Deep Ocean Temperatures through Time

Paul J Valdes[1], Christopher R Scotese[2], Daniel J Lunt[1]

[1] School of Geographical Sciences, University of Bristol, Bristol BS8 1SS, UK

[2] Northwestern University, Dept Earth & Planetary Sci, Evanston, IL USA

Correspondence to Paul Valdes (P.J.Valdes@bristol.ac.uk)

Abstract

Benthic oxygen isotope records are commonly used as a proxy for global mean surface temperatures during the late Cretaceous and Cenozoic, and the resulting estimates have been extensively used in characterising major trends and transitions in the climate system, and for analysing past climate sensitivity. However, some fundamental assumptions governing this proxy have rarely been tested. Two key assumptions are: (a) benthic foraminiferal temperatures are geographically well-mixed and are linked to surface high latitude temperatures, and (b) surface high latitude temperatures are well correlated with global mean temperatures. To investigate the robustness of these assumptions through geological time, we performed a series of 109 climate model simulations using a unique set of paleogeographical reconstructions covering the entire Phanerozoic at the stage-level. The simulations have been run for at least 5000 model years to ensure that the deep ocean is in dynamic equilibrium. We find that the correlation between deep ocean temperatures and global mean surface temperatures is good for the Cenozoic and thus the proxy data are reliable indicators for this time period, albeit with a standard error of 2K. This uncertainty has not normally been assessed and needs to be combined with other sources of uncertainty when, for instance, estimating climate sensitivity based on using $\delta^{18}O$ measurements from benthic foraminifera. The correlation between deep and global mean surface temperature becomes weaker for pre-Cenozoic time periods (when the paleogeography is significantly different than the present-day). The reasons for the weaker correlation include variability in the source region of the deep water (varying hemispheres but also varying latitudes of sinking), the depth of ocean overturning (some extreme warm climates have relatively shallow and sluggish circulations weakening the link between surface and deep ocean), and the extent of polar amplification (e.g. ice albedo feedbacks). Deep ocean sediments prior to the Cretaceous are rare, so extending the benthic foram proxy further into deeper time is problematic, but the model results presented here would suggest that the deep ocean temperatures from such time periods would probably be an unreliable indicator of global mean surface conditions.

1. Introduction

One of the most widely used proxies for estimating global mean surface temperature through the last 100 million years is benthic $\delta^{18}O$ measurements from deep sea foraminifera (Zachos et al., 2001), (Zachos et al., 2008), (Cramer et al., 2009), (Friedrich et al., 2012), (Westerhold et al., 2020). Two key underlying assumptions are that $\delta^{18}O$ from benthic foraminifera represents deep ocean temperature (with a correction for ice volume and any vital effects), and that the deep ocean water masses originate from surface water in polar regions. By further assuming that polar surface temperatures are well correlated with global mean surface temperatures, then deep ocean isotopes can be assumed to track global mean surface temperatures. More specifically, (Hansen et al., 2008), and (Hansen and Sato, 2012) argue that changes in high latitude sea surface temperatures are approximately proportional to global mean surface temperatures because changes are generally amplified at high latitudes but that this is offset because temperature change is amplified over land areas. They therefore directly equate changes in benthic ocean temperatures with global mean surface temperature.

The resulting estimates of global mean surface air temperature have been used to understand past climates (e.g. (Zachos et al., 2008), (Westerhold et al., 2020)). Combined with estimates of atmospheric CO_2 they have also been used to estimate climate sensitivity (e.g. (Hansen et al., 2013)) and hence contribute to the important ongoing debate about the likely magnitude of future climate change.

However, some of the underlying assumptions behind the method remains largely untested, even though we know that there are major changes to paleogeography and consequent changes in ocean circulation and location of deep-water formation in the deep past (e.g. (Lunt et al., 2010; Nunes and Norris, 2006); (Donnadieu et al., 2016); (Farnsworth et al., 2019a); (Ladant et al., 2020)). Moreover, the magnitude of polar amplification is likely to vary depending on the extent of polar ice caps, and changes in cloud cover (Sagoo et al., 2013), (Zhu et al., 2019). These issues are likely to modify the correlation between deep ocean temperatures and global mean surface temperature or, at the very least, increase the uncertainty in reconstructing past global mean surface temperatures.

The aim of this paper is two-fold, (1) we wish to document the setup and initial results from a unique set of 109 climate model simulations of the whole Phanerozoic era (last 540 million years) at the stage level (approximately every 5 million years), and (2) we will use these simulations to investigate

64 the accuracy of the deep ocean temperature proxy in representing global mean surface

65 temperature.

66 The focus of the work is to examine the link between benthic ocean temperatures and surface

67 conditions. However, we evaluate the fidelity of the model by comparing the model predicted ocean

68 temperatures to estimates of the isotopic temperature of the deep ocean during the past 110

69 million years ((Zachos et al., 2008), (Cramer et al., 2009), (Friedrich et al., 2012)), and model

70 predicted surface temperatures to the sea surface temperatures estimates of (O'Brien et al., 2017)

71 and (Cramwinckel et al., 2018). This gives us confidence that the model is behaving plausibly but we

72 emphasise that the fidelity of the simulations is strongly influenced by the accuracy of CO_2 estimates

73 through time. We then use the complete suite of climate simulations to examine changes in ocean

74 circulation, ice formation, and the impact on ocean and surface temperature. Our paper will not

75 consider any issues associated with assumptions regarding the relationship between deep-sea

76 foraminifera $\delta^{18}O$ and various temperature calibrations because our model does not simulate the

77 $\delta^{18}O$ of sea water (or vital effects).

78

79 ## 2. Simulation Methodology

80 ### 2.1 Model Description

81 We use a variant of the Hadley Centre model, HadCM3 ((Pope et al., 2000), (Gordon et al., 2000))

82 which is a coupled atmosphere-ocean-vegetation model. The specific version, HadCM3BL-M2.1aD, is

83 described in detail in (Valdes et al., 2017). The model has a horizontal resolution of 3.75^o x 2.5^o in

84 longitude/latitude (roughly corresponding to an average grid box size of ~300km) in both the

85 atmosphere and the ocean. The atmosphere has 19 unequally spaced vertical levels, and the ocean

86 has 20 unequally spaced vertical levels. To avoid singularity at the poles, the ocean model always has

87 to have land at the poles (90N and 90S), but the atmosphere model can represent the poles

88 correctly (i.e. in the pre-industrial geography, the atmosphere considers there is sea ice covered

89 ocean at the N. Pole but the ocean model has land and hence there is no ocean flow across the

90 pole). Though HadCM3 is a relatively low resolution and low complexity model compared to the

91 current CMIP5/CMIP6 state-of-the-art model, its performance at simulating modern climate is

92 comparable to many CMIP5 models (Valdes et al., 2017). The performance of the dynamic

93 vegetation model compared to modern observations is also described in (Valdes et al., 2017) but the

94 modern deep ocean temperatures are not described in that paper. We therefore include a

95 comparison to present day observed deep ocean temperatures in section 3.1.

96 To perform paleo-simulations, several important modifications to the standard model described in
97 (Valdes et al., 2017) must be incorporated:

98 (a) The standard pre-industrial model uses a prescribed climatological pre-industrial ozone
99   concentration (i.e. prior to the development of the "ozone" hole) which is a function of
100   latitude, atmospheric height, and month of the year. However, we do not know what the
101   distribution of ozone should be in these past climates. (Beerling et al., 2011) modelled small
102   changes in tropospheric ozone for the early Eocene and Cretaceous but no comprehensive
103   stratospheric estimates are available. Hence most paleoclimate model simulations assume
104   unchanging concentrations. However, there is a problem with using a prescribed ozone
105   distribution for paleo-simulations because it does not incorporate ozone feedbacks
106   associated with changes in tropospheric height. During warm climates, the model predicts
107   that the tropopause would rise. In the real world, ozone would track the tropopause rise.
108   However, this rising ozone feedback is not included in our standard model. This leads to
109   substantial extra warming and artificially increases the apparent climate sensitivity.
110   Simulations of future climate change have shown that ozone feedbacks can lead to an over-
111   estimate of climate sensitivity by up to 20% ((Dietmuller et al., 2014), (Nowack et al., 2015))
112   (Hardiman et al., 2019). Therefore, to incorporate some aspects of this feedback, we have
113   changed the ozone scheme in the model. Ozone is coupled to the model predicted
114   tropopause height every model timestep in the following simple way:

115   • 2.0x10^{-8} kg/kg in the troposphere

116   • 2.0x10^{-7} kg/kg at the tropopause

117   • 5.5x10^{-6} kg/kg above the tropopause

118   • 5.5x10^{-6} kg/kg at the top model level.

119   These values are approximate averages of present-day values and were chosen so that the
120   tropospheric climate of the resulting pre-industrial simulation was little altered compared
121   with the standard preindustrial simulations; the resulting global mean surface air
122   temperatures differed by only 0.05 ˚C. These modifications are similar to those used in the
123   FAMOUS model (Smith et al., 2008) except that the values in the stratosphere are greater in
124   our simulation, largely because our model vertical resolution is higher than in FAMOUS.

125   Note that these changes improve upon the scheme used by (Lunt et al., 2016) and
126   (Farnsworth et al., 2019a). They used much lower values of stratospheric ozone and had no
127   specified value at the top of the model. This resulted in their model having ~ 1˚C cold bias

128   for pre-industrial temperatures and may have also affected their estimates of climate
129   sensitivity.

130   (b)   The standard version of HadCM3 conserves the total volume of water throughout the
131   atmosphere and ocean (including in the numerical scheme) but several processes in the
132   model "lose or gain" water:

133   1.   Snow accumulates over ice sheets but there is no interactive loss through iceberg
134   calving resulting in an excess loss of fresh water from the ocean.

135   2.   The model caps salinity at a maximum of 45 PSU (and a minimum of 0 PSU), by
136   artificially adding/subtracting fresh water to the ocean. This mostly affects small,
137   enclosed seas (such as the Red Sea or enclosed Arctic) where the model does not
138   represent the exchanges with other ocean basins.

139   3.   Modelled river runoff includes some river basins which drain internally. These often
140   correspond to relatively dry regions, but any internal drainage simply disappears
141   from the model.

142   4.   The land surface scheme includes evaporation from sub-grid scale lakes (which are
143   prescribed as a lake fraction in each grid box, at the start of the run). The model
144   does not represent the hydrological balance of these lakes, consequently the
145   volume of the lakes does not change. This effectively means that there is a net
146   source/sink of water in the model in these regions.

147   In the standard model, these water sources/sinks are approximately balanced by a flux of
148   water into the surface ocean. This is prescribed at the start of the run and does not vary
149   during the simulations. It is normally set to a pre-calculated estimate based on an old
150   HadCM3-M1 simulation. The flux is strongest around Greenland and Antarctica and is
151   chosen such that it approximately balances the water loss described in (1) i.e. the net snow
152   accumulation over these ice sheets. There is an additional flux covering the rest of the
153   surface ocean which approximately balances the water loss from the remaining three terms
154   (2-4). The addition of this water flux keeps the global mean ocean salinity approximately
155   constant on century time scales. However, depending on the simulation, the drift in average
156   oceanic salinity can be as much as 1PSU per thousand years and thus can have a major
157   impact on ultra-long runs of >5000 years (Farnsworth et al., 2019a).

158   For the paleo-simulations in this paper, we therefore take a slightly different approach.
159   When ice sheets are present in the Cenozoic, we include the water flux (for the relevant

160   hemisphere) described in (1) above, based on modern values of iceberg calving fluxes for

161   each hemisphere. However, to ensure that salinity is conserved, we also interactively

162   calculate an additional globally uniform surface water flux based on relaxing the volume

163   mean ocean salinity to a prescribed value on a 20-year timescale. This ensures that there is

164   no long-term trend in ocean salinity. Tests of this update on the pre-industrial simulations

165   revealed no appreciable impact on the skill of the model relative to the observations. We

166   have not directly compared our simulations to the previous runs of the (Farnsworth et al.,

167   2019a) because they use different CO_2 and different paleogeographies. However in practice,

168   the increase of salinity in their simulations is well mixed and seems to have relatively little

169   impact on the overall climate and ocean circulation.

170   We have little knowledge of whether ocean salinity has changed through time, and so keep

171   the prescribed mean ocean salinity constant across all simulations.

172

173   2.2 Model Boundary Conditions

174 There are several boundary conditions that require modification through time. In this sequence of

175 simulations, we only modify three key time-dependent boundary conditions: 1) the solar constant, 2)

176 atmospheric CO_2 concentrations and, 3) paleogeographic reconstructions. We set the surface soil

177 conditions to a uniform medium loam everywhere. All other boundary conditions (such as orbital

178 parameters, volcanic aerosol concentrations etc.) are held constant at pre-industrial values.

179 The solar constant is based on (Gough, 1981) and increases linearly at an approximate rate of 11.1

180 Wm^{-2} per 100 Ma (0.8% per 100Ma), to 1365Wm^{-2} currently. If we assume a planetary albedo of 0.3,

181 and a climate sensitivity of 0.8 ˚C /Wm^{-2} (approximately equivalent to 3˚C per doubling of CO_2), then

182 this is equivalent to a temperature increase of ~.015˚C per million years (~8˚C over the whole of the

183 Phanerozoic).

184 Estimates of atmospheric CO_2 concentrations have considerable uncertainty. We, therefore, use two

185 alternative estimates (fig. 1a). The first uses the best fit Loess curve from (Foster et al., 2017), which

186 is also very similar to the newer data from (Witkowski et al., 2018). The CO_2 levels have considerable

187 short and long-term variability throughout the time period. Our second estimate removes much of

188 the shorter-term variability in the Foster (2017) curve. It was developed for two reasons. Firstly, a

189 lot of the finer temporal structure in the Loess curve is a product of differing data density of the raw

190 data and does not necessarily correspond to real features. Secondly, the smoother curve was heavily

191 influenced by a previous (commercially confidential) sparser sequence of simulations using non-

192 public paleogeographic reconstructions. The resulting simulations were generally in good agreement

193   with terrestrial proxy datasets (Harris et al., 2017). Specifically, using commercial in confidence

194   paleogeographies, we have performed multiple simulations at different CO_2 values for several stages

195   across the last 440 million years and tested the resulting climate against commercial-in-confidence

196   proxy data (Harris et al., 2017). We then selected the CO_2 that best matched the data. For the

197   current simulations, we linearly interpolated these CO_2 values to every stage. The resulting CO_2

198   curve looks like a heavily smoothed version of the Foster curve and is within the (large) envelope of

199   CO_2 reconstructions. The first-order shapes of the two curves are similar, though they are very

200   different for some time periods (e.g. Triassic and Jurassic). In practice, both curves should be

201   considered an approximation to the actual evolution of CO_2 through time which remains uncertain.

202   We refer to the simulation using the second set of CO_2 reconstructions as the "smooth" CO_2

203   simulations, though it should be recognised that the Foster CO_2 curve has also been smoothed. The

204   Foster CO_2 curve extends back to only 420 Ma, so we have proposed two alternative extensions back

205   to 540 Ma. Both curves increase sharply so that the combined forcing of CO_2 and solar constant are

206   approximately constant over this time period (Foster et al., 2017). The higher CO_2 in the Foster curve

207   relative to the "smooth" curve is because the initial set of simulations showed that the Cambrian

208   simulations were relatively cool compared to data estimates for the period (Henkes et al., 2018).

209   2.3 Paleogeographic Reconstructions

210   The 109 paleogeographic maps used in the HadleyCM3 simulations are digital representations of the

211   maps in the PALEOMAP Paleogeographic Atlas (Scotese, 2016); (Scotese and Wright, 2018). Table 1

212   lists all the time intervals that comprise the PALEOMAP Paleogeographic Atlas. The Paleo Atlas

213   contains one map for nearly every stage in the Phanerozoic. A paleogeographic map is defined as a

214   map that shows the ancient configuration of the ocean basins and continents, as well as important

215   topographic and bathymetric features such as mountains, lowlands, shallow sea, continental

216   shelves, and deep oceans. Paleogeographic reconstructions older than the oldest ocean floor (~Late-

217   Jurassic) have uniform deep ocean floor depth.

218   Once the paleogeography for each time interval has been mapped, this information is then

219   converted into a digital representation of the paleotopography and paleobathymetry. Each digital

220   paleogeographic model is composed of over 6 million grid cells that capture digital elevation

221   information at a 10 km x 10 km horizontal resolution and 40-meter vertical resolution. This

222   quantitative, paleo-digital elevation model, or "paleoDEM", allows us to visualize and analyse the

223   changing surface of the Earth through time using GIS software and other computer modelling

224   techniques. For use with the HadCM3L climate model, the original high-resolution elevation grid was

225   reduced to a ~111 km x ~111 km (1˚ x 1˚) grid.

226 For a detailed description of how the paleogeographic maps and paleoDEMs were produced the

227 reader is referred to (Scotese, 2016); (Scotese and Schettino, 2017); (Scotese and Wright, 2018).

228 (Scotese and Schettino, 2017) includes an annotated bibliography of the more than 100 key sources

229 of paleogeographic information.   Similar paleogeographic paleoDEMs have been produced by

230 (Baatsen et al., 2016) and (Verard et al., 2015).

231 The raw paleogeographic data reconstructs paleo-elevations and paleo-bathymetry at a resolution of

232 1° x 1°. These data were re-gridded to 3.75° x 2.5° resolution that matched the GCM using a simple

233 area (for land sea mask) or volume (for orography and bathymetry) conserving algorithm. The

234 bathymetry was lightly smoothed (using a binomial filter) to ensure that the ocean properties in the

235 resulting model simulations were numerically stable. The high latitudes had this filter applied

236 multiple times. The gridding sometimes produced single grid point enclosed ocean basins,

237 particularly along complicated coastlines, and these were manually removed. Similarly, important

238 ocean gateways were reviewed to ensure that the re-gridded coastlines preserved these structures.

239 The resulting global fraction of land is summarized in fig.1b and examples are shown in figure 2. The

240 original reconstructions can be found at https://www.earthbyte.org/paleodem-resource-scotese-

241 and-wright-2018/. Maps of each HadCM3L paleogeography are included in the supplementary

242 figures.

243 The paleogeographic reconstructions also include an estimate of land ice area ((Scotese and Wright,

244 2018); fig.1c). These were converted to GCM boundary conditions assuming a simple parabolic

245 shape to estimate the ice sheet height. These ice reconstructions suggest small amounts of land ice

246 were present during the early Cretaceous, unlike (Lunt et al., 2016) who used ice-free Cretaceous

247 paleogeographies.

248 2.4 Spin up Methodology

249 The oceans are the slowest evolving part of the modelled climate system and can take multiple

250 millennia to reach equilibrium, depending on the initial condition and climate state. To speed up the

251 convergence of the model, we initialized the ocean temperatures and salinity with the values from

252 previous model simulations from similar time periods using the commercial in confidence

253 paleogeographies. Specifically, we had a set of 17 simulations covering the last 440Ma. We selected

254 the nearest simulation to the time period. For instance, the 10.5 Ma, 14.9 Ma, and the 19.5Ma

255 simulations were initialised from the 13Ma simulation performed using the alternative

256 paleogeographies. Table 2 summarises the simulations performed in this study and shows the

257 initialisation of the model. The Foster CO_2 simulations were initialised from the end point of the

258 smooth CO_2 simulations. In the first set of simulations (smooth CO_2) we also attempted to accelerate

259   the spin up by using the ocean temperature trends at year 500 to linearly extrapolate the bottom 10

260   level temperatures for a further 1000 years. This had limited success and was not repeated. The

261   atmosphere variables were also initialized from the previous model simulations but the spin-up of

262   the atmosphere is much more rapid and did not require further intervention.

263   Simulations were run in parallel so were not initialised from the previous stage results using these

264   paleogeographies. In total, we performed almost 1 million years of model simulation and if we ran

265   simulations in sequence, it would have taken 30 years to complete the simulations. By running these

266   in parallel, initialised from previous modelling studies, we reduced the total run time to 3 months,

267   albeit using a substantial amount of our high-performance computer resources.

268   Although it is always possible that a different initialization procedure may produce different final

269   states, it is impossible to explore the possibility of hysteresis/bistability without performing many

270   simulations for each period, which is currently beyond our computing resources. Previous studies

271   using HadCM3L (not published) with alternative ocean initial states (isothermal at 0C, 8C, and 16C)

272   have not revealed multiple equilibria but this might have been because we did not locate the

273   appropriate part of parameter space that exhibits hysteresis. However, other studies have shown

274   such behaviour (e.g. (Baatsen et al., 2018)). This remains a caveat of our current work and which we

275   wish to investigate when we have sufficient computing resource.

276   The simulations were then run until they reached equilibrium, as defined by:

277   1.   The globally and volume integrated annual mean ocean temperature trend is less than

278            $1^{\circ}C/1000$ year, in most cases considerably smaller than this. We consider the volume

279   integrated temperature because it includes all aspects of the ocean. However, it is

280   dominated by the deep ocean trends and is near identical to the trends at a depth of 2731m

281   (the lowest level that we have archived for the whole simulation).

282   2.   The trends in surface air temperature are less than $0.3^{\circ}C/1000$ year.

283   3.   The net energy balance at the top of the atmosphere, averaged over 100-year period at the

284   end of the simulation, is less than 0.25 Wm^{-2} (in more than 80% of the simulations, the

285   imbalance is less than 0.1 Wm^{-2}). The Gregory plot (Gregory et al., 2004) implies surface

286   temperatures are within $0.3^{\circ}C$ of the equilibrium state.

287   These target trends were chosen somewhat arbitrarily but are all less than typical orbital time scale

288   variability (e.g. temperature changes since the last deglaciation were approximately $5^{\circ}C$ over 10,000

289   years). Most simulations were well within these criteria.   70% of simulations had residual net energy

290   balances at the top of the atmosphere of less than 0.1 Wm^{-2}, but a few simulations were slower to

291   reach full equilibrium. The strength of using multiple constraints is that a simulation may, by chance,

292   pass one or two of these criteria but were unlikely to pass all three tests. For example, all the models

293   that we extended failed at least two of the criteria. The resulting time series of volume integrated

294   global, annual mean ocean temperatures are shown in fig. 3. The supplementary figures also include

295   this for each simulation, as well as the trends at 2731m.

296   The "smooth" CO_2 simulations were all run for 5050 model years and satisfied the criteria. The

297   Foster-CO_2 simulations were initially run for a minimum of 2000 years (starting from the end of the

298   5000-year runs), at which point we reviewed the simulations relative to the convergence criteria. If

299   the simulations had not converged, we extended the runs for an additional 3000 years. If they had

300   not converged at the end of 5000 years, we extended them again for an additional 3000 years. After

301   8000 years, all simulations had converged based on the convergence criteria. In general, the slowest

302   converging simulations corresponded to some of the warmest climates (final temperatures in figure

303   3b and 3c were generally warmer than in figure 3a). It cannot be guaranteed that further changes

304   will not occur; however, we note that the criteria and length of the simulations greatly exceed PMIP-

305   LGM (Kageyama et al., 2017) and PMIP-DeepMIP (Lunt et al., 2017) protocols.

306

3. Results

3.1 Comparison of Deep Ocean Temperatures to Benthic Ocean Data

Before using the model to investigate the linkage of deep ocean temperatures to global mean surface temperatures, it is interesting to evaluate whether the modelled deep ocean temperatures agree with the deep ocean temperatures obtained from the isotopic studies of benthic foraminifera (Friedrich et al., 2012; Zachos et al., 2008). It is important to note that the temperatures are strongly influenced by the choice of CO_2, so we are not expecting complete agreement, but we simply wish to evaluate whether the model is within plausible ranges. If the modelled temperatures were in complete disagreement with data, then it might suggest that the model was too far away from reality to allow us to adequately discuss deep ocean/surface ocean linkages. If the modelled temperatures are plausible, then it shows that we are operating within the correct climate space. A detailed comparison of modelled surface and benthic temperatures to data throughout the Phanerozoic, using multiple CO_2 scenarios, is the subject of a separate ongoing project.

Figure 4a compares the modelled deep ocean temperature to the foraminifera data from the Cenozoic and Cretaceous (115 Ma). The observed isotope data are converted to deep ocean temperature using the procedures described by (Hansen et al., 2013). The modelled deep temperature shown in fig.4a (solid line) is the average temperature at the bottom level of the model, excluding depths less than 1000m (to avoid continental shelf locations which are typically not included in benthic data compilations). The observed benthic data are collected from a range of depths and are rarely at the very deepest levels (e.g. the new cores in (Friedrich et al., 2011) are from current water depths ranging from 1899m to 3192m). Furthermore, large data compilations rarely include how the depth of a particular site changed with time, and thus effectively assume that any differences between basins and through time are entirely due to climate change and not to changes in depth. Hence throughout the rest of the paper we frequently use the modelled 2731m temperatures as a surrogate for the true benthic temperature. This is a pragmatic definition because the area of deep ocean reduces rapidly (e.g. there is typically only 50% of the globe deeper than 3300m). To evaluate whether this procedure gave a reasonable result, we also calculated the global average temperature at the model bottom, and at the model level at a depth of 2731m. The latter is shown by the dashed line in figure 4a. In general, the agreement between model bottom water temperatures and 2731m temperatures is very good. The standard deviation between model bottom water and constant depth of 2731m is 0.7˚C, and the maximum difference is 1.4˚C. Compared to the overall variability, this is a relatively small difference and shows that it is reasonable to assume that the deep ocean has weak vertical gradients.

340   The total change in benthic temperatures over the late Cretaceous and Cenozoic is well reproduced

341   by the model, with the temperatures associated with the "smooth" CO_2 record being particularly

342   good. We do not expect the model to represent sub-stage changes (100,000's of years) such as the

343   PETM excursion or OAEs, but we do expect that the broader temperature patterns should be

344   simulated.

345   Comparison of the two simulations illustrates how strongly CO_2 controls global mean temperature.

346   The Foster-CO_2 driven simulation substantially differs from the estimates of deep-sea temperature

347   obtained from benthic forams and is generally a poorer fit to data. The greatest mismatch between

348   the Foster curve and the benthic temperature curve is during the late Cretaceous and early

349   Paleogene. Both dips in the Foster-CO_2 simulations correspond to relatively low estimates of CO_2

350   concentrations. For these periods, the dominant source of CO_2 values is from paleosols (fig.1) and

351   thus we are reliant on one proxy methodology. Unfortunately, the alternative CO_2 reconstructions of

352   (Witkowski et al., 2018) have a data gap during these periods.

353   A second big difference between the Foster curve and the benthic temperature curve occurs during

354   the Cenomanian-Turonian. This difference is similarly driven by a low estimate of CO_2 in the Foster-

355   CO_2 curve. These low CO_2 values are primarily based on stomatal density indices. As can be seen in

356   figure 1, stomatal indices frequently suggest CO_2 levels lower than estimates obtained by other

357   methods. The CO_2 estimates by (Witkowski et al., 2018) generally supports the higher levels of CO_2

358   (near to 1000 ppmv) that are suggested by the "smooth" CO_2 curve.

359   Both sets of simulations underestimate the warming during the middle Miocene. This issue has been

360   seen before in other models e.g. (You et al., 2009), (Knorr et al., 2011), (Krapp and Jungclaus, 2011)

361   (Goldner et al., 2014) (Steinthorsdottir, 2021). In order to simulate the surface warmth of the middle

362   Miocene (15 Ma), CO_2 concentrations in the range 460–580 ppmv were required, whereas the CO_2

363   reconstructions for this period (Foster et al., 2017) are generally quite low (250-400ppmv). This

364   problem may be either due to the climate models having too low a climate sensitivity or that the

365   estimates of CO_2 are too low (Stoll et al., 2019).

366   The original compilation of (Zachos et al., 2008) represented a relatively small portion of the global

367   ocean and the implicit assumption was made that these results represented the entire ocean basin.

368   (Cramer et al., 2009) examined the data from an ocean basin perspective and suggested that these

369   inter-basin differences were generally small during the Late Cretaceous and early Paleogene (90Ma –

370   35 Ma) and the differences between ocean basins were larger during the late Paleogene and early

371   Neogene. Our model largely also reproduces this pattern. Figure 5 shows the ocean temperature at

372   2731 m during the late Cretaceous (69 Ma), the late Eocene (39 Ma) and the Oligocene (31 Ma) for

373   the "smooth"-CO_2 simulations. In the late Cretaceous, the model temperatures are almost identical

374   in the North Atlantic and Pacific (8˚C – 10˚C). There is warmer deep water forming in the Indian

375   Ocean (deep mixed layer depths, not shown), off the West coast of Australia (10˚C – 12˚C), but

376   otherwise the pattern is very homogeneous. This is in agreement with some paleoreconstructions

377   for the Cretaceous (e.g. (Murphy and Thomas, 2012)).

378   By the time we reach the late Eocene (39 Ma), the North Atlantic and Pacific remain very similar but

379   cooler deep water (6˚C – 8˚C) is now originating in the South Atlantic. The South Atlantic cool

380   bottom water source remains in the Oligocene, but we see a strong transition in the North Atlantic

381   to an essentially modern circulation with the major source of deep, cold water occurring in the high

382   southerly latitudes (3˚C – 5˚C) and strong gradient between the North Atlantic and Pacific.

383   Figure 5 also shows the modelled deep ocean temperatures for present day (Fig. 5d) compared to

384   the World Ocean Atlas Data (fig. 5e). It can be seen that the broad patterns are well reproduced in

385   the model, with good predictions of the mean temperature of the Pacific. The model is somewhat

386   too warm in the Atlantic itself and has a stronger plume from the Mediterranean than is shown in

387   the observations.

388

389   3.2 Comparison of Model Sea Surface Temperature to Proxy Data

390   The previous section focused on benthic temperatures, but it is also important to evaluate whether

391   the modelled sea surface temperatures are plausible (within the uncertainties of the CO_2

392   reconstructions). Figure 4b shows a comparison between the model simulations of sea surface

393   temperature and two published synthesis of proxy SST data. (O'Brien et al., 2017) compiled TEX_{86}

394   and $\delta^{18}O$ for the Cretaceous, separated into tropical and high-latitude (polewards of 48°) regions.

395   (Cramwinckel et al., 2018) compiled early Cenozoic tropical SST data, using Tex_{86}, $\delta^{18}O$, Mg/Ca and

396   clumped isotopes. We compare these to modelled SST for the region 15°S to 15°N, and for the

397   average of Northern and Southern hemispheres between 47.5° and 60°. The proxy data includes

398   sites from all ocean basins and so we also examined the spatial variability within the model. This

399   spatial variability consists of changes along longitude (effectively different ocean basins) and

400   changes with latitude (related to the gradient between equator and pole). We therefore calculated

401   the average standard deviation of SST relative to the zonal mean at each latitude (this is shown by

402   the smaller tick marks) and the total standard deviation of SST relative to the regional average. In

403   practice, the equatorial values are dominated by inter-basin variations and hence the two measures

404   of spatial variability are almost identical. The high latitude variability has a bigger difference

405   between the longitudinal variations and the total variability, because the equator-to-pole

temperature gradient (i.e. the temperatures at the latitude limits of the region are a few degrees warmer/colder than the average). The spatial variability was very similar for the "smooth"-CO_2 and Foster-CO_2 simulations so, for clarity, on figure 4b we only show the results as error bars on the model Foster-CO_2 simulations.

Overall, the comparison between model and data is generally reasonable. The modelled equatorial temperatures largely follow the data, albeit with considerable scatter in the data. Both simulations tend to be towards the warmest equatorial data in the early Cretaceous (Albian). These temperatures largely come from Tex_{86} data. There are many $\delta^{18}O$ based SST which are significantly colder during this period. This data almost exclusively comes from cores 1050/1052 which are in the Gulf of Mexico. It is possible that these data are offset due to a bias in the $\delta^{18}O$ of sea water because of the relatively enclosed region. The Foster-CO_2 simulations are noticeably colder than the data at the Cenomanian peak warmth, which is presumably related to the relatively low CO_2 as discussed for the benthic temperatures. The benthic record also showed a cool (low CO_2) bias in the late Cretaceous. This is not such an obvious feature of the surface temperatures. The Foster simulations are colder than the "smooth"-CO_2 simulations during the late Cretaceous but there is not a strong mismatch between model and data. Both simulations are close to the observations, though the "smooth"-CO_2 simulations better matches the high-latitude data (but is slightly poorer with the tropical data).

The biggest area of disagreement between model and data is at high latitudes in the mid-Cretaceous warm period. In common with previous work with this model in the context of the Eocene (Lunt et al., 2021) the model is considerably cooler than the data, with a 10-15°C mismatch between models and data. The polar sea surface temperature estimates may have a seasonal bias because productivity is likely to be higher during the warmer summer months and, if we select the summer season temperatures from the model, then the mismatch is slightly reduced by about 4°C. The problem of a cool high latitudes in models is seen in many model studies and there is increasing evidence that this is related to the way that the models simulate clouds ((Kiehl and Shields, 2013); (Sagoo et al., 2013); (Upchurch et al., 2015; Zhu et al., 2019)). Of course, in practice deep water is formed during winter so the benthic temperatures do not suffer from a summer bias.

3.3 Correlation of Deep Ocean Temperatures to Polar Sea Surface Temperatures

The previous sections showed that that the climate model was producing a plausible reconstruction of past ocean temperature changes, at least within the uncertainties of the CO_2 estimates. We now use the HadCM3L model to investigate the links between deep ocean temperature and global mean surface temperature.

439   In theory, the deep ocean temperature should be correlated with the sea surface temperature at the

440   location of deep-water formation which is normally assumed to be high latitude surface waters in

441   winter. We therefore compare deep ocean temperatures (defined as the average temperature at the

442   bottom of the model ocean, where the bottom must be deeper than 1000 m) with the average

443   winter sea surface temperature polewards of 60° (fig. 6). Winter is defined as December, January,

444   and February in the northern hemisphere and June, July, and August in the southern hemisphere.

445   Also shown in Figure 6 is the best fit line, which has a slope of 0.40 (+/-0.05 at the 97.5% level), an r^2

446   of 0.59, and a standard error of 1.2°C. We obtained very similar results when we compared the polar

447   sea surface temperatures with the average temperature at 2731m instead of the true benthic

448   temperatures. We also compared the deep ocean temperatures to the mean polar sea surface

449   temperatures when the mixed layer depth exceeded 250 m (poleward of 50°). The results were

450   similar although the scatter was somewhat larger (r^2=0.48).

451   Overall, the relationship between deep ocean temperatures and polar sea surface temperatures is

452   clear (Figure 6) but there is considerable scatter around the best fit line, especially at the high end,

453   and the slope is less steep than perhaps would be expected (Hansen and Sato, 2012). The scatter is

454   less for the Cenozoic and late Cretaceous (up to 100 Ma: green and orange dots and triangles). If we

455   used only Cenozoic and late Cretaceous simulations, then the slope is similar (0.43) but r^2=0.92 and

456   standard error=0.47°C. This provides strong confirmation that benthic data is a robust

457   approximation to polar surface temperatures when the continental configuration is similar to the

458   present.

459   However, the scatter is greater for older time periods, with the largest divergence observed for the

460   warm periods of the Triassic and early Jurassic, particularly for the Foster CO_2 simulations (purple

461   and blue dots). Examination of climate models for these time periods reveals relatively sluggish and

462   shallow ocean circulation, with weak horizontal temperature gradients at depth (though salinity

463   gradients can still be important, (Zhou et al., 2008)). For instance, in the Ladinian stage, mid-Triassic

464   (~240Ma) the overturning circulation is extremely weak (Fig. 7). The maximum strength of the

465   northern hemisphere overturning cell is less than 10 Sv and the southern cell is less than 5 Sv. Under

466   these conditions, deep ocean water does not always form at polar latitudes. Examination of the

467   mixed layer depth (not shown) shows that during these time periods, the deepest mixed layer

468   depths are in the sub-tropics. In subtropics, there is very high evaporation relative to precipitation

469   (due to the low precipitation and high temperatures). This produces highly saline waters that sinks

470   and spread out into the global ocean.

471   The idea that deep water may form in the tropics is in disagreement with early hypothesis (e.g.

472   (Emiliani, 1954)) but they were only considering the Tertiary and our model does not simulate any

473   low latitude deep water formation during this period. We only see significant tropical deep water

474   formation for earlier periods, and this has previously been suggested as a mechanism for warm

475   Cretaceous deep water formation (Brass et al., 1982), (Kennett and Stott, 1991). Deep water

476   typically forms in convective plumes. (Brass et al., 1982) showed that the depth and spreading of

477   these plumes is related to the buoyancy flux with the greatest flux leading to bottom water and

478   plumes of lesser flux leading to intermediate water. (Brass et al., 1982) suggested that this could

479   occur in warm conditions in the tropics, particularly if there was significant epicontinental seaways

480   and hypothesised that it "has been a dominant mechanism of deep-water formation in historical

481   times". It is caused by a strong buoyancy flux linked to strong evaporation at high temperatures.

482

483   Our computer model simulations are partly consistent with this hypothesis. The key aspect for the

484   model is a relatively enclosed seaway in the tropics and warm conditions. The paleogeographic

485   reconstructions (see supplementary figures) suggest an enclosed Tethyan-like seaway starting in the

486   Carboniferous and extending through to the Jurassic and early Cretaceous. However, the colder

487   condition of the Carboniferous prevents strong tropical buoyancy fluxes. When we get into the

488   Triassic and Jurassic, the warmer conditions lead to strong evaporation at low latitudes and bottom

489   water formation in the tropics. This also explains why we see more tropical deep water (and hence

490   poorer correlations between deep and polar surface temperatures in figure 6) when using the Foster

491   CO2 since this is generally higher (and hence warmer) than the smoothed CO_2 record.

492

493   An example of the formation of tropical deep water is shown in fig. 8. This shows a vertical cross-

494   section of temperature and salinity near the equator for the Ladinian stage, mid-Triassic (240Ma).

495   The salinity and temperature cross-section clearly shows high salinity warm waters sinking to the

496   bottom of the ocean and spreading out. This is further confirmed by the water age tracer, fig. 9. This

497   shows the water age (measured as time since it experienced surface conditions, see (England, 1995))

498   at 2731m in the model for the Permian, Triassic, Cretaceous and present day. The present-day

499   simulation shows that the youngest water is in the N. Atlantic and off the coast of Antarctica,

500   indicating that this is where the deep water is forming. By contrast, the Triassic period shows that

501   the youngest water is in the tropical Tethyan region and that it spreads out from there to fill the rest

502   of the ocean basin. There is no young water at high latitudes, confirming that the source of bottom

503   water is tropical only. For the Permian, although there continues to be a Tethyan-like tropical

504   seaway, the colder conditions mean that deep water is again forming at high latitudes only. The

505   Cretaceous is more complicated. It shows younger water in the high latitudes, but also shows some

506   young water in the Tethys which merges with the high latitude waters. Additional indicators of the

507   transitional nature of the Cretaceous are the mixed layer depth (see supplementary figures). This is

508   a measure of where water is mixing to deeper levels. For this time period, there are regions of deep

509   mixed layer in both the tropics and high latitudes, whereas it is only deep in the tropics for the

510   Triassic and at high latitudes for present day.

511

512   This mechanism for warm deep water formation has also been seen in other climate models (e.g.

513   (Barron and Peterson, 1990)). However, (Poulsen et al., 2001) conclude that in his model of the

514   Cretaceous high-latitudes sources of deep water diminish with elevated CO_2 concentrations but did

515   not see the dominance of tropical sources. Other models (e.g. (Ladant et al., 2020)) do not show any

516   significant tropical deep-water formation, suggesting that this feature is potentially a model-

517   dependent result.

518   The correlation between deep ocean temperatures and the temperature of polar surface waters

519   differs between the "smooth" CO_2 simulations and the Foster CO_2 simulations. The slope is only 0.30

520   (r^2=0.57) for the "smooth" CO_2 simulations whereas the slope is 0.48 (r^2=0.65) for the Foster

521   simulations. This is because CO_2 is a strong forcing agent that influences both the surface and deep

522   ocean temperatures. By contrast, if the CO_2 does not vary as much, then the temperature does not

523   vary as much, and the influence of paleogeography becomes more important. These

524   paleogeographic changes generally cause subtle and complicated changes in ocean circulation that

525   affect the location and latitude of deep-water formation.

526   In contrast, the mid-Cretaceous is also very warm but the continental configuration (specifically, land

527   at high southern latitudes) favours the formation of cool, high latitude deep water. Throughout the

528   Cretaceous there is significant southern high latitude source of deep water and hence deep-water

529   temperatures are well correlated with surface high latitude temperatures. The strength of this

530   connection, however, may be over exaggerated in the model. Like many climate models, HadCM3

531   underestimates the reduction in the pole-to-Equator sea surface temperature (Lunt et al., 2012),

532   (Lunt et al., 2021). This means that during the Cretaceous the high latitudes are probably too cold.

533   Consequently, some seasonal sea ice does form which encourages the formation of cold deep-water,

534   via brine rejection.

535   In the late Eocene (~40 Ma), the ocean circulation is similar to the Cretaceous, but the strong

536   southern overturning cell is closer to the South Pole, indicating that the main source of deep water

537   has moved further polewards. The poleward movement of the region of downwelling waters

538   explains some of the variability between deep ocean temperatures and temperature of polar surface

539   waters.

540   For reference, we also include the present-day meridional circulation. The modern southern

541   hemisphere circulation is essentially a strengthening of late Eocene meridional circulation. The

542   Northern hemisphere is dominated by the Atlantic meridional overturning circulation. The Atlantic

543   circulation pattern does not resemble the modern pattern of circulation until the Miocene.

544   3.4 Surface Polar Amplification

545

546   The conceptual model used to connect benthic ocean temperatures to global mean surface

547   temperatures assumes that there is a constant relationship between high latitude sea surface

548   temperatures and global mean annual mean surface air temperature. (Hansen and Sato, 2012)

549   argue that this amplification is partly related to ice-albedo feedback but also includes a factor

550   related to the contrasting amplification of temperatures on land compared to the ocean. To

551   investigate the stability of this relationship, fig. 10 shows the correlation between polar winter sea

552   surface temperatures (60° - 90°) and global mean surface air temperature. The polar temperatures

553   are the average of the two winter hemispheres (i.e. average of DJF polar SSTs in the Northern

554   hemisphere and JJA polar SSTs in the Southern hemisphere). Also shown is a simple linear

555   regression, with an average slope of 1.3 and with an r^2 = 0.79. If we only use Northern polar winter

556   temperatures, the slope is 1.1; if we only use Southern polar winter temperatures, then the slope is

557   0.7. Taken separately, the scatter about the mean is considerably larger (r^2 of 0.5 and 0.6

558   respectively) than the scatter if both data sets are combined (r^2 = 0.79).   The difference between the

559   southern and northern hemisphere response complicates the interpretation of the proxies and leads

560   to potentially substantial uncertainties.

561   As expected, there appears to be a strong non-linear component to the correlation. There are two

562   separate regimes: 1) one with a steeper slope during colder periods (average polar winter

563   temperature less than about 1^oC), and 2) a shallower slope for warmer conditions. This is strongly

564   linked to the extent of sea-ice cover. Cooler periods promote the growth of sea-ice which

565   strengthens the ice-albedo feedback mechanism resulting in a steeper overall temperature gradient

566   (strong polar amplification). Of course, the ocean sea surface temperatures are constrained to be -

567   2^oC but an expansion of sea ice moves this further equatorward. Conversely, the warmer conditions

568   result in less sea ice and hence a weaker sea ice-albedo feedback resulting in a weaker temperature

569   gradient (reduced polar amplification). This suggests that using a simple linear relationship (as in

570   (Hansen et al., 2008)) could be improved upon.

571   Examining the Foster CO_2 and "smooth" CO_2 simulations reveals an additional factor. If we examine

572   the "smooth" CO_2 simulations only, then the best fit linear slope is slightly less than the average

573   slope (1.1 vs 1.3). This can be explained by the fact that we have fewer very cold climates

574   (particularly in the Carboniferous) due to the relatively elevated levels of CO_2. However, the scatter

575   in the "smooth" CO_2 correlation is much larger, with an r^2 of only 0.66. By comparison, correlation

576   between global mean surface temperature and polar sea surface temperature using the Foster CO_2

577   has a similar overall slope to the combined set and a smaller amount of scatter. This suggests that

578   CO_2 forcing and polar amplification response have an important impact on the relationship between

579   global and polar temperatures. The variations of carbon dioxide in the Foster set of simulations are

580   large and they drive large changes in global mean temperature. Conversely significant sea-ice albedo

581   feedbacks characterize times when the polar amplification is important. There are several well

582   studied processes that lead to such changes, including albedo effects from changing ice but also

583   from poleward heat transport changes, cloud cover, and latent heat effects ((Alexeev et al., 2005;

584   Holland and Bitz, 2003; Sutton et al., 2007)). By contrast, the "smooth" CO_2 simulations have

585   considerably less forcing due to CO_2 variability which leads to a larger paleogeographic effect. For

586   instance, when there is more land at the poles, there will be more evaporation over the land areas

587   and hence simple surface energy balance arguments would suggest different temperatures ((Sutton

588   et al., 2007)) .

589   In figure 10, there are a few data points which are complete outliers. These correspond to

590   simulations in the Ordovician; the outliers happen irrespective of the CO_2 model that is used.

591   Inspection of these simulations shows that the cause for this discrepancy is related to two factors:

592   1) a continental configuration with almost no land in the Northern hemisphere and, 2) a

593   reconstruction which includes significant southern hemisphere ice cover (see fig.1 and fig 2).

594   Combined, these factors produced a temperature structure which is highly non-symmetric, with the

595   Southern high latitudes being more than 20°C colder than the Northern high latitudes. This anomaly

596   biases the average polar temperatures shown in figure 10.

597   3.5 Deep Ocean Temperature versus Global Mean Temperature

598

599   The relationships described above help to understand the overall relationship between deep ocean

600   temperatures and global mean temperature. Figure 11 shows the correlation between modelled

601   deep ocean temperatures (> 1000 m) and global mean surface air temperature, and figure 12 shows

602   a comparison of changes in modelled deep ocean temperature compared to model global mean

603   temperature throughout the Phanerozoic.

604   The overall slope is 0.64 (0.59 to 0.69) with an $r^2 = 0.74$. If we consider the last 115 Ma (for which

605   exists compiled benthic temperatures), then the slope is slightly steeper (0.67 with an $r^2 = 0.90$).

606   Similarly, the "smooth"-CO_2 and the Foster-CO_2 simulation results have very different slopes. The

607   "smooth"-CO_2 simulations have a slope of 0.47, whereas the Foster-CO_2 simulations have a slope of

608   0.76. The root mean square departure from the regression line in figure 11 is 1.3^{o}C. Although we

609   could have used a non-linear fit as we might expect such a relationship if the pole-to-equator

610   temperature gradient changes, all use of benthic temperatures as a global mean surface

611   temperature proxy are based on linear relationship.

612

613   The relatively good correlations in the fig.11 are confirmed when examining fig.12a and 12b. On

614   average, the deep ocean temperatures tend to underestimate the global mean change (fig.12b)

615   which is consistent with the regression slope being less than 1. However, the errors are substantial

616   with largest errors occurring during the pre-Cretaceous and can be 4-6 ^{o}C. This is an appreciable

617   error that would have a substantial impact on estimates of climate sensitivity. Even within the late

618   Cretaceous and Cenozoic, the errors can exceed 2°C which can exceed 40% of the total change.

619   The characteristics of the plots can best be understood in terms of figures (6 and 10). For instance,

620   most of the Carboniferous simulations plot below the regression line because the polar SSTs are not

621   well-correlated with the global mean temperature (figure 10). By contrast, the Triassic and Jurassic

622   Foster CO_2 simulations plot above the regression line because the deep ocean temperature is not

623   well-correlated with the polar temperatures (figure 6).

624   4. Discussion and Conclusion

625   The paper has presented the results from two unique sets of paleoclimate simulations covering the

626   Phanerozoic. The focus of the paper has been to use the HadCM3L climate model to evaluate how

627   well we can predict global mean surface temperatures from benthic foram data. This is an important

628   consideration because benthic microfossil data are one of the few datasets used to directly estimate

629   past global mean temperatures. Other methods, such as using planktonic foraminiferal estimates,

630   are more challenging because the sample sites are geographically sparse, so it is difficult to

631   accurately estimate the global mean temperature from highly variable and widely dispersed data.

632   This is particularly an issue for older time periods when fewer isotopic measurements from

633   planktonic microfossils are available and can result in a bias because most of the isotopic

634   temperature sample localities are from tropical latitudes (30˚S – 30˚N) (Song et al., 2019).

635   By contrast, deep ocean temperatures are more spatially uniform. Hence. benthic foram data has

636   frequently been used to estimate past global mean temperatures and climate sensitivity (Hansen et

637   al., 2013). Estimates of uncertainty for deep ocean temperatures incorporate uncertainties from CO_2

638   and from the conversion of $\delta^{18}O$ measurements to temperature but have not been able to assess

639   assumptions about the source regions for deep ocean waters and the importance polar

640   amplification. Of course, in practice, lack of ocean sea floor means that benthic compilations exist

641   only for the last 110Ma.

642   Changes in heat transport also play a potentially important role in polar amplification. In the

643   supplementary figure, we show the change in atmosphere and ocean poleward heat fluxes for each

644   time period. Examination of the modelled poleward heat transport by the atmosphere and ocean

645   shows a very complicated pattern, with all time periods showing the presence of some Bjerknes

646   compensation (Bjerknes, 1964) (see (Outten et al., 2018) for example in CMIP5 models). Bjerknes

647   compensation is where the change in ocean transport is largely balanced by an equal but opposite

648   change in atmospheric transport. For instance, compared to present day, the mid-Cretaceous and

649   Early Eocene warm simulations shows a large increase in northward atmospheric heat transport,

650   linked with enhanced latent heat transport associated with the warmer, moister atmosphere.

651   However, this is partly cancelled by an equal but opposite change in the ocean transport. E.g.

652   compared to present day, the early Eocene northern hemisphere atmospheric heat transport

653   increases by up to 0.5PW, but the ocean   transport is reduced by an equal amount. The net

654   transport from equator to the N. Pole changes by less than 0.1PW (i.e. less than 2% of total). Further

655   back in time, the compensation is still apparent, but the changes are more complicated, especially

656   when the continents are largely in the Southern hemisphere. Understanding the causes of these

657   transport changes will be the subject of another paper.

658   We have shown that although the expected correlation between benthic temperatures and high-

659   latitude surface temperatures exists, the correlation has considerable scatter. This is caused by

660   several factors. Changing paleogeographies results in changing locations for deep water formation.

661   Some paleogeographies result in significant deep-water formation in the Northern hemisphere (e.g.

662   our present-day configuration) although for most of the Phanerozoic, the dominant source of deep-

663   water formation has been southern hemisphere. Similarly, even when deep water is formed in just

664   one hemisphere, there can be substantial regional and latitudinal variations in its location and the

665   corresponding temperatures. Finally, during times of very warm climates (e.g. mid-Cretaceous) the

666   overturning circulation can be very weak and there is a marked decoupling between the surface

667   waters and deep ocean. In the HadCM3 model during hothouse time periods, high temperatures and

668   high rates of evaporation produce hot and saline surface waters which sink to become intermediate

669   and deep waters at low latitudes.

670   Similar arguments can be made regarding the link between global mean temperature and the

671   temperature at high latitudes. Particularly important is the area of land at the poles and the extent

672   of sea ice/land ice. Colder climates and paleogeographic configurations with more land at the pole

673   will result in a steeper latitudinal temperature gradient and hence exhibit a changing relationship

674   between polar and global temperatures. But the fraction of land versus ocean is also important.

675   Finally, the overall relationship between deep ocean temperatures and global mean temperature is

676   shown to be relatively linear, but the slope is quite variable. In the model simulations using the

677   "smooth" CO_2 curve, the slope is substantially shallower (0.48) than slope obtained using the Foster

678   CO_2 curve (0.76). This is related to the different controls that CO_2 and paleogeography exert (as

679   discussed above). In the simulation that uses the "smooth" CO_2 data set, the levels of CO_2 do not

680   vary much, so the paleogeographic controls are more pronounced.

681   This raises the interesting conundrum that when trying to use reconstructed deep ocean

682   temperatures and CO_2 to estimate climate sensitivity, the interpreted global mean temperature also

683   depends, in part, on the CO_2 concentrations. However, if we simply use the combined slope, then

684   the root mean square error is approximately 1.4°C, and the maximum error is over 4°C. The root

685   mean square error is a relatively small compared to the overall changes and hence the resulting

686   uncertainty in climate sensitivity associated with this error is relatively small (~15%) and the CO_2

687   uncertainty dominates. However, the maximum error is potentially more significant.

688   Our work has not addressed other sources of uncertainty. In particular, it would be valuable to use a

689   water isotope-enabled climate model to better address the uncertainties associated with the

690   conversion of the observed benthic $\delta^{18}O$ to temperature. This requires assumptions about the $\delta^{18}O$

691   of sea water. We hope to perform such simulation in future work, though this is a particularly

692   challenging computational problem because the isotope enabled model is significantly slower and

693   the completion of the multi-millennial simulations required for deep ocean estimates would take

694   more than 18 months to complete.

695   Our simulations extend and develop those published by (Lunt et al., 2016), and (Farnsworth et al.,

696   2019a; Farnsworth et al., 2019b). The simulations reported in this paper used the same climate

697   model (HadCM3L) but used an improved ozone concentration and corrected a salinity drift that can

698   lead to substantial changes over the duration of the simulation. Our simulations also use an

699   alternative set of geographic reconstructions that cover a larger time period (540 Ma – Modern).

700   They also include realistic land ice cover estimates, which were not included in the original

701   simulations (except for the late Cenozoic) but generally have a small impact in the Mesozoic.

702 Similarly, the new simulations use two alternative models for past atmospheric CO_2 use more

703 realistic variations in CO_2 through time (compared with idealised constant values in Farnsworth et al

704 and Lunt et al), while at the same time recognizing the levels of uncertainty. Although the Foster CO_2

705 curve is more directly constrained by CO_2 data, it should be noted that this data come from multiple

706 proxies and there are large gaps in the data set. There is evidence that the different proxies have

707 different biases, and it is not obvious that the correct approach is to simply fit a Loess-type curve to

708 the CO_2 data. This is exemplified by the Maastrichtian. The Foster Loess curve shows a minimum in

709 CO_2 during the Maastrichtian which results in the modelled deep ocean temperatures being much

710 too cold. However, detailed examination of the CO_2 data shows most of the Maastrichtian data is

711 based on stomatal index reconstructions which often are lower than other proxies. Thus, the

712 Maastrichtian low CO_2, relative to other periods, is potentially driven by changing the proxy rather

713 than by real temporal changes.

714 Though the alternative, "smooth" CO_2 curve is not the optimum fit to the data, it does pass through

715 the cloud of individual CO_2 reconstructions and hence represents one possible "reality". For the Late

716 Cretaceous and Cenozoic, the "smooth" CO_2 simulation set does a significantly better job simulating

717 the deep ocean temperatures of the Friedrich/Cramer/Zachos curve.

718 Although the focus of the paper has been the evaluation of the modelled relationship between

719 benthic and surface temperatures, the simulations are a potentially valuable resource for future

720 studies. This includes using the simulations for paleoclimate/climate dynamic studies and for climate

721 impact studies, such as ecological niche modelling. We have therefore made available on our

722 website the results from our simulations

723 (https://www.paleo.bristol.ac.uk/ummodel/scripts/papers/Valdes_et_al_2021.html)

724 Data Availability

725 All simulation data is available from:
726 https://www.paleo.bristol.ac.uk/ummodel/scripts/papers/Valdes_et_al_2021.html

727 Author contributions
728 Study was developed by all authors. All model simulations were performed by PJV who also
729 prepared the manuscript with contributions from all co-authors.

730 Competing interests

731 The authors declare that they have no conflict of interest.

732

733

Acknowledgments.

DJL and PJV acknowledge funding from NERC through NE/P013805/1. The production of paleogeographic digital elevation models was funded by the sponsors of the PALEOMAP Project. This work is part of the PhanTASTIC project led by Scott Wing and Brian Huber from the Smithsonian Institution's National Museum of Natural History and was initiated at a workshop supported by Roland and Debra Sauermann. This work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol (http://www.bris.ac.uk/acrc/). The authors declare that they have no competing interests. Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper. Model data can be accessed at www.bridge.bris.ac.uk/resources/simulations.

744

745

746

747

748

749

751   **Figure 1.** Summary of boundary condition changes to model of the Phanerozoic, (a) CO_2

752   reconstructions (from Foster et al. 2017) and the two scenarios used in the models, (b) Land-sea

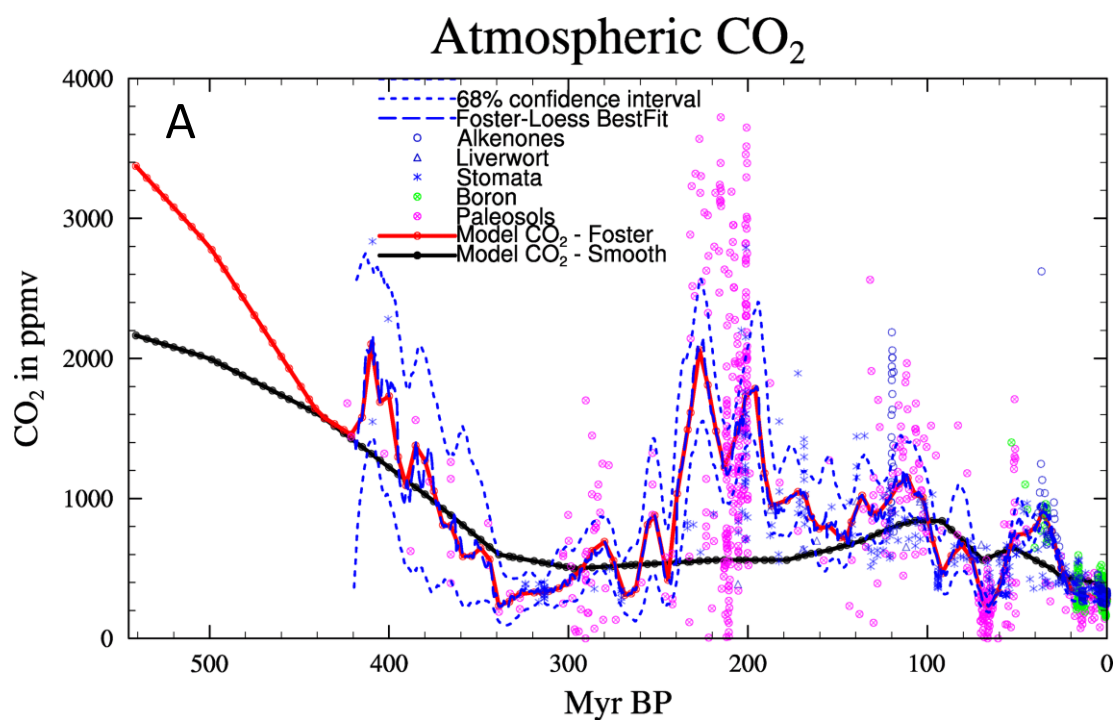
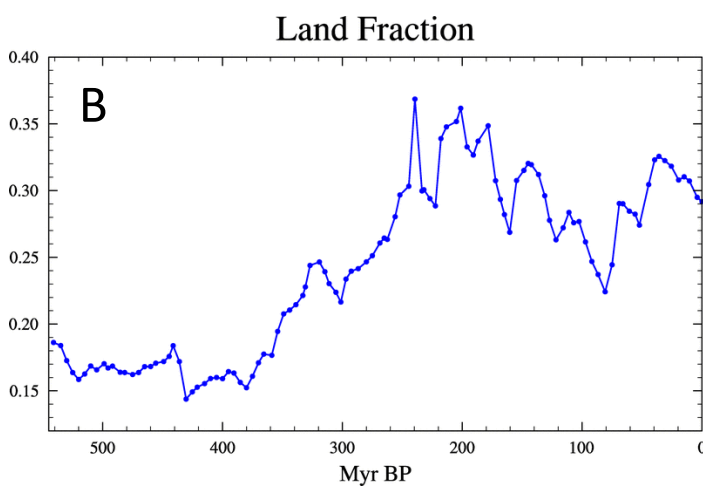
753   fraction from the paleogeographic reconstructions, and (c) land ice area input into model. The

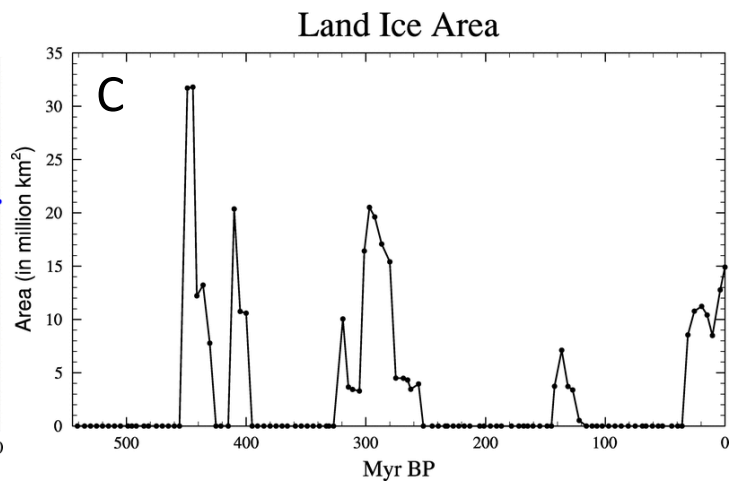
754   paleogeographic reconstructions can be accessed at https://www.earthbyte.org/paleodem-

755   resource-scotese-and-wright-2018/. An animation of the high-resolution (1° x 1°) and model

756   resolution (3.75° longitude x 2.5° latitude) maps can be found here:

757   https://www.paleo.bristol.ac.uk/~ggpjv/scotese/scotese_raw_moll.normal_scotese_moll.normal.ht

758   ml

759

760

761

762   **Figure 2.** A few example paleogeographies, once they have been re-gridded onto the HadCM3L grid.

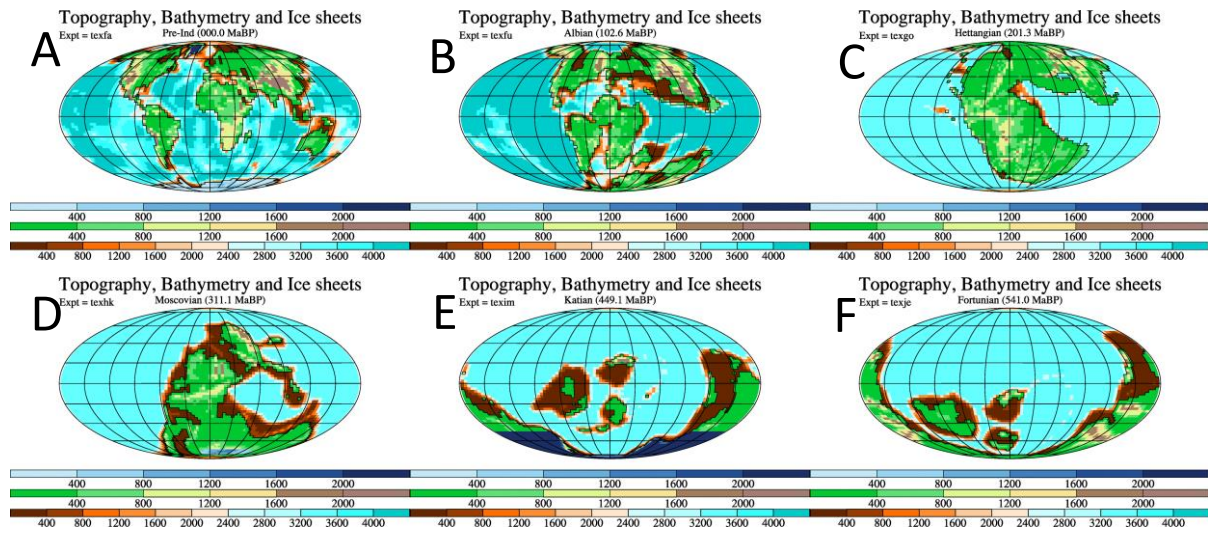
763   The examples are for (a) present day, (b) Albian, 102.6Ma (Lower Cretaceous), (c) Hettangian,

764   201.3Ma (lower Jurassic), (d) Moscovian, 311.1Ma (Pennsylvanian, Carboniferous), (e) Katian,

765   449.1Ma (Upper Ordovician), and (f) Fortunian, 541.0Ma (Cambrian). The top color legend refers to

766   the height of the ice sheets (if they exist), the middle color legend refers to heights on land (except

767   ice), and the lower color legend refers to the ocean bathymetry. All units are meters.

768

769

770

771

772

773

774

775

776

777

778

779

780 **Figure 3.** Time series of the annual, volume mean ocean temperature for all 109 simulations. (a)

781 shows those simulations for which 2000 years was sufficient to satisfy the convergence criteria

782 described in text (these were for all simulations listed in table 1 except those listed in (b) and (c)), (b)

783 those simulation which required 5000 years (these were for all the simulations for 31.0, 35.9, 39.5,

784 55.8, 60.6, 66.0, 69.0, 102.6, 107.0, 121.8, 127.2, 154.7, 160.4, 168.2, 172.2, 178.4, 186.8, 190.8,

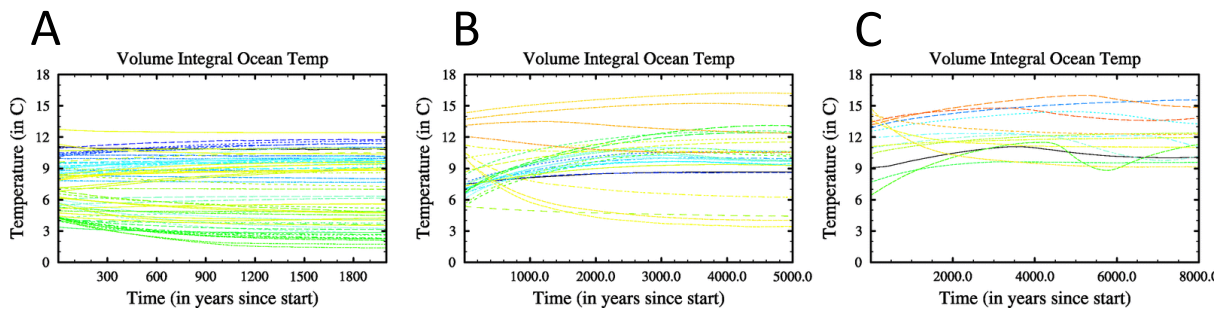
785 196.0, 201.3, 204.9, 213.2, 217.8, 222.4, 227.0, 232.0, and 233.6 Ma BP), and (c) those simulation

786 which required 8000 years (these were simulations for 44.5, 52.2, 86.7, 91.9, 97.2, 111.0, 115.8,

787 131.2, 136.4, 142.4, 145.0, 148.6, 164.8, and 239.5 Ma BP). The different coloured lines show the

788 different runs. The plot simply show the extent to which all runs have reached steady state. For

789 more details about specific simulations, please see the supplementary figures.

790

791

792

793

794   **Figure 4.** (a) Comparison of modelled deep ocean temperatures versus those from (Zachos et al.,

795   2008) and (Friedrich et al., 2012) converted to temperature using the formulation in (Hansen et al.,

796   2013). The model temperatures are global averages over the bottom layer of the model but excludes

797   shallow marine settings (less than 1000m). The dashed lines show the modelled global average

798   ocean temperatures at the model layer centered at 2731m, and (b) Comparison of modelled sea

799   surface temperatures with the compilations of (O'Brien et al., 2017) and (Cramwinckel et al., 2018).

800   The data is a combination of Tex_{86}, $\delta^{18}O$, Mg/Ca, and clumped Isotope data. The model data shows

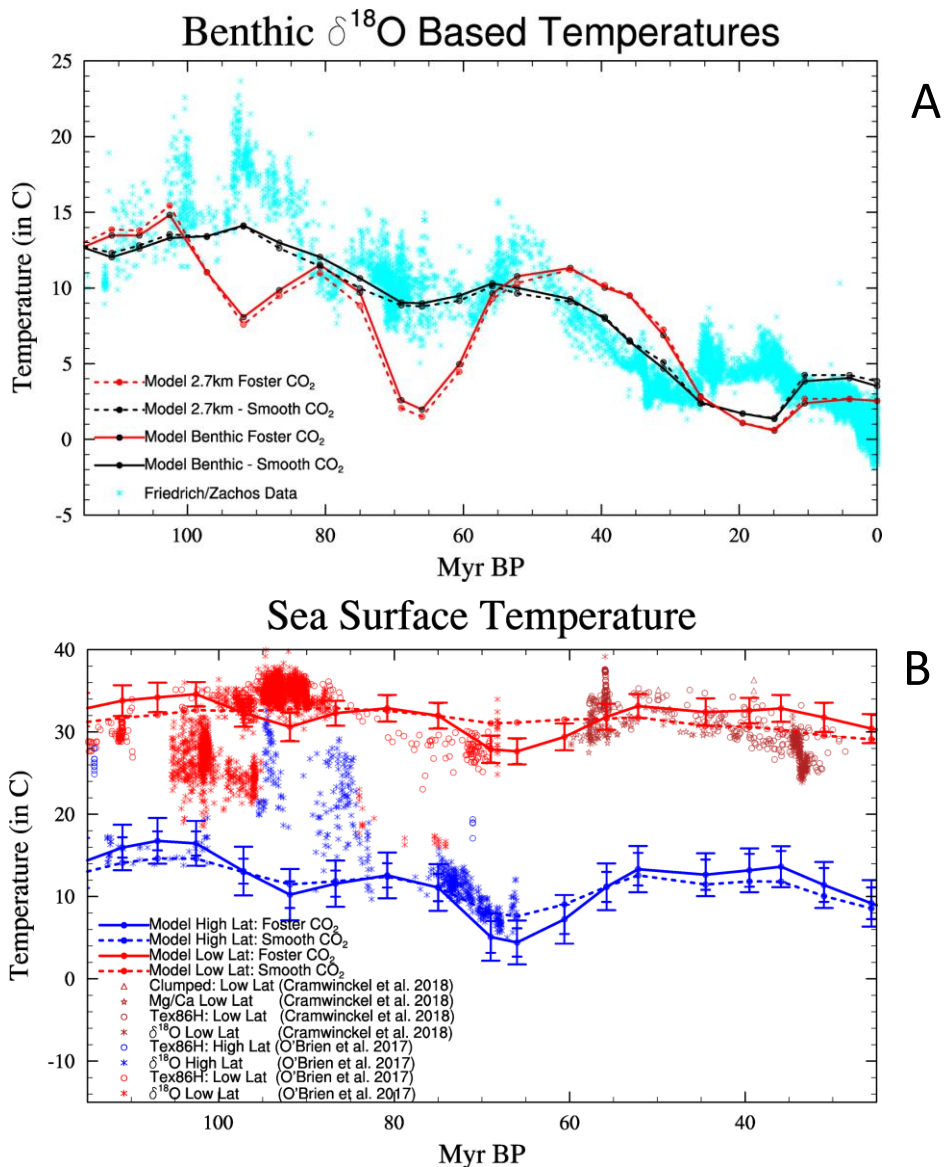
801   low latitude temperatures (averaged from 10S to 10N) and high latitude temperatures (averaged

802   over 47.5N to 65N and 47.5S to 65S). The Foster-CO_2 simulations also show a measure of the spatial

803   variability. The large bars show the spatial standard deviation across the whole region, and the

804   smaller bars shows the average spatial standard deviation along longitudes within the region. Note

805   that the ranges of both the x and y-axis differ between (a) and (b).

806

807 **Figure 5.** Modelled annual mean ocean temperatures are 2731m depth for three example past time

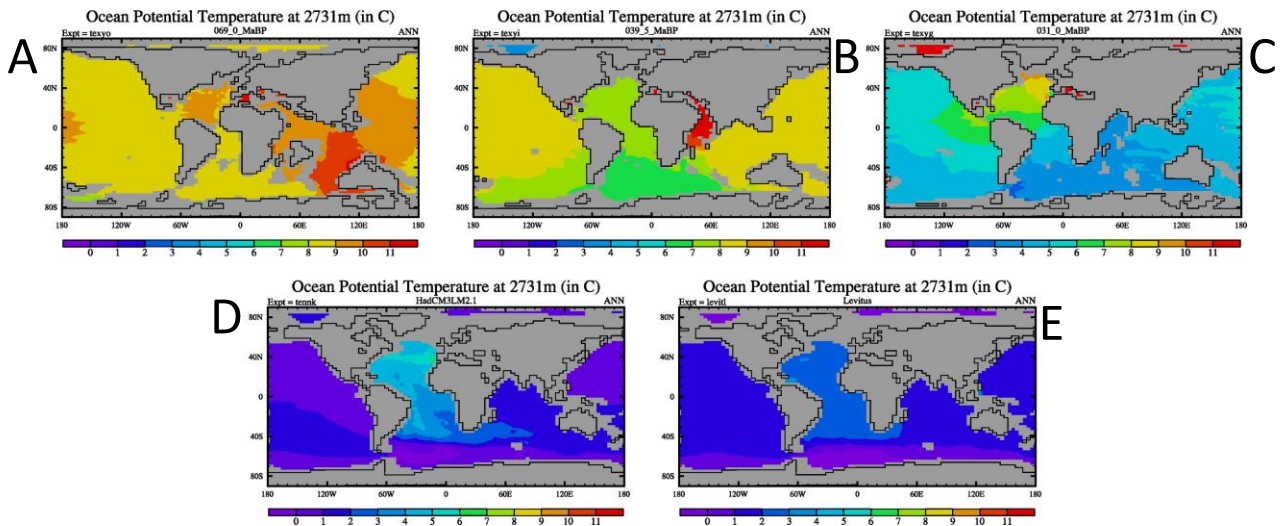
808 periods. The left figure is for the late Cretaceous, the center for the late Eocene (39.5Ma), and the

809 right for the Oligocene (31Ma). These are results from the smooth-CO_2 set of simulations which

810 agree better with the observed benthic temperature data. Also included are the pre-industrial

811 simulation and World Ocean Atlas 1994 observational data, provided by the NOAA-ESRL Physical

812 Sciences Laboratory, Boulder Colorado from their web site at https://psl.noaa.gov/. The thin black

813 lines show the coastlines, and the grey areas are showing where the ocean is shallower than 2731m.

814

815

816   **Figure 6.** Correlations between deep ocean temperatures and surface polar sea surface

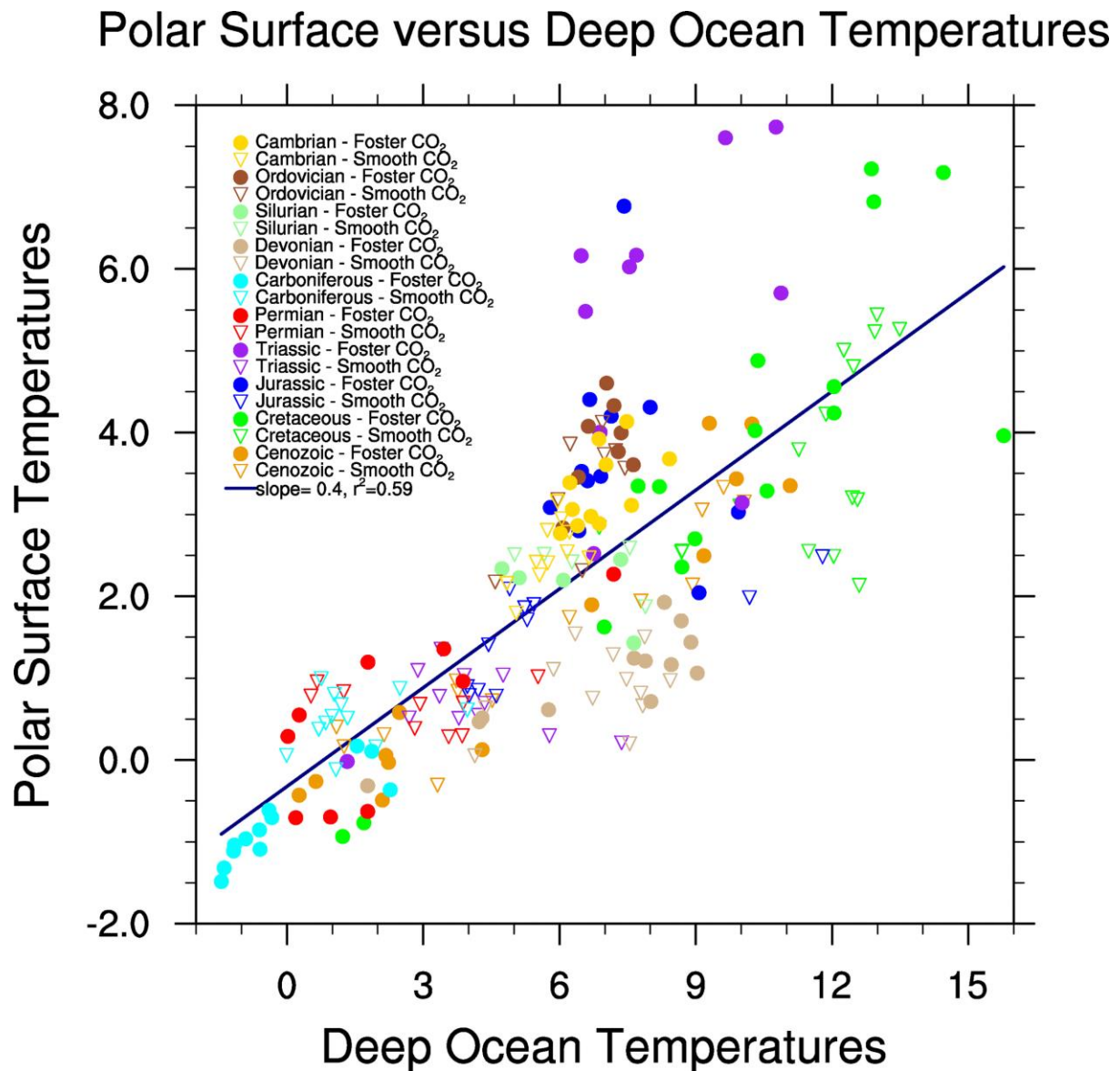
817   temperatures. The deep ocean temperatures are defined as the average temperature at the bottom

818   of the model ocean, where the bottom must be deeper than 1000m. The polar sea surface

819   temperatures are the average winter (i.e. northern polar in DJF and southern polar in JJA) sea

820   surface temperature polewards of 60°. The inverted triangles show the results from the smooth CO_2

821   simulations and the dots refer to the Foster CO_2 simulations. The colors refer to different geological

822   era.

823

824 **Figure 7.** Global Ocean overturning circulation (in Sverdrup) for four different time periods for the

Foster-CO$_2$ simulations. Positive (yellow/red) values correspond to a clockwise circulation, negative

(dark blue/purple) values represent an anti-clockwise circulation. (a) Middle Triassic, Ladinian,

239.5Ma, (b) Lower Cretaceous, Aptian, 121.8 Ma, (c) Late Eocene, Bartonian, 39.5Ma, and (d)

Present Day. Paleogeographic reconstructions older than the oldest ocean floor (~Late-Jurassic) have

uniform deep ocean floor depth.

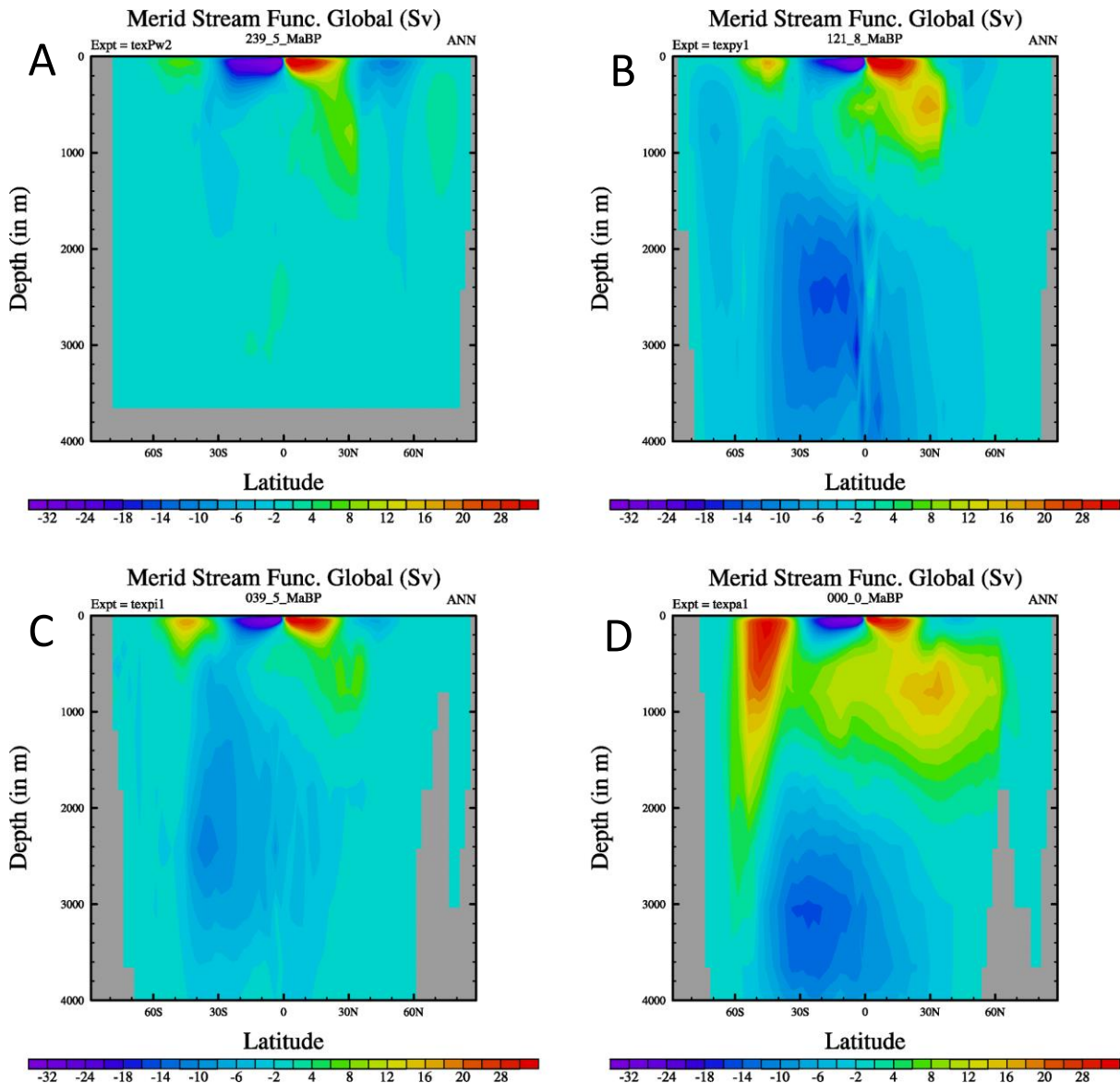
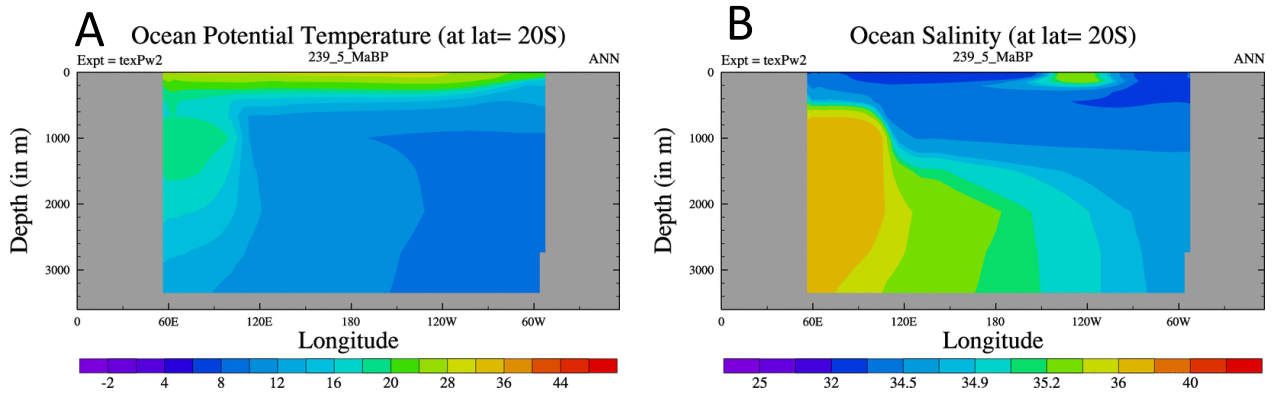
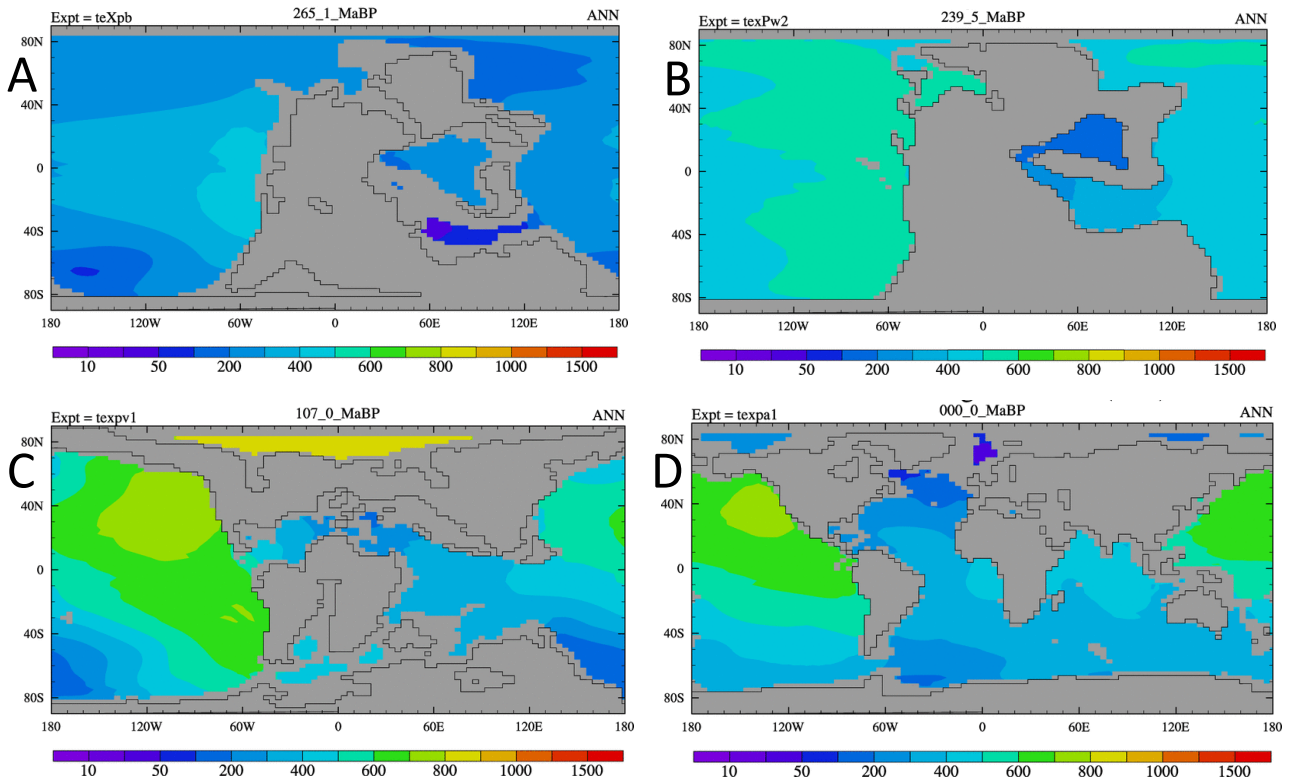



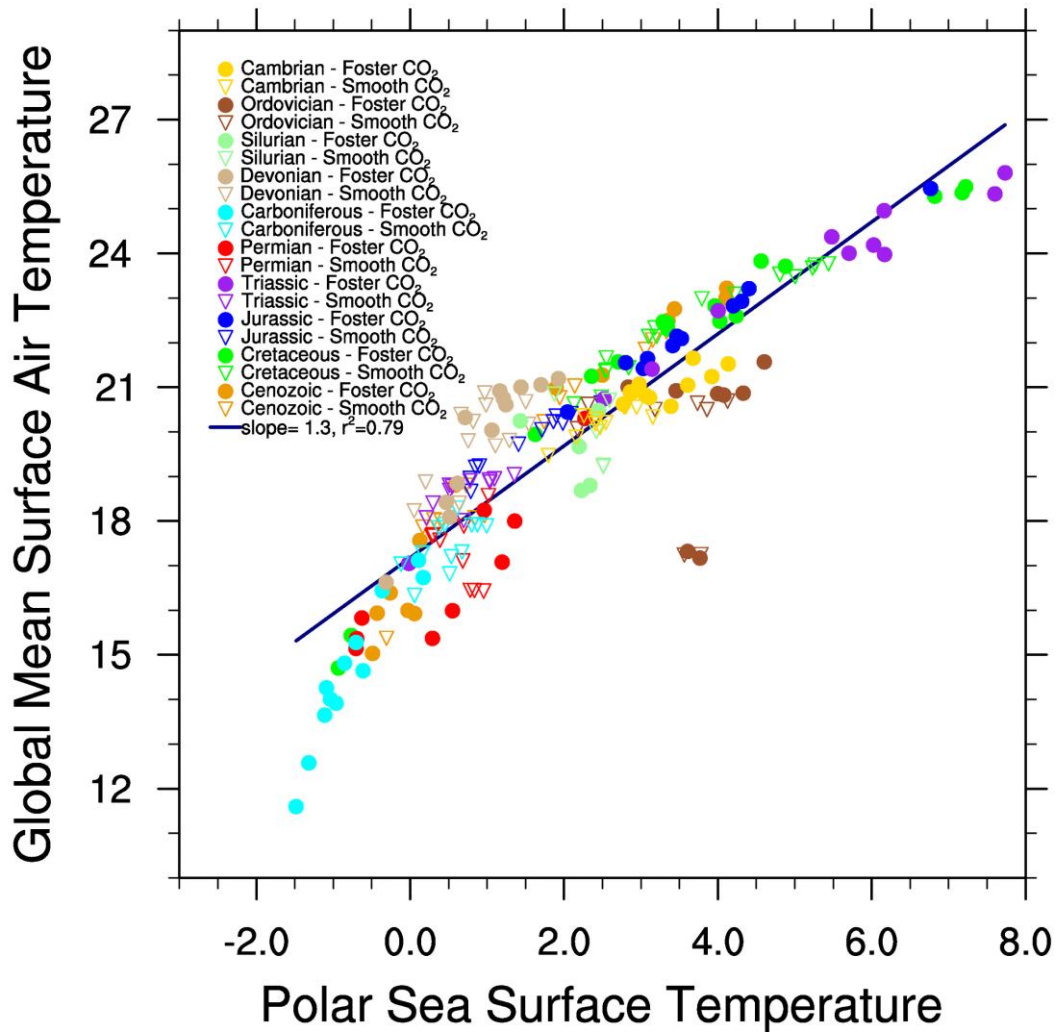
Figure 8. Longitudinal cross section at 20S of (a) ocean potential temperature and (b) salinity for the Ladanian (240Ma). Temperature is in C and salinity is in PSU.

840   **Figure 9.** Modelled age of water tracer at 2731m for 4 different time periods (a) 265Ma, (b) 240Ma,
841   (c) 107Ma, and (d) 0Ma. Units are years.
842

843
844
845
846
847
848

849   **Figure 10**. Correlation between high latitude ocean temperatures (polewards of 60°) and the annual

850   mean, global mean surface air temperature. The polar temperatures are the average of the two

851   winter hemispheres (i.e. northern DJF and southern JJA). Other details as in figure 6.

852

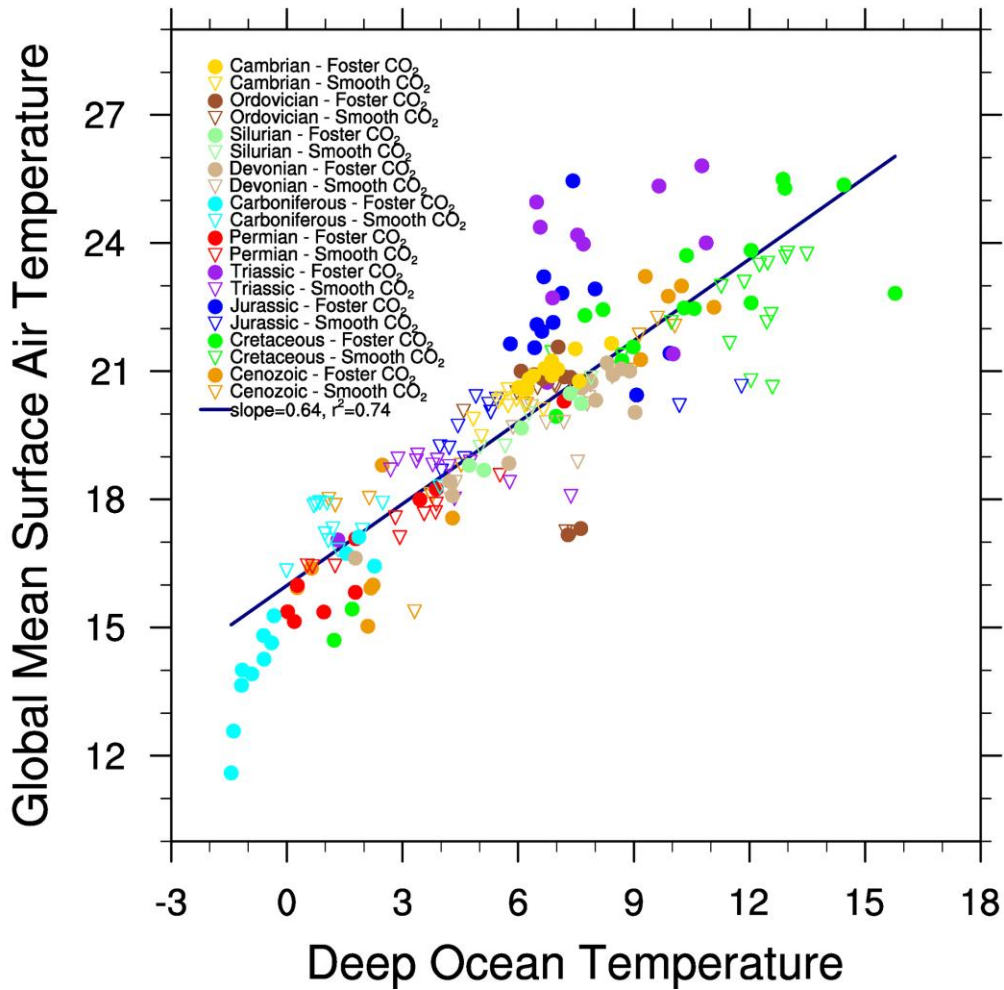
853

854

855   **Figure 11.** Correlation between the global mean, annual mean surface air temperature and the deep

856   ocean temperature. The deep ocean temperatures are defined as the average temperature at the

857   bottom of the model ocean, where the bottom must be deeper than 1000m. Other details as in

858   figure 6.

859

860

861

862   **Figure 12.** Phanerozoic Time series of modelled temperature change (relative to pre-Industrial) for

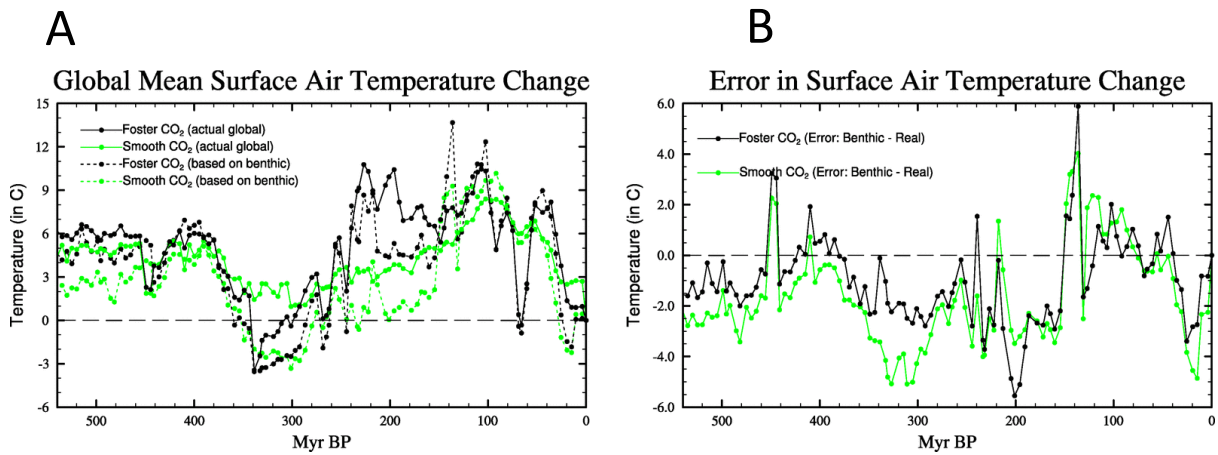
863   the smooth (green lines) and Foster-CO_2 (black) simulations (a) shows the actual modelled global

864   mean surface air temperature (solid lines) whereas the dashed line shows the estimate based on

865   deep ocean temperatures, and (b) error in the estimate of global mean temperature change if based

866   on deep ocean temperatures (i.e. deep ocean – global mean surface temperatures).

867

868   References

869   Alexeev, V.A., Langen, P.L., Bates, J.R., 2005. Polar amplification of surface warming on an
870   aquaplanet in "ghost forcing" experiments without sea ice feedbacks. Clim. Dyn. 24, 655-666.
871   Baatsen, M., van Hinsbergen, D.J.J., von der Heydt, A.S., Dijkstra, H.A., Sluijs, A., Abels, H.A., Bijl, P.K.,
872   2016. Reconstructing geographical boundary conditions for palaeoclimate modelling during the
873   Cenozoic. Clim. Past. 12, 1635-1644.
874   Baatsen, M.L.J., von der Heydt, A.S., Kliphuis, M., Viebahn, J., Dijkstra, H.A., 2018. Multiple states in
875   the late Eocene ocean circulation. Glob. Planet. Change 163, 18-28.
876   Barron, E.J., Peterson, W.H., 1990. MID-CRETACEOUS OCEAN CIRCULATION: RESULTS FROM MODEL
877   SENSITIVITY STUDIES. Paleoceanography 5, 319-337.
878   Beerling, D.J., Fox, A., Stevenson, D.S., Valdes, P.J., 2011. Enhanced chemistry-climate feedbacks in
879   past greenhouse worlds. P Natl Acad Sci USA 108, 9770-9775.
880   Bjerknes, J., 1964. Atlantic Air-Sea Interaction, in: Landsberg, H.E., Van Mieghem, J. (Eds.), 0065-
881   2687. Elsevier, pp. 1-82.
882   Brass, G.W., Southam, J.R., Peterson, W.H., 1982. Warm Saline Bottom Water in the Ancient Ocean.
883   Nature 296, 620-623.
884   Cramer, B.S., Toggweiler, J.R., Wright, J.D., Katz, M.E., Miller, K.G., 2009. Ocean overturning since the
885   Late Cretaceous: Inferences from a new benthic foraminiferal isotope compilation.
886   Paleoceanography 24.
887   Cramwinckel, M.J., Huber, M., Kocken, I.J., Agnini, C., Bijl, P.K., Bohaty, S.M., Frieling, J., Goldner, A.,
888   Hilgen, F.J., Kip, E.L., Peterse, F., van der Ploeg, R., Rohl, U., Schouten, S., Sluijs, A., 2018.
889   Synchronous tropical and polar temperature evolution in the Eocene. Nature 559, 382-+.
890   Dietmuller, S., Ponater, M., Sausen, R., 2014. Interactive ozone induces a negative feedback in CO2-
891   driven climate change simulations. J. Geophys. Res.-Atmos. 119, 1796-1805.
892   Donnadieu, Y., Puceat, E., Moiroud, M., Guillocheau, F., Deconinck, J.F., 2016. A better-ventilated
893   ocean triggered by Late Cretaceous changes in continental configuration. Nat Commun 7, 12.
894   Emiliani, C., 1954. Temperatures of Pacific Bottom Waters and Polar Superficial Waters during the
895   Tertiary. Science 119, 853-855.
896   England, M.H., 1995. The Age Of Water And Ventilation Timescales In A Global Ocean Model. J Phys
897   Oceanogr 25, 2756-2777.
898   Farnsworth, A., Lunt, D.J., O'Brien, C.L., Foster, G.L., Inglis, G.N., Markwick, P., Pancost, R.D.,
899   Robinson, S.A., 2019a. Climate Sensitivity on Geological Timescales Controlled by Nonlinear
900   Feedbacks and Ocean Circulation. Geophys. Res. Lett. 46, 9880-9889.
901   Farnsworth, A., Lunt, D.J., Robinson, S.A., Valdes, P.J., Roberts, W.H.G., Clift, P.D., Markwick, P., Su,
902   T., Wrobel, N., Bragg, F., Kelland, S.J., Pancost, R.D., 2019b. Past East Asian monsoon evolution
903   controlled by paleogeography, not CO2. Sci Adv 5.
904   Foster, G.L., Royer, D.L., Lunt, D.J., 2017. Future climate forcing potentially without precedent in the
905   last 420 million years. Nat Commun 8.
906   Friedrich, O., Norris, R.D., Erbacher, J., 2011. Evolution of Cretaceous oceans: A 55 million year
907   record of Earth's temperature and carbon cycle. Grzyb Found Spec Pub 17, 85-85.
908   Friedrich, O., Norris, R.D., Erbacher, J., 2012. Evolution of middle to Late Cretaceous oceans-A 55
909   m.y. record of Earth's temperature and carbon cycle. Geology 40, 107-110.
910   Goldner, A., Herold, N., Huber, M., 2014. The challenge of simulating the warmth of the mid-
911   Miocene climatic optimum in CESM1. Clim. Past. 10, 523-536.
912   Gordon, C., Cooper, C., Senior, C.A., Banks, H., Gregory, J.M., Johns, T.C., Mitchell, J.F.B., Wood, R.A.,
913   2000. The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley
914   Centre coupled model without flux adjustments. Clim. Dyn. 16, 147-168.
915   Gough, D.O., 1981. Solar Interior Structure and Luminosity Variations. Sol Phys 74, 21-34.

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

Hansen, J., Sato, M., Kharecha, P., Beerling, D., Berner, R., Masson-Delmotte, V., Pagani, M., Raymo, M., Royer, D.L., Zachos, J.C., 2008. Target atmospheric CO_2: Where should humanity aim? Open Atmos. Sci. J. 2, 217-231.

Hansen, J., Sato, M., Russell, G., Kharecha, P., 2013. Climate sensitivity, sea level and atmospheric carbon dioxide. Philos T R Soc A 371.

Hansen, J.E., Sato, M., 2012. Paleoclimate Implications for Human-Made Climate Change. Springer Vienna, Vienna, pp. 21-47.

Hardiman, S.C., Andrews, M.B., Andrews, T., Bushell, A.C., Dunstone, N.J., Dyson, H., Jones, G.S., Knight, J.R., Neininger, E., O'Connor, F.M., Ridley, J.K., Ringer, M.A., Scaife, A.A., Senior, C.A., Wood, R.A., 2019. The Impact of Prescribed Ozone in Climate Projections Run With HadGEM3-GC3.1. J Adv Model Earth Sy 11, 3443-3453.

Harris, J., Ashley, A., Otto, S., Valdes, P., Crossley, R., Preston, R., Watson, J., Goodrich, M., Team., M.P., 2017. Paleogeography and Paleo-Earth Systems in the Modeling of Marine Paleoproductivity: A Prerequisite for the Prediction of Petroleum Source Rocks, in: Mahdi A. AbuAli, I.M., and Hege M. Nordgård Bolås (Ed.), Petroleum Systems Analysis—Case Studies. AAPG Memoir, pp. 27-60.

Henkes, G.A., Passey, B.H., Grossman, E.L., Shenton, B.J., Yancey, T.E., Perez-Huerta, A., 2018. Temperature evolution and the oxygen isotope composition of Phanerozoic oceans from carbonate clumped isotope thermometry. Earth Planet. Sci. Lett. 490, 40-50.

Holland, M.M., Bitz, C.M., 2003. Polar amplification of climate change in coupled models. Clim. Dyn. 21, 221-232.

Kageyama, M., Albani, S., Braconnot, P., Harrison, S.P., Hopcroft, P.O., Ivanovic, R.F., Lambert, F., Marti, O., Peltier, W.R., Peterschmitt, J.Y., Roche, D.M., Tarasov, L., Zhang, X., Brady, E.C., Haywood, A.M., LeGrande, A.N., Lunt, D.J., Mahowald, N.M., Mikolajewicz, U., Nisancioglu, K.H., Otto-Bliesner, B.L., Renssen, H., Tomas, R.A., Zhang, Q., Abe-Ouchi, A., Bartlein, P.J., Cao, J., Li, Q., Lohmann, G., Ohgaito, R., Shi, X.X., Volodin, E., Yoshida, K., Zhang, X., Zheng, W.P., 2017. The PMIP4 contribution to CMIP6-Part 4: Scientific objectives and experimental design of the PMIP4-CMIP6 Last Glacial Maximum experiments and PMIP4 sensitivity experiments. Geosci Model Dev 10, 4035-4055.

Kennett, J.P., Stott, L.D., 1991. Abrupt Deep-Sea Warming, Palaeoceanographic Changes and Benthic Extinctions at the End of the Paleocene. Nature 353, 225-229.

Kiehl, J.T., Shields, C.A., 2013. Sensitivity of the Palaeocene-Eocene Thermal Maximum climate to cloud properties. Philos T R Soc A 371.

Knorr, G., Butzin, M., Micheels, A., Lohmann, G., 2011. A warm Miocene climate at low atmospheric CO_2 levels. Geophys. Res. Lett. 38, 5.

Krapp, M., Jungclaus, J.H., 2011. The Middle Miocene climate as modelled in an atmosphere-ocean-biosphere model. Clim. Past. 7, 1169-1188.

Ladant, J.B., Poulsen, C.J., Fluteau, F., Tabor, C.R., MacLeod, K.G., Martin, E.E., Haynes, S.J., Rostami, M.A., 2020. Paleogeographic controls on the evolution of Late Cretaceous ocean circulation. Clim. Past. 16, 973-1006.

Lunt, D.J., Bragg, F., Chan, W.L., Hutchinson, D.K., Ladant, J.B., Morozova, P., Niezgodzki, I., Steinig, S., Zhang, Z., Zhu, J., Abe-Ouchi, A., Anagnostou, E., de Boer, A.M., Coxall, H.K., Donnadieu, Y., Foster, G., Inglis, G.N., Knorr, G., Langebroek, P.M., Lear, C.H., Lohmann, G., Poulsen, C.J., Sepulchre, P., Tierney, J.E., Valdes, P.J., Volodin, E.M., Dunkley Jones, T., Hollis, C.J., Huber, M., Otto-Bliesner, B.L., 2021. DeepMIP: model intercomparison of early Eocene climatic optimum (EECO) large-scale climate features and comparison with proxy data. Clim. Past 17, 203-227.

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

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

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

Lunt, D.J., Valdes, P.J., Dunkley Jones, T., Ridgwell, A., Haywood, A.M., Schmidt, D.N., Marsh, R., Maslin, M., 2010. CO2-driven ocean circulation changes as an amplifier of Paleocene-Eocene thermal maximum hydrate destabilization. Geology 38, 875-878.

Murphy, D.P., Thomas, D.J., 2012. Cretaceous deep-water formation in the Indian sector of the Southern Ocean. Paleoceanography 27.

Nowack, P.J., Abraham, N.L., Maycock, A.C., Braesicke, P., Gregory, J.M., Joshi, M.M., Osprey, A., Pyle, J.A., 2015. A large ozone-circulation feedback and its implications for global warming assessments. Nat Clim Change 5, 41-45.

Nunes, F., Norris, R.D., 2006. Abrupt reversal in ocean overturning during the Palaeocene/Eocene warm period. Nature 439, 60-63.

O'Brien, C.L., Robinson, S.A., Pancost, R.D., Damste, J.S.S., Schouten, S., Lunt, D.J., Alsenz, H., Bomemann, A., Bottini, C., Brassell, S.C., Farnsworth, A., Forster, A., Huber, B.T., Inglis, G.N., Jenkyns, H.C., Linnert, C., Littler, K., Markwick, P., McAnena, A., Mutterlose, J., Naafs, B.D.A., Puttmann, W., Sluijs, A., van Helmond, N.A.G.M., Vellekoop, J., Wagner, T., Wrobel, N.E., 2017. Cretaceous sea-surface temperature evolution: Constraints from TEX86 and planktonic foraminiferal oxygen isotopes. 0012-8252 172, 224-247.

Outten, S., Esau, I., Ottera, O.H., 2018. Bjerknes Compensation in the CMIP5 Climate Models. J. Clim. 31, 8745-8760.

Pope, V.D., Gallani, M.L., Rowntree, P.R., Stratton, R.A., 2000. The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3. Clim. Dyn. 16, 123-146.

Poulsen, C.J., Barron, E.J., Arthur, M.A., Peterson, W.H., 2001. Response of the mid-Cretaceous global oceanic circulation to tectonic and CO2 forcings. Paleoceanography 16, 576-592.

Sagoo, N., Valdes, P., Flecker, R., Gregoire, L.J., 2013. The Early Eocene equable climate problem: can perturbations of climate model parameters identify possible solutions? Philos T R Soc A 371.

Scotese, C.R., 2016. PALEOMAP PaleoAtlas for GPlates and the PaleoData Plotter Program, PALEOMAP Project, http://www.earthbyte.org/paleomap-paleoatlas-for-gplates/.

Scotese, C.R., Schettino, A., 2017. Late Permian – Early Jurassic Paleogeography of Western Tethys and the World, in: Soto, J.I., Flinch, J., Tari, G. (Eds.), Permo-Triassic Salt Provinces of Europe, North Africa and the Atlantic Margins. Elsevier, pp. 57-95.

Scotese, C.R., Wright, N., 2018. PALEOMAP Paleodigital Elevation MOdels (PaleoDEMS) for the Phanerozoic, PALEOMAP Project, https://www.earthbyte.org/paleodem-resource-scotese-and-wright-2018/.

Smith, R.S., Gregory, J.M., Osprey, A., 2008. A description of the FAMOUS (version XDBUA) climate model and control run. Geosci Model Dev 1, 53-68.

Song, H.J., Wignall, P.B., Song, H.Y., Dai, X., Chu, D.L., 2019. Seawater Temperature and Dissolved Oxygen over the Past 500 Million Years. J Earth Sci-China 30, 236-243.

Steinthorsdottir, M., H. K. Coxall, A. M. de Boer, M. Huber, N. Barbolini, C. D. Bradshaw, N. J. Burls, S. J. Feakins, E. Gasson, J. Henderiks, A. Holbourn, S. Kiel, M. J. Kohn, G. Knorr, W. M. Kurschner, C. H. Lear, D. Liebrand, D. J. Lunt, T. Mors, P. N. Pearson, M. J. Pound, H. Stoll, C. A. E. Stromberg, 2021. The Miocene: the Future of the Past. Paleoceonagraphy and Paleoclimatology. In press.

1018   Stoll, H.M., Guitian, J., Hernandez-Almeida, I., Mejia, L.M., Phelps, S., Polissar, P., Rosenthal, Y.,
1019   Zhang, H.R., Ziveri, P., 2019. Upregulation of phytoplankton carbon concentrating mechanisms
1020   during low CO2 glacial periods and implications for the phytoplankton pCO(2) proxy. Quat. Sci. Rev.
1021   208, 1-20.
1022   Sutton, R.T., Dong, B.W., Gregory, J.M., 2007. Land/sea warming ratio in response to climate change:
1023   IPCC AR4 model results and comparison with observations. Geophys. Res. Lett. 34.
1024   Upchurch, G.R., Kiehl, J., Shields, C., Scherer, J., Scotese, C., 2015. Latitudinal temperature gradients
1025   and high-latitude temperatures during the latest Cretaceous: Congruence of geologic data and
1026   climate models. Geology 43, 683-686.
1027   **Valdes, P.J.**, Armstrong, E., Badger, M.P.S., Bradshaw, C.D., Bragg, F., Crucifix, M., Davies-Barnard, T.,
1028   Day, J.J., Farnsworth, A., Gordon, C., Hopcroft, P.O., Kennedy, A.T., Lord, N.S., Lunt, D.J., Marzocchi,
1029   A., Parry, L.M., Pope, V., Roberts, W.H.G., Stone, E.J., Tourte, G.J.L., Williams, J.H.T., 2017. The
1030   BRIDGE HadCM3 family of climate models: HadCM3@Bristol v1.0. Geosci Model Dev 10, 3715-3743.
1031   Verard, C., Hochard, C., Baumgartner, P.O., Stampfli, G.M., 2015. 3D palaeogeographic
1032   reconstructions of the Phanerozoic versus sea-level and Sr-ratio variations. J Palaeogeog-English 4,
1033   64-84.
1034   Westerhold, T., Marwan, N., Drury, A.J., Liebrand, D., Agnini, C., Anagnostou, E., Barnet, J.S.K.,
1035   Bohaty, S.M., De Vleeschouwer, D., Florindo, F., Frederichs, T., Hodell, D.A., Holbourn, A.E., Kroon,
1036   D., Lauretano, V., Littler, K., Lourens, L.J., Lyle, M., Pälike, H., Röhl, U., Tian, J., Wilkens, R.H., Wilson,
1037   P.A., Zachos, J.C., 2020. An astronomically dated record of Earth's climate and its predictability over
1038   the last 66 million years. Science 369, 1383.
1039   Witkowski, C.R., Weijers, J.W.H., Blais, B., Schouten, S., Damste, J.S.S., 2018. Molecular fossils from
1040   phytoplankton reveal secular Pco(2)trend over the Phanerozoic. Sci Adv 4.
1041   You, Y., Huber, M., Muller, R.D., Poulsen, C.J., Ribbe, J., 2009. Simulation of the Middle Miocene
1042   Climate Optimum. Geophys. Res. Lett. 36.
1043   Zachos, J., Pagani, M., Sloan, L., Thomas, E., Billups, K., 2001. Trends, rhythms, and aberrations in
1044   global climate 65 Ma to present. Science 292, 686-693.
1045   Zachos, J.C., Dickens, G.R., Zeebe, R.E., 2008. An early Cenozoic perspective on greenhouse warming
1046   and carbon-cycle dynamics. Nature 451, 279-283.
1047   Zhou, J., Poulsen, C.J., Pollard, D., White, T.S., 2008. Simulation of modern and middle Cretaceous
1048   marine delta O-18 with an ocean-atmosphere general circulation model. Paleoceanography 23, 11.
1049   Zhu, J., Poulsen, C.J., Tierney, J.E., 2019. Simulation of Eocene extreme warmth and high climate
1050   sensitivity through cloud feedbacks. Sci Adv 5.

1051

1052

Table I. List of Paleogeographic Maps and PaleoDEMs

Map Number	Stratigraphic Age Description	Plate Model Age
1	Present-day (Holocene, 0 Ma)	0
2	*Last Glacial Maximum (Pleistocene, 21 ky) **	0
3	*Late Pleistocene (122 ky) **	0
4	*Middle Pleistocene (454 ky) **	0
5	*Early Pleistocene (Calabrian, 1.29 Ma) **	0
6	*Early Pleistocene (Gelasian, 2.19) **	0
7	Late Pliocene (Piacenzian, 3.09)	5
8	*Early Pliocene (Zanclean, 4.47 Ma) **	5
9	*latest Miocene (Messinian, 6.3 Ma) **	5
10	Middle/Late Miocene (Serravallian&Tortonian, 10.5 Ma)	10
11	Middle Miocene (Langhian, 14.9 Ma)	15
12	Early Miocene (Aquitanian&Burdigalian, 19.5 Ma)	20
13	Late Oligocene (Chattian, 25.6 Ma)	25
14	Early Oligocene (Rupelian, 31 Ma)	30
15	Late Eocene (Priabonian, 35.9 Ma)	35
16	late Middle Eocene (Bartonian, 39.5 Ma)	40
17	early Middle Eocene (Lutetian, 44.5 Ma)	45
18	Early Eocene (Ypresian, 51.9 Ma)	50
19	Paleocene/Eocene Boundary (PETM, 56 Ma)	55
20	Paleocene (Danian&Thanetian, 61 Ma)	60
21	KT Boundary (latest Maastrichtian, 66 Ma)	65
22	Late Cretaceous (Maastrichtian, 69 Ma)	70
23	Late Cretaceous (Late Campanian, 75 Ma)	75
24	Late Cretaceous (Early Campanian, 80.8 Ma)	80
25	Late Cretaceous (Santonian&Coniacian, 86.7 Ma)	85
26	Mid-Cretaceous (Turonian, 91.9 Ma)	90

27	Mid-Cretaceous (Cenomanian, 97.2 Ma)	95
28	Early Cretaceous (late Albian, 102.6 Ma)	100
29	Early Cretaceous (middle Albian, 107 Ma)	105
30	Early Cretaceous (early Albian, 111 Ma)	110
31	Early Cretaceous (late Aptian, 115.8 Ma)	115
32	Early Cretaceous (early Aptian, 121.8 Ma)	120
33	Early Cretaceous (Barremian, 127.2 Ma)	125
34	Early Cretaceous (Hauterivian, 131.2 Ma)	130
35	Early Cretaceous (Valanginian, 136.4 Ma)	135
36	Early Cretaceous (Berriasian, 142.4 Ma)	140
37	Jurassic/Cretaceous Boundary (145 Ma)	145
38	Late Jurassic (Tithonian, 148.6 Ma)	150
39	Late Jurassic (Kimmeridgian, 154.7 Ma)	155
40	Late Jurassic (Oxfordian, 160.4 Ma)	160
41	Middle Jurassic (Callovian, 164.8 Ma)	165
42	Middle Jurassic (Bajocian&Bathonian, 168.2)	170
43	Middle Jurassic (Aalenian, 172.2 Ma)	175
44	Early Jurassic (Toarcian, 178.4 Ma)	180
45	Early Jurassic (Pliensbachian, 186.8 Ma)	185
46	Early Jurassic (Sinemurian/Pliensbachian, 190.8 Ma)	190
47	Early Jurassic (Hettangian&Sinemurian, 196 Ma)	195
48	Late Triassic (Rhaetian/Hettangian, 201.3 Ma)	200
49	Late Triassic (Rhaetian, 204.9 Ma)	205
50	Late Triassic (late Norian, 213.2 Ma)	210
51	Late Triassic (mid Norian, 217.8 Ma)	215
52	Late Triassic (early Norian, 222.4 Ma)	220
53	Late Triassic (Carnian/Norian 227 Ma)	225
54	Late Triassic (Carnian, 232 Ma)	230

82	Late Devonian (early Famennian, 370 Ma)	370
83	Late Devonian (late Frasnian, 375 Ma)	375
84	Late Devonian (early Frasnian, 380 Ma)	380
85	Middle Devonian (Givetian, 385.2 Ma)	385
86	Middle Devonian (Eifelian, 390.5 Ma)	390
87	Early Devonian (late Emsian, 395 Ma)	395
88	Early Devonian (middle Emsian, 400 Ma)	400
89	Early Devonian (early Emsian, 405 Ma)	405
90	Early Devonian (Pragian, 409.2 Ma)	410
91	Early Devonian (Lochkovian, 415 Ma)	415
92	Late Silurian (Pridoli, 421.1 Ma)	420
93	Late Silurian (Ludlow, 425.2 Ma)	425
94	Middle Silurian (Wenlock, 430.4 Ma)	430
95	Early Silurian (late Llandovery, 436 Ma)	435
96	Early Silurian (early Llandovery, 441.2 Ma)	440
97	Late Ordovician (Hirnantian, 444.5 Ma)	445
98	Late Ordovician (Katian, 449.1 Ma)	450
99	Late Ordovician (Sandbian, 455.7 Ma)	455
100	Middle Ordovician (late Darwillian,460 Ma)	460
101	Middle Ordovician (early Darwillian,465 Ma)	465
102	Early Ordovician (Floian/Dapingianboundary, 470 Ma)	470
103	Early Ordovician (late Early Floian, 475 Ma)	475
104	Early Ordovician (Tremadoc, 481.6 Ma)	480
105	Cambro-Ordovician Boundary (485.4 Ma)	485
106	Late Cambrian (Jiangshanian, 491.8 Ma)	490
107	Late Cambrian (Pabian, 495.5 Ma)	495
108	late Middle Cambrian (Guzhangian, 498.8 Ma)	500
109	late Middle Cambrian (early Epoch 3, 505 Ma)	505

110	early Middle Cambrian (late Epoch 2, 510 Ma)	510
111	early Middle Cambrian (middle Epoch 2, 515 Ma)	515
112	Early/Middle Cambrian boundary (520 Ma)	520
113	Early Cambrian (late Terreneuvian, 525 Ma)	525
114	Early Cambrian (middle Terreneuvian, 530 Ma)	530
115	Early Cambrian (early Terreneuvian, 535 Ma)	535
116	Cambrian/Precambrian boundary (541 Ma)	540

1053   * *Simulations were not run for the time intervals highlighted in italics.*

1054

1055

1056 Table 2. Summary of Model Simulations

1057 The table summarises the simulations performed in this study. The 6th and 7th column refers to how

1058 the model was initialised. The smooth CO_2 simulations were initialised from existing simulations

1059 using paleogeographies which were provided commercial in confidence. The time period for which

1060 these paleogeographies correspond to are listed in column 6, and the CO_2 value used in the runs is in

1061 column 7. The Foster CO_2 simulations were initialised from the end point of the smooth CO_2

1062 simulations.

Time Period (in Ma)	CO_2 (in ppmv) for the Smooth CO_2	CO_2 (in ppmv) for the Foster CO_2	Length of Run (in years) (Smooth CO_2)	Length of Run (in years) (Foster CO_2)	Time period (in Ma) from existing simulations used for the initial condition for smooth CO2 simulation	CO_2 (in ppmv) used for the initial conditions
0.0	280	276	5000	5000	0	280
3.1	384	298	5000	2000	3	401
10.5	410	299	5000	2000	13	280
14.9	423	310	5000	2000	13	280
19.5	430	338	5000	2000	13	280
25.6	439	502	5000	2000	26	560
31.0	500	764	5000	5000	26	560
35.9	533	901	5000	5000	26	560
39.5	557	796	5000	5000	26	560
44.5	594	751	5000	8000	26	560
51.9	649	736	5000	8000	52	560

56.0	630	570	5000	5000	52	560
61.0	604	335	5000	5000	52	560
66.0	576	229	5000	5000	69	560
69.0	560	262	5000	5000	69	560
75.0	633	559	5000	2000	69	560
80.8	704	667	5000	2000	69	560
86.7	775	590	5000	8000	69	560
91.9	839	466	5000	8000	92	560
97.2	840	707	5000	8000	92	560
102.6	840	1008	5000	5000	92	560
107.0	840	1028	5000	5000	92	560
111.0	827	1148	5000	8000	92	560
115.8	811	1103	5000	8000	92	560
121.8	784	986	5000	5000	92	560
127.2	752	898	5000	5000	92	560
131.2	728	896	5000	5000	92	560
136.4	699	1020	5000	8000	136	840
142.4	677	832	5000	8000	136	840
145.0	667	713	5000	8000	136	840
148.6	654	721	5000	8000	136	840
154.7	631	802	5000	5000	155	560
160.4	617	785	5000	5000	155	560
164.8	606	868	5000	8000	155	560
168.2	596	1019	5000	5000	167	840
172.2	581	1046	5000	5000	167	840
178.4	560	986	5000	5000	178	1120

186.8	560	949	5000	5000	178	1120
190.8	560	1181	5000	5000	178	1120
196.0	560	1784	5000	5000	178	1120
201.3	560	1729	5000	5000	178	1120
204.9	560	1503	5000	5000	218	560
213.2	560	1223	5000	5000	218	560
217.8	560	1481	5000	5000	218	560
222.4	557	1810	5000	5000	218	560
227.0	553	2059	5000	5000	218	560
232.0	549	1614	5000	5000	218	560
233.6	548	1492	5000	5000	218	560
239.5	543	1034	5000	8000	218	560
244.6	540	419	5000	2000	218	560
252.0	534	879	5000	2000	257	1120
256.0	531	811	5000	2000	257	1120
262.5	526	352	5000	2000	257	1120
265.1	524	321	5000	2000	257	1120
268.7	521	311	5000	2000	257	1120
275.0	517	556	5000	2000	257	1120
280.0	513	690	5000	2000	257	1120
286.8	508	626	5000	2000	257	1120
292.6	503	495	5000	2000	297	280
297.0	500	445	5000	2000	297	280
301.3	510	393	5000	2000	297	280
305.4	520	358	5000	2000	297	280
311.1	534	338	5000	2000	297	280

314.6	542	327	5000	2000	297	280
319.2	553	328	5000	2000	297	280
327.0	571	317	5000	2000	297	280
330.9	581	296	5000	2000	339	420
333.0	586	263	5000	2000	339	420
338.8	600	233	5000	2000	339	420
344.0	653	565	5000	2000	339	420
349.0	705	645	5000	2000	339	420
354.0	758	589	5000	2000	339	420
358.9	809	587	5000	2000	339	420
365.6	880	806	5000	2000	339	420
370.0	926	811	5000	2000	339	420
375.0	979	1052	5000	2000	339	420
380.0	1029	1269	5000	2000	339	420
385.2	1079	1377	5000	2000	377	1680
390.5	1131	1093	5000	2000	377	1680
395.0	1174	1297	5000	2000	377	1680
400.0	1223	1731	5000	2000	377	1680
405.0	1271	1689	5000	2000	377	1680
409.2	1319	2102	5000	2000	377	1680
415.0	1368	1579	5000	2000	377	1680
421.1	1427	1457	5000	2000	377	1680
425.2	1466	1490	5000	2000	377	1680
430.4	1517	1531	5000	2000	377	1680
436.0	1571	1576	5000	2000	377	1680
441.2	1614	1643	5000	2000	439	1877

444.5	1636	1708	5000	2000	439	1877
449.1	1666	1799	5000	2000	439	1877
455.7	1710	1929	5000	2000	439	1877
460.0	1738	2013	5000	2000	439	1877
465.0	1770	2111	5000	2000	439	1877
470.0	1803	2210	5000	2000	439	1877
475.0	1836	2308	5000	2000	439	1877
481.6	1879	2438	5000	2000	439	1877
485.4	1904	2513	5000	2000	439	1877
491.8	1946	2639	5000	2000	439	1877
495.5	1970	2711	5000	2000	439	1877
498.8	1992	2776	5000	2000	439	1877
505.0	2020	2870	5000	2000	439	1877
510.0	2040	2940	5000	2000	439	1877
515.0	2060	3010	5000	2000	439	1877
520.0	2080	3080	5000	2000	439	1877
525.0	2100	3150	5000	2000	439	1877
530.0	2120	3220	5000	2000	439	1877
535.0	2140	3290	5000	2000	439	1877
541.0	2164	3374	5000	2000	439	1877

1063