



# A model–data comparison for a multi-model ensemble of early Eocene atmosphere–ocean simulations: EoMIP

D. J. Lunt<sup>1</sup>, T. Dunkley Jones<sup>2,\*</sup>, M. Heinemann<sup>3</sup>, M. Huber<sup>4</sup>, A. LeGrande<sup>5</sup>, A. Winguth<sup>6</sup>, C. Loptson<sup>1</sup>, J. Marotzke<sup>7</sup>, C. D. Roberts<sup>8</sup>, J. Tindall<sup>9</sup>, P. Valdes<sup>1</sup>, and C. Winguth<sup>6</sup>

<sup>1</sup>School of Geographical Sciences, University of Bristol, UK

<sup>2</sup>Department of Earth Science and Engineering, Imperial College London, UK

<sup>3</sup>International Pacific Research Center, University of Hawaii, USA

<sup>4</sup>Department of Earth and Atmospheric Sciences, Purdue University, USA

<sup>5</sup>NASA/Goddard Institute for Space Studies, USA

<sup>6</sup>Department of Earth and Environmental Science, University of Texas at Arlington, USA

<sup>7</sup>Max Planck Institute for Meteorology, Hamburg, Germany

<sup>8</sup>The Met Office, Exeter, UK

<sup>9</sup>School of Earth and Environment, University of Leeds, UK

\*now at: School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK

Correspondence to: D. J. Lunt (d.j.lunt@bristol.ac.uk)

Received: 29 March 2012 – Published in Clim. Past Discuss.: 16 April 2012

Revised: 19 September 2012 – Accepted: 20 September 2012 – Published: 29 October 2012

**Abstract.** The early Eocene ( $\sim 55$  to 50 Ma) is a time period which has been explored in a large number of modelling and data studies. Here, using an ensemble of previously published model results, making up “EoMIP” – the Eocene Modelling Intercomparison Project – and syntheses of early Eocene terrestrial and sea surface temperature data, we present a self-consistent inter-model and model–data comparison. This shows that the previous modelling studies exhibit a very wide inter-model variability, but that at high CO<sub>2</sub>, there is good agreement between models and data for this period, particularly if possible seasonal biases in some of the proxies are considered. An energy balance analysis explores the reasons for the differences between the model results, and suggests that differences in surface albedo feedbacks, water vapour and lapse rate feedbacks, and prescribed aerosol loading are the dominant cause for the different results seen in the models, rather than inconsistencies in other prescribed boundary conditions or differences in cloud feedbacks. The CO<sub>2</sub> level which would give optimal early Eocene model–data agreement, based on those models which have carried out simulations with more than one CO<sub>2</sub> level, is in the range of 2500 ppmv to 6500 ppmv. Given the spread of model results, tighter bounds on proxy estimates

of atmospheric CO<sub>2</sub> and temperature during this time period will allow a quantitative assessment of the skill of the models at simulating warm climates. If it is the case that a model which gives a good simulation of the Eocene will also give a good simulation of the future, then such an assessment could be used to produce metrics for weighting future climate predictions.

## 1 Introduction

Making robust predictions of future climate change is a major challenge, which has environmental, societal, and economic relevance. The numerical models which are used to make these predictions are normally tested over time periods for which there are extensive instrumental records of climate available, typically over the last  $\sim 100$  yr (Hegerl et al., 2007). However, the variations in climate over these timescales are small relative to the variations predicted for the next 100 yr or more (Meehl et al., 2007), and so likely provide only a weak constraint on future predictions. As such, proxy indicators of climate from older time periods are increasingly being used to test models. On the timescale of

~ 100 000 yr, the Paleoclimate Modelling Intercomparison Project (PMIP, Braconnot et al., 2007), now in its third phase, is focusing on three main time periods: the mid-Holocene (6000 yr ago, 6 k), the Last Glacial Maximum (LGM, 21 k), and the Last Interglacial (LIG, 125 k). However, these time periods are either colder than modern (LGM), or their warmth is primarily caused not by enhanced greenhouse gases, but by orbital forcing (mid-Holocene, LIG). As such, their use for testing models used for future climate prediction is also limited. On the timescale of millions of years, several time periods show potential for model evaluation, being characterised by substantial warmth which is thought to be driven primarily by enhanced atmospheric CO<sub>2</sub> concentrations. An example is the mid-Pliocene (3 million years ago, 3 Ma), when global annual temperature was ~ 3 °C greater than pre-industrial (Dowsett et al., 2009). However, the latest estimates of mid-Pliocene CO<sub>2</sub> (Pagani et al., 2010; Seki et al., 2010) range from ~ 360 to ~ 420 ppmv, which is similar to that of modern (~ 390 ppmv in 2010 according to the Scripps CO<sub>2</sub> program, <http://scrippsco2.ucsd.edu/>), and substantially less than typical IPCC scenarios for CO<sub>2</sub> concentration at the end of this century (~ 1000 ppmv in the A1F1 scenario, > 1370 ppmv CO<sub>2</sub>-equivalent in the RCP8.5 scenario, Meehl et al., 2007; Moss et al., 2010). The time period which shows possibly the most similarity to projections of the end of the 21st century and beyond is the early Eocene, ~ 55 to ~ 50 Ma. A recent compilation of Cenozoic atmospheric CO<sub>2</sub> is relatively data-sparse during the early Eocene, with large uncertainty range, meaning that values more than 2000 ppmv cannot be ruled out (Beerling and Royer, 2011). Relatively high values for the early Eocene are consistent with recent CO<sub>2</sub> reconstructions for the latest Eocene of the order 1000 ppmv (Pearson et al., 2009; Pagani et al., 2011). Proxy indicators have been interpreted as showing tropical temperatures at this time ~ 5 °C warmer than modern (e.g. Pearson et al., 2001), and high latitude terrestrial temperatures more than 20 °C warmer (e.g. Huber and Caballero, 2011). Recently, due at least in part to interest associated with this time period as a possible future analogue, there have been a number of new sea surface temperature (SST) and terrestrial temperature data published, using a range of proxy reconstruction methods. There have also been several models recently configured for the early Eocene and attempts made to understand the mechanisms of Eocene warmth. Most of these studies have carried out some form of model–data comparison; however, the models have not been formally intercompared in a consistent framework, and new data now allows a more robust and extensive evaluation of the models.

The aims of this paper are:

- To present an intercomparison of five models, all recently used to simulate the early Eocene climate.
- To carry out a consistent and comprehensive comparison of the model results with the latest proxy

temperature indicators, taking full account of uncertainties in the reconstructions.

- By analysing the energy balance and fluxes in the models, to gain an understanding of the reasons behind the differences in the model results.

Section 2 describes the model simulations, Sect. 3 presents the datasets used to evaluate the models, and Sect. 4 presents the model results and model–data comparison. Section 5 quantifies the reasons for the differences between the model results, and Sect. 6 discusses, concludes, and proposes directions for future research.

## 2 Model simulation descriptions

Many model simulations have been carried out over the last two decades with the aim of representing the early Eocene. Here, we present and discuss results from a selection of these. We present all simulations of which we are aware that (a) are published in the peer-reviewed literature, and (b) are carried out with fully dynamic atmosphere–ocean general circulation models (GCMs), with primitive equation atmospheres. This makes a total of 4 models – HadCM3L (Lunt et al., 2010), ECHAM5/MPI-OM (Heinemann et al., 2009), CCSM3 (Winguth et al., 2010, 2012; Liu et al., 2009; Huber and Caballero, 2011), and GISS ModelE-R (Roberts et al., 2009). Criterion (b) is chosen to select the models which are most similar to those used in future climate change projection (i.e. we exclude models with energy balance atmospheres such as GENIE; Panchuk et al., 2008). There are two sets of CCSM3 simulations, which we name CCSM\_W (Winguth et al., 2010, 2012) and CCSM\_H (Liu et al., 2009; Huber and Caballero, 2011). All the models and simulations are summarised in Table 1. Together they make up the “Eocene Modelling Intercomparison Project”, EoMIP. EoMIP differs from more formal model intercomparisons, such as those carried out under the auspices of PMIP, in that the groups have carried out their own experimental design and simulations in isolation, and the comparison is being carried out post hoc, rather than being planned from the outset. As such, the groups have used different paleogeographical boundary conditions and CO<sub>2</sub> levels to simulate their Eocene climates. This has advantages and disadvantages compared to the more formal approach with a single experimental design: The main disadvantage is that a direct comparison between models is impossible due to even subtle differences in imposed boundary conditions; the main advantage is that in addition to uncertainties in the models themselves, the model ensemble also represents the uncertainties in the paleoenvironmental conditions, and therefore more fully represents the uncertainty in our climatic predictions for that time period.

**Table 1.** Summary of model simulations in EoMIP. Some models have irregular grids in the atmosphere and/or ocean, or have spectral atmospheres. The atmospheric and ocean resolutions are given in number of grid boxes,  $X \times Y \times Z$  where  $X$  is the effective number of grid boxes in the zonal,  $Y$  in the meridional, and  $Z$  in the vertical. See the original references for more details.

Name	Eocene simulation reference	Model name and reference	Atmosphere resolution	Ocean resolution
HadCM	Lunt et al. (2010)	HadCM3L, Cox et al. (2001)	$96 \times 73 \times 19$	$96 \times 73 \times 20$
ECHAM	Heinemann et al. (2009)	ECHAM5/MPI-OM, Roeckner et al. (2003)	$96 \times 48 \times 19$	$142 \times 82 \times 40$
CCSM_W	Winguth et al. (2010, 2012)	CCSM3, Collins et al. (2006); Yeager et al. (2006)	$96 \times 48 \times 26$	$100 \times 116 \times 25$
CCSM_H	Liu et al. (2009) Huber and Caballero (2011)	CCSM3, Collins et al. (2006); Yeager et al. (2006)	$96 \times 48 \times 26$	$100 \times 122 \times 25$
GISS	Roberts et al. (2009)	GISS ModelE-R, Schmidt et al. (2006)	$72 \times 45 \times 20$	$72 \times 45 \times 13$

Name	Paleogeography	Sim. length (years)	CO <sub>2</sub> levels	Vegetation	Aerosols
HadCM	proprietary	> 3400	$\times 2, 4, 6$	homogenous shrubland	as control
ECHAM	Bice and Marotzke (2001)	2500	$\times 2$	homogenous woody savanna	as control
CCSM_W	Sewall et al. (2000) with marginal sea parameterisation	1500	$\times 4, 8, 16$	Shellito and Sloan (2006)	as control
CCSM_H	Sewall et al. (2000)	> 3500	$\times 2, 4, 8, 16$	Sewall et al. (2000)	reduced aerosol
GISS	Bice and Marotzke (2001)	2000	$\times 2$	Sewall et al. (2000)	as control

## 2.1 HadCM

Lunt et al. (2010) investigated the potential role of hydrate destabilisation as a mechanism for the Paleocene–Eocene Thermal Maximum (PETM,  $\sim 55$  Ma), using the HadCM3L model. They found a switch in modelled ocean circulation which occurred between  $\times 2$  and  $\times 4$  pre-industrial concentrations of atmospheric CO<sub>2</sub>, which resulted in a non-linear warming of intermediate ocean depths. They hypothesised that this could be a triggering mechanism for hydrate release. For the three Eocene simulations carried out ( $\times 2$ ,  $\times 4$ , and  $\times 6$ ), vegetation was set globally to a “shrub” plant functional type. The paleogeography is proprietary but is illustrated in Supplementary Information of Lunt et al. (2010). An additional simulation at  $\times 6$  CO<sub>2</sub> was carried out with the same model by Tindall et al. (2010), which incorporated oxygen isotopes into the hydrological cycle. The  $\delta^{18}\text{O}$  of seawater from the Tindall et al. (2010) simulation is used in our SST compilation to inform the uncertainty range of the proxies based on  $\delta^{18}\text{O}$  measurements (see Sect. 3).

## 2.2 ECHAM

Heinemann et al. (2009) presented an ECHAM5/MPI-OM Eocene simulation and compared it to a pre-industrial simulation, diagnosing the reasons for the Eocene warmth by making use of a simple 1-D energy balance model (which

we use in this paper in Sect. 5). They reported a larger polar warming than many previous studies, which they attributed to local radiative forcing changes rather than modified poleward heat transport. The Eocene simulation was carried out under  $\times 2$  CO<sub>2</sub> levels, and a globally homogeneous vegetation was prescribed, with characteristics similar to present-day woody savanna.

## 2.3 CCSM\_W and CCSM\_H

Huber and Caballero (2011) presented a set of Eocene CCSM3 simulations, originally published by Liu et al. (2009), with the main aim of comparing these with a new compilation of proxy terrestrial temperature data. They found that at high CO<sub>2</sub> ( $\times 16$ ) they obtained good agreement with data from mid and high latitudes. We use this same proxy dataset in this paper, including estimates of uncertainty, for evaluating all the EoMIP simulations.

Winguth et al. (2010) and Winguth et al. (2012) carried out an independent set of CCSM3 simulations motivated by investigating the role of hydrates as a possible cause of the PETM. They found evidence of non-linear ocean warming and enhanced stratification in response to increasing atmospheric CO<sub>2</sub> concentrations, and a shift of deep water formation from northern and southern sources to a predominately southern source.

The CCSM\_W and CCSM\_H simulations differ mainly in the treatment of aerosols. In the CCSM\_W simulation, a high aerosol load is applied, whereas the CCSM\_H simulation considers a lower-than-present-day aerosol distribution following the approach by Kump and Pollard (2008), possibly justified by a reduced ocean productivity and thus reduced DMS (Dimethyl sulfide) emissions. A globally reduced productivity is supported by the recent study of Winguth et al. (2012). However, it remains uncertain to what extent intensified volcanism near the PETM might have increased the aerosol load (Storey et al., 2007).

## 2.4 GISS

Roberts et al. (2009) carried out an investigation into the role of the geometry of Arctic gateways in determining Eocene climate with the GISS ModelE-R, configured with  $\times 4$  CO<sub>2</sub> and  $\times 7$  CH<sub>4</sub> compared with pre-industrial. They estimated the change to the total forcing to be about  $\times 4.3$  of CO<sub>2</sub>-equivalent, but for the purposes of this paper we assume their simulations were at  $\times 4$  CO<sub>2</sub>. They found that restricting Arctic gateways led to warming of the North Atlantic and freshening of the Arctic Ocean, similar to data associated with the “Azolla” event (Brinkhuis et al., 2006). They incorporated oxygen isotopes into the hydrological cycle in their model (Roberts et al., 2011), and used the predicted isotopic concentrations of seawater to more directly compare with proxy temperature estimates. The  $\delta^{18}\text{O}$  of seawater from the Roberts et al. (2011) simulation is used in our SST compilation to inform the uncertainty range of the proxies based on  $\delta^{18}\text{O}$  measurements (see Sect. 3).

## 3 Early Eocene SST and land temperature datasets

To evaluate the various climate model simulations, we make use of both terrestrial and marine temperature datasets. The marine dataset has been compiled for this paper, and the terrestrial data is identical to that presented in Huber and Caballero (2011). In both cases we take as full account as possible of the various uncertainties associated with each proxy.

The purpose of these compilations is not to provide a tightly constrained “time-slice” reconstruction of any point in the early Eocene against which the ensemble or individual model runs can be compared; instead, we include data spanning the entire early Eocene. This approach is consistent with the EoMIP simulations themselves, in which models have not been run with the same specific set of simulation boundary conditions, such as paleogeography or atmospheric greenhouse gas forcings, but can be considered to reflect a possible range of time periods within the early Eocene.

### 3.1 Marine dataset

For the purposes of model–data comparison, we have compiled (see Supplement) available paleotemperature

estimates for sea surface (archaeal-derived isoprenoid glycerol dibiphytanyl glycerol tetraether (GDGT) paleothermometry), near-sea surface (mixed layer dwelling planktonic foraminifera) and shallow inner-shelf bottom waters (bivalve oxygen isotopes) from across the early Eocene (Ypresian stage;  $\sim 55.9$  to 49 Ma). Also included in the compilation are some data from the very latest Paleocene, within the interval immediately before but not including the Paleocene–Eocene Thermal Maximum (PETM). These data are included to increase the geographical coverage of data, especially in the mid to low latitudes.

Long-term paleotemperature records through the early Eocene indicate the presence of a significant warming trend in both oceanic intermediate waters of  $\sim 4^\circ\text{C}$  (Zachos et al., 2008) and high-latitude sea surface temperatures of up to  $\sim 10^\circ\text{C}$  (Bijl et al., 2009; Hollis et al., 2012) in the lead-up to the early Eocene Climatic Optimum (EECO) (Zachos et al., 2008). Although tropical sea surface temperatures may have been more stable (Pearson et al., 2007) through this interval, attention should be paid to the age of the compiled proxy records relative to this warming trend. Notably, comparison of EECO records from the southern high latitudes with basal early Eocene records from the mid and low latitudes, or non-EECO proxy records from the high latitudes, could introduce a spurious reduction in the latitudinal temperature gradient or false high latitude proxy–proxy disagreements, respectively.

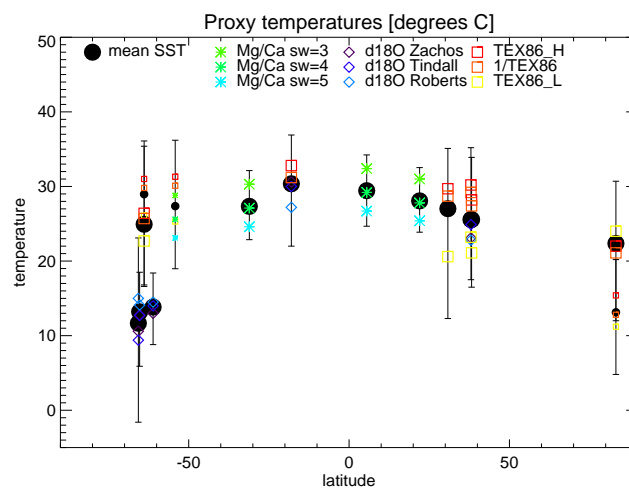
The relative paucity of the available data however, which is taken from a small number of locations, many of which have limited time series and/or poor age control, prohibits a narrowly focused time-slice reconstruction of SSTs within the early Eocene. Instead, we have chosen to divide the data into two broad categories, those from the period of peak Cenozoic warmth during the EECO, and the remainder, assigned to a generally cooler “background” early Eocene climate state. Pre-PETM records are included in this “background” early Eocene compilation. Given there is some evidence for warming between pre- and post-PETM conditions in the high latitudes (Bijl et al., 2009; Hollis et al., 2012), as well as the general early Eocene warming trend, these pre-PETM records are likely to have a slight cool bias relative to other estimates of early Eocene temperatures.

The identification of EECO records is only possible where there is either good age control and/or a clear temperature-trend across a long-term early Eocene record. Only three marine SST proxy datasets are identified as representing EECO conditions for all or part of the associated SST time series – ODP Site 1172D, Waipara River and the Arctic IODP Site M0004. The EECO in the ODP Site 1172D record is identified following Hollis et al. (2012) as spanning  $\sim 53.1$  to  $\sim 49$  Ma (588.85 to 562.70 m b.s.f.), with the pre-EECO interval from 54.9 to 53.3 Ma (611.0 to 591.15 m b.s.f.). All of the Waipara River data used here are identified as representing EECO conditions (Hollis et al., 2009, 2012). Although the early Eocene age model for the Arctic IODP M0004 is poorly constrained between early Eocene hyperthermal

events and the termination of the “Azolla” phase, there is a distinct *cooling* in GDGT-derived proxy temperature estimates between the stratigraphically lower (core 27X) and upper (cores 23X to 19X) parts of this section (Sluijs et al., 2008). For the purposes of a more refined proxy–proxy and proxy–model comparison, we have labelled the (warmer) SST data from the lower part of the M0004 as pre-EEOCO and the (cooler) SST data from the upper part as EEOCO. This is in line with speculation in Sluijs et al. (2008) that while there is a global trend of warming through the early Eocene, Arctic SSTs reduced during this interval. Data from M0004 Core 27X between 369 to 367.9 mcd has also been excluded, which, based on carbon isotope stratigraphy, likely represents the early Eocene hypothermal event ETM2 (Sluijs et al., 2008).

The remaining sites have either relatively poorly constrained age models *and* have no discernable SST trend through the dataset used (ODP Sites 690 and 738), are spot samples within the early Eocene (Tanzania, Hatchetigbee Bluff) or are well-constrained pre-PETM records. Seymour Island is the exception to this, where there remains uncertainty about even the gross age of this succession. The data used in this compilation is sourced from Telms 3 to 5, which, based on strontium isotope stratigraphy and sparse biostratigraphic data, were thought to span the early Eocene, extending just across the early/middle Eocene boundary (Ivany et al., 2008). A revised age assessment, based on dinoflagellate biostratigraphy, suggests that the lower part of this sequence is middle and not early Eocene in age (Douglas et al., 2011). This new biostratigraphic data remains somewhat tentative, and while awaiting the publication of a fully revised age model for these successions, the Seymour Island data is provisionally included in the SST compilation. It is however, assigned to the background “pre-EEOCO” category, as it appears to be more likely to be representative of middle Eocene post-EEOCO cooling. It is also noted that although Tanzanian Drilling Project Site 2 (TDP2) was originally reported as extending down into the early Eocene (Nicholas et al., 2006; Pearson et al., 2004), with the resolution of planktic foraminifera–nannofossil biostratigraphic mismatches around the early/middle Eocene boundary (Payros et al., 2007), TDP2 is now considered to be entirely within the basal middle Eocene (P. Pearson, personal communication, July 2012) and is excluded from this compilation.

For each location the primary geochemical proxy data were first collated and then used to generate a range of SST estimates based on a set of plausible assumptions about the underlying paleotemperature methodology. All of the paleotemperature estimation methods are subject to uncertainty arising from their present-day calibrations, necessary assumptions about ancient seawater chemistry and potential non-analogue behaviour between modern and ancient systems. Although positive steps are being made with deep-time proxy inter-comparison studies (Hollis et al., 2012), potential



**Fig. 1.** SST proxy dataset, compiled for this study. Coloured symbols show the various median estimates from the literature, with various assumptions about Mg/Ca of seawater,  $\delta^{18}\text{O}_{\text{sw}}$ , and TEX<sub>86</sub> calibration. Error bars indicate the maximum and minimum range at each site including temporal variability and calibration error. The filled black circles represent the mean SST at each site, averaged over the various assumptions. Larger symbols represent “background” early Eocene state, smaller symbols represent the EEOCO. See Sect. 3.1 for more details.

non-analogue behaviour is very difficult to assess and we do not try to quantify this directly in our uncertainty analysis. We do, however, attempt to quantify uncertainty associated with both paleotemperature calibrations and the estimates of ancient seawater chemistry. This is achieved by (1) applying the standard error determined from the modern calibration dataset to paleotemperature estimates; (2) the use of multiple alternate calibrations where there is ongoing debate about the most appropriate calibration or proxy method (GDGT paleothermometry) or where modern calibrations vary with environmental conditions (oxygen isotopes); and (3) applying multiple estimates (Mg/Ca) or different estimation methodologies (oxygen isotopes) for seawater chemistry. Where distinct proxies (GDGT paleothermometry) or distinct parameters for seawater chemistry are used (Mg/Ca and oxygen isotopes), the derived temperature ranges are calculated separately at each site. This leads to the following sets of proxy data: TEX<sub>86</sub><sup>H</sup>, TEX<sub>86</sub><sup>L</sup>, 1/TEX<sub>86</sub>, oxygen isotope paleothermometry with modelled and latitudinal corrected  $\delta^{18}\text{O}_{\text{sw}}$  estimates, and Mg/Ca paleothermometry with assumed Mg/Ca<sub>sw</sub> values of 3, 4 and 5 mol mol<sup>−1</sup>. Details of all these methods and calculations are described below, and illustrated in Fig. 1.

### 3.1.1 Oxygen isotopes

For planktic foraminifera-derived  $\delta^{18}\text{O}$  temperature estimates, we applied the temperature– $\delta^{18}\text{O}$  calibrations of Bemis et al. (1998) for the symbiotic planktic foraminifera,

*Orbulina universa*, under both high and low light conditions (their Eqs. 1 and 2). Together, these two equations bracket most of the natural variability in planktic foraminifera temperature– $\delta^{18}\text{O}$  space within modern plankton tow data (Bemis et al., 1998). Unlike the multiple GDGT paleotemperature equations, which seem to vary in their accuracy with geographical location or paleoenvironment, the two equations of Bemis et al. (1998) represent the natural variability at a single location (high/low light conditions). The derived SST estimates from these two equations are thus combined into a single range representing the potential environmental variability at any location. The standard errors on Eq. (1) (low light) and 2 (high light) are  $\pm 0.7$  and  $0.5^\circ\text{C}$ , respectively.

Three sets of temperature estimate are, however, plotted in Fig. 1 for each location with planktic foraminifera  $\delta^{18}\text{O}$  data based on three estimates of  $\delta^{18}\text{O}_{\text{sw}}$ : the latitudinal correction of Zachos et al. (2008) and the modelled mixed-layer  $\delta^{18}\text{O}_{\text{sw}}$  of Tindall et al. (2010) and of Roberts et al. (2011). The latitudinal correction is a first-order approximation of the effects of the global hydrological cycle on seawater  $\delta^{18}\text{O}_{\text{sw}}$  and is a widely used improvement on an “ice-free” globally uniform estimate of  $\delta^{18}\text{O}_{\text{sw}}$ . This empirical relationship does not, however, include zonal deviations from this general pattern. In the early Paleogene world, these zonal deviations are likely to have been significantly different to the modern due to the closure of the major Southern Ocean gateways and the resulting high-latitude isolation of the Atlantic and Pacific basins. The isotope-enabled versions of HadCM3L and GISS, however, reproduce both the expected latitudinal gradients in  $\delta^{18}\text{O}_{\text{sw}}$  and an estimation of early Paleogene basin-to-basin zonal gradients (Tindall et al., 2010; Roberts et al., 2011). These modelled  $\delta^{18}\text{O}_{\text{sw}}$  values, taken from modelled ocean depths of 50 m, are a potential refinement to early Paleogene  $\delta^{18}\text{O}_{\text{sw}}$  estimation and are included here to provide a more comprehensive assessment of the range of possible  $\delta^{18}\text{O}$  SST estimates.

For the latitudinal correction, we assume a global average  $\delta^{18}\text{O}$  composition of early Eocene seawater of  $-1\text{‰}$  to be consistent with the value used by Tindall et al. (2010). This is comparable to the  $-0.96\text{‰}$  used in a recent early Eocene paleotemperature inter-comparison (Hollis et al., 2012), and the value of  $-0.81\text{‰}$  used in Roberts et al. (2011). A SMOW to VPDB conversion of  $-0.27\text{‰}$  was applied to all estimates of  $\delta^{18}\text{O}_{\text{sw}}$ . The maximum difference in paleotemperature estimates between the three  $\delta^{18}\text{O}_{\text{sw}}$  assumptions is for Seymour Island, where the median value using the modelled  $\delta^{18}\text{O}_{\text{sw}}$  of Roberts et al. (2011) is  $5.6^\circ\text{C}$  warmer than that using the  $\delta^{18}\text{O}_{\text{sw}}$  of Tindall et al. (2010).

For the *Eurhormalea* and *Cucullaea* bivalve  $\delta^{18}\text{O}$  data, we used the biogenic aragonite  $\delta^{18}\text{O}$ –temperature calibration of Grossman and Ku (1986) as modified by Kobashi et al. (2003) with both the latitude-corrected and modelled  $\delta^{18}\text{O}_{\text{sw}}$  noted above. We calculated the standard error on the Grossman and Ku (1986) calibration, based on the original dataset

for biogenic aragonite, to be  $\pm 1.4^\circ\text{C}$ . A SMOW to VPDB correction of  $-0.2\text{‰}$  is already implicit within the Kobashi et al. (2003) form of this temperature equation.

### 3.1.2 Mg/Ca ratios of planktonic foraminifera

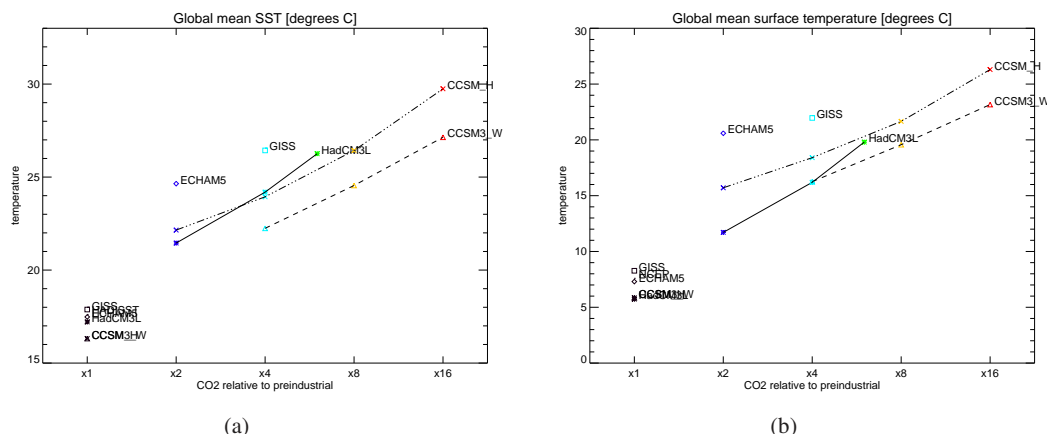
To estimate calcification temperature, we used the multi-species sediment trap calibration of Anand et al. (2003), which has a calibration standard deviation of  $\pm 1.13^\circ\text{C}$ . This paleotemperature estimation relies strongly upon the assumed value of the Mg/Ca ratio in early Eocene seawater, which is still poorly constrained. Values in the range of  $3\text{--}4\text{ mol mol}^{-1}$  are typically used within paleoceanographic studies, and these produce tropical (Sexton et al., 2006) and mid-latitude (Creech et al., 2010) surface ocean temperatures that are broadly consistent with independent paleotemperature estimates. Here, we separately calculate and plot (in Fig. 1) paleotemperature estimates based on three values of seawater Mg/Ca across a wide range, namely 3, 4 and  $5\text{ mol mol}^{-1}$ . Distinguishing temperature estimates based on these three values allows for (1) a clear representation of the sensitivity of temperature to the assumed value of Mg/Ca<sub>sw</sub>, and (2) the future use of the most appropriate temperature range if/when more robust constraints on early Paleogene Mg/Ca<sub>sw</sub> become available.

For reference, an estimate of  $\sim 3.5\text{ mol mol}^{-1}$  is obtained using the Lear et al. (2002) calibration for *Oridorsalis umbonatus*, the paired foraminifera Mg/Ca value of  $2.78\text{ mol mol}^{-1}$  and  $\delta^{18}\text{O}$ -derived bottom water temperature of  $12.4^\circ\text{C}$  they quote for  $\sim 49\text{ Ma}$ . A lower,  $\sim 3\text{ mol mol}^{-1}$  value is obtained by the same method but using the revised calibrations for *O. umbonatus* (Rathmann et al., 2004; Rathmann and Kuhnert, 2008). Higher values of  $\sim 4$  to  $5\text{ mol mol}^{-1}$  are indicated by  $\delta^{18}\text{O}$ –Mg/Ca paleotemperature inter-comparisons with well-preserved early Eocene foraminifera (Sexton et al., 2006), whilst recent modelling of trace metal fluxes and assessments of the long-term benthic foraminifera  $\delta^{18}\text{O}$ –Mg/Ca record suggest values of  $\sim 3\text{ mol mol}^{-1}$  or less (Cramer et al., 2011; Farkaš et al., 2007). These lower values are more consistent with estimates based on ridge flank hydrothermal carbonate veins, at around  $2\text{ mol mol}^{-1}$  (Coggon et al., 2010). There remains a pressing need to understand the causes of these discrepancies and establish robust estimates of the Mg/Ca ratio of ancient seawater (see discussion in Coggon et al., 2011).

### 3.1.3 TEX<sub>86</sub>

Determining the appropriate proxy index and calibration for deep-time paleotemperature estimates based on the relative abundances of archaeal-derived isoprenoid glycerol dibiphytanyl glycerol tetraethers (GDGTs) is problematic. Three methods have recently been proposed: separate “low” and “high” temperature proxies based on different ratios of GDGTs, TEX<sub>86</sub><sup>L</sup> and TEX<sub>86</sub><sup>H</sup> (Kim et al., 2010),





**Fig. 2.** Global annual mean (a) SST ( $\langle \text{SST} \rangle$ ) and (b) continental 2 m air temperature ( $\langle \text{LAT} \rangle$ ), as a function of  $\text{CO}_2$  for all simulations, and for observational datasets. The simulations at  $\times 1 \text{ CO}_2$  are pre-industrial reference simulations.

and a non-linear calibration of the original  $\text{TEX}_{86}$  index, “ $1/\text{TEX}_{86}$ ” (Liu et al., 2009), revised by Kim et al. (2010).  $\text{TEX}_{86}^{\text{H}}$  and  $1/\text{TEX}_{86}$  are based on the same underlying ratio of GDGTs – the original  $\text{TEX}_{86}$  proxy – but differ in the form of their calibration equations (logarithmic versus reciprocal). The fundamentally different ratio of GDGT isomers within  $\text{TEX}_{86}^{\text{L}}$  results in a proxy that can, in certain instances, produce temperature trends contrary to  $\text{TEX}_{86}^{\text{H}}$  and  $1/\text{TEX}_{86}$ . It is, as yet, unclear which of these proxies is the most appropriate for early Eocene paleotemperature estimation. There are indications that their suitability may vary with both the temperature range and paleoenvironment of GDGT formation (Hollis et al., 2012).

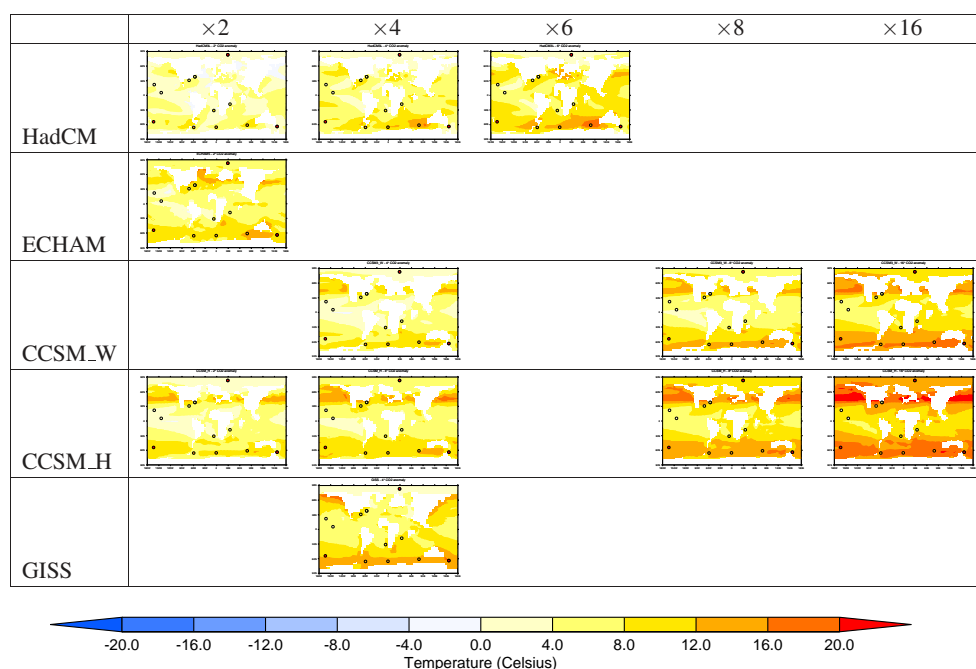
For the purposes of this study, we separately calculate and plot (in Fig. 1) paleotemperatures at each site using all three measures:  $\text{TEX}_{86}^{\text{H}}$ ,  $\text{TEX}_{86}^{\text{L}}$  and  $1/\text{TEX}_{86}$ . This illustrates the full range of temperature estimates produced by these GDGT-based proxies at a given location but also allows for a more refined use of this data as the behaviour of these proxies becomes better understood. The calculation of all three proxies at all sites may be considered by some to be an erroneous application of, for example, a “low temperature” proxy,  $\text{TEX}_{86}^{\text{L}}$ , to the mid and low latitudes of a warm climate state. We must stress that we do not intend to imply that all three are equally applicable at all sites. Rather, by showing all three proxies at all sites alongside other proxy temperature estimates, we hope to contribute to the ongoing discussion about the behaviour of GDGT proxies in deep-time paleoenvironments (Hollis et al., 2012).

Recent development of good practice suggests the exclusion of paleotemperature estimates from samples with a BIT index in excess of 0.3 (Kim et al., 2010). Although we accept this as a recommendation, in the existing published data compiled here it would result in the exclusion of all early Eocene data from Tanzania and Hatchetigbee Bluff, which both have BIT indices in the range 0.3 to 0.5. We choose

to include this published data but note these higher BIT indices. In some cases, as for the early Eocene data from Tanzania,  $\text{TEX}_{86}^{\text{L}}$  temperature estimates are clearly erroneous and are excluded. Due to the greater availability of data, samples from the Arctic Ocean IODP Site M0004 with BIT indices  $> 0.3$  were excluded. The standard  $\text{TEX}_{86}$  proxies discussed above can be applied to this early Eocene Arctic data rather than the  $\text{TEX}_{86}$  proxy used through the PETM by Sluijs et al. (2006). The errors ( $^{\circ}\text{C}$ ) on the GDGT-based proxies are  $\pm 2.5$  for  $\text{TEX}_{86}^{\text{H}}$  (GDGT index-2),  $\pm 4.0$  for  $\text{TEX}_{86}^{\text{L}}$  (GDGT index-1) and  $\pm 5.4$  for  $1/\text{TEX}_{86}$  (Kim et al., 2010).

From the arrays of time-varying temperature estimates at each site, and for all assumptions of seawater composition and  $\text{TEX}_{86}$  calibration, we calculated the median, maximum and minimum values from the time series as the basis for the model–data comparisons. There is an important caveat to this approach that relates to the effect of data quantity and stratigraphic range on the temperature envelopes plotted. Where there are reasonably extensive time series, natural temporal variability can result in a larger envelope of temperature estimates than at sites where data is limited to a few spot samples. As a result, these envelopes should not be taken to solely represent uncertainty in paleotemperature estimation, but to also include a measure of the temporal variability at individual sites.

Finally, in order to provide a single “zero-order estimate” for SST at each location, we averaged the median estimates at each site across the various assumptions of seawater chemistry and  $\text{TEX}_{86}$  calibration. The resulting estimates are shown as filled black circles in Fig. 1, separately for EECO and non-EECO estimates. We are not suggesting that these are “best” estimates – interpretation of proxies is undergoing rapid development at present – instead their main use is to provide single-value estimates in the maps in Figs. 3 and 9a, and to provide a zero-order target for the model error scores in Table 2. By providing the raw data in the Supplement,



**Fig. 3.** SST anomaly in the model simulations ( $SST_e^m - \overline{SST_p^m}$ ), as a function of model and fractional CO<sub>2</sub> increase from pre-industrial. Also shown for the proxies are  $SST_e^d - \overline{SST_p^d}$ .

others will be able to provide their own interpretation if necessary as data is re-interpreted.

### 3.2 Terrestrial dataset

For the terrestrial, we make use of the data compilation presented in Huber and Caballero (2011). This is based largely on macrofloral assemblages, with mean annual temperatures being reconstructed primarily by leaf-margin analysis and/or CLAMP (physiognomic analysis of leaf fossils). Other proxies are also incorporated, such as isotopic estimates, organic geochemical indicators, and palynoflora. The error bars associated with each data point incorporate uncertainty in calibration, topography, and dating. More information on the data themselves, and the estimates of uncertainty, can be found in Huber and Caballero (2011).

Both marine and terrestrial datasets are provided in the Supplement, and are plotted geographically in Figs. 3 and 4, and latitudinally in Figs. 5 and 7.

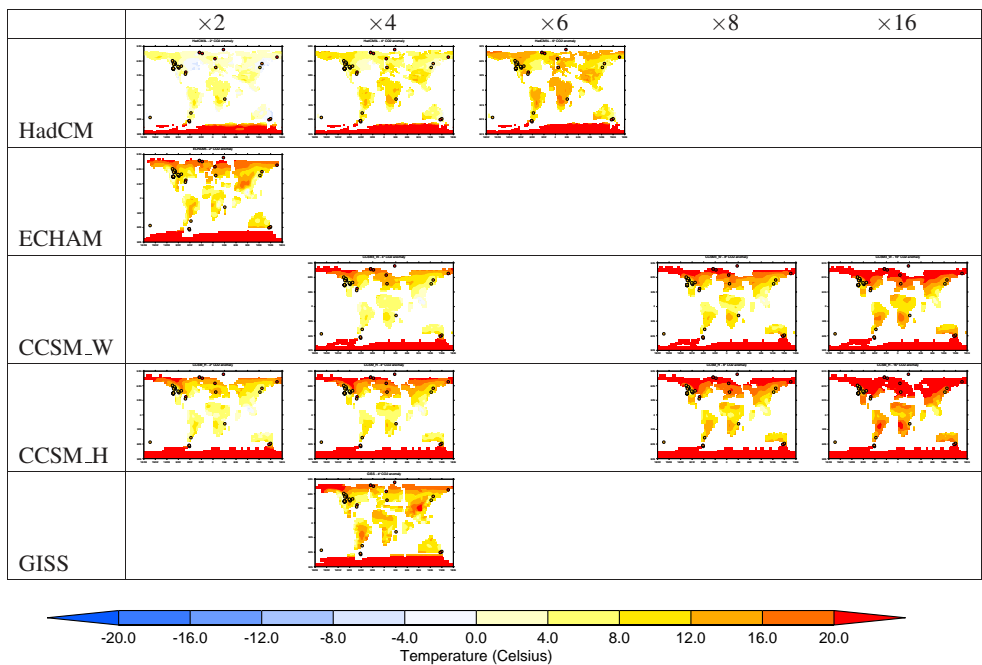
The SST plots have error bars which include the contributions from the two sources of uncertainty we have considered, related to calibration and temporal trends. This approach to the data aims to include a wide range of potential uncertainties in order to highlight both the regions of potential model–data agreement, but more importantly where there appear to be genuine discrepancies that cannot realistically be explained by the uncertainties in the proxy temperature estimations.

## 4 Results and model–data comparison

In this section, we present results from the EoMIP model ensemble (early Eocene simulations and pre-industrial controls), as described in Sect. 2, and compare them with the data described in Sect. 3. The reasons for the different model results are explored in more detail in Sect. 5.

It is useful at this stage to define some nomenclature. To represent the distribution of temperature, we write SST for sea surface temperature (only defined over ocean), or LAT for land near-surface ( $\sim 1.5$  m) air temperature (only defined over continents), or GAT for near-surface air temperature (defined globally), or GST for surface temperature (defined globally), or just  $T$  for a generic temperature. Global means are denoted by angled brackets, so that, for example, the global mean sea surface temperature is  $\langle \text{SST} \rangle$ . Zonal means are denoted by overbars, so that the zonal mean sea surface temperature is  $\overline{\text{SST}}$ . In the case of model output, ensemble means are denoted by square brackets, such as  $[\text{LAT}]$ . Eocene quantities are given a subscript e, and present/pre-industrial (i.e. modern) quantities are given a subscript p. Model values are given a superscript m, and proxy or observed data are given a superscript d. Because the modern observed data has global coverage (albeit interpolated, or assimilated with models in some regions), but the Eocene proxy data is sparse, the modern observed global or zonal means  $\langle T_p^d \rangle$  and  $\overline{T_p^d}$  are defined, but the Eocene equivalents are not.





**Fig. 4.** Continental surface air temperature anomaly in the model simulations ( $\text{LAT}_e^m - \overline{\text{GAT}}_p^m$ ), as a function of model and fractional  $\text{CO}_2$  increase from pre-industrial. Also shown for the proxies are  $\text{LAT}_e^d - \overline{\text{GAT}}_p^d$ .

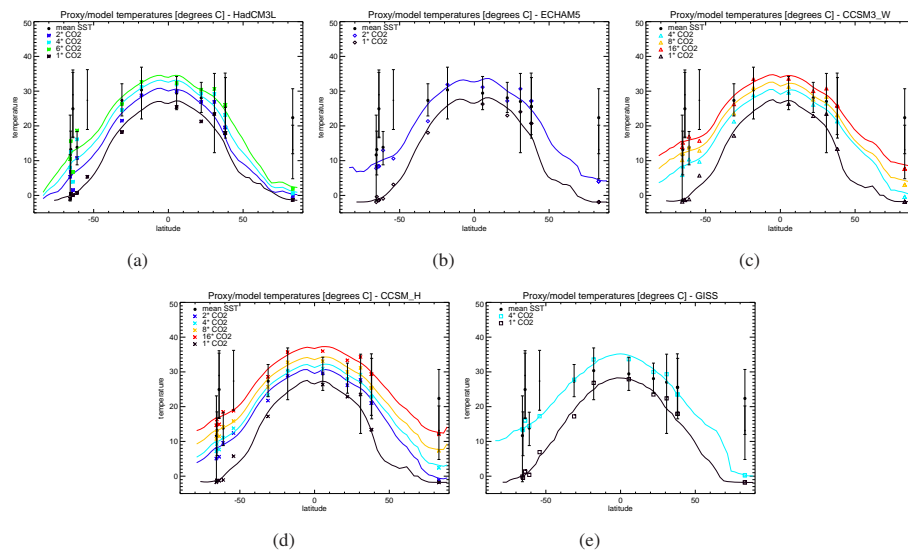
**Table 2.** Global mean temperatures and model mean error scores for each simulation. Scores are calculated based on the SST ( $\sigma_{\text{SST}}$ ) and land surface air temperature ( $\sigma_{\text{lat}}$ ) data. Definitions of the scores are given in Eqs. (1) and (2). Rows in bold indicate the best (i.e. lowest  $\sigma$ )  $\text{CO}_2$  level for each model.

Model	$\text{CO}_2$	$\langle \text{SST} \rangle$	$\langle \text{LAT} \rangle$	$\langle \text{GST} \rangle$	$\sigma_{\text{SST}} (^{\circ}\text{C})$	$\sigma_{\text{lat}} (^{\circ}\text{C})$
HadCM	2×	21.45	11.71	18.54	8.8	15.5
	4×	24.19	16.20	21.95	6.0	11.4
	<b>6×</b>	26.25	19.80	24.56	<b>3.9</b>	<b>7.7</b>
ECHAM	2×	24.65	20.59	24.03	<b>5.8</b>	<b>9.7</b>
CCSM.W	4×	22.24	16.26	20.95	6.7	10.3
	8×	24.45	19.57	23.59	4.0	7.2
	<b>16×</b>	27.14	23.16	26.46	<b>0.9</b>	<b>3.7</b>
CCSM.H	2×	22.15	15.71	21.12	8.6	11.5
	4×	23.94	18.41	23.17	6.6	8.5
	8×	26.43	21.66	25.79	3.6	5.1
	<b>16×</b>	29.75	26.30	29.47	<b>0.0</b>	<b>0.4</b>
GISS	<b>4×</b>	26.43	21.97	23.25	<b>3.8</b>	<b>6.9</b>

4.1 Inter-model comparison

Figure 2 shows the global annual mean sea surface temperature,  $\langle \text{SST} \rangle$ , and global annual mean near-surface land air temperature,  $\langle \text{LAT} \rangle$ , from all the GCM simulations in the EoMIP ensemble, and for modern observations; the Eocene values are also given in Table 2. The observed modern datasets are HadISST for SSTs (pre-industrial; 1850–1890) and NCEP (Kalnay et al., 1996) for near-surface air

temperatures (present; 1981–2010). For any given  $\text{CO}_2$  level, there is a wide range of modelled Eocene global mean values; for example, at 560 ppmv, there is a 8.9 °C range in  $\langle \text{LAT}_e^m \rangle$  and a 3.2 °C range in  $\langle \text{SST}_e^m \rangle$ . This range is larger than the range of simulated modern global means, which themselves agree well with the observed modern global means. The spread in Eocene results is due to (a) differences in the way the Eocene boundary conditions have been implemented in different models, and (b) different climate sensitivities in the different models. These differences are explored in Sect. 5. The clustering of the pre-industrial results is likely a result of tuning of the pre-industrial simulations to best match observations. For those models with more than one Eocene simulation, the Eocene climate sensitivity ( $\Delta \langle \text{GAT} \rangle$  per  $\text{CO}_2$  doubling) can also be seen to vary, both between models and also within one model as a function of  $\text{CO}_2$ . The variation of climate sensitivity between models is well documented in the context of future climate simulations (e.g. IPCC, 2007). The increase in climate sensitivity with  $\text{CO}_2$  (for example in the CCSM.H model) is due to the non-linear behaviour of climate system feedbacks, for example, associated with water vapour (see Sect. 5); however, there is also some non-linearity in the forcing itself as  $\text{CO}_2$  increases (Colman and McAveney, 2009). For HadCM, it is also related to a switch in ocean circulation which occurs between  $\times 2$  and  $\times 4$   $\text{CO}_2$  and is associated with a non-linear increase in surface ocean temperature (Lunt et al., 2010). The HadCM model also carried out an Eocene simulation with



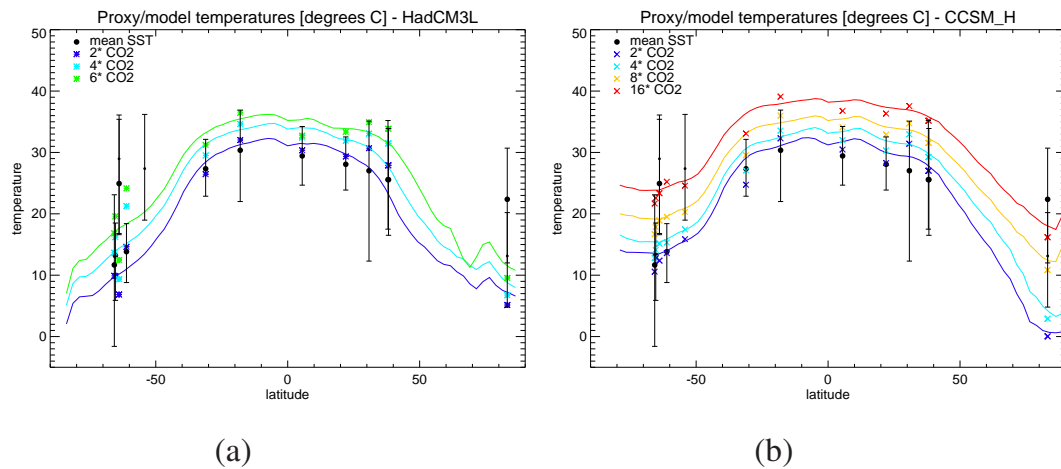
**Fig. 5.** Comparison of modelled SST with proxy-derived temperatures, SST vs. latitude. The simulations at  $\times 1$   $\text{CO}_2$  are pre-industrial reference simulations. For the model results, the continuous lines represent the zonal mean, and the open symbols represent the modelled temperature at the same location (longitude, latitude) as the proxy data. For the proxy data, the filled symbols represent the mean proxy temperature, and the error bars represent the range. The smaller filled symbols are EECO temperatures. See Sect. 3 and the Supplement for more details on the range calculations.

$\times 1$   $\text{CO}_2$  (not shown). Comparison of that simulation with its pre-industrial control shows that changing the non- $\text{CO}_2$  boundary conditions to those of the Eocene (i.e. topographic, bathymetric, vegetation, and solar constant changes) results in a  $1.8^\circ\text{C}$  increase in global mean surface air temperature, by comparison with a  $3.3^\circ\text{C}$  increase for a  $\text{CO}_2$  doubling from  $\times 1$  to  $\times 2$  under Eocene conditions. At a given  $\text{CO}_2$  level, the CCSM\_W and CCSM\_H models give quite different global means. This difference in mean Eocene climate state between the two similar models is mostly due to differences in the assumed Eocene atmospheric aerosol loading; CCSM\_W includes modern aerosols, whereas CCSM\_H includes no aerosol loading (see Sect. 2 and Table 1). Both these models share the same pre-industrial simulation. For all models, the  $\langle \text{LAT} \rangle$  and  $\langle \text{SST} \rangle$  means share similar characteristics, albeit with  $\langle \text{SST} \rangle$  varying over a smaller temperature range.

Figure 3 shows the simulated annual mean SST anomaly from each model and for the proxy reconstructions. A simple anomaly  $\text{SST}_e - \text{SST}_p$  would not be particularly informative because many regions would be undefined due to the difference in continental positions between the Eocene and present. Instead, we show  $\text{SST}_e - \overline{\text{SST}_p}$ , which is only undefined over Eocene continental regions and latitudes at which there is no ocean in the modern. The figures show that some features of temperature change are simulated consistently across models, such as the greatest ocean warming occurring in the mid latitudes. This mid-latitude maximum is due to reduced SST warming in the high latitudes due to the presence of seasonal sea ice anchoring the temperatures

close to  $0^\circ\text{C}$  combined with reduced warming in the tropics due to a lack of snow and sea ice albedo feedbacks. However, other patterns are not consistent. For example, GISS at  $\times 4$  and HADCM at  $\times 6$  have similar values of  $\langle \text{SST} \rangle$  relative to their controls ( $8.6$  and  $9.0^\circ\text{C}$ , respectively), but the warming in GISS is greatest in the northeast Pacific and the Southern Ocean, and the warming in HADCM is greatest in the North Atlantic and west of Australia. Similarly, ECHAM at  $\times 2$  and CCSM\_H at  $\times 4$  have similar global mean SST anomalies ( $7.2$  and  $7.6^\circ\text{C}$ , respectively), but the greatest Northern Hemisphere warming is in the Atlantic for ECHAM and in the Pacific for CCSM\_H. The two CCSM models exhibit similar patterns of warming, correcting for their offset in absolute Eocene temperature – i.e. the patterns of warming in CCSM\_H at  $\times 8$  are similar to those in CCSM\_W at  $\times 16$  (with anomalies of  $10.1$  and  $10.8^\circ\text{C}$ , respectively).

Figure 4 shows the simulated annual mean LAT anomaly from each model and for the proxy reconstructions. The anomaly is calculated relative to the pre-industrial (or modern in the case of the proxies) global (land plus ocean) zonal mean air temperature for each model, i.e.  $\text{LAT}_e - \overline{\text{GAT}_p}$ . The global (as opposed to land-only) zonal mean is used for calculating the anomaly in order to avoid undefined points (for example in the latitudes of the Southern Ocean where there is no land in the modern). Similar to SST, there are some consistent features between models – greatest warming is in the Antarctic (due to the lower topography via the lapse-rate effect and the change in albedo), and there is substantial boreal polar amplification. Again, there are also differences between models. For example, GISS at  $\times 4$  and ECHAM at



**Fig. 6.** As Fig. 5a and d, but the HadCM and CCSM\_H modelled zonal mean represents the warm month mean SST as opposed to annual mean.

$\times 2$  have similar values of  $\langle \text{LAT} \rangle$  relative to their controls (8.5 and 7.3 °C, respectively), but GISS has a substantially greater warming over Southeast Asia. These differences cannot be explained solely by differences in topography – the GISS and ECHAM models both use the Eocene topography of Bice and Marotzke (2001).

#### 4.2 Model–data comparison

Figure 5 shows a zonal SST model–data comparison for each model. The longitudinal locations of the SST data can be seen in Fig. 3. Each model is capable of simulating Eocene SSTs which are within the uncertainty estimates of the majority of the data points. The data points which lie furthest from the model simulations are the ACEX TEX<sub>86</sub> Arctic SST estimate (Sluijs et al., 2006), and the  $\delta^{18}\text{O}$  and TEX<sub>86</sub> estimates from the southwest Pacific (Bijl et al., 2009). The Arctic temperature reconstructions have uncertainty estimates which mean that at high CO<sub>2</sub>, the CCSM\_H ( $\times 8$ –16) and CCSM\_W ( $\times 16$ ) model simulations are just within agreement. At this CO<sub>2</sub> level, these models are also consistent with most of the tropical temperature estimates. From Fig. 2a, it is likely that other models could also obtain similarly high Arctic temperatures if they were run at sufficiently high CO<sub>2</sub> or low aerosol forcing. Also, given that some of these models (e.g. HadCM) have a higher climate sensitivity than CCSM\_H, this model–data consistency could be potentially obtained at a lower CO<sub>2</sub> than in CCSM.

TEX<sub>86</sub> is a relatively new proxy, which, as discussed in Sect. 3, is currently undergoing a process of rapid development. In this context, it has been suggested that the proxy could be recording the paleotemperature anomaly of the bloom season of the marine archaeota as opposed to a true annual mean. If this is the case, then it is likely that a more appropriate comparison is with the modelled summer temperature. This is illustrated in Fig. 6, for the HadCM and

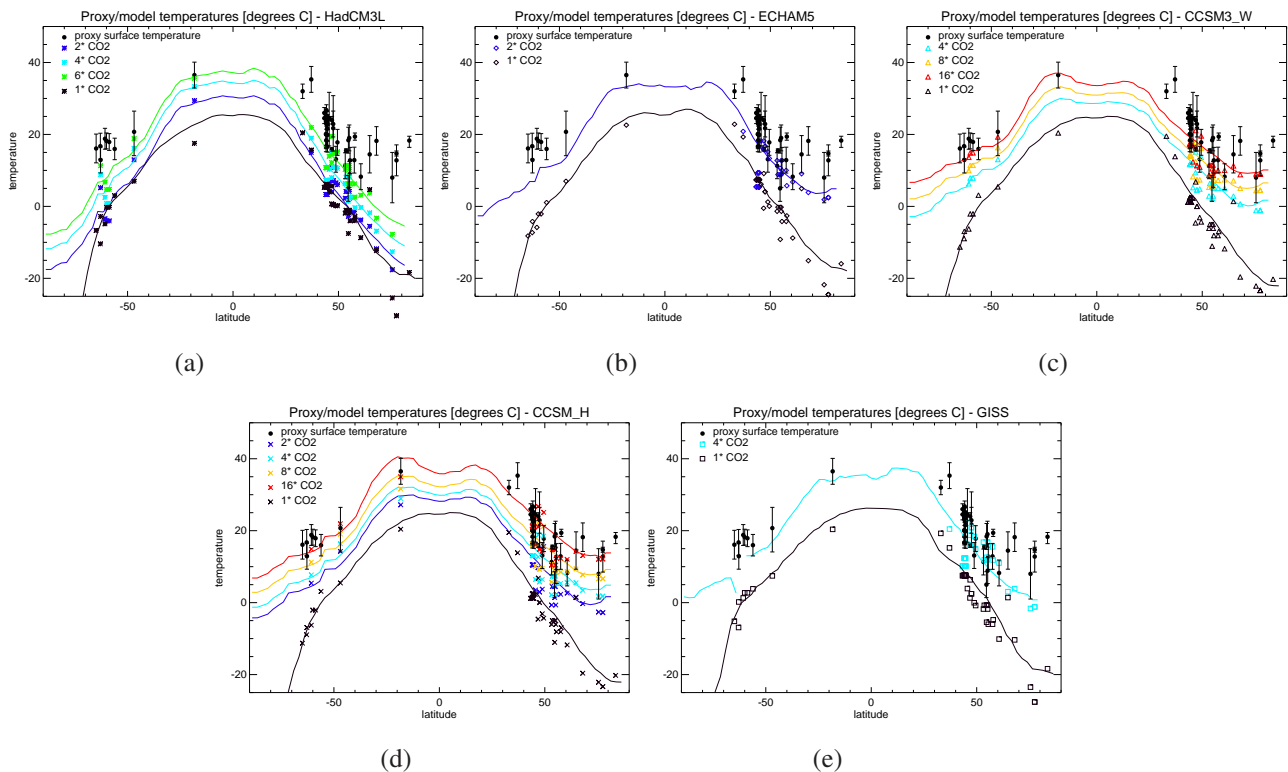
CCSM\_H models. In this case, the modelled warm month mean temperature is within the uncertainty range of the Arctic TEX<sub>86</sub> temperatures for both models.

Figure 7 shows the terrestrial temperature model–data comparison for each model. Those models which have been run at high CO<sub>2</sub> (both CCSM models), show good agreement with the data across all latitudes. The other models do not simulate such high temperatures, but, as with SST, it does appear that if they had been run at higher CO<sub>2</sub>, the model–data agreement would have been better. The HadCM model appears to be somewhat of an outlier in the Northern Hemisphere high latitudes, as it shows less polar amplification than the other models (see Sect. 4.3), an effect also seen in SST.

A quantitative indication of the model–data comparison for each simulation cannot currently be used to rank the models themselves, because the actual CO<sub>2</sub> forcing is not well constrained by data. However, it could give an indication of the range of CO<sub>2</sub> concentrations which are most consistent with the data. Given the sparseness of the SST and terrestrial data, any score should be treated with some caution. This is confounded by the uneven spread of the data; for example, there is a relatively high concentration of terrestrial data in North America. There are also issues associated with the different land–sea masks in the different models, which mean that the number of proxy data locations at which there are defined modelled values differs between the models. Therefore, we generate a simple mean error score for each simulation,  $\sigma$ , for both SST ( $\sigma_{\text{sst}}$ ) and land air temperature ( $\sigma_{\text{lat}}$ ) by averaging the error in temperature anomaly at the location of each of  $N$  data points:

$$\sigma_{\text{sst}} = \frac{1}{N} \sum (\text{SST}_e^m - \overline{\text{SST}_p^m} - \text{SST}_e^d + \overline{\text{SST}_p^d}), \quad (1)$$

$$\sigma_{\text{lat}} = \frac{1}{N} \sum (\text{LAT}_e^m - \overline{\text{LAT}_p^m} - \text{LAT}_e^d + \overline{\text{LAT}_p^d}), \quad (2)$$



**Fig. 7.** Comparison of modelled SAT with proxy-derived temperatures, SAT vs. latitude. The simulations at  $\times 1$  CO<sub>2</sub> are pre-industrial reference simulations. For the model results, the continuous lines represent the zonal mean, and the symbols represent the modelled temperature at the same location (longitude, latitude) as the proxy data. For the proxy data, the symbols represent the proxy temperature, and the error bars represent the range, as given by Huber and Caballero (2011).

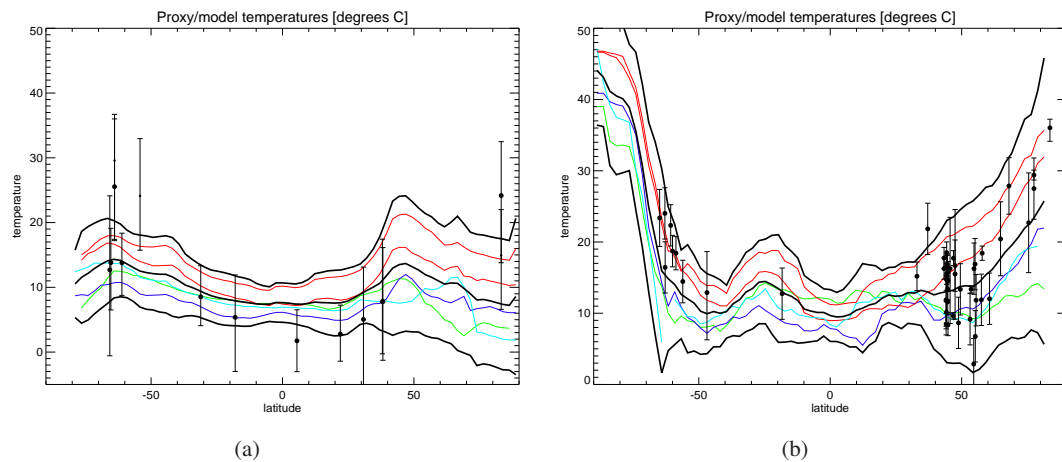
but proceed with caution, being mindful that there is a considerable uncertainty in the score itself. Values of  $\sigma$  for each model simulation are given in Table 2. For each model, the best results are obtained for the highest CO<sub>2</sub> level which was simulated (a result which also applies if an RMS score is used in place of a mean error score). The CCSM.H model at  $16\times$  CO<sub>2</sub> has the best (i.e. lowest absolute) values of  $\sigma$ . However, as noted before, it appears that other models would also obtain good  $\sigma$  scores if they had been run at sufficiently high CO<sub>2</sub>. A “best-case” multi-model ensemble can be created by averaging the simulations from each model which have the lowest values of  $\sigma$  (it turns out that those models with the best  $\sigma_{\text{lat}}$  also have the best  $\sigma_{\text{sst}}$ ). These are the models highlighted in bold in Table 2. The model–data comparison for this multi-model ensemble is shown in Figs. 8 and 9. The 2 standard-deviation width of the “best-case” ensemble overlaps the uncertainty estimates of every terrestrial and ocean proxy data point. However, the high latitude southwest Pacific SST estimates are right at the boundary of consistency. The terrestrial data shows very good agreement with the model ensemble, and both data and models show a similar degree of polar amplification (see Sect. 4.3).

By regressing the CO<sub>2</sub> levels and  $\sigma$  values in Table 2, it is possible (for those models with more than one Eocene

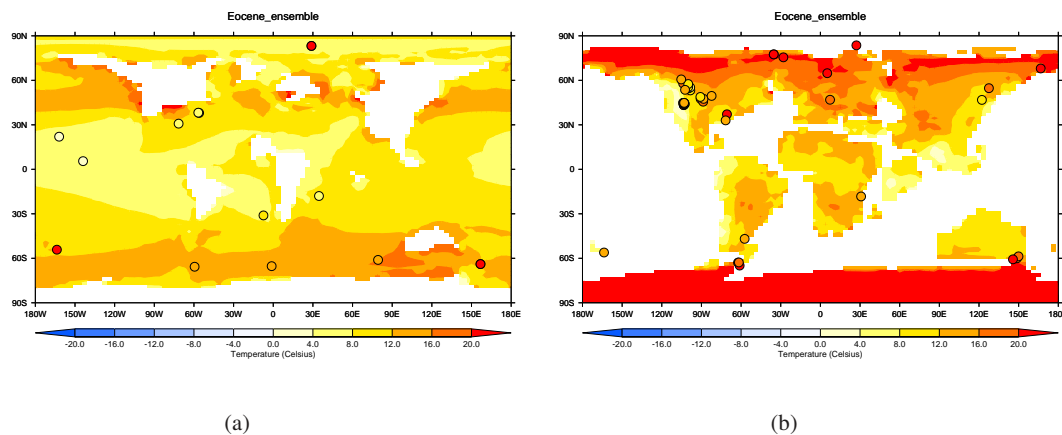
simulation) to provide a first-order estimate of the CO<sub>2</sub> level, for each model, which could give the best agreement with the proxy estimates. For HadCM, CCSM.H, and CCSM.W, using  $\sigma_{\text{sst}}$  this is 2600 ppmv, 4300 ppmv, and 4900 ppmv, respectively, and using  $\sigma_{\text{lat}}$  this is 2800 ppmv, 4500 ppmv, and 6300 ppmv, respectively. These estimates come with many caveats, not least that the uneven and sparse data spread means that the absolute minimum mean error,  $\sigma$ , is not necessarily a good indicator of the correct global mean temperature. However, they do indicate the magnitude of the range of CO<sub>2</sub> values that could be considered consistent with model results. These values are significantly higher than those presented for this time period in the compilation of Beerling and Royer (2011).

### 4.3 Meridional gradients and polar amplification

The changes in meridional temperature gradient are summarised in Fig. 10, which shows the surface temperature difference between the low latitudes ( $|\phi| < 30^\circ$ ) and the high latitudes ( $|\phi| > 60^\circ$ ) as a function of global mean temperature, and how this is partitioned between land and ocean warming (Fig. 10b,  $\phi$  is latitude in degrees). All the Eocene simulations have a reduced meridional surface temperature



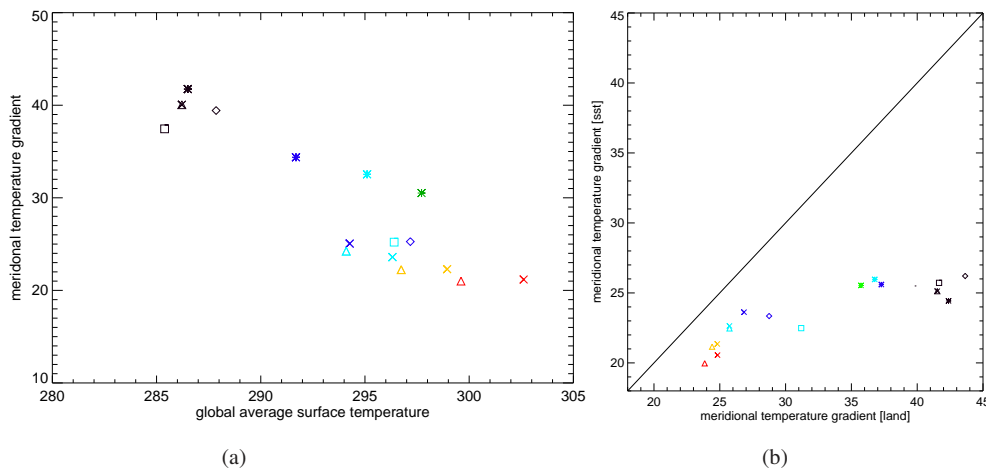
**Fig. 8.** Zonal ensemble mean model (middle thick black line), and data, presented as an anomaly relative to present/pre-industrial. Outer thick black lines indicate  $\pm 2$  standard deviations in the models. Coloured lines represent each individual model simulation in the ensemble, with the colour indicating  $CO_2$  level as in Figs. 5 and 7. (a)  $[SST_e - SST_p]$ . (b)  $[LAT_e - GAT_p]$ . For this figure, the ensemble consists of the best simulation from each model, as highlighted in bold in Table 2. Descriptions of the proxy error bars are given in the captions to Figs. 5 and 7.



**Fig. 9.** Ensemble mean modelled Eocene warming, presented as an anomaly relative to present/pre-industrial. (a)  $[SST_e - SST_p]$ . (b)  $[LAT_e - GAT_p]$ . The ensemble consists of the best simulation from each model, as highlighted in bold in Table 2.

gradient compared with the pre-industrial, and the gradient reduces further as  $CO_2$  increases, i.e. polar amplification increases (Fig. 10a). However, there is a high degree of inter-model variability in the absolute Eocene gradient, with HadCM appearing to be an outlier with a relatively high Eocene gradient. There is some indication that the models are trending towards a minimum gradient of about  $20^\circ C$ . This, along with our energy flux analysis (see Sect. 5), supports previous work (Huber et al., 2003) that implied that meridional temperature gradients of the order  $20^\circ C$  were physically realistic, even without large changes to meridional heat transport. Compared with pre-industrial, the meridional surface temperature gradient reduces more on land than over ocean (Fig. 10b). For HadCM, this applies also to the Eocene simulations as  $CO_2$  increases. However, for the two CCSM

models, the meridional temperature gradient is reduced by a similar amount over land and ocean as a function of  $CO_2$ , with some indication, at maximum ( $\times 16$ )  $CO_2$ , that the SST gradient starts reducing more over ocean than over land. This implies that when considering changes relative to the modern, it is possible to have substantially different temperature changes over land compared with over ocean at the same latitude. This is also clear from comparing Fig. 3 with Fig. 4, and shows the importance of differentiating terrestrial and oceanic signals when considering the consistency between different proxy data, and between data and models.



**Fig. 10.** (a) Meridional surface temperature gradient  $GST_{|\phi|>60} - GST_{|\phi|<30}$ , where  $|\phi|$  is the absolute value of the latitude in degrees, as a function of global mean surface temperature,  $\langle GST \rangle$ , for all the simulations presented in this paper. (b) Meridional surface temperature gradient over land/ocean,  $SST_{|\phi|>60} - SST_{|\phi|<30}$  vs.  $LAT_{|\phi|>60} - LAT_{|\phi|<30}$ . Symbols and colours correspond to those in Fig. 2.

## 5 Reasons for inter-model variability: an energy flux analysis

It is interesting up to a point to simply intercompare model results, and to compare with data, but also of interest is to know *why* different models behave differently. Given the complexity of climate models, this can be problematic, and traditionally, groups such as PMIP have not often diagnosed in detail the differences. Here, we attempt to diagnose some aspects of the differences between the model results, building on a 1-D energy-balance approach as outlined by Heinemann et al. (2009). Here, the causes of the zonal mean temperature response of a model are diagnosed from the top of the atmosphere and surface radiative fluxes, including their clear-sky values, assuming simple energy balance. Any difference between the meridional temperature profile in the GCM and that estimated from the energy-balance approach is attributed to meridional heat transport. As such, the change in meridional temperature profile between two simulations (such as a pre-industrial control and an Eocene simulation) can be attributed to a combination of (1) changes in emissivity due to changes in clouds, (2) changes in emissivity due to changes in the greenhouse effect (i.e.  $CO_2$  and water vapour concentration changes, and lapse-rate effects), (3) changes in albedo due to changes in clouds, (4) changes in albedo due to Earth surface and atmospheric aerosol changes, and (5) changes in meridional heat transport.

Following Heinemann et al. (2009), the 1-D energy balance model (EBM) is formulated by equating the incoming solar radiation with outgoing long wave radiation, with any local imbalance attributed to local meridional heat transport:

$$SW_t^\downarrow (1 - \alpha) + H = \epsilon \sigma \tau^4, \quad (3)$$

where  $SW_t^\downarrow$  is the incoming solar radiation at the top of the atmosphere,  $\alpha$  is the planetary albedo,  $H$  is the net meridional heat transport convergence,  $\epsilon$  is the atmospheric emissivity,  $\sigma$  is the Stephan–Boltzmann constant, and  $\tau$  is the surface temperature, to be diagnosed by the EBM. All variables are functions of latitude apart from  $\sigma$ .

The planetary albedo is given by

$$\alpha = \frac{SW_t^\uparrow}{SW_t^\downarrow}, \quad (4)$$

and the atmospheric emissivity is given by

$$\epsilon = \frac{LW_t^\uparrow}{LW_s^\uparrow}, \quad (5)$$

where  $SW_t^\uparrow$  and  $SW_t^\downarrow$  are the outgoing and incoming top of the atmosphere shortwave radiation, respectively, and  $LW_t^\uparrow$  and  $LW_s^\uparrow$  are the upwelling top of the atmosphere and surface longwave radiation, respectively. Given that the surface emits long wave radiation according to

$$LW_s^\uparrow = \sigma \tau^4, \quad (6)$$

it follows that the meridional heat transport convergence,  $H$ , is given by

$$H = -(SW_t^{\text{net}} + LW_t^{\text{net}}), \quad (7)$$

where the superscript net represents net flux (positive downwards). Equation (7) reflects the necessity that, in equilibrium, any net downward shortwave plus longwave heat flux has to be compensated by a negative meridional heat flux convergence (note that there is a typo in the equivalent equation of Heinemann et al. (2009), the minus sign in their Eq. 4



is erroneous). All the radiative fluxes are output directly from the GCMs, and used as input into the energy balance model. From Eq. (3), it follows that

$$\tau = \left( \frac{1}{\epsilon \sigma} (\text{SW}_t^\downarrow (1 - \alpha) + H)^{0.25} \right) \equiv E(\epsilon, \alpha, H). \quad (8)$$

The difference in temperature between two simulations,  $\Delta T = \tau - \tau'$ , is given by  $E(\epsilon, \alpha, H) - E(\epsilon', \alpha', H')$ , where the prime,  $'$ , represents values in the second simulation. In order to diagnose the reasons for the temperature differences in two simulations, we consider changes to the diagnosed emissivity, planetary albedo, and heat transport, and write

$$\Delta T_{\text{emm}} = E(\epsilon, \alpha, H) - E(\epsilon', \alpha, H) \quad (9)$$

$$\Delta T_{\text{alb}} = E(\epsilon, \alpha, H) - E(\epsilon, \alpha', H) \quad (10)$$

$$\Delta T_{\text{tran}} = E(\epsilon, \alpha, H) - E(\epsilon, \alpha, H'), \quad (11)$$

where  $\Delta T_{\text{emm}}$ ,  $\Delta T_{\text{alb}}$ , and  $\Delta T_{\text{tran}}$  are the components of  $\Delta T$  due to emissivity, planetary albedo, and heat transport changes, respectively. Because the changes in emissivity, albedo, and heat transport are relatively small compared to their magnitude,

$$\Delta T \simeq \Delta T_{\text{emm}} + \Delta T_{\text{alb}} + \Delta T_{\text{tran}}. \quad (12)$$

We further partition the  $\Delta T_{\text{emm}}$  and  $\Delta T_{\text{alb}}$  terms by considering the clear-sky radiative fluxes, also output directly from the GCMs. Using  $\text{cs}$  as a subscript to denote clear-sky fluxes, we can estimate the contribution due to the greenhouse effect ( $\text{CO}_2$  and water vapour and lapse rate) changes,  $\Delta T_{\text{gg}}$ , and the contribution due to surface albedo and aerosol changes,  $\Delta T_{\text{salb}}$ :

$$\Delta T_{\text{gg}} = E(\epsilon_{\text{cs}}, \alpha_{\text{cs}}, H_{\text{cs}}) - E(\epsilon'_{\text{cs}}, \alpha_{\text{cs}}, H_{\text{cs}}) \quad (13)$$

$$\Delta T_{\text{salb}} = E(\epsilon_{\text{cs}}, \alpha_{\text{cs}}, H_{\text{cs}}) - E(\epsilon_{\text{cs}}, \alpha'_{\text{cs}}, H_{\text{cs}}), \quad (14)$$

because the emissivity change in the clear-sky case is solely due to greenhouse effect changes, and the albedo change in the clear-sky case is mainly due to surface albedo and aerosols. Considering the remaining temperature difference as due to clouds, we can then write

$$\Delta T_{\text{lwc}} = \Delta T_{\text{emm}} - \Delta T_{\text{gg}} \quad (15)$$

$$\Delta T_{\text{swc}} = \Delta T_{\text{alb}} - \Delta T_{\text{salb}}, \quad (16)$$

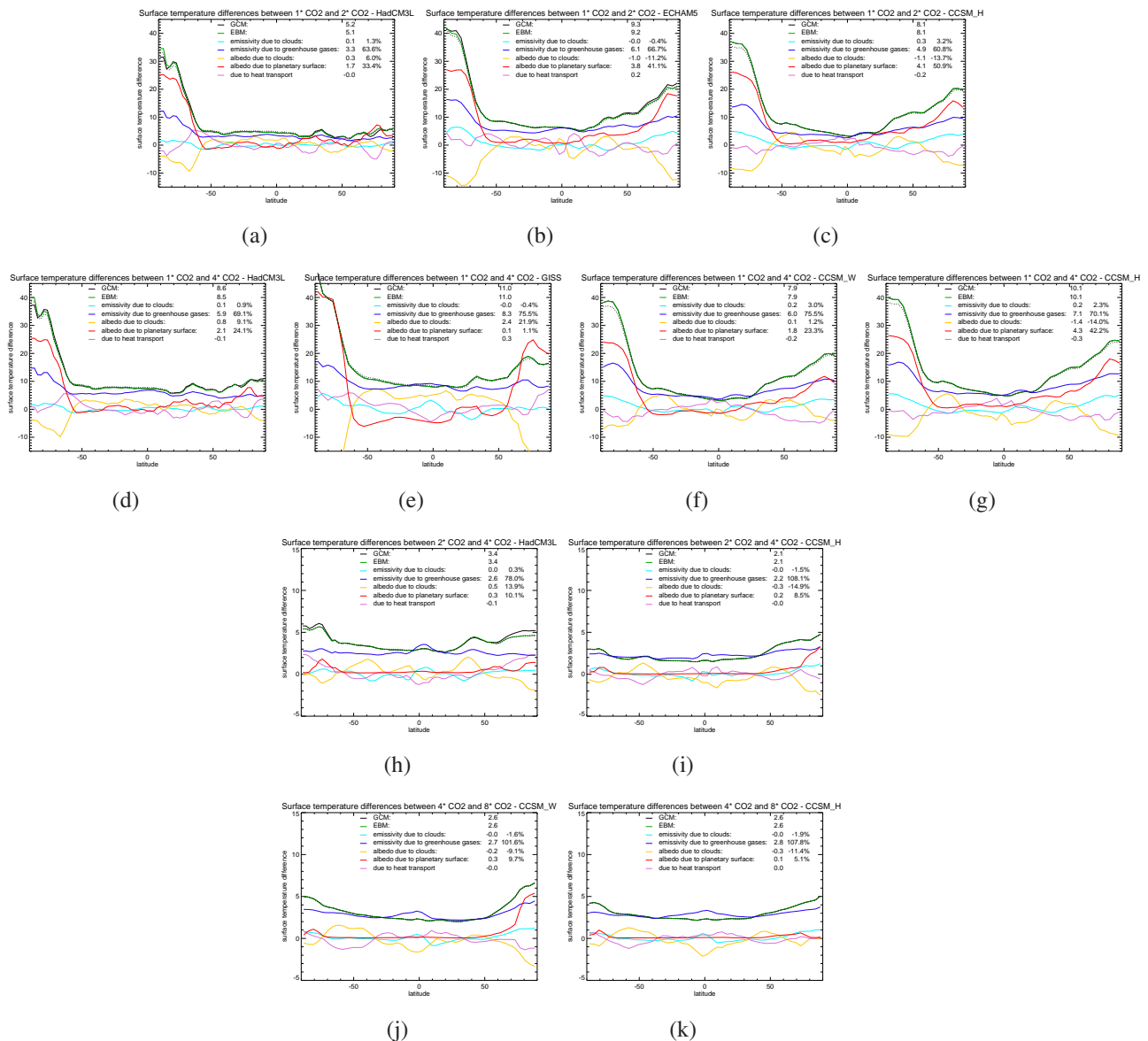
where  $\Delta T_{\text{lwc}}$  and  $\Delta T_{\text{swc}}$  are the components of  $\Delta T$  due to long-wave cloud changes and short-wave cloud changes, respectively. In this way, a temperature difference between two simulations can be partitioned into 5 components, given by Eqs. (9)–(11) and (13)–(16).

Figure 11 shows the results from this energy balance analysis for a number of simulations. Figure 11a–c shows the models which simulate a transition from pre-industrial to Eocene at  $\times 2 \text{ CO}_2$ . ECHAM and CCSM\_H show similar results in terms of the reasons for this change. They show a high latitude warming in both hemispheres caused mainly

by non-cloud albedo changes, with a significant contribution also from emissivity changes. In both these models, short-wave cloud albedo changes act to reduce the polar amplification in both hemispheres. The greater global temperature change in ECHAM compared with CCSM\_H is due to the greater change in greenhouse effect. However, the energy balance analysis does not allow us to diagnose if this is due to a greater radiative forcing given the same  $\text{CO}_2$  increase, or due to greater water vapour feedbacks or lapse-rate changes in ECHAM. HadCM exhibits quite different behaviour. In the Southern Hemisphere, the zonal mean temperature increase is due predominantly to non-cloud albedo changes, and is reduced relative to the other two models. In the Northern Hemisphere, the increase in temperature is much reduced relative to the other two models due to a lack of non-cloud albedo feedbacks and changes in emissivity. Abbot and Tziperman (2008) suggested that the lack of sea ice in the Arctic can lead to stronger convection over the relatively warm Arctic sea surface during winter, leading to more convective clouds and increased water vapour concentrations, and thereby causing polar amplification via both albedo and emissivity effects. The largely decreased (versus unchanged) surface albedo in northern high latitudes in CCSM\_H and ECHAM (versus HadCM), increased (versus virtually unchanged) longwave cloud radiative forcing, and reduced (versus hardly changed) clear-sky emissivity indicate that this sea ice/convection feedback is active for  $\times 1$  to  $\times 2$  in CCSM\_H and ECHAM, but absent in HadCM. The reduced strength of this feedback in HadCM may be related to the relatively strong Eocene seasonality in HadCM compared with the other models (as can be seen by comparing Fig. 6 with Fig. 5), which suppresses Arctic convection in HadCM in winter. Changes in heat transport are playing a relatively minor role in determining the latitudinal temperature profile in ECHAM, CCSM\_H, and HadCM, which supports previous findings using ECHAM alone (Heinemann et al., 2009).

Figure 11d–g shows the models which simulate a transition from pre-industrial to Eocene at  $\times 4 \text{ CO}_2$ . For HadCM and CCSM\_H, the results are very similar to  $\times 2 \text{ CO}_2$ , but with greater magnitude; for both models each component contributes the same fraction to the total warming under  $\times 2$  as to under  $\times 4$ , to within  $\sim 10\%$ . CCSM\_W is very similar to CCSM\_H, except that it has reduced warming due to decreased change in non-cloud albedo. This is most likely a direct result of the different aerosol fields applied in the these two models for the Eocene (see Table 1). The model which exhibits the greatest warming is the GISS model. This high sensitivity relative to the other models is due to greater greenhouse gas effect changes and greater cloud albedo feedbacks. The warming over Antarctica is particularly large in the GISS model, and is due to a greater local change in non-cloud albedo. However, the GISS model also has strong negative cloud forcing at high latitudes in both hemispheres.

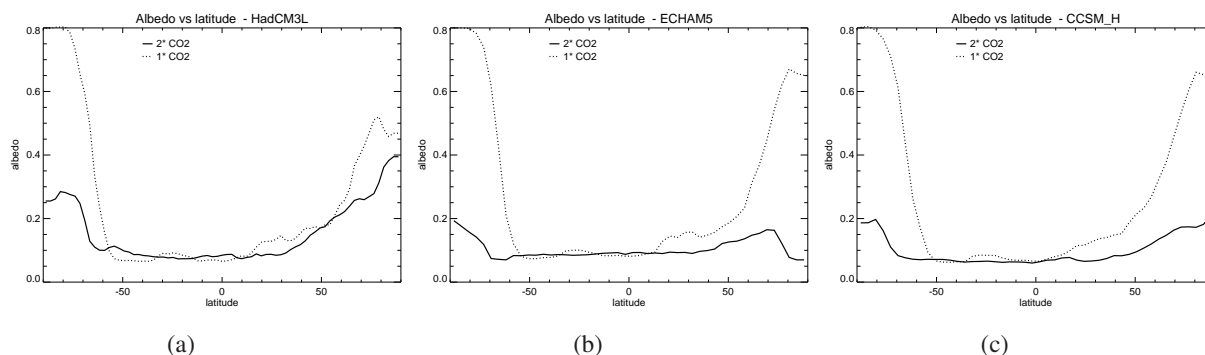
Figure 11h–i shows the models which simulate a transition from  $\times 2$  to  $\times 4 \text{ CO}_2$  under Eocene conditions. HadCM



**Fig. 11.** The zonal mean surface temperature change under a range of CO<sub>2</sub> transitions, and energy balance analysis of the reasons for the changes. (a–c)  $\times 1$  to  $\times 2$  CO<sub>2</sub>, (d–g)  $\times 1$  to  $\times 4$  CO<sub>2</sub>, (h–i)  $\times 2$  to  $\times 4$  CO<sub>2</sub>, (j–k)  $\times 4$  to  $\times 8$  CO<sub>2</sub>. The simulations at  $\times 1$  CO<sub>2</sub> are pre-industrial reference simulations. Note the difference in vertical scale in panels (a–g) compared with (h–k). The dotted lines in the plots show the sum of the various components, which in each case should be very close to the GCM line (i.e. the “actual” temperature change from the model) and the EBM line (i.e.  $\Delta T$  as calculated from Eq. 8 for the two climate states).

has a greater climate sensitivity than CCSM\_H, and this is due to greater changes in greenhouse gas emissivity, and a positive as opposed to negative cloud albedo feedback. The relative lack of polar amplification in both models compared to the results discussed above is due to the lack of Antarctic ice sheet in the Eocene. The small amount of polar amplification in HadCM is due to changes in heat transport; in CCSM it is due to non-cloud albedo changes in the Northern Hemisphere.

Figure 11j–k shows the models which simulate a transition from  $\times 4$  to  $\times 8$  CO<sub>2</sub> under Eocene conditions. Similar to the transition from  $\times 2$  to  $\times 4$ , the polar amplification is relatively small. The warming is due almost entirely to the changes in emissivity (direct CO<sub>2</sub> forcing and water vapour feedbacks and lapse-rate changes), and unsurprisingly has a similar latitudinal distribution in the two models. However, in the Northern Hemisphere high latitudes the CCSM\_H model shows strong opposing effects of cloud and surface changes, which are not present in CCSM\_W. This is most



**Fig. 12.** Zonal mean albedo in the  $\times 1$  and  $\times 2$   $\text{CO}_2$  simulations using the (a) HadCM, (b) ECHAM, and (c) CCSM\_H models. The simulations at  $\times 1$   $\text{CO}_2$  are pre-industrial reference simulations.

likely due to the remnants of Arctic sea ice in CCSM\_W at  $\times 8$   $\text{CO}_2$  which are not present in the warmer CCSM\_H model. Comparison of Fig. 11k with Fig. 11i shows that the increase in climate sensitivity in CCSM\_H as a function of background  $\text{CO}_2$  is due almost entirely to increased non-cloud emissivity changes; the framework does not allow us to determine if this is due to increasing radiative effects due to  $\text{CO}_2$ , or increasing water vapour feedbacks or lapse-rate changes. However, it is clear that it is not due to increased albedo feedbacks, or cloud processes.

Given that the models prescribe Eocene vegetation in quite different ways, it is interesting to assess how much this affects inter-model variability. Figure 12 shows the surface albedo in the pre-industrial control and the  $\times 2$   $\text{CO}_2$  simulations for HadCM, CCSM\_H, and ECHAM. At the high latitudes, this is affected by snow and sea ice cover and prescribed changes in ice sheets, but at low latitudes this is purely a result of the imposed vegetation and open-ocean albedos. The fact that all the models have a low latitude albedo which is similar to their control, and similar to each other, indicates that this aspect of experimental design is likely not playing an important role in determining the differences in results between the models.

## 6 Conclusions and outlook

We have carried out an intercomparison and model–data comparison of the results from 5 early Eocene modelling studies, using 4 different climate models. The model results show a large spread in global mean temperatures, for example a  $\sim 9^\circ\text{C}$  range in surface air temperature under a single  $\text{CO}_2$  value, and are characterised by warming in different regions. The models which have been run at sufficiently high  $\text{CO}_2$  show very good agreement with the terrestrial data. The comparison with SST data is also good, but the model and data uncertainty only just overlap for the Arctic and southwest Pacific  $\delta^{18}\text{O}$  and TEX<sub>86</sub> proxies. However, if a possible seasonality bias in the proxies is taken into account, then

the model–data agreement improves further. We have interrogated the reasons for the differences between the models, and found differences in climate sensitivity to be due primarily to a combination of greenhouse effect and surface albedo feedbacks, rather than differences in heat transport or cloud feedbacks.

There are several issues which have emerged from this study which should be addressed in future work aimed at reconciling model simulations and proxy data reconstructions of the early Eocene (many of which also apply to other time periods).

Firstly, modelling groups should aim to carry out simulations over a wider range of atmospheric  $\text{CO}_2$  levels. In particular, the results of CCSM\_H indicate that at high prescribed atmospheric  $\text{CO}_2$  and low aerosol forcing, the models and data come close together. Some of this work is in progress (e.g. simulations at  $\times 3$   $\text{CO}_2$  are currently being analysed for the ECHAM model). However, it should be noted that this is not always possible. For example, the Eocene HadCM model has been run at  $\times 8$   $\text{CO}_2$ , but after about 2700 yr the model developed a runaway greenhouse, and the model eventually crashed (Lunt et al., 2007). A similar effect has been observed in the ECHAM model at  $\times 4$   $\text{CO}_2$  (Heinemann, 2009). Whether such an effect is “real”, i.e. whether the real world would also develop a runaway greenhouse, is completely unknown. In any case, it is clear that modelling the early Eocene climate pushes the climate model parameterisations to the boundaries within which they were designed to operate, if not beyond these boundaries.

Some of the differences between the model results can be attributed to differences in the experimental design. In particular, some models apply a very generic Eocene vegetation, which is not particularly realistic. A slightly more coordinated study could provide guidelines for ways to better represent Eocene vegetation, for example by making use of palynological data or by using dynamic vegetation models where available. This would provide an ensemble of model results which better represented the true uncertainty in our model simulations. Other inconsistencies between model

simulations should not necessarily be eliminated – for example, different models using different paleogeographical reconstructions may be more representative of the true spread of model results than if all groups used a single paleogeography.

On the data side, better understanding of the temperature proxies and their associated uncertainties, in particular seasonal effects, is a clear goal for future work, as is greater geographical and finer temporal coverage. More tropical Eocene terrestrial data would be especially beneficial for assessing the terrestrial meridional temperature gradient.

Perhaps most crucial of all, better CO<sub>2</sub> constraints from proxies would be of huge benefit to model–data comparison exercises such as this. Recently, much work is being undertaken in this area, but this should be intensified wherever possible. We note that at high CO<sub>2</sub>, due to the logarithmic nature of the CO<sub>2</sub> forcing, proxies which may have relatively coarse precision at low CO<sub>2</sub> can actually provide very strong constraints on the CO<sub>2</sub> forcing itself. Such constraints on CO<sub>2</sub>, combined with proxy temperature reconstructions with well defined uncertainty ranges, could provide a strong constraint on model simulations, providing quantitative metrics for assessing model performance, and could ultimately provide relative weightings for model simulations of future climates.

**Supplementary material related to this article is available online at:** <http://www.clim-past.net/8/1717/2012/cp-8-1717-2012-supplement.zip>.

**Acknowledgements.** This paper resulted from discussions at the “CO<sub>2</sub> at the zoo” event held in Bristol in 2010, and from a Royal Society Kavli meeting held in 2011. DJL acknowledges funding as an RCUK research fellow. NCEP Reanalysis Derived data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at <http://www.esrl.noaa.gov/psd/>. M. H. was funded by NSF P2C2 grant OCE 0902882. This is PCCRC paper number 1220.

Edited by: A. Haywood

## References

- Abbot, D. S. and Tziperman, E.: Sea Ice, High Latitude Convection, and Equable Climates, *Geophysical Research Letters*, 35, L03702, doi:10.1029/2007GL032286, 2008.
- Anand, P., Elderfield, H., and Conte, M.: Calibration of Mg/Ca thermometry in planktonic foraminifera from a sediment trap time series, *Paleoceanography*, 18, 1050, doi:10.1029/2002PA000846, 2003.
- Beerling, D. and Royer, D.: Convergent Cenozoic CO<sub>2</sub> history, *Nat. Geosci.*, 4, 418–420, 2011.
- Bemis, B., Spero, H., Bijma, J., and Lea, D.: Reevaluation of the oxygen isotopic composition of planktonic foraminifera: Experimental results and revised paleotemperature equations, *Paleoceanography*, 13, 150–160, 1998.
- Bice, K. L. and Marotzke, J.: Numerical evidence against reversed thermohaline circulation in the warm Paleocene/Eocene ocean, *J. Geophys. Res.*, 106, 11529–11542, 2001.
- Bijl, P. K., Schouten, S., Sluijs, A., Reichert, G. J., Zachos, J. C., and Brinkhuis, H.: Early Palaeogene temperature evolution of the southwest Pacific Ocean, *Nature*, 461, 776–779, 2009.
- Braconnot, P., Otto-Bliesner, B., Harrison, S., Joussaume, S., Peterchmitt, J.-Y., Abe-Ouchi, A., Crucifix, M., Driesschaert, E., Fichefet, Th., Hewitt, C. D., Kageyama, M., Kitoh, A., Laîné, 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:10.5194/cp-3-261-2007, 2007.
- Brinkhuis, H., Schouten, S., Collinson, M. E., Sluijs, A., Damsté, J. S. S., Dickens, G. R., Huber, M., Cronin, T. M., Onodera, J., Takahashi, K., Bujak, J. P., Stein, R., van der Burgh, J., Eldrett, J. S., Harding, I. C., Lotter, A. F., Sangiorgi, F., van Konijnenburg-van Cittert, H., de Leeuw, J. W., Matthiessen, J., Backman, J., Moran, K., and the Expedition 302 Scientists: Episodic fresh surface waters in the Eocene Arctic Ocean, *Nature*, 441, 606–609, 2006.
- Coggon, R. M., Teagle, D. A. H., Smith-Duque, C. E., Alt, J. C., and Cooper, M. J.: Reconstructing past seawater Mg/Ca and Sr/Ca from mid-ocean ridge flank calcium carbonate veins, *Science*, 327, 1114–1117, 2010.
- Coggon, R. M., Teagle, D. A. H., and Dunkley Jones, T.: Comment on What do we know about the evolution of Mg to Ca ratios in seawater? by Wally Broecker and Jimin Yu, *Paleoceanography*, 26, PA3224, doi:10.1029/2011PA002186, 2011.
- Collins, W. D., Bitz, C. M., Blackmon, M. L., Bonan, G. B., Bretherton, C. S., Carton, J. A., Chang, P., Doney, S. C., Hack, J. J., Henderson, T. B., Kiehl, J. T., Large, W. G., McKenna, D. S., Santer, B. D., and Smith, R. D.: The Community Climate System Model Version 3 (CCSM3), *J. Climate*, 19, 2122–2143, 2006.
- Colman, R. and McAvaney, B.: Climate feedbacks under a very broad range of forcing, *Geophysical Research Letters*, 36, L01702, doi:10.1029/2008GL036268, 2009.
- 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*, edited by: Pearce, R., Academic Press, 259–279, 2001.
- Cramer, B. S., Miller, K. G., Barrett, P. J., and Wright, J. D.: Late Cretaceous–Neogene trends in deep ocean temperature and continental ice volume: Reconciling records of benthic foraminiferal geochemistry ( $\delta^{18}\text{O}$  and Mg/Ca) with sea level history, *J. Geophys. Res.*, 116, C12023, doi:10.1029/2011JC007255, 2011.
- Creech, J. B., Baker, J. A., Hollis, C. J., Morgans, H. E. G., and Smith, E. G. C.: Eocene sea temperatures for the mid-latitude southwest Pacific from Mg/Ca ratios in planktonic and benthic foraminifera, *Earth Planet. Sci. Lett.*, 299, 483–495, 2010.
- Douglas, P. M., Ivany, L., Pagani, M., Hollis, C. J., Beu, A. G., and Affek, H. P.: Eocene Southern high latitude sea surface temperatures: new constraints from clumped isotope paleothermometry, AGU fall meeting, San Francisco, PP33B–1924, 2011.
- Dowsett, H. J., Robinson, M. M., and Foley, K. M.: Pliocene three-dimensional global ocean temperature reconstruction, *Clim. Past*, 5, 769–783, doi:10.5194/cp-5-769-2009, 2009.

- Farkaš, J., B'ohm, F., Wallmann, K., Blenkinsop, J., Eisenhauer, A., van Geldern, R., Munnecke, A., Voigt, S., and Veizer, J.: Calcium isotope record of Phanerozoic oceans: Implications for chemical evolution of seawater and its causative mechanisms, *Geochem. Cosmochem. Acta*, 71, 5117–5134, 2007.
- Grossman, E. and Ku, T. L.: Oxygen and carbon isotope fractionation in biogenic aragonite: Temperature effects, *Chem. Geol.*, 59, 59–74, 1986.
- Hegerl, G. C., Zwiers, F. W., Braconnot, P., Gillett, N. P., Luo, Y., Orsini, J. A. M., Nicholls, N., Penner, J. E., and Stott, P. A.: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, chap. Understanding and Attributing Climate Change, Cambridge University Press, 2007.
- Heinemann, M.: Warm and sensitive Paleocene-Eocene climate, Max Planck Institute for Meteorology, Hamburg, Germany, Reports on Earth System Science 70, <http://www.mpimet.mpg.de/fileadmin/staff/heinemannmalte/publications/MPI.Rep70.pdf> (last access: 18 October 2012), 2009.
- Heinemann, M., Jungclaus, J. H., and Marotzke, J.: Warm Paleocene/Eocene climate as simulated in ECHAM5/MPI-OM, *Clim. Past*, 5, 785–802, doi:10.5194/cp-5-785-2009, 2009.
- Hollis, C. J., Handley, L., Crouch, E. M., Morgans, H. E. G., Baker, J. A., Creech, J., Collins, K. S., Gibbs, S. J., Huber, M., Schouten, S., Zachos, J. C., and Pancost, R. D.: Tropical sea temperatures in the high-latitude South Pacific during the Eocene, *Geology*, 37, 99–102, 2009.
- 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 Planet. Sci. Lett.*, 349–350, 53–66, 2012.
- Huber, M. and Caballero, R.: The early Eocene equable climate problem revisited, *Clim. Past*, 7, 603–633, doi:10.5194/cp-7-603-2011, 2011.
- Huber, M., Sloan, L. C., and Shellito, C.: Causes and consequences of globally warm climates in the early Paleogene, chap. Early Paleogene oceans and climate: A fully coupled modeling approach using the NCAR CCSM, Geological Society of America Special Paper 369, 2003.
- Ivany, L. C., Lohmann, K. C., Hasiuk, F., Blake, D. B., Glass, A., Aronson, R. B., and Moody, R. M.: Eocene climate record of a high southern latitude continental shelf: Seymour Island, Antarctica, *Geol. Soc. Am. Bull.*, 150, 659–678, 2008.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, *Bull. Amer. Meteor. Soc.*, 77, 437–470, 1996.
- Kim, J.-H., Meer, J. V. D., Schouten, S., Helmke, P., Willmott, V., Sangiorgi, F., Koç, N., Hopmans, E. C., and Damsté, J. S. S.: New indices and calibrations derived from the distribution of crenarchaeal isoprenoid tetraether lipids: Implications for past sea surface temperature reconstructions, *Geochim. Cosmochim. Acta*, 74, 4639–4654, 2010.
- Kobashi, T., Grossman, E., Yancey, T. E., and Dockery, D. T.: The oxygen isotopic record of seasonality in *Conus* shells and its applications to understanding the late middle Eocene (38 Ma) climate, *Paleontol. Res.*, 7, 343–355, 2003.
- Kump, L. R. and Pollard, D.: Amplification of Cretaceous Warmth by Biological Cloud Feedbacks, *Science*, 320, p. 195, 2008.
- Lear, C., Rosenthal, Y., and Slowey, N.: Benthic foraminiferal Mg/Ca-paleothermometry: A revised core-top calibration, *Geochimica et Cosmochimica Acta*, 66, 3375–3387, 2002.
- Liu, Z., Pagani, M., Zinniker, D., DeConto, R. M., Huber, M., Brinkhuis, H., Shah, S. R., Leckie, R. M., and Pearson, A.: Global Cooling During the Eocene-Oligocene Climate Transition, *Science*, 323, 1187–1190, 2009.
- Lunt, D. J., Valdes, P. J., and Ridgwell, A.: On the linearity of Climate Sensitivity: A case study for the future and the PETM, AGU poster, available at: [http://www.paleo.bris.ac.uk/~ggdjl/conferences/agu2007\\_climsens.pdf](http://www.paleo.bris.ac.uk/~ggdjl/conferences/agu2007_climsens.pdf) (last access: 18 October 2012), 2007.
- Lunt, D. J., Valdes, P. J., Dunkley Jones, T., Ridgwell, A., Haywood, A. M., Schmidt, D. N., Marsh, R., and Maslin, M.: CO<sub>2</sub>-driven ocean circulation changes as an amplifier of Paleocene-Eocene thermal maximum hydrate destabilization, *Geology*, 38, 875–878, 2010.
- Meehl, G. A., Stocker, T. F., Collins, W. D., Friedlingstein, P., Gaye, A. T., Gregory, J. M., Kitoh, A., Knutti, R., Murphy, J. M., Noda, A., Raper, S. C. B., Watterson, I. G., Weaver, A. J., and Zhao, Z.-C.: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, chap. Global climate projections, Cambridge University Press, 2007.
- Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., Kram, T., Meehl, G. A., Mitchell, J. F. B., Nakicenovic, N., Riahi, K., Smith, S. J., Stouffer, R. J., Thomson, A. M., Weyant, J. P., and Wilbanks, T. J.: The next generation of scenarios for climate change research and assessment, *Nature*, 463, 747–756, 2010.
- Nicholas, C., Pearson, P. N., Bown, P. R., Dunkley Jones, T., Huber, B. T., Karega, A., Lees, J. A., McMillan, I. K., O'Halloran, A., Singano, J. M., and Wade, B. S.: Stratigraphy and sedimentology of the Upper Cretaceous to Paleogene Kilwa Group, southern coastal Tanzania, *J. Afr. Earth Sci.*, 45, 431–466, 2006.
- Pagani, M., Liu, Z., LaRiviere, L., and Ravelo, A. C.: High Earth-system climate sensitivity determined from Pliocene carbon dioxide concentrations, *Nat. Geosci.*, 3, 27–30, 2010.
- Pagani, M., Huber, M., Liu, Z., Bohaty, S. M., Henderiks, J., Sijp, W., Krishnan, S., and DeConto, R. M.: The role of carbon dioxide during the onset of Antarctic glaciation, *Science*, 334, 1261–1264, 2011.
- Panchuk, K., Ridgwell, A., and Kump, L. R.: Sedimentary response to Paleocene-Eocene Thermal Maximum carbon release: A model-data comparison, *Geology*, 36, 315–318, 2008.
- Payros, A., Bernaola, G., Orue-Etxebarria, X., Tosquella, J. D.-T., and Apellaniz, E.: Reassessment of the Early-Middle Eocene biomagnetochronology based on evidence from the Gorrondaxte section (Basque Country, western Pyrenees), *Lethaia*, 40, 183–195, 2007.
- Pearson, P. N., Ditchfield, P. W., Singano, J., Harcourt-Brown, K. G., Nicholas, C. J., and Olsson, R. K.: Warm tropical sea surface temperatures in the Late Cretaceous and Eocene epochs,

- Nature, 413, 481–487, 2001.
- Pearson, P. N., Nicholas, C. J., Singano, J. M., Bown, P. R., Coxall, H. K., van Dongen, B. E., Huber, B. T., Karega, A., and E. Msaky, J. A. L., Pancost, R. D., Pearson, M., and Roberts, A. P.: Paleogene and cretaceous sediment cores from the Kilwa and Lindi areas of coastal Tanzania: Tanzania Drilling Project Sites 1–5, *J. Afr. Earth Sci.*, 39, 25–62, 2004.
- Pearson, P. N., van Dongen, B. E., Nicholas, C. J., Pancost, R. D., Schouten, S., Singano, J. M., and Wade, B. S.: Stable warm tropical climate through the Eocene Epoch, *Geology*, 35, 211–214, 2007.
- Pearson, P. N., Foster, G. L., and Wade, B. S.: Atmospheric carbon dioxide through the Eocene-Oligocene climate transition, *Nature*, 461, 1110–1113, 2009.
- Rathmann, S. and Kuhnert, H.: Carbonate ion effect on Mg/Ca, Sr/Ca and stable isotopes on the benthic foraminifera *Oridorsalis umbonatus* off Namibia, *Mar. Micropaleontol.*, 66, 120–133, 2008.
- Rathmann, S., Hess, S., Kuhnert, H., and Mulitza, S.: Mg/Ca ratios of the benthic foraminifera *Oridorsalis umbonatus* obtained by laser ablation from core top sediments: Relationship to bottom water temperatures, *Geochem., Geophys., Geosyst.*, 5, Q12013, doi:10.1029/2004GC000808, 2004.
- Roberts, C. D., LeGrande, A. N., and Tripathi, A. K.: Climate sensitivity to Arctic seaway restriction during the early Paleogene, *Earth Planet. Science Lett.*, 286, 576–585, 2009.
- Roberts, C. D., LeGrande, A. N., and Tripathi, A. K.: Sensitivity of seawater oxygen isotopes to climatic and tectonic boundary conditions in an early Paleogene simulation with GISS ModelE-R, *Paleoceanography*, 126, PA4203, doi:10.1029/2010PA002025, 2011.
- Roegner, E., Bauml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kirchner, I., Kornbluh, L., Manzini, E., Rhodin, A., Schlese, U., Schulzweida, U., and Tompkins, A.: The atmospheric general circulation model ECHAM5, Tech. rep., Max Planck Institute for Meteorology, Hamburg, Germany, available at: <http://www.mpimet.mpg.de> (last access: 18 October 2012), 2003.
- Schmidt, G. A., Ruedy, R., Hansen, J. E., Aleinov, I., Bell, N., Bauer, M., Bauer, S., Cairns, B., Canuto, V., Cheng, Y., Genio, A. D., Faluvegi, G., Friend, A. D., Hall, T. M., Hu, Y., Kelley, M., Kiang, N. Y., Koch, D., Lacis, A. A., Lerner, J., Lo, K. K., Miller, R. L., Nazarenko, L., Oinas, V., Perlwitz, J. P., Perlwitz, J., Rind, D., Romanou, A., Russell, G. L., Sato, M., Shindell, D. T., Stone, P. H., Sun, S., Tausnev, N., Thresher, D., and Yao, M. S.: Present day atmospheric simulations using GISS ModelE: Comparison to in-situ, satellite and reanalysis data, *J. Climate*, 19, 153–192, 2006.
- Seki, O., Foster, G. L., Schmidt, D. N., Mackenden, A., Kawamura, K., and Pancost, R. D.: Alkenone and boron-based Pliocene pCO<sub>2</sub> records, *Earth Planet. Sci. Lett.*, 292, 201–211, 2010.
- Sewall, J. O., Sloan, L. C., Huber, M., and Wing, S.: Climate sensitivity to changes in land surface characteristics, *Glob. Planet. Change*, 26, 445–465, 2000.
- Sexton, P. F., Wilson, P. A., and Pearson, P. N.: Microstructural and geochemical perspectives on planktic foraminiferal preservation: Glassy versus Frosty, *Geochem. Geophys. Geosyst.*, 7, Q12P19, doi:10.1029/2006GC001291, 2006.
- Shellito, C. J. and Sloan, L. C.: Reconstructing a lost Eocene paradise: Part I. Simulating the change in global floral distribution at the initial Eocene thermal maximum, *Glob. Planet. Change*, 50, 1–17, 2006.
- Sluijs, A., Schouten, S., Pagani, M., Woltering, M., Brinkhuis, H., Damsté, J. S., Dickens, G. R., Huber, M., Reichert, G. J., Stein, R., and L. J. Lourens, J. M., Pedentchouk, N., Backman, J., and Moran, K.: Subtropical Arctic Ocean temperatures during the Palaeocene/Eocene thermal maximum, *Nature*, 441, 610–613, 2006.
- Sluijs, A., Rohl, U., Schouten, S., Sangiorgi, H. J. B. F., Sinninghe Damsté, J. S., and Brinkhuis, H.: Arctic late Paleocene-early Eocene paleoenvironments with special emphasis on the Paleocene-Eocene thermal maximum (Lomonosov Ridge, Integrated Ocean Drilling Program Expedition 302), *Paleoceanography*, 23, PA1S11, doi:10.1029/2007PA001495, 2008.
- Storey, M., Duncan, R. A., and Swisher, C. C.: Paleocene-Eocene thermal maximum and the opening of the northeast Atlantic, *Science*, 316, 587–589, 2007.
- Tindall, J., Flecker, R., Valdes, P., Schmidt, D. N., Markwick, P., and Harris, J.: Modelling oxygen isotope distribution of ancient seawater using a coupled ocean-atmosphere GCM: Implications for reconstructing early Eocene climate, *Earth Planet. Sci. Lett.*, 292, 265–273, 2010.
- 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, *J. Climate*, 23, 2562–2584, 2010.
- Winguth, A. M. E., Thomas, E., and Winguth, C.: Global decline in ocean ventilation, oxygenation, and productivity during the Paleocene-Eocene Thermal Maximum: Implications for the benthic extinction, *Geology*, 40, 263–266, doi:10.1130/G32529.1, 2012.
- Yeager, S. G., Shields, C. A., Large, W. G., and Hack, J. J.: The low-resolution CCSM3, *J. Climate*, 19, 2545–2566, 2006.
- Zachos, J. C., Dickens, G. R., and Zeebe, R. E.: An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics, *Nature*, 451, 279–283, 2008.