



Kennedy-Asser, A. T., Lunt, D. J., Farnsworth, A., & Valdes, P. J. (2019). Assessing Mechanisms and Uncertainty in Modeled Climatic Change at the EoceneOligocene Transition. *Paleoceanography and Paleoclimatology*, 34(1), 16-34. <https://doi.org/10.1029/2018PA003380>

Publisher's PDF, also known as Version of record

License (if available):  
CC BY

Link to published version (if available):  
[10.1029/2018PA003380](https://doi.org/10.1029/2018PA003380)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via Wiley at DOI: 10.1029/2018PA003380. Please refer to any applicable terms of use of the publisher.

## **University of Bristol - Explore Bristol Research**

### **General rights**

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/pure/about/ebr-terms>

# Paleoceanography and Paleoclimatology



## RESEARCH ARTICLE

10.1029/2018PA003380

### Special Section:

Climatic and Biotic Events of the Paleogene: Earth Systems and Planetary Boundaries in a Greenhouse World

### Key Points:

- Uncertainties in boundary conditions for deep-time model simulations lead to substantial uncertainties in modeled climate
- Lack of model spin-up can lead to errors in model simulations that are as large as the uncertainties due to boundary conditions
- Comparison with proxy data shows that the modeled climate change at the EOT is best explained by a drop in  $p\text{CO}_2$

### Supporting Information:

- Supporting Information S1
- Table S1

### Correspondence to:

A. T. Kennedy-Asser,  
alan.kennedy@bristol.ac.uk

### Citation:

Kennedy-Asser, A. T., Lunt, D. J., Farnsworth, A., & Valdes, P. J. (2019). Assessing mechanisms and uncertainty in modeled climatic change at the Eocene-Oligocene transition. *Paleoceanography and Paleoclimatology*, 34, 16–34. <https://doi.org/10.1029/2018PA003380>

Received 1 APR 2018

Accepted 10 DEC 2018

Accepted article online 21 DEC 2018

Published online 16 JAN 2019

©2018. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

## Assessing Mechanisms and Uncertainty in Modeled Climatic Change at the Eocene-Oligocene Transition

A. T. Kennedy-Asser<sup>1,2</sup> , D. J. Lunt<sup>1,2</sup> , A. Farnsworth<sup>1,2</sup> , and P. J. Valdes<sup>1,2</sup>

<sup>1</sup>BRIDGE, School of Geographical Sciences, University of Bristol, Bristol, UK, <sup>2</sup>Cabot Institute for the Environment, University of Bristol, Bristol, UK

**Abstract** The Earth system changed dramatically across the Eocene-Oligocene Transition (EOT) on a variety of spatial and temporal scales. Understanding the many complex and interacting factors affecting the Earth's atmosphere and oceans at the EOT requires the combination of both data and modeling approaches and an understanding of the uncertainty in both of these elements. Here uncertainty in the Earth system response to various imposed forcings typical of changes at the EOT is assessed. By using an ensemble of simulations from the fully coupled general circulation model, HadCM3L, the uncertainty due to differences in the boundary conditions and insufficient model spin-up is quantified. The surface temperature response in high-latitude ocean regions, particularly where deep water formation occurs, is found to be highly sensitive to differences in boundary conditions (i.e., have the greatest magnitude of uncertainty), while low-latitude oceans are the most insensitive to differences in boundary conditions (i.e., have the lowest magnitude of uncertainty). The length of spin-up (or how far the model is from equilibrium) can have a significant effect on the response to some forcings and on the magnitude of uncertainty due to differences in boundary conditions. These findings are important to consider for future modeling work and for interpreting previous published simulations.

**Plain Language Summary** Around 34 million years ago, the first major ice sheet formed over Antarctica. Understanding why the Earth cooled at this time and how this would have impacted on global climate requires the use of geologic data and computer models. Here multiple coupled climate-ocean model simulations are used to assess changes in modeled climate around this period in the Earth's history. Specifically, the climatic sensitivity to subtle changes in the model setup and how long the simulations are run for (the model spin-up) are evaluated. The temperature response in certain regions is found to be particularly sensitive to changes in the model setup and lack of spin-up. This is the case particularly for high-latitude ocean regions, making it harder to build up a consistent picture of how the climate of these regions changed at this time. Lower latitude oceans are much less sensitive, giving us higher confidence in model results for these regions. In these simulations, the global changes that occurred around 34 million years ago are best described by declining atmospheric  $p\text{CO}_2$ .

## 1. Introduction

The Eocene-Oligocene Transition (EOT) occurred approximately 34 million years ago and was a time of major global environmental change (Zachos et al., 2001). The EOT marks an abrupt end to gradual cooling during the late Eocene (Inglis et al., 2015), with significant ice sheet growth over Antarctica (Galeotti et al., 2016) and upheaval in some marine ecosystems (Houben et al., 2013). Other major changes occur around the same time as, but not necessarily coincident with, the EOT, for example, changes in ocean biogeochemistry (Pälike et al., 2012) and ocean circulation (Scher et al., 2015), suggesting possible but unclear linkages and the presence of complex feedback in the Earth system.

Arguably, the most significant event to occur at the EOT is global cooling followed by the growth of a semipermanent Antarctic ice sheet (AIS) to near modern-day proportions (Kennett, 1977; Zachos et al., 2001). This is usually observed in oxygen isotope proxy records as a 1.5‰ increase in two steps, the first representing benthic cooling and the second representing ice growth (Coxall et al., 2005; Lear et al., 2008). This oxygen isotope shift could represent a benthic ocean cooling of 2–3 °C and an expansion of the AIS to 60–130% of its modern volume (Bohaty et al., 2012), with uncertainties due to the unknown isotopic composition of the AIS, for example. Evidence of Antarctic glaciation is further supported by changes in clay mineralogy of sediments around Antarctica and appearances of ice rafted debris at the EOT (Carter et al., 2017; Galeotti et al., 2016).

Not all changes that occurred around the time of the EOT, however, are so clearly marked. Some changes are more gradual, starting during the late Eocene and continuing through the EOT. This is the case for oceanographic changes such as the initiation of the Antarctic Circumpolar Current (ACC) and the formation of deep water in the North Atlantic. Traditionally, the former was cited as the cause of Antarctic glaciation, through the thermal isolation of the continent (Kennett, 1977). However, revised proxy estimates of the opening of Southern Ocean gateways (the Tasman Seaway and Drake Passage) and proxies of ocean circulation suggest changes in the ACC occurred gradually, with some changes during the Eocene but the ACC not reaching its full strength until the mid-late Oligocene (Lyle et al., 2007; Scher et al., 2015; Stickley et al., 2004), suggesting that other mechanisms were also fundamentally important. The formation of North Atlantic deep water shows changes broadly around the EOT but with some uncertainty on the timing (Katz et al., 2011; Via & Thomas, 2006). Particularly, recent studies show the deep water formation started occurring in advance of the EOT (Coxall et al., 2018; Hohbein et al., 2012).

Additionally, terrestrial studies show a range of responses during this period of time, although these records are subject to weaker age control than ocean records. On one hand, there is strong evidence for continental aridification in central Asia (Sun & Windley, 2015) and the western interior of North America (Sheldon & Retallack, 2004; Zanazzi et al., 2007), likely due to uplift of mountains and changing moisture supply. In contrast, compiled records of vegetation ecosystem change show few clear changes neither globally over the EOT (Pound & Salzmann, 2017) nor regionally through the Oligocene (Li et al., 2018). This suggests changes to the terrestrial system occurred at a different rate to other climatic changes at the EOT with potentially complex driving mechanisms.

Numerical climate modeling has been very important for understanding the mechanisms behind these changes occurring at the EOT. DeConto and Pollard (2003) provided the first strong evidence that declining atmospheric  $p\text{CO}_2$  was more important for glaciation than changes in ocean gateways, a finding that has been echoed by a number of studies since (Huber & Nof, 2006; Ladant et al., 2014; Sijp et al., 2011). However, processes and feedback within the Earth system could mean that even a small change in the ocean gateways around this time could have an effect on climate in other ways (Elsworth et al., 2017). Meanwhile, Gasson et al. (2014) showed that the  $p\text{CO}_2$  threshold to induce glaciation is highly model dependent, while small changes in the setup and paleogeography of a single model can affect the modeled climate (Kennedy et al., 2015) and ocean circulation (Baatsen, von der Heydt, Kliphuis, et al., 2018). Goldner et al. (2014) showed that the Earth system may respond in a similar way to different forcings (in that case opening of the Southern Ocean gateways and formation of the AIS) and this could have impacts for how proxy records should be interpreted.

These modeling studies offered useful findings but cannot be taken without scrutiny. It is not uncommon to use idealized simulations to test paleogeographic sensitivity to specific changes, such as open and closed ocean gateways to isolate their impact on climate (Hill et al., 2013; Huber & Nof, 2006). In these cases, paleogeography outside the immediate region of interest is possibly assumed to have little effect on the climate, but this may not be the case. Currently, no studies have assessed the sensitivity in modeled climate that can arise from a range of dissimilar global paleogeographies in a general circulation model (GCM). Here we aim to provide an estimate of the degree of uncertainty that would be introduced to climate model simulations if subtle changes were made to the global paleogeography and other boundary conditions.

Another typical issue faced in paleoclimate modeling is insufficient spin-up. Spin-up is the period of time over, which a model simulation adjusts to the imposed boundary conditions before the climate of the model is calculated. Ideally, this should be longer than the adjustment time of the slowest responding component of the system, usually the deep ocean. In general, the further the model initial conditions are from the true equilibrium state for a given climate variable, the longer the required spin-up. However, computational constraints mean that often the spin-up period is stopped before the model reaches equilibrium (which can be defined in terms of net radiative imbalance or deep ocean temperature and/or salinity drift over a given period of time). Furthermore, the large number of potential model parameters (e.g., paleogeography, greenhouse gas levels, and orbital configurations) can require a number of boundary condition sets to be used, further limiting computational resources available for spin-up. As a result, usually, a compromise must be made between sampling the parameter space and the length of the spin-up (e.g., Lunt et al., 2016, used a large ensemble of simulations but a relatively short spin-up). This insufficient spin-up could produce misleading results.

While many proxy data studies include error bars or other measures of uncertainty in their results, often, the meaning of uncertainty in model simulations is not clearly defined. Models will produce different answers with different boundary conditions, and different models (or identical models with different internal parameters) will produce different answers with the same boundary conditions. Furthermore, differing periods of spin-up will produce different results from the same model with the same boundary conditions. Here efforts are made to understand the uncertainty due to boundary conditions and the uncertainty due to length of spin-up in the response to three idealized model forcings at the EOT: Antarctic glaciation, paleogeographic change, and atmospheric  $p\text{CO}_2$  reduction. We focus on three research questions: What are the impacts and relative importance of these three forcings on global and Antarctic climate at the EOT? For each of these forcings, what is the range of uncertainty in the magnitude of certain modeled variables resulting from differences in the model boundary conditions? How does the simulation spin-up time affect the modeled impacts associated with these forcings?

The methods described in section 2 outline the model, the ensemble of simulations, and the statistical methods used. Results addressing the research questions follow in section 3. Discussions, including a model-data comparison, are in section 4 and conclusions in section 5.

## 2. Methods

### 2.1. Model

All simulations carried out here used the fully coupled GCM HadCM3BL-M2.1aE (Valdes et al., 2017). This is a specific setup used at the University of Bristol of the unified model, HadCM3L, created by the UK Met Office. The model runs at  $2.5 \times 3.75^\circ$  horizontal resolution in both the ocean and atmosphere, with 19 vertical levels in the atmosphere and 20 in the ocean. Included in the model are representations of sea ice and land surface processes, and it is coupled with the land surface scheme MOSES and the vegetation model TRIFFID. Ice sheets in the model are not interactive, meaning they cannot melt or expand. They are prescribed by raising the topography, changing the TRIFFID land fraction to *ice* (which affects the surface albedo and roughness length, for example) and covering it with a layer of snow (which is applied as a thick layer so that it does not completely melt).

### 2.2. Boundary Condition Ensemble

To investigate the intramodel uncertainty due to boundary conditions, results were used from a collaboration between the University of Bristol and Getech Group Plc., modeling multiple paleogeographic reconstructions with a consistent but relatively short spin-up procedure (Lunt et al., 2016). Rather than modeling an idealized change in paleogeography, here a range of realistic paleogeographic uncertainty is tested based on the assumption that the Priabonian, Rupelian, and Chattian geologic stage reconstructions (shown in Figures S1a–S1c in the supporting information) are all reasonable representations of the paleogeography around the time of the EOT or early Oligocene, given dating uncertainties in the reconstructions. All of these reconstructions are qualitatively similar, in that they all have the same ocean gateways and similar continental positions, but they are subtly different globally. Along with different paleogeographic boundary conditions, simulations also vary in terms of their  $p\text{CO}_2$  level (either 1,120 or 560 ppmv) and AIS state (either Antarctica is ice free, has an East AIS, or has a full AIS covering the whole continent). In Kennedy et al. (2015) it was noted that some simulations had inconsistencies in their island definition (required for the HadCM3BL ocean barotropic solver). Here, we have corrected those simulations and ensured that all of the simulations in this analysis have consistent islands.

In total this gives a boundary condition model ensemble of 15 simulations, detailed in Table 1. Each simulation is run with a consistent four-phase spin-up procedure lasting 1,422 years, outlined in Lunt et al. (2016). Briefly summarized, the models are initialized for each paleogeography with preindustrial  $p\text{CO}_2$  levels, uniform vegetation, and a zonally symmetric sea surface temperature (SST) distribution. The ocean is initially stationary with a constant salinity of 35 ppt and an idealized zonal mean temperature structure that is a cosine function of latitude and depth (Lunt et al., 2016). The spin-up procedure then involves three initial phases totaling 422 years, in which the  $p\text{CO}_2$  and ozone levels, lakes, ice sheets, and (interactive) vegetation are adjusted, before there is a final 1,000 year phase of spin-up in which the model and boundary conditions stay constant. According to Lunt et al. (2016) by the end

**Table 1**  
*Model Simulation Details*

Epoch	Age (Ma)	Stage (paleogeog.)	$p\text{CO}_2$ (ppmv)	AIS state	Drake Passage	Total spin-up (years)	Simulation			
Oligocene	28.1–23.0	Chattian	1,120	No ice	Open	1,422	TDLUY			
				EAIS			TDLUX			
				Full AIS			TDZSE			
	33.9–28.1	Rupelian	1,120	560	No ice	Open	1,422	TDLUU		
					EAIS			TDLUQ		
					Full AIS			TDWQF		
				1,120	560	No ice	Open	1,422	TDLUV	
						EAIS			TDLUW	
						Full AIS			TDLUT	
		Rupelian (alternate paleogeog.)	840	Full AIS	Open	6,921	TECQS			
							Closed	13,557	TECQT	
								10,385	TECQU	
				560	Full AIS	Open	6,493	TECQV		
								Closed	9,350	TECQN
										TECQO
Eocene	38.0–33.9	Priabonian	1,120	No ice	Open	1,422	TECQP			
				EAIS			TECQQ			
				Full AIS			TDZSC			
			560	No ice	Open	1,422	TDZSD			
				EAIS			TDWQK <sup>a</sup>			
				Full AIS			TDWQV <sup>a</sup>			

Note. The *Rupelian (alternate paleogeog.)* simulations are those in the spin-up ensemble, and all others are in the boundary condition ensemble.

<sup>a</sup>These simulations are updated versions of the simulations in Kennedy et al. (2015) with corrected islands.

of the fourth phase the properties of the land and surface ocean are *approaching equilibrium*, and as can be seen in Figure 4c of Lunt et al., the short early period of spin-up with preindustrial  $p\text{CO}_2$  levels has little effect on the long-term ocean temperature trends. Some of these simulations will have coincidentally been initialized closer to their equilibrium state and hence will be exhibiting more gradual trends. The range in distances these model simulations are from equilibrium introduces another element of uncertainty to the simulations that is investigated as discussed in section 2.3.

### 2.3. Spin-Up Ensemble

A further set of eight model simulations outlined in Table 1 assess the model uncertainty due to insufficient spin-up. These simulations use an alternative paleogeographic reconstruction of the Rupelian provided by Robertsons, shown in Figure S1d, which has some notable differences to the Getech Rupelian reconstruction in terms of its connection between the Atlantic and Arctic oceans, the Paratethys, and in West Antarctica. These simulations were run independently (Li et al., 2018) to those detailed in section 2.2 and as a result have a different spin-up procedure. The simulations were continuations of older paleoclimate simulations (themselves thousands of years long) that originally were initialized with a preindustrial ocean, with ocean variables extrapolated to fill gaps whenever the paleogeography was changed. These latest continuations of the simulations were run until the volume integrated ocean temperature was nearly stable ( $<0.1$  °C drift over 1,000 years) with negligible top of the atmosphere energy imbalance. Given the length of these simulations, the specific initial conditions are expected to have little effect on the final climate, assuming that there is not bistability in the model (a condition which we are unaware of ever being found in HadCM3BL). Due to the long simulation times, only one paleogeography is used, but there are variations in the AIS (either ice free or full AIS),  $p\text{CO}_2$  level (either 840 or 560 ppmv), and idealized changes in the Drake Passage (either open or closed). 100-year climatologies are taken 1,000 years into the simulation as well as at the end, allowing changes in the mean response due to a more complete spin-up to be assessed.

Because these simulations were run independently to the boundary condition ensemble (i.e., with different forcings and a different spin-up procedure), the two ensembles are not directly compared. Although the forcings are qualitatively similar between the ensembles, they are not identical and so should be thought of as

independent experiments for each part of the analysis. It is still useful, however, to see how the uncertainty varies between the two ensembles.

Unlike the boundary condition uncertainty, the *uncertainty* due to an unequilibrated climate state could be seen as model *error*, with the fully equilibrated model taken as the correct answer that could be achieved if only more time was available. Regardless of how this uncertainty is defined, it is still useful to have an approximate measure of how much a lack of spin-up might affect results.

#### 2.4. Analysis and Assessment of Uncertainty

Analysis of the climate forcing mechanisms (Antarctic ice growth, paleogeographic change, and  $p\text{CO}_2$  reduction) compares all pairs of simulations that keep all boundary conditions the same, except for those relating to the forcing being analyzed. For example, if analyzing the effect of  $p\text{CO}_2$  reduction, pairs of simulations that have the same paleogeography and ice sheet state but different  $p\text{CO}_2$  levels will be compared. The Antarctic ice growth forcing refers to the difference between either the EAIS or full AIS simulations and the ice free simulations. Both EAIS and full AIS simulations were counted as *glaciated*, given the uncertainty in the state of the ice sheet present after the EOT (Bohaty et al., 2012). Finally, the paleogeographic change forcing is taken as the difference between any stage and the stage immediately previous, that is, Rupelian-Priabonian or Chattian-Rupelian. The Chattian and Priabonian are not differenced as this would constitute an unrealistically large jump in paleogeography.

It should be noted that in the real Earth system, the growth of the AIS is actually a response to a forcing, and not a forcing in its own right. However, the presence of the ice would likely have a large impact on the Earth's climate that is important to understand when interpreting proxy records (i.e., it is a major feedback mechanism; Goldner et al., 2014). The GCM used here has no interactive ice sheet module, meaning the ice sheet must be prescribed and so in model terms Antarctic ice growth acts like a forcing. It is worth bearing in mind the distinction between model and true Earth system forcing when interpreting these results and those of similar studies (Goldner et al., 2014; Kennedy et al., 2015).

The boundary conditions used in any model simulation will always contain some uncertainty. For example, the paleogeographies will contain uncertainties primarily relating to proxy reconstructions used in their construction (e.g., dating uncertainty and calibration error). Furthermore, there is also uncertainty in the assumptions driving different plate models used to reconstruct paleogeographies, such as the reference frame, plate spreading rates, and mantle convection rates (Baatsen et al., 2016). Uncertainties in ice sheet volume reconstructions and atmospheric  $p\text{CO}_2$  levels will also come about through proxy dating and assumptions in calibration calculations (Bohaty et al., 2012; Pearson et al., 2009).

In the model, each of these potential sets of boundary conditions will have its own realization in climate space, and together, all of the potential realized climates can be thought of in statistical terms as the *population*. Assuming that the climate predicted by the model is reasonably linear in response to each of the changes in boundary conditions, all of these idealized simulations making up the population should be expected to fall on a normal distribution, that is, holding with central limit theorem. In reality, it is not practical to model every single possible combination of boundary conditions, so instead a limited number of simulations must act as a representative sample of the population. This sample of model simulations can be used to infer the underlying nature of the population from which they are drawn, in terms of the mean and spread of the normal distribution if the sample size is large enough (typically greater than 30).

The mean response,  $\bar{x}$ , of all pairs of simulations for each forcing (where  $n$  is the number of pairs) can be calculated. An estimate of the uncertainty of this sampled mean value,  $\bar{x}$ , compared to the true population mean can then be calculated based upon a  $t$ -distribution. A  $t$ -distribution, which has a similar shape to a normal distribution but is slightly wider, is used here because the number of simulations is less than 30 (Burt et al., 2009). This uncertainty,  $U_t$ , is defined as such:

$$U_t = \frac{s}{\sqrt{n}} \times t_{(\alpha/2, n-1)} \quad (1)$$

where  $s$  is the standard deviation of the sample simulation pairs and  $t$  is the value that defines the width of the  $t$ -distribution and its confidence intervals, taken from a look-up table for the desired confidence level

$(1 - \alpha)$  with  $n - 1$  degrees of freedom (Burt et al., 2009). Throughout this work, we use  $\pm 95\%$  confidence intervals as standard.

Theoretically, this kind of analysis could be done for many parameters or combination of parameters in the model to better describe the model climate space. However, it is important to consider when the assumptions made in this analysis do not hold. A key assumption is the linearity of the climate response to the boundary condition uncertainty, which would ensure that the theoretical climate space is normally distributed. This may not be the case for aspects of the model where there are tipping points and different modal states (Lucarini & Bódai, 2017; von der Heydt et al., 2016). If, for example, deep water formation has an *off* and an *on* state in a specific ocean basin, the potential climate space could be expected to be bimodal, with potentially nonlinear transitions between the states. In the case of a bimodal or multimodal climate space, the uncertainty range provided by this analysis will likely be an underestimation, and the concept of  $\bar{x}$  will not be valid as the true population should be described in terms of modal, not mean, states. In the case of variables that are particularly nonlinear, it could be more appropriate to describe the uncertainty in terms of the range in the model responses. This produces a much larger uncertainty margin but makes no assumptions about the linearity of the response.

Although in certain cases (such as deep water formation) it seems likely that the climate response is nonlinear, because we only have a limited sample size available, it is not possible to definitively state whether the climate state is linear with a normally distributed response to the various perturbations or not. It is important to bear in mind the assumptions behind the methods used and where potential nonlinearities could be causing an underestimation of the uncertainty. Additionally, given that only three boundary condition parameters are partially sampled, this uncertainty analysis is by no means exhaustive. For example, the uncertainty due to differences in the paleogeographic boundary conditions could be found to be much greater if simulations with random variation around the reconstructions used here, or with entirely different paleogeographic reconstructions (Baatsen et al., 2016), were also carried out. The potential impacts of nonlinearity on this analysis are discussed further in section 4.1.

### 3. Results

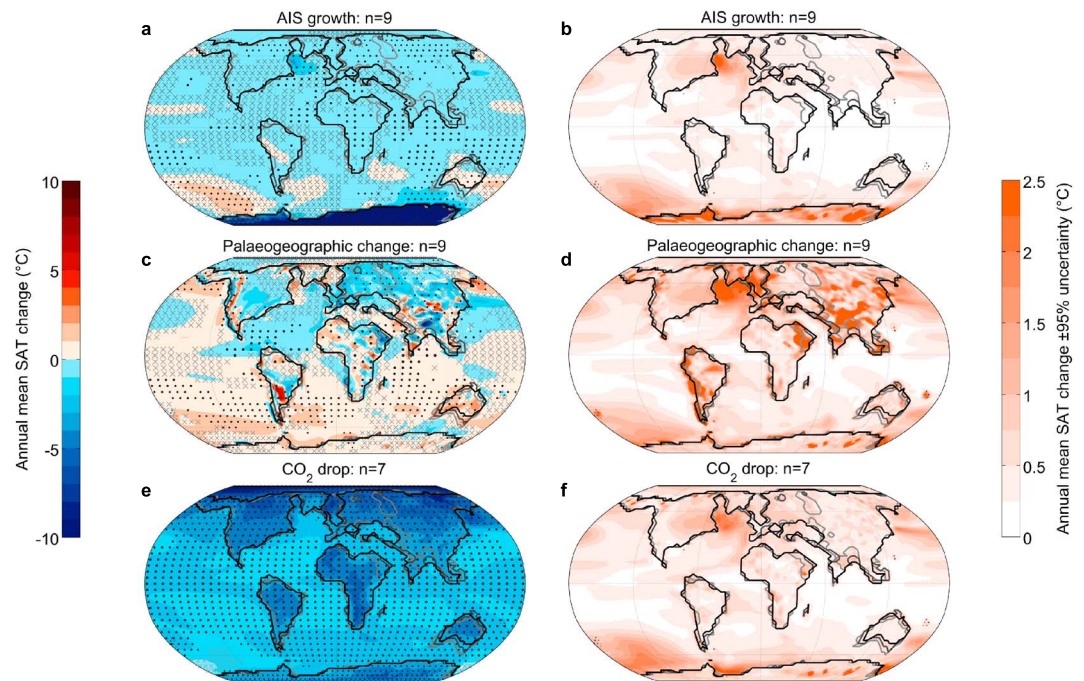
#### 3.1. Boundary Condition Sensitivity

The effect on mean annual surface air temperature (SAT) of Antarctic ice growth, paleogeographic change, and  $p\text{CO}_2$  reduction for the multiple simulations with varying boundary conditions is shown in Figure 1. The mean change for the three forcing mechanisms is shown in Figures 1a, 1c, and 1e, and the  $\pm 95\%$   $U_t$  of each mean response is shown in Figures 1b, 1d, and 1f.

Figure 2 shows a range of climatic variables (in terms of both absolute values from individual simulations and the mean changes for each forcing) for all the simulations used in this study. The climatic variables assessed here are annual mean Drake Passage throughflow (Figures 2a and 2b), annual mean SST in the Southern Ocean (south of  $60^\circ\text{S}$ ; Figures 2c and 2d), mean annual precipitation over Antarctic land areas (Figures 2e and 2f), and summer (December-January-February) mean SAT approximately over the Gamburtsev mountains in Antarctica (Figures 2g and 2h). The absolute values from all simulations for these variables and several others (e.g., global mean SAT) are available for reference in the supplementary spreadsheet.

Likewise, Figure 3 shows absolute values and mean changes for variables relating to ocean overturning, specifically, the maximum depth of the mixed layer in the Southern Ocean (Figures 3a and 3b) and North Atlantic (Figures 3c and 3d) and maximum overturning strength in the Southern (Figures 3e and 3f) and Northern (Figures 3g and 3h) Hemispheres. Unlike the results of Hutchinson et al. (2018), none of the simulations here suggest overturning in the North Pacific. The simulations here generally have lower sea surface salinities and a much shallower mixed layer depth (MLD) in the Pacific than in the Atlantic, suggesting that overturning is absent from this region (figures of North Pacific MLD and salinity not shown). Given that oceanic overturning may be more susceptible to nonlinear behavior, the range of all responses for these simulations is also shown, as  $U_t$  could potentially be underestimating uncertainty.

Antarctic ice growth causes a strong cooling of the SAT over Antarctica due to the imposed change in albedo and topographic height, but generally a low cooling response globally that is often less than the uncertainty



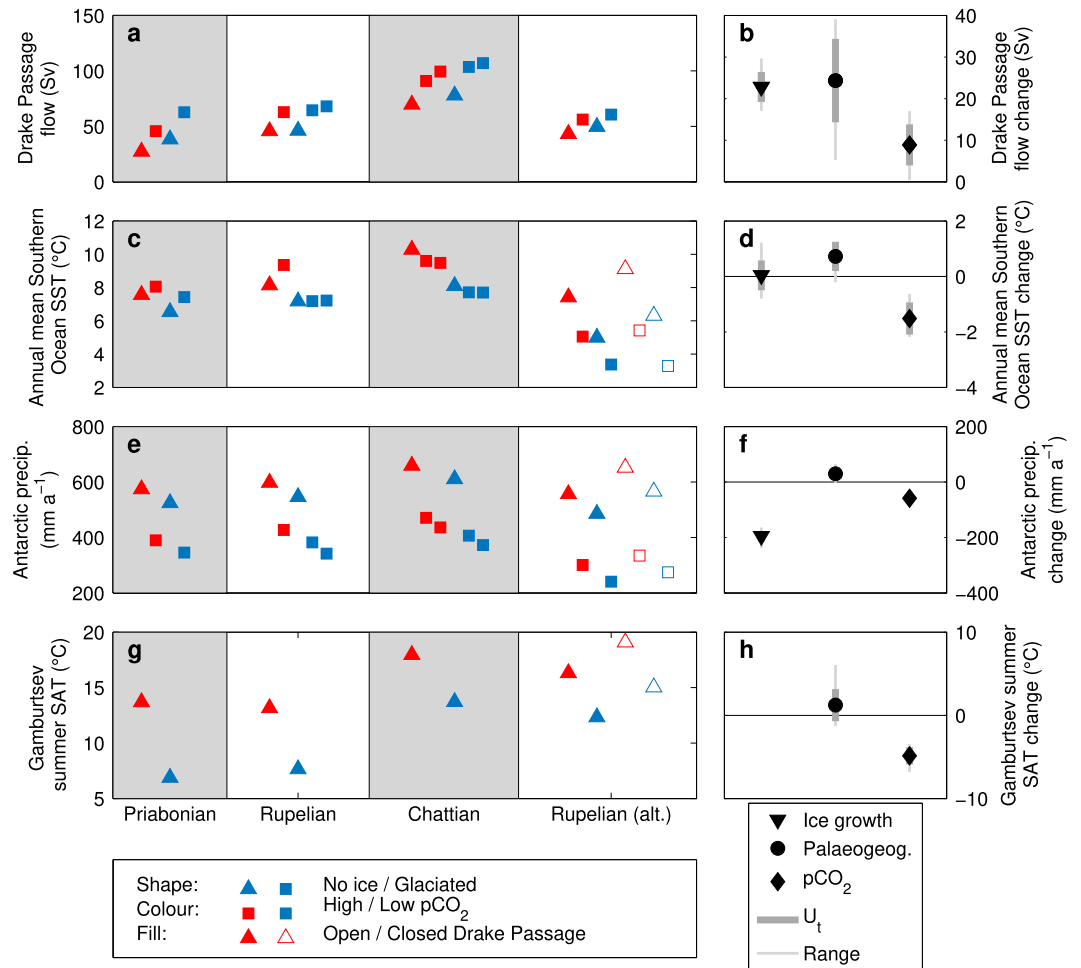
**Figure 1.** The modeled annual mean surface air temperature (SAT) response to various forcings and its  $\pm 95\%$  uncertainty due to boundary condition differences: (a) mean response to Antarctic ice growth, (b) uncertainty in (a), (c) mean response to palaeogeographic change, (d) uncertainty in (c), (e) mean response to  $p\text{CO}_2$  reduction (1,120–560 ppmv), and (f) uncertainty in (e). The black stippling indicates where all  $n$  pairs of simulations agree on the direction of change, and the grey crosses indicate where the mean change is less than the uncertainty. The land-sea masks are shown for each stage (Priabonian = palest; Chattian = darkest).

(i.e., many regions have grey hatching; Figure 1a). Globally, the resultant change in planetary albedo accounts for approximately half of the total temperature change based upon a simple energy budget calculation. The only areas with mean changes greater than  $\pm 1^\circ\text{C}$  are in the Southern Ocean (where there is warming) and North Atlantic (where there is cooling) that are associated with changes in deep water formation. These same areas, however, have higher uncertainties of up to  $\pm 2^\circ\text{C}$ , which could potentially be an underestimation if there is nonlinear behavior in the system here. There is relatively limited agreement within the models on the direction of change, with black stippling over 27% of the globe. The main regions where there is agreement between the models (i.e., there is stippling) are in the North Atlantic, Africa, and in the Indian Ocean, South Pacific, and Southern Ocean around Australia.

In terms of the other variables shown in Figures 2 and 3, Antarctic ice growth causes an increase in the strength of the ACC through the Drake Passage for all pairs of simulations. The average increase is  $22.8 \pm 3.6$  Sv (Figure 2b). This increase is related to increased pressure gradients around  $60^\circ\text{S}$  (not shown) that intensify the westerly winds and ACC flow in the Southern Ocean (Kennedy et al., 2015). The effect on mean Southern Ocean SST varies, with the earlier stages (Priabonian and Rupelian) showing a warming in response to ice growth and the Chattian showing cooling in response to ice growth (Figure 2c), resulting in a mean change close to zero (Figure 2d). Precipitation decreases significantly over Antarctica with ice growth, due to reduced transfer of moisture over the continent from its increase in topographic height and cooling.

The effect of ice growth on overturning is pronounced in the Southern Hemisphere, with both maximum Southern Ocean MLD and overturning strength increasing, even when considering the larger uncertainty from the range (Figures 3b and 3f, respectively). This is likely due to the resulting heat loss from the presence of the ice sheet demanding greater meridional heat transport, with increased temperature and pressure gradients in the region also resulting in strengthened westerly winds, more vigorous ocean circulation and colder waters at the Antarctic margin contributing to more intense deep water formation. Changes in the

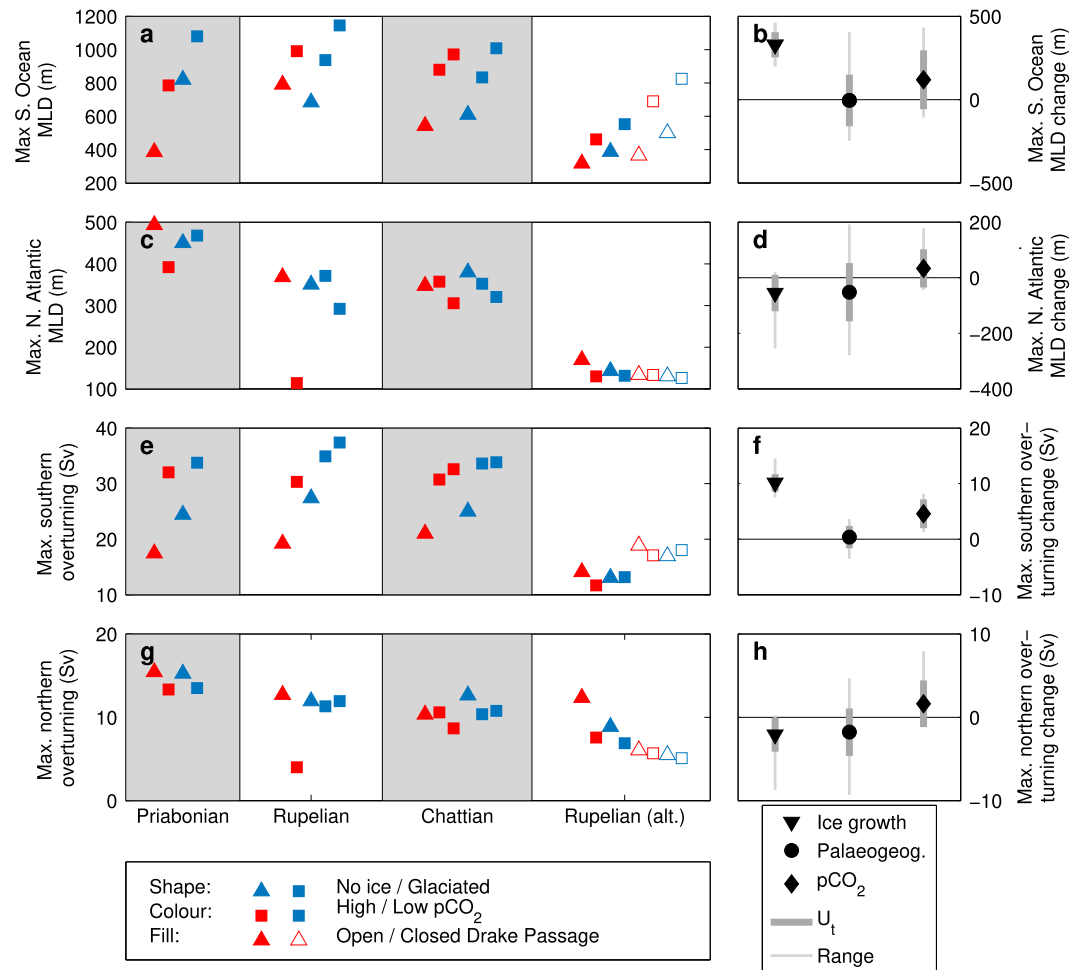




**Figure 2.** Absolute values for individual model simulations and mean changes in response to forcings of four climatic/oceanic variables: (a) annual mean Drake Passage throughflow (Sv), (b) mean changes in (a) due to forcings, (c) annual mean Southern Ocean sea surface temperature (SST; °C), (d) mean changes in (c) due to forcings, (e) annual Antarctic precipitation (mm/a), (f) mean changes in (e) due to forcings, (g) summer mean surface air temperature (SAT; °C) over the Gamburtsev mountains, and (h) mean changes in (g) due to forcings. The Priabonian, Rupelian and Chattian columns show simulations from the boundary condition ensemble; the Rupelian (alt.) column shows simulations from the spin-up ensemble. The mean changes in (b), (d), (f) and (h) are only for the boundary condition ensemble. The shape, color, and fill of markers in (a), (c), (e), and (g) detail the model simulation, while the grey error bars in (b), (d), (f), and (h) highlight  $\pm 95\%$   $U_t$  and the range in responses to forcings. The absolute summer SAT over the Gamburtsev mountains (which are highlighted in Figure S1) is only shown for ice free simulations.

Northern Hemisphere MLD and overturning are more varied, with  $U_t$  and ranges that cover zero change (Figures 3d and 3h).

Paleogeographic changes have a hemispheric mean annual SAT response with greater uncertainty than the Antarctic ice growth response (Figures 1c and 1d). Generally, paleogeographic changes result in an intensification of Southern Ocean circulation (by  $24.3 \pm 10.0$  Sv; Figure 2b) and a warming of the Southern Hemisphere (in the Southern Ocean by  $0.72 \pm 0.53$  °C; Figure 2d) at the expense of cooling in the Northern Hemisphere, potentially due to a different mode of ocean circulation. Once again, there is relatively limited agreement between the models in the mean SAT response, with stippling over 28% of the globe. All simulation pairs agree on this oceanic warming in the Southern Hemisphere over midlatitudes as well as some cooling in the Caribbean and North Atlantic. Over land, patterns of change are more varied and highly dependent on changes in paleoelevation, which can change a lot between reconstructions where there is horizontal movement of mountain ranges and coastlines. The Rocky Mountains, for example, show a



**Figure 3.** As Figure 2 for oceanic overturning variables: (a and b) maximum annual mean mixed layer depth (MLD) in the Southern Ocean (m), (c and d) maximum annual mean MLD in the North Atlantic (m), (e and f) maximum Southern Hemisphere overturning (Sv), and (g and h) maximum Northern Hemisphere overturning (Sv).

temperature change dipole running north-south due to the westward movement of the mountain range with successive stages. There is some agreement on cooling over central Asia and warming over Australia as it moves equatorward. Uncertainties in the modeled temperatures are generally up to  $\pm 1.5$  °C, with some terrestrial regions and areas of deep water formation well exceeding  $\pm 2.5$  °C. There is a subtle increase in precipitation over Antarctica (Figure 2f), and mean summer SAT over the Gamburtsev mountains is greatest for the Chattian, while the Priabonian and Rupelian are more similar (Figure 2g). This is mainly due to elevation changes between the Stages, as can be seen in Figure S1.

None of the ocean overturning metrics in Figure 3 shows a coherent response to paleogeographic change, with the  $U_t$  estimates and range in responses being very large. The largest North Atlantic MLD and Northern Hemisphere overturning values are generally found for the Priabonian simulations (Figures 3c and 3g), with the Rupelian and Chattian simulations being more similar. The Southern Ocean MLD and overturning are similar for all three Stages, with no significant change in response to the forcing. The high  $p\text{CO}_2$ , glaciated Rupelian simulation shows a near-complete shutdown of North Atlantic deep water formation, suggesting potential nonlinearity in the system here. The reasons behind the lack of deep water formation for this specific simulation are not fully understood.

Finally,  $p\text{CO}_2$  reduction has a global cooling effect. Generally, it is greater over land and shows polar amplification, particularly in the Northern Hemisphere (Figure 1e). This is largely associated with increased sea ice coverage in the Northern Hemisphere with reduced  $p\text{CO}_2$ , which reflects more shortwave radiation

across the region (figure not shown). Model agreement is generally good, with stippling over 95% of the globe. The only regions where the model simulation pairs diverge in their direction of response are deep water formation regions or areas of major ocean currents at high latitudes, namely, the North Atlantic, Weddell Sea, Ross Sea/South Pacific, and Kuroshio Current region. These areas also have an elevated uncertainty of  $\pm 2$  °C, compared to other regions that are generally less than  $\pm 1$  °C (Figure 1f). The Ross Sea is the only area where the modeled change is less than the uncertainty (i.e., it is hatched), once again showing that this region is highly sensitive. It is possible that this variability is due to the averaging period over which the climate is calculated, as deep water formation areas can show multidecadal variability even in fully equilibrated simulations (Armstrong et al., 2017).

The other climatic variables suggest that declining  $p\text{CO}_2$  has a small effect of increasing ACC flow through the Drake Passage (by  $8.8 \pm 5.0$  Sv; Figure 2b). Predictably, it causes a cooling of the mean Southern Ocean SST and of the Gamburtsev summer SAT, by  $1.5 \pm 0.6$  and  $4.9 \pm 1.0$  °C, respectively (Figures 2d and 2h), and causes a slight decrease in Antarctic precipitation (Figure 2f). In the ocean, a  $p\text{CO}_2$  reduction modestly increases Southern Hemisphere overturning (Figure 3f), while MLD in both hemispheres and Northern Hemisphere overturning are more variable with large  $U_i$  and ranges.

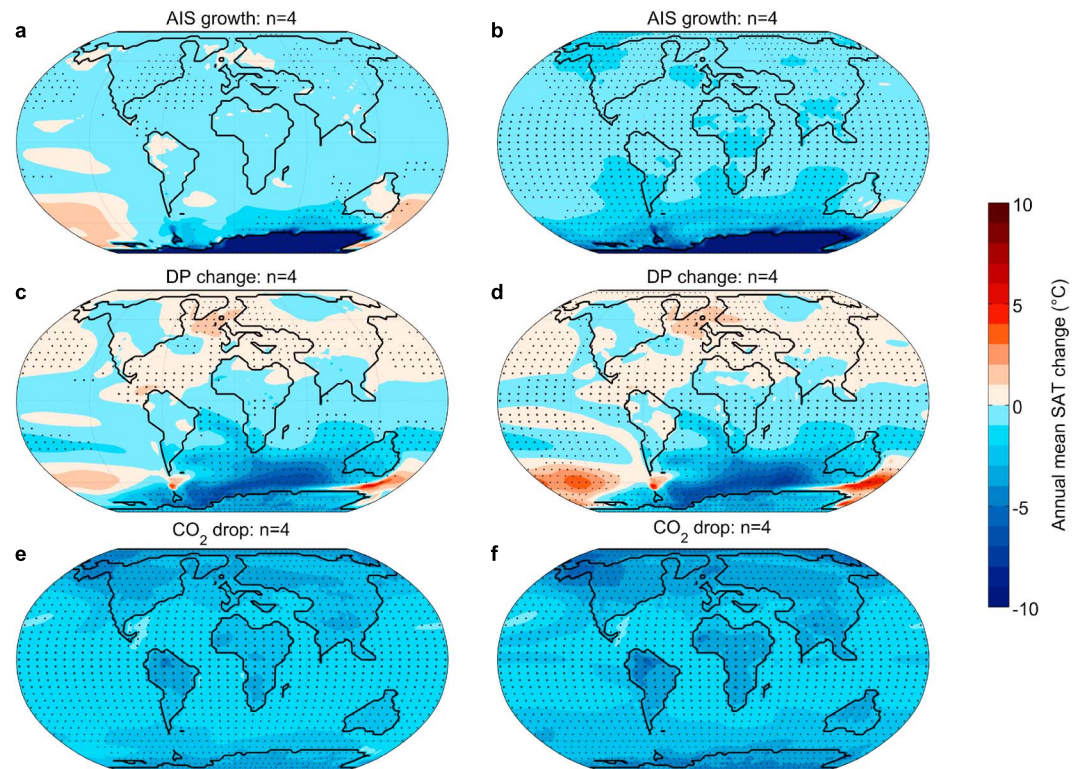
Figure S2 shows the mean SAT response for each individual model pairing averaged in Figure 1, allowing deeper analysis of the results described here. Under the Antarctic ice growth forcing, the Chattian shows the least Southern Ocean warming, while the Priabonian and Rupelian simulation pairs at 1,120 ppmv  $p\text{CO}_2$  show greater Southern Ocean warming. For the paleogeographic change forcing, there is less consistency as to where some simulations warm and others cool in the same areas and by how much. Under the  $p\text{CO}_2$  forcing, the Chattian simulations are the only ones not to show any areas of warming; however, the Priabonian and Rupelian simulations show pockets of warming in different areas where major ocean currents are or deep water formation occurs.

### 3.2. Spin-Up Uncertainty

One possible reason for the differences between individual simulation pairs shown in Figure S2 could be how close the simulations are to reaching equilibrium, requiring further investigation into the effect of spin-up. Certain combinations of boundary conditions could coincidentally be initialized in greater balance between the surface and deep oceans, having further implications for deep water formation and the spatial patterns of SAT. The effect of spin-up on the modeled mean annual SAT response to Antarctic ice growth, opening of the Drake Passage, and  $p\text{CO}_2$  reduction is shown in Figure 4. The increasing prevalence of stippling for all forcings at the end of the spin-up (Figures 4b, 4d, and 4f) shows that model ensemble agreement generally improves once the simulations are fully spun-up. Although these simulations are independent from those in Figures 1 and S2, it is possible that the boundary condition ensemble simulations could also converge with increased spin-up. Besides improved model agreement with increased spin-up, there are also important implications for the spatial patterns of SAT change for some forcings.

The Antarctic ice growth simulation pairs at 1,000 years (Figure 4a) have a similar spatial response to the boundary condition ensemble (Figure 1a) and the results of Kennedy et al. (2015), both of which have qualitatively the same forcing. However, by the end of the spin-up, the area of 1–3 °C warming over the South Pacific sector of the Southern Ocean ceases to exist, instead showing 1–3 °C of cooling, with good model agreement (Figure 4b). This warming midway through the spin-up could be due to a larger imbalance between the surface and deep ocean at this point in the simulation, compared to the equilibrium state. Early in the spin-up, there are major differences in the deep water formation between the ice-free and the glaciated simulations (similar to that described in Kennedy et al., 2015); however, the overturning converges later in the spin-up (figure not shown) causing the change in response to glaciation shown in Figure 4. Additionally, this suggests that the results described in Kennedy et al. (2015) could be an artifact of insufficient spin-up and that under a complete spin-up, Antarctic ice growth causes surface cooling over most regions due its effect on the global net radiation through increased reflectance of solar radiation (not shown).

In contrast to the reversal of the response to Antarctic ice growth in this region with increasing spin-up, with opening the Drake Passage, increased spin-up causes an amplification of the warming shown in

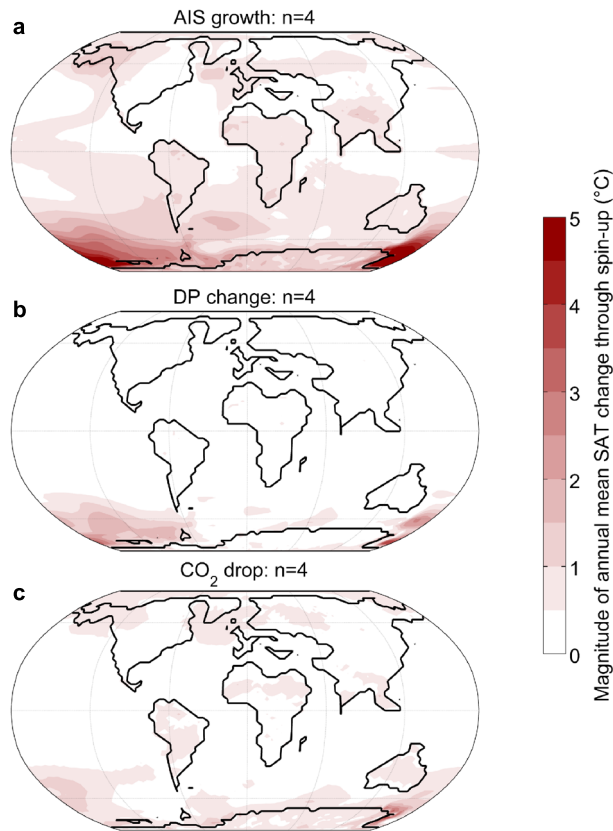


**Figure 4.** The modeled annual mean surface air temperature (SAT) response to forcings midway through the spin-up (after 1,000 years) and at the end of the spin-up (once the model is fully equilibrated). Shown are the responses to Antarctic ice growth (a) midway through and (b) at the end of the spin-up; opening of the Drake Passage (c) midway through and (d) at the end of the spin-up, and  $p\text{CO}_2$  reduction from 840 to 560 ppmv (e) midway through and (f) at the end of the spin-up. The black stippling indicates where all  $n$  pairs of simulations agree on the direction of change.

the high-latitude South Pacific early in the spin-up. There is 1–2 °C warming at 1,000 years with poor model agreement (Figure 4c), increasing to 2–4 °C warming with good model agreement at the end of the spin-up (Figure 4d). In this case, there is a zonal redistribution of heat in the Southern Ocean when the gateway is opened with a shift in deep water formation region from the Weddell to the Ross Sea, which takes some time to reach its full magnitude. All other regions outside the high Southern Hemisphere show very little change (<1 °C) between the middle and end of the spin-up, with ensemble agreement improving globally.

The effect of  $p\text{CO}_2$  reduction changes the least with increasing spin-up (Figures 4e and 4f) and, like the boundary condition ensemble (Figure 1e), already has good model agreement early in the simulation. This is possibly due to the global nature of this forcing requiring less reorganization of oceanic flow from the initial state. The model agreement in the direction of change improves in the Ross Sea (the only region where there was disagreement early in the spin-up). The global mean SAT cooling in response to the  $p\text{CO}_2$  reduction increases from 2.06 °C midway through the spin-up to 2.38 °C at the end of the simulation (global mean SAT values not shown).

For all forcings, outside of deep water formation regions, areas where there is model agreement in the mean SAT early in the spin-up tend to persist once fully equilibrated. However, in deep water formation regions it remains challenging to generalize how the model response changes with increased spin-up, with different responses showing stability, intensification, and reversal depending on the region and forcing. This again highlights the potential nonlinear nature of the behavior of the model in these regions. The magnitude of changes between the midpoints and end-points of the spin-up for each of these forcings is shown in Figure 5, highlighting that the uncertainty due to spin-up is greatest for the Antarctic ice growth forcing, least for the  $p\text{CO}_2$  reduction and most localized for the Drake Passage



**Figure 5.** The magnitude of mean annual surface air temperature (SAT) change between the middle and end of the spin-up for the three forcings shown in Figure 4.

opening. Like the boundary condition uncertainty plots shown in Figure 1, the areas with the greatest spin-up uncertainty also tend to be found at higher latitudes.

The *Rupelian (alt.)* columns of Figures 2 and 3 show the absolute values of the climate variables for these simulations once they are fully spun-up. These simulations have similar absolute Drake Passage throughflow values compared to the Priabonian and Rupelian simulations from the boundary condition ensemble (Figure 2a). The mean responses to Antarctic ice growth and  $p\text{CO}_2$  reduction are slightly lower than the boundary condition ensemble, causing an increase in flow of 11.9 and 5.4 Sv, respectively (mean change from spin-up ensemble not shown). Absolute annual mean Southern Ocean SSTs (Figure 2c) are generally lower for these simulations than those in the boundary condition ensemble due to having reduced Southern Ocean overturning, differences in the paleogeographic reconstructions (with the Robertsons reconstruction having more ocean area at higher latitude) and due to being better equilibrated (with the boundary condition ensemble simulations generally showing cooling trends at the end of their spin-up periods; not shown). Also, due to being closer to equilibrium, there is a slightly stronger cooling in this ensemble, of 2.7 and 2.3 °C in response to Antarctic ice growth and  $p\text{CO}_2$  reduction, respectively (not shown). Finally, Antarctic precipitation and Gamburgtsev summer mean SAT have similar absolute values and responses to the boundary condition ensemble (Figures 2e and 2g, respectively), both decreasing with glaciation and reduction of  $p\text{CO}_2$ .

Absolute maximum MLD and overturning values are generally less for the fully equilibrated simulations compared to those in the boundary condition ensemble (Figures 3a, 3c, 3e, and 3g). Additionally, North Atlantic MLD and overturning in both hemispheres are relatively unaffected by either Antarctic ice growth or  $p\text{CO}_2$  reduction. For both forcings, North Atlantic MLD shoals by <20 m, southern overturning reduces by <1 Sv and northern overturning reduces by <2 Sv (mean changes from spin-up ensemble not shown). Southern Ocean MLD shows the greatest changes in response to the forcings, deepening by 240 m in response to Antarctic ice growth and 108 m in response to  $p\text{CO}_2$  reduction (not shown). For certain model simulations, the MLD changes significantly with increased spin-up, for example, in the ice free setups it takes some time to deepen in the Ross Sea (i.e., for overturning to increase) and to converge with the equivalent glaciated simulation (figure not shown). A similar increase in MLD is found for the ice-free simulations in the boundary condition ensemble that exhibit deep-water formation in the Ross Sea (figure not shown). Although the two ensembles used here were run independently with different spin-up procedures and are not directly comparable, it is possible that some of the spread in responses for these variables in the boundary condition ensemble will reduce with increased spin-up as the simulations approach steadier equilibrium states.

Similar to Figures 1b, 1d, and 1f, the  $U_t$  for each of the forcings in the spin-up ensemble can be calculated, shown for the middle and end of the spin-up in Figure S3. For this ensemble,  $U_t$  is generally large as it is calculated based upon a very small sample size of only four simulation pairs for each forcing, which increases the value of  $t$  in equation (1). With the exception of the North Atlantic, generally,  $U_t$  decreases with spin-up for all forcings and in some cases the reduction is significant, suggesting that the lack of spin-up can be a major contributor to the magnitude of  $U_t$ . However, the fact that a similar spatial pattern persists between the middle and end of the spin-up suggests that not all of the  $U_t$  in the responses is due to lack of spin-up and the uncertainties in boundary conditions are still important to consider. Work is currently ongoing to run the boundary condition ensemble simulations further to properly quantify how much of the  $U_t$  in each response in Figure 1 is due to the lack of spin-up.

## 4. Discussion

### 4.1. Generalizing Spatial Patterns of Model Uncertainty ( $U_t$ )

Although the forcings assessed here in both simulation ensembles are very different in terms of their regional versus global extent and how they impact on the climate system, they have similar characteristics in their spatial patterns of SAT  $U_t$ . The exact magnitude of the  $U_t$  depends on the forcing, how close the model is to equilibrium and the ensemble size, but these generalized spatial patterns appear robust for the changes typical of the EOT. Low and midlatitude oceans are the most consistent regions with low  $U_t$ , typically  $<1$  °C (i.e., we have higher confidence in modeled results in these regions). Most continental regions have a moderate degree of  $U_t$ , typically 1–2 °C (i.e., we have medium confidence in modeled results in these regions). Finally, high-latitude oceans, particularly deep water formation regions, and mountainous regions are the most sensitive with high  $U_t$ , typically  $>2$  °C (i.e., we have low confidence in modeled results in these regions).

These high-latitude ocean regions are the most uncertain as they generally have high variability (e.g., on inter-annual, decadal, and longer timescales) and are sensitive to small differences in the forcings due to multiple feedback mechanisms. In the case of the Ross Sea and South Pacific sector of the Southern Ocean, subtle changes in ocean gateways of only one model grid cell can significantly alter the flow of the ACC and the subsequent mixing between Indian and Pacific Ocean basins in the Southern Ocean (Kennedy et al., 2015). The resulting changes in salinity can influence the strength of deep water formation in the region which then has further impacts on oceanic heat transfer, particularly if model spin-up is short and it is further from equilibrium. Finally, sea ice feedback with the atmosphere can amplify the changes so the resulting  $U_t$  is large. Likewise, changes in ice sheet state can affect the zonal winds over the Southern Ocean and  $p\text{CO}_2$  reduction can affect the heat transfer, interacting further with this network of feedbacks. In the Northern Hemisphere, although there are no differences in terms of ice sheets, differences in the runoff basins in each paleogeography may be having an effect on the salinity of the North Atlantic and subsequently affecting the surface density and deep water formation. It is also important to note that the Getech reconstructions have an isolated Arctic Ocean, whereas the Robertsons Rupelian reconstruction has a connection the North Atlantic and the Arctic (see Figure S1). This results in a much lower salinity and reduced overturning (shown in Figure 3) in the North Atlantic in the simulations with the Robertsons reconstruction.

Because of the multiple processes and feedbacks involved in deep water formation, this aspect of the Earth system response is likely to be particularly nonlinear and could be bistable (Baatsen, von der Heydt, Kliphuis, et al., 2018). In that regard, the assumptions behind the  $t$ -distribution method used here may be invalid for variables directly affected by deep water formation. It is important to consider the evidence for and implications of nonlinearity. If one ocean state can be definitively ruled out by other evidence (i.e., if there was or was not deep water formation without doubt), then all simulations that exhibit the incorrect state should be excluded from the analysis. The remaining simulations (showing the correct state) however could be expected to behave linearly, and so the methods used here could be appropriate for that selective subsample.

If there is no strong evidence to confirm if either one state or another is correct and no simulations can be ruled out, then true quantification of the uncertainty would require a different method, and portraying the mean response becomes inappropriate. The method used here would result in an underestimated uncertainty around a mean that is between the two modal states or subsamples. A possible alternative method is to show the range between the maximum and minimum responses, as shown in Figure S4 for the annual mean SAT of the boundary condition ensemble. This simplified method makes no assumptions about linearity but produces a much larger estimation of the uncertainty. It is important to note that the spatial patterns of the magnitude of range match those of  $U_t$ , suggesting that the generalized regions of low, medium, or high confidence discussed above are robust.

No statistical test will be without assumptions, which are invalidated in certain circumstances, and we argue that the  $U_t$  method is still appropriate to use for many variables in many regions. It is not possible to exhaustively validate the individual responses for all variables on the required range of spatial and temporal scales, so it is important to be aware of the potential issues that nonlinearity could cause and that the uncertainties reported here may be an underestimation.

Although it may not be possible for all studies to use an ensemble like this to quantify the  $U_t$ , consideration should be given to the broad spatial  $U_t$  (confidence) patterns. When interpreting model output for deep-time periods it is worth considering how much the model simulations could vary with only minor differences in the boundary conditions or increased spin-up, particularly if there is interest in these highly uncertain regions. In this regard, the results shown in Kennedy et al. (2015) (their Figure 3) should not be seen as surprising; the large area of warming in the Southern Ocean identified in that study is shown here to be the artifact of a sensitive region of the model, exacerbated by a lack of spin-up. Although outside the scope of this study it is also worth considering intermodel uncertainty. Different model constructs will have a dissimilar array of model physics, parameterizations, and resolutions that can also add further uncertainty, which could compound the uncertainty derived from the model boundary conditions.

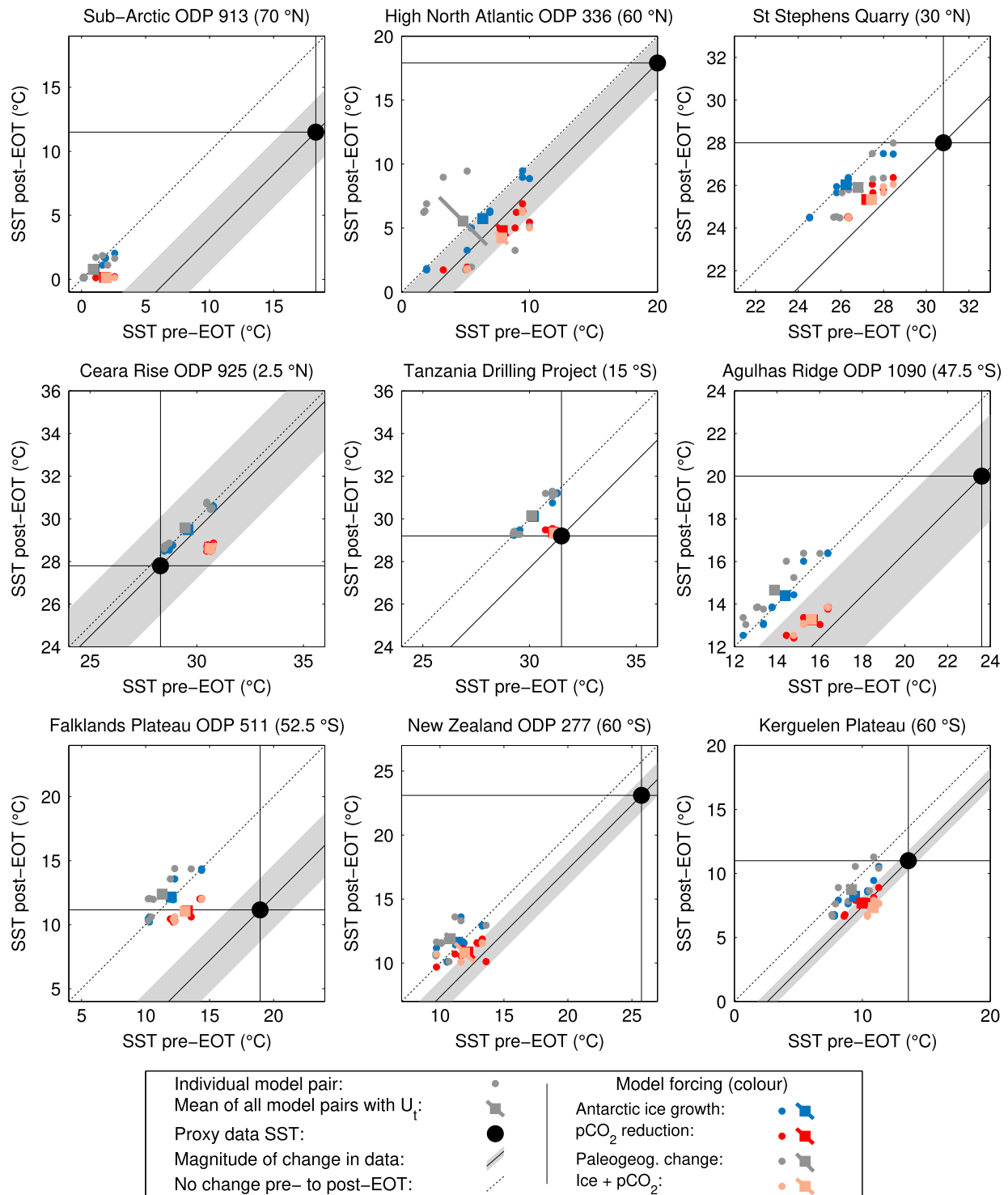
#### 4.2. Comparison to Data

In addressing the first research question into the impacts and relative importance of these forcings on global and Antarctic climate at the EOT, it is worthwhile comparing the GCM simulations with proxy records for the period. To this end, all of the model simulation pairs that have gone into the boundary condition uncertainty analysis are compared with SST records from nine locations for the EOT (Bohaty et al., 2012; Liu et al., 2009; Pearson et al., 2007; Wade et al., 2012). In addition to the three main forcings described so far, the combined effect of ice growth and  $p\text{CO}_2$  reduction (for which  $n = 5$ ) is also compared to the data. Figure 6 plots the pre-EOT proxy SST against the post-EOT proxy SST at each location. Plotting this way shows if the individual model simulations have a general warm or cold bias (a cold bias will shift the point below and left of the proxy data, whereas a warm bias will shift the point above and right of the proxy data), as well as if the forcing produces the correct magnitude of change (which should fall on or close to the solid diagonal line if the magnitude is the same as the proxy records, even if there is a warm or cold bias). If simulations fall above and to the left of the zero change (dotted) line, it suggests that they are simulating the wrong direction of change (i.e., warming over the EOT) and are performing particularly poorly.

The proxy records of change across the EOT used here are taken from papers or their supporting information with uncertainty in the change plotted where available. There are some notable uncertainties in these records. In some cases, multiple proxies have been averaged; for example, the St Stephens Quarry record (Wade et al., 2012) is an average of  $\text{TEX}_{86}$  and Mg/Ca records, while the Kerguelen Plateau data use absolute values from Petersen and Schrag (2015), with the change calculated using Mg/Ca and  $\delta^{18}\text{O}$  from Bohaty et al. (2012). Additionally, the time definition of pre-EOT and post-EOT varies slightly by site. To account for some uncertainty in the proxy location, the average ocean temperature is taken over a 3 by 3 grid cell area in longitude and latitude, which encompasses both the published paleolocations and the possible range of locations computed by back rotation of the Getech plate model. Temperature is averaged in the top 55 m of the ocean (the top five ocean layers), except for the Kerguelen Plateau, which is taken deeper in the water column (~100–500 m) to account for the deeper habitat of the foraminifera species used (Bohaty et al., 2012). Full details of the sites and proxy data used can be found in Table S1 and proxy locations are shown in Figure S1a.

In terms of absolute values, many data locations (e.g., sub-Arctic, high North Atlantic, St Stephen's Quarry, Agulhas Ridge, New Zealand, and Kerguelen Plateau) show a cold bias of up to 15–20 °C, particularly with increasing latitude of the sites. The Falklands Plateau is found to be too cold before the EOT but close to the proxies after the EOT, while the Ceara Rise and Tanzania sites are close to the proxies both before and after the EOT, with Ceara Rise having a slight warm bias (<3 °C). The reasonable agreement with the low-latitude Ceara Rise and Tanzania sites highlights that the model is producing a steeper latitudinal SST gradient than is found in the data, with high latitudes being too cold as a result, a common issue in climate models (Huber & Caballero, 2011; Lunt et al., 2012).

In terms of magnitude of change, the model is generally found to be slightly conservative relative to the proxy records. Of the individual forcings tested here,  $p\text{CO}_2$  halving is found to be the closest to the proxy records for most sites, only lying outside of the proxy uncertainty for the Falklands Plateau and sub-Arctic sites. The combined forcing of both ice growth and  $p\text{CO}_2$  reduction (which could be expected to be the most realistic forcing for the EOT) also does a comparable job to  $p\text{CO}_2$  reduction in isolation in terms of explaining the changes. Antarctic ice growth and paleogeographic change generally result in very small changes (sometimes of opposite sign to the proxies), suggesting that these forcings in isolation do a poor job of simulating



**Figure 6.** Comparison of the modeled sea surface temperature (SST) pre-Eocene-Oligocene Transition (EOT) and post-EOT to proxy data for nine sites (Bohaty et al., 2012; Liu et al., 2009; Pearson et al., 2007; Wade et al., 2012). Individual pairs of model simulations are shown for the Antarctic ice growth, paleogeographic change,  $pCO_2$  reduction, and combined ice growth and  $pCO_2$  reduction forcings, with the mean for each of these forcings indicated by the squares with error bars of the same respective colors. The grey shading highlights the published uncertainty in the magnitude of change in the proxy data (where available), and the dotted line indicates no change across the EOT. Note that the axis ranges are different for low- and high-latitude sites.



the observed changes at the EOT; however, it is possible that these responses could change with increased spin-up. The only site at which ice growth and paleogeographic change are closer to the change suggested by the proxy data is Ceara Rise; however, this site has very high uncertainty in the proxy itself and all forcings lie within the potential error.

In line with the generalized spatial patterns of  $U_t$  discussed in section 4.1, the sites with the largest uncertainty in the mean model responses are New Zealand, the high North Atlantic, St Stephen's Quarry, and the Kerguelen Plateau, all regions close to where the model has some deep water formation. For all of these sites there is overlap in the uncertainty in the magnitude of proxy change (where available) and the  $U_t$  of the mean  $p\text{CO}_2$  reduction and combined ice growth and  $p\text{CO}_2$  reduction forcings. In contrast, even with the wide  $U_t$  margin, paleogeographic change does not have an effect similar to the proxy records at three of the sites, suggesting that this  $U_t$  method does not simply increase tolerance so any model forcing can be described as *good*. Again, however, these responses could change if the model simulations here were further spun-up. The New Zealand site in particular is close to the region that showed a reversal in the temperature anomaly in response to Antarctic ice growth for the spin-up ensemble (Figures 4a and 4b). Assuming a similar process could occur here with the boundary condition ensemble if the simulations were run for longer, it is possible that ice growth forcing could produce a more realistic change relative to the proxy records.

This analysis is reasonably simplistic as it looks only at idealized forcings and as such is not an attempt at a best guess of the actual climate transition that occurred at the EOT. In that regard, it is unsurprising that the modeled SSTs differ from the proxy data. In reality a combination of these forcings (and other processes that are not modeled here) would have been occurring at the EOT. Missing or incorrectly represented processes must be responsible for at least some of the absolute model discrepancy from the data. It is possible that the  $p\text{CO}_2$  could have been higher both before and after the EOT than the 1,120 and 560 ppmv values used here (Pearson et al., 2009; Zhang et al., 2013). This could account for some of the cold bias in the high-latitude sites relative to the proxies; however, it could not account for over 10 °C of warming without making the Ceara Rise and Tanzania sites much too hot relative to the proxies. Additionally, proxy bias (including issues resolving high temperatures or bias toward summer time conditions at high latitudes; Schouten et al., 2013) debatably could widen the uncertainties beyond the ranges given, while single proxy sites could be biased to localized conditions, which are not resolved in the model, making like-for-like comparisons difficult.

Notwithstanding these potential sources of bias in proxy records, it is worth bearing in mind that HadCM3BL is shown to have extremely high seasonality over Southern Hemisphere high latitudes (a seasonal range of 30–60 °C), to the point that it would inhibit AIS growth even under favorable orbits unless  $p\text{CO}_2$  is reduced considerably (Gasson et al., 2014). Assuming that this large seasonality in the model is the result of some unknown missing or incorrectly represented processes, it is plausible that similar unknown processes could account for at least part of the model-data discrepancy. Inclusion of such processes that could alter the modeled climate enough to fit the data would also likely have a huge impact on the modeled uncertainty looked at in this study. In that regard, the uncertainty results shown here are far from exhaustive.

## 5. Conclusions

Using two ensembles of fully coupled GCM simulations, we generalize the Earth's atmospheric and oceanic response to various forcings typical of the EOT. The size of the ensembles used here allows the uncertainty in these responses due to boundary condition uncertainty,  $U_b$ , to be assessed using a  $t$ -distribution. Additionally, long spin-up simulations highlight which regions and forcings are particularly sensitive to a lack of spin-up (and so are less robust in shorter simulations).

For most climatic and oceanic variables, there is found to be significant variability in the impact of each of the forcings due to differences in boundary conditions. The growth of the AIS causes a mixed SAT response (which is particularly uncertain due to a lack of spin-up), a minor global cooling, an increase in ACC flow through the Drake Passage, and an increase in overturning and MLD in the Southern Ocean. Paleogeographic change also increases ACC flow, with wider Southern Ocean gateways having greater flow through the Drake Passage, associated with a slight hemispheric redistribution of heat, with the ensemble agreeing on the Southern Hemisphere warming slightly particularly over the oceans. There is poor agreement as to how paleogeographic change will affect ocean overturning in either hemisphere (although

with increased spin-up the model agreement could possibly improve). Finally, atmospheric  $p\text{CO}_2$  reduction has a clear cooling effect on global SAT, with good model agreement, and it has a modest effect of intensifying ACC flow through the Drake Passage. Of the three forcings, this is the most consistent between model simulations and the least susceptible to change with increased spin-up.

For mean annual SAT, the spatial patterns of  $U_t$  have similar characteristics for each forcing from both ensembles. Therefore, although the exact magnitude of  $U_t$  varies with the number of simulations used, the nature of the climate forcing and how far the model simulations are away from equilibrium (and will therefore vary for future studies), we can generalize where we have higher, medium, and lower confidence in our modeling results. Low-latitude to midlatitude oceans generally show the lowest uncertainties, most continental regions generally have slightly higher uncertainties, while high-latitude oceans (particularly deep water formation regions) and mountainous regions have the highest uncertainties. The range in model responses to each forcing (an alternative method for portraying the uncertainty that does not make an assumption about the linearity of the modeled climate response) also shows similar spatial patterns of low-high uncertainty.

The modeled SAT response to various forcings can be profoundly affected by a lack of spin-up. Here it is shown that the Antarctic ice growth forcing in particular has a reversal in the response in the Southern Ocean. In general, the absolute changes due to a lack of spin-up are of a similar magnitude (or greater) to the uncertainties due to different boundary conditions. However, it is harder to generalize how the model response (particularly in high-latitude ocean regions) might change with a more complete spin-up. Again, high-latitude oceans are found to be most susceptible to change due to a more complete spin-up, while responses in low-latitude to midlatitude ocean regions where there is agreement early in the spin-up tend to be more stable, with the responses persisting through to the end of the spin-up.

In terms of the relative importance of each of the forcings on climate at the EOT, the boundary condition ensemble of model simulations shows  $p\text{CO}_2$  reduction gives the best match to the magnitude of change recorded by proxy records of SST, compared to Antarctic ice growth or paleogeographic change forcings. A combined forcing of Antarctic ice growth and  $p\text{CO}_2$  reduction also performs well, albeit with a smaller sample size. However, this analysis is reasonably simplistic as only idealized forcings are assessed and further work needs to be done to improve the match with data, as absolute biases in the temperature are generally found to be very high. It would also be of interest in future work to expand this analysis to other oceanic and terrestrial records (Baatsen, von der Heydt, Huber, et al., 2018; Li et al., 2018; Pound & Salzmann, 2017).

These uncertainty patterns may be similar for other Cenozoic time periods when the paleogeography had similar features to those assessed here for the late Eocene and early Oligocene (i.e., similar ocean gateways and continental positions), but this should be tested further. Deeper time periods when the paleogeography was radically different may or may not behave in a similar way, requiring further study and assessment. The method used here with a medium-sized ensemble and the uncertainty calculated with a  $t$ -distribution is one possible way to do this. If there are particular concerns about the linearity of the modeled climate response, other methods such as the range in responses can also be used to compliment this method. The generalized uncertainty highlighted by these model simulations should be considered when interpreting paleoclimate research when single or very few model simulations are used.

#### Acknowledgments

The authors declare no conflicts of interests. In addition to the model mean data in the supplementary spreadsheet, upon publication, the model data used in all of the analysis will be uploaded to a freely accessible BRIDGE server (<https://www.paleo.bristol.ac.uk/ummodel/scripts/papers/>), as is the standard protocol at the University of Bristol. A. T. Kennedy-Asser was funded by NERC (NE/L002434/1). A. Farnsworth was funded by NERC (NE/K014757/1, NE/I005722/1, and NE/I005714/1). Climate simulations were carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol (<http://www.bris.ac.uk/acrc; Bluecrystal>). We thank the three anonymous reviewers for their helpful comments in developing the manuscript.

#### References

- Armstrong, E., Valdes, P., House, J., & Singarayer, J. (2017). Investigating the impact of  $\text{CO}_2$  on low-frequency variability of the AMOC in HadCM3. *Journal of Climate*, 30(19), 7863–7883. <https://doi.org/10.1175/JCLI-D-16-0767.1>
- Baatsen, M., van Hinsbergen, D. J. J., von der Heydt, A. S., Dijkstra, H. A., Sluijs, A., Abels, H. A., & Bijl, P. K. (2016). Reconstructing geographical boundary conditions for palaeoclimate modelling during the Cenozoic. *Climate of the Past*, 12(8), 1635–1644. <https://doi.org/10.5194/cp-12-1635-2016>
- Baatsen, M., von der Heydt, A. S., Huber, M., Kliphuis, M. A., Bijl, P. K., Sluijs, A., & Dijkstra, H. A. (2018). Equilibrium state and sensitivity of the simulated middle-to-late Eocene climate. *Climate of the Past Discussions*, 1–49. <https://doi.org/10.5194/cp-2018-43>
- Baatsen, M., von der Heydt, A. S., Kliphuis, M., Viebahn, J., & Dijkstra, H. A. (2018). Multiple states in the late Eocene ocean circulation. *Global and Planetary Change*, 163, 18–28. <https://doi.org/10.1016/j.gloplacha.2018.02.009>
- Bohaty, S., Zachos, J., & Delaney, M. (2012). Foraminiferal Mg/Ca evidence for Southern Ocean cooling across the Eocene-Oligocene transition. *Earth and Planetary Science Letters*, 317–318, 251–261. <https://doi.org/10.1016/j.epsl.2011.11.037>
- Burt, J., Barber, G., & Rigby, D. (2009). *Elementary statistics for geographers* (3rd ed.). New York: Guildford Press.

- Carter, A., Riley, T., Hillenbrand, C., & Rittner, M. (2017). Widespread Antarctic glaciation during the Late Eocene. *Earth and Planetary Science Letters*, 458, 49–57. <https://doi.org/10.1016/j.epsl.2016.10.045>
- Coxall, H., Huck, C. E., Huber, M., Lear, C. H., Legarda-Lisarrri, A., O'Regan, M., Sliwinski, K. K., et al. (2018). Export of nutrient rich Northern Component Water preceded early Oligocene Antarctic glaciation. *Nature Geoscience*, 11(3), 190–196. <https://doi.org/10.1038/s41561-018-0069-9>
- Coxall, H., Wilson, P. A., Pälike, H., Lear, C. H., & Backman, J. (2005). Rapid stepwise onset of Antarctic glaciation and deeper calcite compensation in the Pacific Ocean. *Nature*, 433(7021), 53–57. <https://doi.org/10.1038/nature03135>
- DeConto, R. M., & Pollard, D. (2003). Rapid Cenozoic glaciation of Antarctica induced by declining atmospheric CO<sub>2</sub>. *Nature*, 421(6920), 245–249. <https://doi.org/10.1038/nature01290>
- Elsworth, G., Galbraith, E., Halverson, G., & Yang, S. (2017). Enhanced weathering and CO<sub>2</sub> drawdown caused by latest Eocene strengthening of the Atlantic meridional overturning circulation. *Nature Geoscience*, 10(3), 213–216. <https://doi.org/10.1038/ngeo2888>
- Galeotti, S., DeConto, R., Naish, T., Stocchi, P., Florindo, F., Pagani, M., Barrett, P., et al. (2016). Antarctic ice sheet variability across the Eocene-Oligocene boundary climate transition. *Science*, 352(6281), 76–80. <https://doi.org/10.1126/science.aab0669>
- Gasson, E., Lunt, D. J., DeConto, R., Goldner, A., Heinemann, M., Huber, M., LeGrande, A. N., et al. (2014). Uncertainties in the modelled CO<sub>2</sub> threshold for Antarctic glaciation. *Climate of the Past*, 10(2), 451–466. <https://doi.org/10.5194/cp-10-451-2014>
- Goldner, A., Herold, N., & Huber, M. (2014). Atlantic glaciation caused ocean circulation changes at the Eocene-Oligocene transition. *Nature*, 511(7511), 574–577. <https://doi.org/10.1038/nature13597>
- Hill, D., Haywood, A. M., Valdes, P. J., Francis, J. E., Lunt, D. J., Wade, B. S., & Bowman, V. C. (2013). Paleogeographic controls on the onset of the Antarctic circumpolar current. *Geophysical Research Letters*, 40, 5199–5204. <https://doi.org/10.1002/grl.50941>
- Hohbein, M., Sexton, P., & Cartwright, J. (2012). Onset of North Atlantic Deep Water production coincident with inception of the Cenozoic global cooling trend. *Geology*, 40(3), 255–258. <https://doi.org/10.1130/G32461.1>
- Houben, A., Bijl, P. K., Pross, J., Bohaty, S. M., Passchier, S., Stickley, C. E., Röhl, U., et al. (2013). Reorganization of Southern Ocean plankton ecosystem at the onset of Antarctic glaciation. *Science*, 340(6130), 341–344. <https://doi.org/10.1126/science.1223646>
- Huber, M., & Caballero, R. (2011). The early Eocene equable climate problem revisited. *Climate of the Past*, 7(2), 603–633. <https://doi.org/10.5194/cp-7-603-2011>
- Huber, M., & Nof, D. (2006). The ocean circulation in the southern hemisphere and its climatic impacts in the Eocene. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 231(1–2), 9–28. <https://doi.org/10.1016/j.palaeo.2005.07.037>
- Hutchinson, D., de Boer, A. M., Coxall, H. K., Caballero, R., Nilsson, J., & Baatsen, M. (2018). Climate sensitivity and meridional overturning circulation in the late Eocene using GFDL CM2.1. *Climate of the Past*, 14(6), 789–810. <https://doi.org/10.5194/cp-14-789-2018>
- Inglis, G., Farnsworth, A., Lunt, D., Foster, G. L., Hollis, C. J., Pagani, M., Jardine, P. E., et al. (2015). Descent toward the icehouse: Eocene sea surface cooling inferred from GDGT distributions. *Paleoceanography*, 30, 1000–1020. <https://doi.org/10.1002/2014PA002723>
- Katz, M., Cramer, B. S., Toggweiler, J. R., Esmay, G., Liu, C., Miller, K. G., Rosenthal, Y., et al. (2011). Impact of Antarctic Circumpolar Current development on late Paleogene ocean structure. *Science*, 332(6033), 1076–1079. <https://doi.org/10.1126/science.1202122>
- Kennedy, A., Farnsworth, A., Lunt, D., Lear, C. H., & Markwick, P. (2015). Atmospheric and oceanic impacts of Antarctic glaciation across the Eocene–Oligocene transition. *Philosophical Transactions of the Royal Society A*, 373(2054), 20140419. <https://doi.org/10.1098/rsta.2014.0419>
- Kennett, J. (1977). Cenozoic evolution of Antarctic glaciation, circum Antarctic ocean, and their impact on global paleoceanography. *Journal of Geophysical Research*, 82(27), 3843–3860. <https://doi.org/10.1029/JC082i027p03843>
- Ladant, J., Donnadieu, Y., Lefebvre, V., & Dumas, C. (2014). The respective role of atmospheric carbon dioxide and orbital parameters on ice sheet evolution at the Eocene-Oligocene transition. *Paleoceanography*, 29, 810–823. <https://doi.org/10.1002/2013PA002593>
- Lear, C., Bailey, T. R., Pearson, P. N., Coxall, H. K., & Rosenthal, Y. (2008). Cooling and ice growth across the Eocene–Oligocene transition. *Geology*, 36(3), 251–254. <https://doi.org/10.1130/G24584A.1>
- Li, S., Xing, Y., Valdes, P. J., Huang, Y., Su, T., Farnsworth, A., Lunt, D. J., et al. (2018). Oligocene climate signals and forcings in Eurasia revealed by plant macrofossil and modelling results. *Gondwana Research*, 61, 115–127. <https://doi.org/10.1016/j.gr.2018.04.015>
- Liu, Z., Pagani, M., Zinniker, D., DeConto, R., Huber, M., Brinkhuis, H., Shah, S. R., et al. (2009). Global cooling during the Eocene–Oligocene climate transition. *Science*, 323(5918), 1187–1190. <https://doi.org/10.1126/science.1166368>
- Lucarini, V., & Bódai, T. (2017). Edge states in the climate system: Exploring global instabilities and critical transitions. *Nonlinearity*, 30(7), R32–R66. <https://doi.org/10.1088/1361-6544/aa6b11>
- Lunt, D. J., Dunkley Jones, T., Heinemann, M., Huber, M., LeGrande, A., Winguth, A., et al. (2012). A model–data comparison for a multi-model ensemble of early Eocene atmosphere–ocean simulations: EoMIP. *Climate of the Past*, 8, 1717–1736. <https://doi.org/10.5194/cp-8-1717-2012>
- Lunt, D. J., Farnsworth, A., Loptson, C., Foster, G. L., Markwick, P., O'Brien, C. L., Pancost, R. D., et al. (2016). Palaeogeographic controls on climate and proxy interpretation. *Climate of the Past*, 12(5), 1181–1198. <https://doi.org/10.5194/cp-12-1181-2016>
- Lyle, M., Gibbs, S., Moore, T., & Rea, D. (2007). Late Oligocene initiation of the Antarctic Circumpolar Current: Evidence from the South Pacific. *Geology*, 35(8), 691–694. <https://doi.org/10.1130/G23806A.1>
- Pälike, H., Lyle, M. W., Nishi, H., Raffi, I., Ridgwell, A., Gamage, K., Klaus, A., et al. (2012). A Cenozoic record of the equatorial Pacific carbonate compensation depth. *Nature*, 488(7413), 609–614. <https://doi.org/10.1038/nature11360>
- Pearson, P., Foster, G., & Wade, B. (2009). Atmospheric carbon dioxide through the Eocene–Oligocene climate transition. *Nature*, 461(7267), 1110–1113. <https://doi.org/10.1038/nature08447>
- Pearson, P., van Dongen, B. E., Nicholas, C. J., Pancost, R. D., Schouten, S., Singano, J. M., & Wade, B. S. (2007). Stable warm tropical climate through the Eocene epoch. *Geology*, 35(3), 211–214. <https://doi.org/10.1130/G23175A.1>
- Petersen, S., & Schrag, D. (2015). Antarctic ice growth before and after the Eocene-Oligocene transition: New estimates from clumped isotope paleothermometry. *Paleoceanography*, 30, 1305–1317. <https://doi.org/10.1002/2014PA002769>
- Pound, M., & Salzmann, U. (2017). Heterogeneity in global vegetation and terrestrial climate change during the late Eocene to early Oligocene transition. *Scientific Reports*, 7(1), 43,386. <https://doi.org/10.1038/srep43386>
- Scher, H., Whittaker, J. M., Williams, S. E., Latimer, J. C., Kordesch, W. E. C., & Delaney, M. L. (2015). Onset of Antarctic Circumpolar Current 30 million years ago as Tasmanian Gateway aligned with westerlies. *Nature*, 523(7562), 580–583. <https://doi.org/10.1038/nature14598>
- Schouten, S., Hopmans, E., & Sinninghe Damsté, J. (2013). The organic geochemistry of glycerol dialkyl glycerol tetraether lipids: A review. *Organic Geochemistry*, 54, 19–61. <https://doi.org/10.1016/j.orggeochem.2012.09.006>
- Sheldon, N., & Retallack, G. (2004). Regional paleoprecipitation records from the Late Eocene and Oligocene of North America. *The Journal of Geology*, 112(4), 487–494. <https://doi.org/10.1086/421076>

- Sijp, W., England, M., & Huber, M. (2011). Effect of deepening the Tasman Gateway on the global ocean. *Paleoceanography*, *34*, PA4207. <https://doi.org/10.1029/2011PA002143>
- Stickley, C., Brinkhuis, H., Schellenberg, S. A., Sluijs, A., Röhl, U., Fuller, M., Grauert, M., et al. (2004). Timing and nature of the deepening of the Tasmanian Gateway. *Paleoceanography*, *19*, PA4027. <https://doi.org/10.1029/2004PA001022>
- Sun, J., & Windley, B. (2015). Onset of aridification by 34 Ma across the Eocene-Oligocene transition in Central Asia. *Geology*, *43*(11), 1015–1018. <https://doi.org/10.1130/G37165.1>
- Valdes, P., Armstrong, E., Badger, M. P. S., Bradshaw, C. D., Bragg, F., Crucifix, M., Davies-Barnard, T., et al. (2017). The BRIDGE HadCM3 family of climate models: HadCM3@Bristol v1.0. *Geoscientific Model Development*, *10*(10), 3715–3743. <https://doi.org/10.5194/gmd-10-3715-2017>
- Via, R., & Thomas, D. (2006). Evolution of Atlantic thermohaline circulation: Early Oligocene onset of deep-water production in the North Atlantic. *Geology*, *34*(6), 441–444. <https://doi.org/10.1130/G22545.1>
- von der Heydt, A., Dijkstra, H., van de Wal, R. S. W., Caballero, R., Crucifix, M., Foster, G. L., Huber, M., et al. (2016). Lessons on climate sensitivity from past climate changes. *Current Climate Change Reports*, *2*(4), 148–158. <https://doi.org/10.1007/s40641-016-0049-3>
- Wade, B., Houben, A. J. P., Quaijtaal, W., Schouten, S., Rosenthal, Y., Miller, K. G., Katz, M. E., et al. (2012). Multiproxy record of abrupt sea-surface cooling across the Eocene-Oligocene transition in the Gulf of Mexico. *Geology*, *40*(2), 159–162. <https://doi.org/10.1130/G32577.1>
- Zachos, J., Pagani, M., Sloan, L., Thomas, E., & Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, *292*(5517), 686–693. <https://doi.org/10.1126/science.1059412>
- Zanazzi, A., Kohn, M., MacFadden, B., & Terry, D. (2007). Large temperature drop across the Eocene–Oligocene transition in central North America. *Nature*, *445*(7128), 639–642. <https://doi.org/10.1038/nature05551>
- Zhang, Y., Pagani, M., Liu, Z., Bohaty, S. M., & DeConto, R. (2013). 40-million year history of atmospheric CO<sub>2</sub>. *Philosophical Transactions of the Royal Society A*, *371*, 20130096.