section 2.2, some models develop instabilities at higher CO_2 than those shown here.

    The limited range of CO_2 concentrations explored by some models, coupled with the relatively large uncertainties in EECO CO_2 from proxies, means that a model-data comparison of individual model simulations with the proxy data can be misleading. As such, here we only carry out a detailed model-data comparison for those models that have carried out simulations under

405   more than one CO_2 concentration. For those models (CESM, GFDL, HadCM3B, IPSL, and NorESM), we apply a global-mean scaling factor to the modelled SST anomalies such that they best fit the SST proxy data. We then compare the spatial patterns in the scaled model outputs with the spatial patterns in the proxies. We provide a quantitative metric for the model-data fit, and compare this with some idealised temperature distributions to put these metrics in context.

    The scaling factor that we apply is such that the RMS difference between the model outputs and the proxy data is minimised.

410   In other words, for each model we find a single value of a scaling factor, s, such that an RMS skill score, σ, is minimised:

$$\sigma = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(^{L}T_i^m + s(^{H}T_i^m - {}^{L}T_i^m) - T_i^d)^2} \tag{1}$$

where N is the number of SST proxy data points, T_i^d is the proxy SST at location i, $^{L}T_i^m$ is the modelled SST at location i for a lower CO_2 concentration simulation, and $^{H}T_i^m$ is the modelled SST at location i for a higher CO_2 concentration simulation. We also extrapolate above the highest CO_2 concentration simulation by assuming a constant climate sensitivity appropriate for

415   that model. However, given that some of the models exhibit strong non-linearity, for each model we only extend this by one CO_2-multiple of the highest two CO_2 concentrations (e.g. for IPSL, which has CO_2 concentrations of $\times 1.5$ and $\times 3$, we extend up to $3 \times 3/1.5 = \times 6$, and for HadCM3B, which has CO_2 concentrations at $\times 1$, $\times 2$ and $\times 3$, we extend up to $3 \times 3/2 = \times 4.5$).

    The scaling factor allows us to infer the CO_2 concentration for each model which, if the model scaled linearly as a function of CO_2 between the simulations, would give the best fit to the data. This inferred CO_2 is given by

420   $$CO_2 = 280 \times {}^{L}CO_2 \left(\frac{^{H}CO_2}{^{L}CO_2}\right)^s, \tag{2}$$

where $^{H}CO_2$ and $^{L}CO_2$ are the CO_2 concentrations that correspond to the high and low simulations in Equation 1.

    For CESM, the inferred CO_2 is 2320 ppmv ($\times 8.3$), which is within the range of the CO_2 proxy estimates of the Eocene (Anagnostou et al., 2016). For GFDL, IPSL, HadCM3B, and NorESM, the best fit to the proxy SST data occurs at the maximum of the extrapolated range, which is 2520 ppmv ($\times 9$) for GFDL, 1680 ppmv ($\times 6$) for IPSL, 1260 ppmv ($\times 4.5$) for HadCM3B,

425   and 2240 ppmv ($\times 8$) for NorESM. As such, these are not inferred best estimates for CO_2, but inferred minimum estimates. The scaled SSTs for CESM, GFDL, IPSL, HadCM3B, and NorESM, along with the proxy SST data from Hollis et al. (2019), are shown in Figure 4. In general, the models agree reasonably well with the tropical and mid-latitude SST data, but there is a large model-data inconsistency in the southwest Pacific sites around New Zealand and south of Australia, where the models are colder than proxy estimates by 5–10 °C. See also Figure S1 in Supp Info for the modelled absolute SSTs and absolute SST
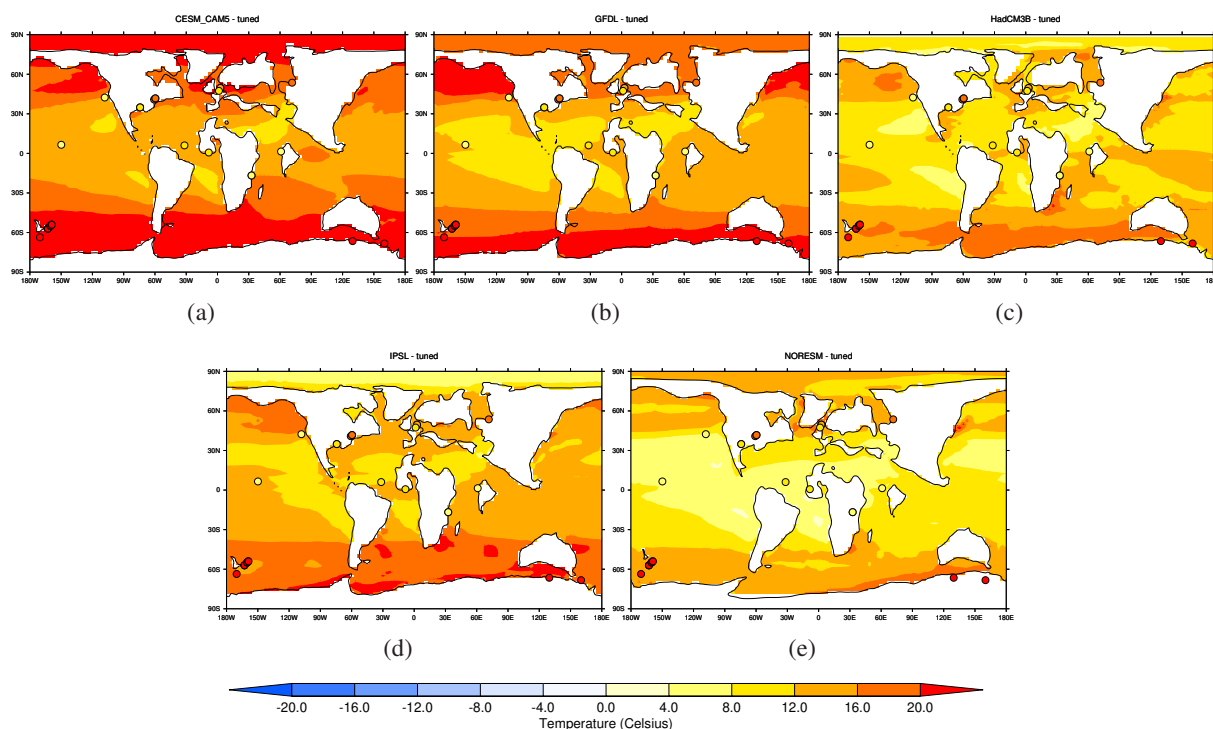
430   proxy data.

Figure 4. Modelled SST anomalies for the Eocene, relative to the zonal mean of the associated preindustrial simulation. The variable plotted is $SST_e^m - \overline{SST_e^m}$ in the nomenclature of Lunt et al. (2012). The Eocene simulations have been scaled using a global tuning factor, as described in the text, to that they best fit the SST proxy data according to an RMS skill-score. As such, only models that carried out simulations with more then one CO_2 concentration are shown. (a) CESM, (b) GFDL, (c) HadCM3B, (d) IPSL, (e) NorESM. Also shown are the proxy SST estimates from Hollis et al. (2019) for the EECO, excluding those sites that they identified as being affected by diagenesis.

The skill-score, σ, from equation 1 is 7.1 for CESM, 7.6 for GFDL, 8.8 for IPSL, 9.6 for HadCM3B, and 10.3 for NorESM. To put these numbers in context, we also calculate the same skill score for some idealised temperature distributions. These are (i) a constant value of zero (i.e. no change from the zonal mean of the pre-industrial), (ii) a constant value C tuned to minimise the skill score, and (ii) a function $A + B(1 - \cos\phi)$ tuned to minimise the skill score. These idealised functions are shown in Figure 5(e-g) as zonal means, along with the tuned DeepMIP models (Figure 5(a-d)). The global anomaly of zero relative to the zonal mean of the preindustrial is associated with a skill score σ=20.1, a global mean constant temperature anomaly is associated with σ=9.4, and the $\cos\phi$ temperature profile is associated with σ=6.5. This means that the CESM, GFDL, and IPSL DeepMIP simulations can be considered as having some skill in capturing the latitudinal gradient (because the skill score of those models is better than that of the global constant), but none of the models have skill in capturing the regional temperature patterns (because the skill score of all models is worse than that of the $\cos\phi$ distribution).

So far this analysis has focussed on SSTs, but we also compare with terrestrial near-surface air temperature data (SAT), even though it is in general less well constrained in age than SSTs, and as such likely represents a wider range of climate states. The
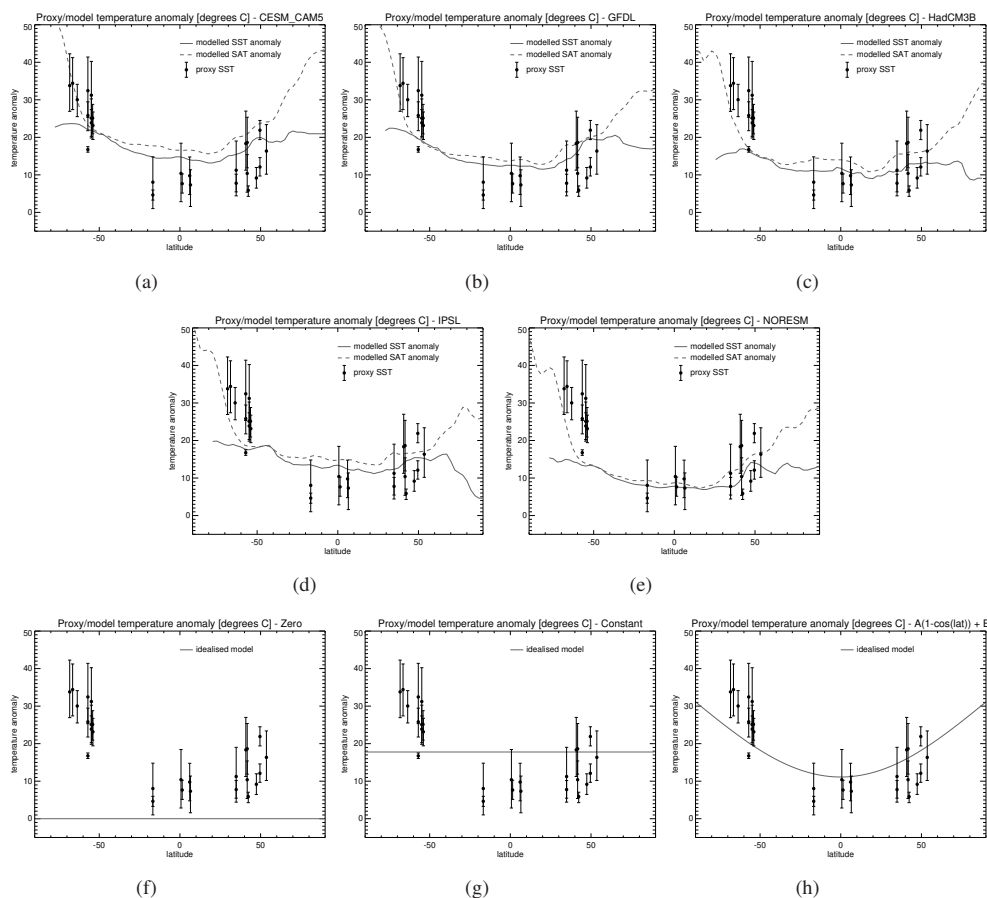
Figure 5. (a,b,c,d) Zonal mean SST (solid lines) and near-surface air temperature (dashed line) anomalies, relative to the zonal mean of the associated preindustrial simulation, for the tuned version of the (a) CESM, (b) GFDL, (c) HadCM3B, (d) IPSL, andn (e) NorESM models. Also shown are the EECO SSTs and error bars from Hollis et al. (2019), also expressed as a difference relative to the zonal mean of the preindustrial. (f,g,h) As (a,b,c,d,e) but instead of a model we show idealised temperature distributions of (f) 0, (g) C and (h) $A + B\cos\phi$. All plots also show the proxy SST estimates from Hollis et al. (2019) for the EECO, excluding those sites that they identified as being affected by diagenesis.

absolute SAT model-data comparison for each DeepMIP simulation is shown in Figure S2 in Supp Info. For those models that carried out more than one CO_2 simulation (CESM, GFDL, HadCM3B, IPSL, and NorESM), Figure S3 and S4 in Supp Info
445 show the SATs from the scaled models in comparison with proxy data. Note that the scaling is calculated to best match the SST data, so Figures 4 and 5 are directly comparable and consistent with Figures S3 and S4. These figures show that the models are capturing the polar amplification indicated in the SAT proxies. In particular, the southwest Pacific site SATs are well simulated in general by the models. Given that the modelled SSTs are substantially colder than the SST proxies in this region implies an inconsistency between marine and terrestrial temperatures in either the proxies or models in this region. Factors that have been

450 previously suggested to cause such a discrepancy include a potential summer bias to all mid and high latitude SST proxies (Hollis et al., 2012), and Red Sea-like features of GDGT distributions in high SST samples from the southwest Pacific and Wilkes Land, that seem to amplify proxy SSTs where isoGDGT$_{RS}$ > 30 (Inglis et al., 2015). If the models are are scaled to best fit the global SAT dataset instead of SSTs, using Equation 1, then the inferred atmospheric CO_2 concentration (Equation 2) is between 900 and 1100 ppmv for CESM, GFDL, IPSL, and HadCM3B, and 1700 ppmv for NorESM, for comparison with

455 the range of CO_2 estimates from Anagnostou et al. (2016) of 900-2500 ppmv.

4 Conclusions

We have presented an ensemble of model simulations of the Eocene, carried out in the framework of DeepMIP. Focus has been on documenting the annual mean modelled temperatures, and comparing with proxy data. Compared with previous model simulations, the results show reduced spread across the ensemble, likely a result of consistent boundary conditions. The

460 contribution to Eocene warmth from non-CO_2 boundary conditions (paleogeography and vegetation and aerosols) is between 3.1°C (HadCM3B) and 5.1°C (CESM). Due to the uncertainties in proxy reconstructed CO_2, we interpolate and extrapolate between simulations at multiple CO_2 concentrations to infer the concentration that gives the best fit to the proxy SSTs, and then compare the model inferred temperatures to the proxy SSTs. This shows that CESM and GFDL give the best fit to the proxy SSTs (at an inferred CO_2 of 2320 ppm and 2520 ppmv respectively), and have "skill" in representing the polar amplification

465 in that they agree better with the data than a tuned global constant warming. It is worth noting that these two models are the only models that implemented modified aerosols in their Eocene simulations (see Section 2.2); this remains a topic for further investigation. However, they do not reproduce the exceptional warmth in the southwest Pacific Australian and New Zealand proxy SSTs, although the modelled and proxy SATs are in agreement in this region, pointing to an inconsistency between the marine and terrestrial paleo temperatures in either the models or the proxies. Despite the regional limitations in the SST

470 model-proxy consistency, the GFDL and CESM models can simulate polar amplification and global mean temperature better than previous simulations, and, in contrast to previous simulations, for the first time achieve this at a CO_2 concentration that is also consistent with the proxies, without prescribing changes to clouds. Future work will explore the model simulations and model-data comparisons in more detail, in particular the hydrological cycle and ocean circulation.

Author contributions. All authors contributed to the writing of the paper. DL wrote the first draft and carried out the analysis. The model simulations were carried out by FB and SS (HadCM3B), WLC (MIROC), DH (GFDL), JBL,PS and YD (IPSL), IN (COSMOS), ZZ (NORESM), and JZ (CESM). TDJ and CH led the compilation of the DeepMIP database. DL, MH, and BOB coordinated the study.

Climate
of the Past
Discussions

References

Abelmann, A., Gersonde, R., Knorr, G., Zhang, X., Chapligin, B., Maier, E., Esper, O., Friedrichsen, H., Lohmann, G., Meyer,
490   H., and Tiedemann, R.: The seasonal sea-ice zone in the glacial Southern Ocean as a carbon sink, Nature communications, 6,
       https://doi.org/10.1038/ncomms9136, 2015.

Anagnostou, E., John, E., Edgar, K., Foster, G., Ridgwell, A., Inglis, G., Pancost, R., Lunt, D., and Pearson, P.: Changing atmospheric CO2
       concentration was the primary driver of early Cenozoic climate, Nature, https://doi.org/10.1038/nature17423, 2016.

Baatsen, M., van Hinsbergen, D. J. J., von der Heydt, A. S., Dijkstra, H. A., Sluijs, A., Abels, H. A., and Bijl, P. K.: Reconstructing geographi-
495   cal boundary conditions for palaeoclimate modelling during the Cenozoic, Climate of the Past, 12, 1635–1644, https://doi.org/10.5194/cp-
       12-1635-2016, https://www.clim-past.net/12/1635/2016/, 2016.

Barron, E. J.: Eocene equator-to-pole surface ocean temperatures: A significant climate problem?, Paleoceanography, 2, 729–739,
       https://doi.org/10.1029/PA002i006p00729, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/PA002i006p00729, 1987.

Beerling, D. and Royer, D.: Convergent Cenozoic CO$_2$ history, Nature Geoscience, 4, 418–420, 2011.

500   Burke, K. D., Williams, J. W., Chandler, M. A., Haywood, A. M., Lunt, D. J., and Otto-Bliesner, B. L.: Pliocene and
       Eocene provide best analogs for near-future climates, Proceedings of the National Academy of Sciences, 115, 13 288–13 293,
       https://doi.org/10.1073/pnas.1809600115, https://www.pnas.org/content/115/52/13288, 2018.

Caballero, R. and Huber, M.: State-dependent climate sensitivity in past warm climates and its implications for future climate projections,
       Proc Natl Acad Sci, 110, 14 162–14 167, 2013.

505   Carlson, H. and Caballero, R.: Atmospheric circulation and hydroclimate impacts of alternative warming scenarios for the Eocene, Climate
       of the Past, 13, 1037–1048, https://doi.org/10.5194/cp-13-1037-2017, https://www.clim-past.net/13/1037/2017/, 2017.

Carmichael, M. J., Lunt, D. J., Huber, M., Heinemann, M., Kiehl, J., LeGrande, A., Loptson, C. A., Roberts, C. D., Sagoo, N., Shields, C.,
       Valdes, P. J., Winguth, A., Winguth, C., and Pancost, R. D.: A model-model and data-model comparison for the early Eocene hydrological
       cycle, Climate of the Past, 12, 455–481, https://doi.org/10.5194/cp-12-455-2016, http://www.clim-past.net/12/455/2016/, 2016.

510   Chan, W.-L., Abe-Ouchi, A., and Ohgaito, R.: Simulating the mid-Pliocene climate with the MIROC general circulation model: experimental
       design and initial results, Geoscientific Model Development, 4, 1035–1049, 2011.

Cramwinckel, M. J., Huber, M., , I. J. K., Agnini, C., Bijl, P. K., Bohaty, S. M., Frieling, J., Hilgen, A. G. F. J., Kip, E. L., Peterse, F.,
       van der Ploegand U. Röhl, R., Schouten, S., and Sluijs, A.: Synchronous tropical and polar temperature evolution in the Eocene, Nature,
       559, 382–386, 2018.

515   de Lavergne, C., Falahat, S., Madec, G., Roquet, F., Nycander, J., and Vic, C.: Toward global maps of internal tide energy sinks, Ocean
       Modelling, 137, 52 – 75, https://doi.org/https://doi.org/10.1016/j.ocemod.2019.03.010, http://www.sciencedirect.com/science/article/pii/
       S1463500318302890, 2019.

Delworth, T. L., Broccoli, A. J., Rosati, A., Stouffer, R. J., Balaji, V., Beesley, J. A., Cooke, W. F., Dixon, K. W., Dunne, J., Dunne, K. A.,
       Durachta, J. W., Findell, K. L., Ginoux, P., Gnanadesikan, A., Gordon, C. T., Griffies, S. M., Gudgel, R., Harrison, M. J., Held, I. M.,
520   Hemler, R. S., Horowitz, L. W., Klein, S. A., Knutson, T. R., Kushner, P. J., Langenhorst, A. R., Lee, H.-C., Lin, S.-J., Lu, J., Malyshev,
       S. L., Milly, P. C. D., Ramaswamy, V., Russell, J., Schwarzkopf, M. D., Shevliakova, E., Sirutis, J. J., Spelman, M. J., Stern, W. F., Winton,
       M., Wittenberg, A. T., Wyman, B., Zeng, F., and Zhang, R.: GFDL's CM2 Global Coupled Climate Models. Part I: Formulation and
       Simulation Characteristics, Journal of Climate, 19, 643–674, https://doi.org/10.1175/JCLI3629.1, http://dx.doi.org/10.1175/JCLI3629.1,
       2006.

525   Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., de Noblet, N., Duvel, J.-P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J.-Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M.-P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio,
530   C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N., and Vuichard, N.: Climate Change Projections Using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5, Climate Dynamics, 40, 2123–2165, https://doi.org/10.1007/s00382-012-1636-1, 2013.

Emile-Geay, J. and Madec, G.: Geothermal heating, diapycnal mixing and the abyssal circulation, Ocean Science, 5, 203–217, https://doi.org/10.5194/os-5-203-2009, https://www.ocean-sci.net/5/203/2009/, 2009.

535   Farnsworth, A., Lunt, D. J., O'Brien, C. L., Foster, G. L., Inglis, G. N., Markwick, P., Pancost, R. D., and Robinson, S. A.: Climate Sensitivity on Geological Timescales Controlled by Nonlinear Feedbacks and Ocean Circulation, Geophysical Research Letters, 46, 9880–9889, https://doi.org/10.1029/2019GL083574, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL083574, 2019.

Feng, R., Otto-Bliesner, B. L., Xu, Y., Brady, E., Fletcher, T., and Ballantyne, A.: Contributions of aerosol-cloud interactions to mid-Piacenzian seasonally sea ice-free Arctic Ocean, Geophysical Research Letters, 46, 9920–9929, https://doi.org/10.1029/2019GL083960,
540   2019.

Galbraith, E. D., Gnanadesikan, A., Dunne, J. P., and Hiscock, M. R.: Regional impacts of iron-light colimitation in a global biogeochemical model, Biogeosciences, 7, 1043–1064, https://doi.org/10.5194/bg-7-1043-2010, http://www.biogeosciences.net/7/1043/2010/, 2010.

Gasson, E., Lunt, D. J., DeConto, R., Goldner, A., Heinemann, M., Huber, M., LeGrande, A. N., Pollard, D., Sagoo, N., Siddall, M., Winguth, A., and Valdes, P. J.: Uncertainties in the modelled CO_2 threshold for Antarctic glaciation, Climate of the Past, 10, 451–466,
545   https://doi.org/10.5194/cp-10-451-2014, http://www.clim-past.net/10/451/2014/, 2014.

Gettelman, A., Kay, J. E., and Shell, K. M.: The Evolution of Climate Sensitivity and Climate Feedbacks in the Community Atmosphere Model, Journal of Climate, 25, 1453–1469, https://doi.org/10.1175/JCLI-D-11-00197.1, https://doi.org/10.1175/JCLI-D-11-00197.1, 2012.

Gierz, P., Werner, M., and Lohmann, G.: Simulating climate and stable water isotopes during the Last Interglacial using a coupled climate-
550   isotope model, Journal of Advances in Modeling Earth Systems, 9, 2027–2045, https://doi.org/10.1002/2017MS001056, 2017.

Gong, X., Knorr, G., Lohmann, G., and Zhang, X.: Dependence of abrupt Atlantic meridional ocean circulation changes on climate background states, Geophysical Research Letters, 40, 3698–3704, https://doi.org/10.1002/grl.50701, 2013.

Green, J. A. M. and Huber, M.: Tidal dissipation in the early Eocene and implications for ocean mixing, Geophysical Research Letters, 40, 2707–2713, https://doi.org/10.1002/grl.50510, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/grl.50510, 2013.

555   Guo, C., Bentsen, M., Bethke, I., Ilicak, M., Tjiputra, J., Toniazzo, T., Schwinger, J., and Otterå, O. H.: Description and evaluation of NorESM1-F: a fast version of the Norwegian Earth System Model (NorESM), Geoscientific Model Development, 12, 343–362, https://doi.org/10.5194/gmd-12-343-2019, https://www.geosci-model-dev.net/12/343/2019/, 2019.

Hasumi, H.: CCSR Ocean Component Model (COCO) Version 2.1, Tech. rep., The University of Tokyo, 2000.

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.,
560   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, Climate of the Past, 9, 191–209, https://doi.org/10.5194/cp-9-191-2013, http://www.clim-past.net/9/191/2013/, 2013.

Heavens, N. G., Shields, C. A., and Mahowald, N. M.: A paleogeographic approach to aerosol prescription in simulations of deep time climate, Journal of Advances in Modeling Earth Systems, 4, https://doi.org/10.1029/2012MS000166, https://doi.org/10.1029/2012MS000166, 2012.

Heinemann, M., Jungclaus, J. H., and Marotzke, J.: Warm Paleocene/Eocene climate as simulated in ECHAM5/MPI-OM, Climate of the Past, 5, 785–802, 2009.

Herold, N., Buzan, J., Seton, M., Goldner, A., Green, J. A. M., Müller, R. D., Markwick, P., and Huber, M.: A suite of early Eocene (55 Ma) climate model boundary conditions, Geoscientific Model Development, 7, 2077–2090, https://doi.org/10.5194/gmd-7-2077-2014, http://www.geosci-model-dev.net/7/2077/2014/, 2014.

Hollis, C. J., Taylor, K. W. R., Handley, L., Pancost, R. D., Huber, M., Creech, J. B., Hines, B. R., Crouch, E. M., Morgans, H. E. G., Crampton, J. S., Gibbs, S., Pearson, P. N., and Zachos, J. C.: Early Paleogene temperature history of the Southwest Pacific Ocean: Reconciling proxies and models, Earth and Planetary Science Letters, 349-350, 53–66, 2012.

Hollis, C. J., Dunkley Jones, T., Anagnostou, E., Bijl, P. K., Cramwinckel, M. J., Cui, Y., Dickens, G. R., Edgar, K. M., Eley, Y., Evans, D., Foster, G. L., Frieling, J., Inglis, G. N., Kennedy, E. M., Kozdon, R., Lauretano, V., Lear, C. H., Littler, K., Lourens, L., Meckler, A. N., Naafs, B. D. A., Pälike, H., Pancost, R. D., Pearson, P. N., Röhl, U., Royer, D. L., Salzmann, U., Schubert, B. A., Seebeck, H., Sluijs, A., Speijer, R. P., Stassen, P., Tierney, J., Tripati, A., Wade, B., Westerhold, T., Witkowski, C., Zachos, J. C., Zhang, Y. G., Huber, M., and Lunt, D. J.: The DeepMIP contribution to PMIP4: methodologies for selection, compilation and analysis of latest Paleocene and early Eocene climate proxy data, incorporating version 0.1 of the DeepMIP database, Geoscientific Model Development, 12, 3149–3206, https://doi.org/10.5194/gmd-12-3149-2019, https://www.geosci-model-dev.net/12/3149/2019/, 2019.

Huang, X., Stärz, M., Gohl, K., Knorr, G., and Lohmann, G.: Impact of Weddell Sea shelf progradation on Antarctic bottom water formation during the Miocene, Paleoceanography, 32, 304–317, https://doi.org/10.1002/2016PA002987, 2017.

Huber, M. and Caballero, R.: The early Eocene equable climate problem revisited, Climate of the Past, 7, 603–633, 2011.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J. F., Large, W. G., Lawrence, D., Lindsay, K., Lipscomb, W. H., Long, M. C., Mahowald, N., Marsh, D. R., Neale, R. B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W. D., Hack, J. J., Kiehl, J. T., and Marshall, S.: The community earth system model: A framework for collaborative research, Bulletin of the American Meteorological Society, 94, 1339–1360, https://doi.org/10.1175/BAMS-D-12-00121.1, http://dx.doi.org/10.1175/BAMS-D-12-00121.1, 2013.

Hutchinson, D. K., de Boer, A. M., Coxall, H. K., Caballero, R., Nilsson, J., and Baatsen, M.: Climate sensitivity and meridional overturning circulation in the late Eocene using GFDL CM2.1, Climate of the Past, 14, 789–810, https://doi.org/10.5194/cp-14-789-2018, https://www.clim-past.net/14/789/2018/, 2018.

Hutchinson, D. K., Coxall, H. K., O'Regan, M., Nilsson, J., Caballero, R., and de Boer, A. M.: Arctic closure as a trigger for Atlantic overturning at the Eocene-Oligocene Transition, Nature Communications, 10, 3797, https://doi.org/10.1038/s41467-019-11828-z, https://doi.org/10.1038/s41467-019-11828-z, 2019.

Inglis, G. N., Farnsworth, A., Lunt, D., Foster, G., Hollis, C., Pagani, M., Jardine, P., Pearson, P., Markwick, P., Raynman, L., Galsworthy, A., and Pancost, R.: Descent towards the Icehouse: Eocene sea surface cooling inferred from GDGT distributions, Paleoceanography, 30, doi:10.1002/2014PA002 723, 2015.

Jungclaus, J. H., Keenlyside, N., Botzet, M., Haak, H., Luo, J.-J., Latif, M., Marotzke, J., Mikolajewicz, U., and Roeckner, E.: Ocean circulation and tropical variability in the coupled model ECHAM5/MPI-OM, Journal of climate, 19, 3952–3972, https://doi.org/10.1175/JCLI3827.1, 2006.

Climate
of the Past
Discussions

Open Access

EGU

K-1 model developers: K-1 coupled model (MIROC) description, Tech. rep., Center for Climate System Research (University of Tokyo), National Institute for Environmental Studies, Frontier Research Center for Global Change, 2004.

Kageyama, M., Braconnot, P., Bopp, L., Caubel, A., Foujols, M.-A., Guilyardi, E., Khodri, M., Lloyd, J., Lombard, F., Mariotti, V., Marti, O., Roy, T., and Woillez, M.-N.: Mid-Holocene and Last Glacial Maximum climate simulations with the IPSL model—part I: comparing IPSL_CM5A to IPSL_CM4, Climate Dynamics, 40, 2447–2468, https://doi.org/10.1007/s00382-012-1488-8, https://doi.org/10.1007/s00382-012-1488-8, 2013.

Kay, J. E., Hillman, B. R., Klein, S. A., Zhang, Y., Medeiros, B., Pincus, R., Gettelman, A., Eaton, B., Boyle, J., Marchand, R., and Ackerman, T. P.: Exposing Global Cloud Biases in the Community Atmosphere Model (CAM) Using Satellite Observations and Their Corresponding Instrument Simulators, Journal of Climate, 25, 5190–5207, https://doi.org/10.1175/JCLI-D-11-00469.1, https://doi.org/10.1175/JCLI-D-11-00469.1, 2012.

Kiehl, J. T. and Shields, C. A.: Sensitivity of the PalaeoceneEocene Thermal Maximum climate to cloud properties, Phil. Trans. R. Soc. A, 371, https://doi.org/10.1098/rsta.2013.0093, 2013.

Knorr, G. and Lohmann, G.: Climate warming during Antarctic ice sheet expansion at the Middle Miocene transition, Nature Geoscience, 7, 376–381, https://doi.org/10.1038/NGEO2119, 2014.

Knorr, G., Butzin, M., Micheels, A., and Lohmann, G.: A warm Miocene climate at low atmospheric $CO2$ levels, Geophyscial Research Letters, 38, https://doi.org/10.1029/2011GL048873, 2011.

Koch-Larrouy, A., Madec, G., Bouruet-Aubertot, P., Gerkema, T., Bessières, L., and Molcard, R.: On the transformation of Pacific Water into Indonesian Throughflow Water by internal tidal mixing, Geophysical Research Letters, 34, https://doi.org/10.1029/2006GL028405, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006GL028405, 2007.

Lohmann, G., Pfeiffer, M., Laepple, T., Leduc, G., and Kim, J.-H.: A model-data comparison of the Holocene global sea surface temperature evolution, Climate of the Past, 9, 1807–1839, https://doi.org/10.5194/cp-9-1807-2013, 2013.

Lott, F.: Alleviation of Stationary Biases in a GCM through a Mountain Drag Parameterization Scheme and a Simple Representation of Mountain Lift Forces, Monthly Weather Review, 127, 788–801, https://doi.org/10.1175/1520-0493(1999)127<0788:AOSBIA>2.0.CO;2, 1999.

Lott, F. and Miller, M. J.: A new subgrid-scale orographic drag parametrization: Its formulation and testing, Quarterly Journal of the Royal Meteorological Society, 123, 101–127, https://doi.org/10.1002/qj.49712353704, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712353704, 1997.

Lunt, D. J., Foster, G. L., Haywood, A. M., and Stone, E. J.: Late Pliocene Greenland glaciation controlled by a decline in atmospheric CO_2 levels, Nature, 454, 1102–1105, 2008.

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, Nature Geoscience, 3, 60–64, 2010a.

Lunt, D. J., Valdes, P. J., Dunkley Jones, T., Ridgwell, A., Haywood, A. M., Schmidt, D. N., Marsh, R., and Maslin, M.: CO_2-driven ocean circulation changes as an amplifier of Paleocene-Eocene thermal maximum hydrate destabilization, Geology, 38, 875–878, 2010b.

Lunt, D. J., Dunkley Jones, T., Heinemann, M., Huber, M., LeGrande, A., Winguth, A., Loptson, C., Marotzke, J., Roberts, C. D., Tindall, J., Valdes, P., and Winguth, C.: A model-data comparison for a multi-model ensemble of early Eocene atmosphere-ocean simulations: EoMIP, Climate of the Past, 8, 1717–1736, 2012.

Lunt, D. J., Farnsworth, A., Loptson, C., Foster, G. L., Markwick, P., O'Brien, C. L., Pancost, R. D., Robinson, S. A., and Wrobel, N.: Palaeogeographic controls on climate and proxy interpretation, Climate of the Past, 12, 1181–1198, 2016.

Lunt, D. J., Huber, M., Anagnostou, E., Baatsen, M. L. J., Caballero, R., DeConto, R., Dijkstra, H. A., Donnadieu, Y., Evans, D., Feng, R.,
640   Foster, G. L., Gasson, E., von der Heydt, A. S., Hollis, C. J., Inglis, G. N., Jones, S. M., Kiehl, J., Kirtland Turner, S., Korty, R. L., Koz-
       don, R., Krishnan, S., Ladant, J.-B., Langebroek, P., Lear, C. H., LeGrande, A. N., Littler, K., Markwick, P., Otto-Bliesner, B., Pearson, P.,
       Poulsen, C. J., Salzmann, U., Shields, C., Snell, K., Stärz, M., Super, J., Tabor, C., Tierney, J. E., Tourte, G. J. L., Tripati, A., Upchurch,
       G. R., Wade, B. S., Wing, S. L., Winguth, A. M. E., Wright, N. M., Zachos, J. C., and Zeebe, R. E.: The DeepMIP contribution to PMIP4:
       experimental design
645   for model simulations of the EECO, PETM, and
       pre-PETM (version 1.0), Geoscientific Model Development, 10, 889–901, https://doi.org/10.5194/gmd-10-889-2017, https://www.
       geosci-model-dev.net/10/889/2017/, 2017.

Marsland, S. J., Haak, H., Jungclaus, J. H., Latif, M., and Röske, F.: The Max-Planck-Institute global ocean/sea ice model with orthogonal
       curvilinear coordinates, Ocean Modelling, 5, 91–127, https://doi.org/10.1016/S1463-5003(02)00015-X, 2003.

650   Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., Rouco, J. G., Jansen, E., Lambeck, K., Luterbacher, J., Naish,
       T., Osborn, T., Otto-Bliesner, B., Quinn, T., Ramesh, R., Rojas, M., Shao, X., and Timmermann, A.: Information from Paleoclimate
       Archives, 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, edited by Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels,
       A., Xia, Y., Bex, V., and Midgley, P., Cambridge University Press, 2013.

655   Müller, R. D., Sdrolias, M., Gaina, C., and Roest, W. R.: Age, spreading rates, and spreading asymmetry of the world's ocean crust, Geochem.
       Geophys. Geosyst., 9, Q04 006, https://doi.org/10.1029/2007GC001743, 2008.

Niezgodzki, I., Knorr, G., Lohmann, G., Tyszka, J., and Markwick, P. J.: Late Cretaceous climate simulations with different CO_2 levels and
       subarctic gateway configurations: A model–data comparison, Paleoceanography, 32, https://doi.org/10.1002/2016PA003055, 2017.

Niezgodzki, I., Tyszka, J., Knorr, G., and Lohmann, G.: Was the Arctic Ocean ice free during the latest Cretaceous? The role of CO_2 and
660   gateway configurations, Global and Planetary Change, 177, 201–212, https://doi.org/10.1016/j.gloplacha.2019.03.011, 2019.

Otto-Bliesner, B. L., Brady, E. C., Fasullo, J., Jahn, A., Landrum, L., Stevenson, S., Rosenbloom, N., Mai, A., and Strand, G.: Climate
       Variability and Change since 850 CE: An Ensemble Approach with the Community Earth System Model, Bulletin of the American
       Meteorological Society, 97, 735–754, https://doi.org/10.1175/BAMS-D-14-00233.1, 2015.

Pfeiffer, M. and Lohmann, G.: Greenland Ice Sheet influence on Last Interglacial climate: global sensitivity studies performed with an
665   atmosphere–ocean general circulation model, Climate of the Past, 12, 1313–1338, https://doi.org/10.5194/cp-12-1313-2016, 2016.

Roberts, C. D., LeGrande, A. N., and Tripati, A. K.: Climate sensitivity to Arctic seaway restriction during the early Paleogene, Earth and
       Planetary Science Letters, 286, 576–585, 2009.

Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kirchner, I., Kornbleuh, L., Manzini, E.,
       Rhodin, A., Schlese, U., Schulzweida, U., and Tompkins, A.: The atmospheric general circulation model ECHAM 5. PART I: Model de-
670   scription, Report 349, Max-Planck Institut fűr Meteorologie, Hamburg, http://www.mpimet.mpg.de/fileadmin/models/echam/mpi_report_
       349.pdf, 2003.

Sagoo, N., Valdes, P., Flecker, R., and Gregoire, L.: The Early Eocene equable climate problem: can perturbations of climate model parame-
       ters identify possible solutions?, Phil. Trans. R. Soc. A, 371, https://doi.org/10.1098/rsta.2013.0123, 2013.

Sepulchre, P., Caubel, A., Ladant, J.-B., Bopp, L., Boucher, O., Braconnot, P., Brockmann, P., Cozic, A., Donnadieu, Y., Estella-Perez, V.,
675   Ethé, C., Fluteau, F., Foujols, M.-A., Gastineau, G., Ghattas, J., Hauglustaine, D., Hourdin, F., Kageyama, M., Khodri, M., Marti, O., Meur-
       desoif, Y., Mignot, J., Sarr, A.-C., Servonnat, J., Swingedouw, D., Szopa, S., and Tardif, D.: IPSL-CM5A2. An Earth System Model de-

signed for multi-millennial climate simulations, Geoscientific Model Development Discussions, 2019, 1–57, https://doi.org/10.5194/gmd-2019-332, https://www.geosci-model-dev-discuss.net/gmd-2019-332/, 2019.

680 Simmons, H. L., Jayne, S. R., Laurent, L. C., and Weaver, A. J.: Tidally driven mixing in a numerical model of the ocean general circulation, Ocean Modelling, 6, 245 – 263, https://doi.org/https://doi.org/10.1016/S1463-5003(03)00011-8, http://www.sciencedirect.com/science/article/pii/S1463500303000118, 2004.

Sloan, L. and Barron, E. J.: A comparison of Eocene climate model results to quantified paleoclimatic interpretations, Palaeogeography, Palaeoclimatology, Palaeoecology, 93, 183 – 202, https://doi.org/https://doi.org/10.1016/0031-0182(92)90096-N, http://www.sciencedirect.com/science/article/pii/003101829290096N, 1992.

685 Smith, R., Jones, P., Briegleb, B., Bryan, F., Danabasoglu, G., Dennis, J., Dukowicz, J., Eden, C., Fox-Kemper, B., Gent, P., and Others: The parallel ocean program (POP) reference manual ocean component of the community climate system model (CCSM) and community earth system model (CESM), Rep. LAUR-01853, 141, 1–140, 2010.

Stärz, M., Jokat, W., Knorr, G., and Lohmann, G.: Threshold in North Atlantic-Arctic Ocean circulation controlled by the subsidence of the Greenland-Scotland Ridge, Nature Communications, 8, https://doi.org/10.1038/ncomms15681, 2017.

690 Stein, C. A. and Stein, S.: A model for the global variation in oceanic depth and heat flow with lithospheric age, Nature, 359, 123–129, https://doi.org/10.1038/359123a0, https://doi.org/10.1038/359123a0, 1992.

Stepanek, C. and Lohmann, G.: Modelling mid-Pliocene climate with COSMOS, Geoscientific Model Development, 5, 1221–1243, https://doi.org/10.5194/gmd-5-1221-2012, 2012.

Takata, K., Watanabe, T., and Emori, S.: Development of the minimal advanced treatments of surface interaction and runoff, Global and

695 Planetary Change, 38, 209–222, 2003.

Thibodeau, B., Not, C., Zhu, J., Schmittner, A., Noone, D., Tabor, C., Zhang, J., and Liu, Z.: Last Century Warming Over the Canadian Atlantic Shelves Linked to Weak Atlantic Meridional Overturning Circulation, Geophysical Research Letters, 45, 12,376–12,385, https://doi.org/10.1029/2018GL080083, 2018.

Valdes, P. J., Armstrong, E., Badger, M. P. S., Bradshaw, C. D., Bragg, F., Crucifix, M., Davies-Barnard, T., Day, J. J., Farnsworth, A.,

700 Gordon, C., Hopcroft, P. O., Kennedy, A. T., Lord, N. S., Lunt, D. J., Marzocchi, A., Parry, L. M., Pope, V., 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, Geoscientific Model Development, 10, 3715–3743, https://doi.org/10.5194/gmd-10-3715-2017, 2017.

Walliser, E. O., Lohmann, G., Niezgodzki, I., Tűtken, T., and Schőne, B. R.: Response of Central European SST to atmospheric pCO_2 forcing during the Oligocene–A combined proxy data and numerical climate model approach, Palaeogeography, Palaeoclimatology, Palaeoecol-

705 ogy, 459, 552–569, https://doi.org/10.1016/j.palaeo.2016.07.033, 2016.

Walliser, E. O., Lohmann, G., Niezgodzki, I., and Schőne, B. R.: Inter-annual climate variability in Europe during the Oligocene icehouse, Palaeogeography, Palaeoclimatology, Palaeoecology, 475, 140–153, https://doi.org/10.1016/j.palaeo.2017.03.020, 2017.

Wei, W. and Lohmann, G.: Simulated Atlantic multidecadal oscillation during the Holocene, Journal of Climate, 25, 6989–7002, https://doi.org/10.1175/JCLI-D-11-00667.1, 2012.

710 Wei, W., Lohmann, G., and Dima, M.: Distinct modes of internal variability in the global meridional overturning circulation associated with the Southern Hemisphere westerly winds, Journal of Physical Oceanography, 42, 785–801, https://doi.org/10.1175/JPO-D-11-038.1, 2012.

Winguth, A., Shellito, C., Shields, C., and Winguth, C.: Climate Response at the Paleocene-Eocene Thermal Maximum to Greenhouse Gas Forcing - A Model Study with CCSM3, Journal of Climate, 23, 2562–2584, 2010.

715 Zhang, X., Lohmann, G., Knorr, G., and Xu, X.: Different ocean states and transient characteristics in Last Glacial Maximum simulations and implications for deglaciation, Climate of the Past, 9, 2319–2333, https://doi.org/10.5194/cp-9-2319-2013, 2013.

Zhang, X., Lohmann, G., Knorr, G., and Purcell, C.: Abrupt glacial climate shifts controlled by ice sheet changes, Nature, 512, 290–294, https://doi.org/10.1038/nature13592, 2014.

Zhang, X., Knorr, G., Lohmann, G., and Barker, S.: Abrupt North Atlantic circulation changes in response to gradual CO_2 forcing in a glacial
720 climate state, Nature Geoscience, 10, 518–523, https://doi.org/10.1038/ngeo2974, 2017.

Zhang, Z. S., Nisancioglu, K., Bentsen, M., Tjiputra, J., Bethke, I., Yan, Q., Risebrobakken, B., Andersson, C., and Jansen, E.: Pre-industrial and mid-Pliocene simulations with NorESM-L, Geoscientific Model Development, 5, 523–533, https://doi.org/10.5194/gmd-5-523-2012, https://www.geosci-model-dev.net/5/523/2012/, 2012.

Zhu, J., Liu, Z., Brady, E., Otto-Bliesner, B., Zhang, J., Noone, D., Tomas, R., Nusbaumer, J., Wong, T., Jahn, A., and Tabor, C.: Re-
725 duced ENSO variability at the LGM revealed by an isotope-enabled Earth system model, Geophysical Research Letters, 44, 6984–6992, https://doi.org/10.1002/2017GL073406, 2017a.

Zhu, J., Liu, Z., Brady, E. C., Otto-Bliesner, B. L., Marcott, S. A., Zhang, J., Wang, X., Nusbaumer, J., Wong, T. E., Jahn, A., and Noone, D.: Investigating the direct meltwater effect in terrestrial oxygen-isotope paleoclimate records using an isotope-enabled Earth system model, Geophysical Research Letters, 44, 12 501–12 510, https://doi.org/10.1002/2017GL076253, 2017b.

730 Zhu, J., Poulsen, C. J., and Tierney, J. E.: Simulation of Eocene extreme warmth and high climate sensitivity through cloud feedbacks, Science Advances, 5, https://doi.org/10.1126/sciadv.aax1874, 2019.