Simulation of the mid-Pliocene Warm Period using HadGEM3: Experimental design and results from model-model and model-data comparison

Charles J. R. Williams[1,6], Alistair A. Sellar[2], Xin Ren[1], Alan M. Haywood[3], Peter Hopcroft[4], Stephen J. Hunter[3], William H. G. Roberts[5], Robin S. Smith[6], Emma J. Stone[1], Julia C. Tindall[3], Daniel J. Lunt[1]

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

[2]Met Office Hadley Centre, UK

[3]School of Earth and Environment, University of Leeds, UK

[4]School of Geography, Earth and Environmental Sciences, University of Birmingham, UK

[5]Department of Geography and Environmental Sciences, Northumbria University, UK

[6]NCAS, Department of Meteorology, University of Reading, UK

* Corresponding author address:

School of Geographical Sciences,

University of Bristol,

University Road, Bristol, BS8 1SS

United Kingdom

Email: C.J.R.Williams@bristol.ac.uk

Short title: HadGEM3 simulates a warmer Pliocene than proxy data and other climate models

Keywords: Palaeoclimate, Pliocene, model-data comparisons

ABSTRACT

Here we present the experimental design and results from a new mid-Pliocene simulation using the latest version of the UK's physical climate model, HadGEM3-GC31-LL, conducted under the auspices of CMIP6/PMIP4/PlioMIP2. Although two other paleoclimate simulations have been recently run using this model, they both focused on more recent periods within the Quaternary and therefore this is the first time this version of the UK model has been run this far back in time. The mid-Pliocene Warm Period, ~3 Ma, is of particular interest because it represents a time period when the Earth was in equilibrium with CO_2 concentrations roughly equivalent to those of today, providing a possible analogue for current and future climate change.

The implementation of the Pliocene boundary conditions is firstly described in detail, based on the PRISM4 dataset, including CO_2, ozone, orography, ice mask, lakes, vegetation fractions and vegetation functional types. These were incrementally added into the model, to change from a preindustrial setup to a Pliocene setup.

The results of the simulation are then presented, which are firstly compared with the model's pre-industrial simulation, secondly with previous versions of the same model and with available proxy data, and thirdly with all other models included in PlioMIP2. Firstly, the comparison with preindustrial suggests that the Pliocene simulation is consistent with current understanding and existing work, showing warmer and wetter conditions, and with the greatest warming occurring over high latitude and polar regions. The global mean surface air temperature anomaly at the end of the Pliocene simulation is 5.1°C, which is the 2nd highest of all models included in PlioMIP2 and is consistent with the fact that HadGEM3-GC31-LL has one of the highest Effective Climate Sensitivities of all CMIP6 models. Secondly, the comparison with previous generation models and with proxy data suggests a clear increase in global sea surface temperatures as the model has undergone development. Up to a certain level of warming, this results in a better agreement with available proxy data, and the "sweet spot" appears to be the previous CMIP5 generation of the model, HadGEM2-AO. The most recent simulation presented here, however, appears to show poorer agreement with the proxy data compared with HadGEM2, and may be overly sensitive to the Pliocene boundary conditions resulting in a climate that is too warm. Thirdly, the comparison with other models from PlioMIP2 further supports this conclusion, with HadGEM3-GC31-LL being one of the warmest and wettest models in all of PlioMIP2 and, if all the models are ordered according to agreement with proxy data, HadGEM3-GC31-LL ranks approximately halfway among them. A caveat to these results is the relatively short run length of the simulation, meaning the model is not in full equilibrium. Given the computational cost of the model it was not possible to run for longer; a Gregory plot analysis indicates that had it been allowed to come to full equilibrium, the final global mean surface temperature could have been approximately 1.5°C higher.

1. INTRODUCTION

Model simulations of past climate states are useful because, among other aspects, they allow us to interrogate the mechanisms that have caused past climate change (Haywood *et al.* 2020, Lunt *et al.* 2021). They also give us a global picture of past climate variables (such as sea surface temperature, SST) that can only be reconstructed by geological data at specific locations, and of variables (such as upper atmospheric winds) that cannot be reconstructed by geological data at all. However, before models can be used in this way, it is important to validate them by comparing with geological data, where available, from the time periods of interest. Such model-data comparisons can also be useful for evaluating the model outside of the modern climate states that it was likely tuned to, thereby providing an independent assessment of the model that can be important for interpreting any future climate projections arising from the model (e.g. Zhu *et al.* 2020).

The mid-Pliocene Warm Period (mPWP, ~3 million years ago, hereafter referred to as the Pliocene) is an ideal climate state for such a model-data comparison because: i) there has recently been a concerted community effort to provide a synthesis of proxy SST reconstructions (McClymont *et al.* 2020); ii) community-endorsed boundary conditions exist which can be used to configure climate model simulations (Haywood *et al.* 2016); and iii) there is a wealth of previous model intercomparison projects (MIPs), with which model simulations can be compared and contrasted, that have been carried out with these recent boundary conditions (PlioMIP2, Dowsett *et al.* 2016 and Haywood *et al.* 2020) and with previous versions of the boundary conditions (PlioMIP1, Haywood *et al.* 2013). The Pliocene is also a relatively warm period compared to both preindustrial conditions and those of today, with comparable CO_2 levels to today (McClymont *et al.* 2020, Salzmann *et al.* 2013), and so provides a climate state with similarities to those that might be expected in the future (Burke *et al.* 2018, Tierney *et al.* 2020).

PlioMIP2 was a community effort to carry out and analyse coordinated model simulations to explore mechanisms associated with Pliocene climate, and to evaluate multiple models with Pliocene proxy data. To date, 16 models have participated in PlioMIP2, all of which used boundary conditions from the US Geological Survey's PRISM4 (Pliocene Research, Interpretation and Synoptic Mapping v4; see Dowsett *et al.* 2016) and the results of this intercomparison and evaluation are described in Haywood *et al.* 2020 (hereafter abbreviated to H20). H20 first explored the large-scale features (global means, polar amplification and land-sea contrast) of temperature and precipitation in the simulations, finding a global ensemble mean warming of 3.2°C relative to preindustrial and a 7% increase in precipitation. There was a clear signal of polar amplification, but tropical zonal gradients remained largely unchanged compared with preindustrial. Compared with proxies from Foley and Dowsett (2019), the SSTs in the tropics were broadly consistent in the models and data, and in the Atlantic the polar amplification was better represented by the models compared with previous model-

109   data comparisons such as those from PlioMIP1. Recent studies using the PlioMIP2 ensemble have

110   explored other aspects of the model simulations, such as ocean circulation (Zhang *et al.* 2021) and the

111   African monsoon (Berntell *et al.*, in review). It is of interest to evaluate simulations from additional

112   models as they become available, and that is what we do here, presenting results from a new model,

113   HadGEM3-GC31-LL, for the Pliocene. This is of particular interest because HadGEM3-GC31-LL is

114   a Coupled Model Intercomparison Project Phase 6 (CMIP6) "high Effective Climate Sensitivity

115   (ECS)" model (Zelinka *et al.* 2020), with a climate sensitivity to CO_2 doubling of more than 5°C

116   (Andrews *et al.* 2019). Only one other model in CMIP6, CanESM5, has a higher climate sensitivity

117   (5.64°C compared with 5.55°C). HadGEM3-GC31-LL is also of interest because it represents the

118   third generation of UK Met Office model that has participated in PlioMIP (Bragg *et al.* 2012, Tindall

119   and Haywood 2020, Hunter *et al.* 2019), allowing us to assess how much, if any, progress has been

120   made in simulating the Pliocene with the UK family of models.

121

122   In this paper we address 3 main questions:

123   1) What are the large-scale features of the Pliocene climate produced by HadGEM3-GC31-LL?

124   2) To what extent has the development of new boundary conditions and more complex models

125   led to improvements in the simulation of the Pliocene by UK Met Office models?

126   3) How does HadGEM3-GC31-LL compare with other models participating in PlioMIP2?

127

128   Section 2 of this paper describes HadGEM3-GC31-LL, how the PlioMIP2 boundary conditions were

129   implemented in the model, and the experimental design of the model. Section 3 presents the large-

130   scale features of the Pliocene in HadGEM3-GC31-LL, and Section 4 compares the HadGEM3-GC31-

131   LL simulation with proxy data and previous generations of the same UK model, and with other

132   PlioMIP2 models.

133

134   **2. MODEL AND EXPERIMENT DESIGN**

135   **2.1. Naming conventions and terminology**

136   Consistent with CMIP nomenclature, when the simulation is spinning up towards atmospheric and

137   oceanic equilibrium, with initially incomplete boundary conditions, it is referred to as the 'Spin-up

138   phase' and is only briefly presented here. In contrast, once all required boundary conditions were

139   implemented, the results themselves are taken from the end of the simulation, referred to here as the

140   'Production run'. Here, results are based on the final 50-year climatology of this production run.

141   Concerning geological intervals, the preindustrial and mid-Pliocene Warm Period are referred to as

142   the PI and Pliocene, respectively. In contrast, concerning the model simulations using HadGEM3-

143   GC31-LL, consistent with CMIP6 they are referred to as the *piControl* and *mPWP* simulations,

144   respectively. We also make use of the naming convention of Haywood *et al.* 2016, hereafter

145   abbreviated to H16), including the nomenclature Ex^c (where c is the concentration of CO_2 in ppmv,

146 and x are any boundary conditions which are Pliocene as opposed to PI, which can be any or none of

147 o = orography, v = vegetation and i = ice sheets). So, for example, Eov^{500} would be an experiment

148 using Pliocene orography and vegetation and with CO_2 at 500 ppmv, but with preindustrial ice sheets.

149

150 **2.2. Model description**

151 The model presented here is the Global Coupled (GC) 3.1 configuration of the UK's physical climate

152 model, HadGEM3-GC31-LL, which is the "CMIP6-class" UK Met Office physical climate model.

153 The *piControl* simulation for this model was conducted elsewhere as part of CMIP6, and is used here

154 for comparative purposes; see Williams *et al*. (2017), Kuhlbrodt *et al*. (2018) and Menary *et al*.

155 (2018) for further details on HadGEM3-GC31-LL and its *piControl* simulation. The *mPWP*

156 simulation presented here was run with identical components to those used in other CMIP6/PMIP4

157 simulations using this model, namely the *midHolocene* and *lig127k* simulations (Williams *et al*.

158 2020). The full title for this configuration is HadGEM3-GC31-LL N96ORCA1 UM10.7 NEMO3.6

159 (hereafter referred to as HadGEM3). The model was run using the Unified Model (UM), version

160 10.7, and included the following components: i) Global Atmosphere (GA) version 7.1, with an N96

161 atmospheric spatial resolution (approximately 1.875° longitude by 1.25° latitude) and 85 vertical

162 levels; ii) NEMO ocean version 3.6, including Global Ocean (GO) version 6.0 (ORCA1), with an

163 isotropic Mercator grid which, despite varying in both meridional and zonal directions, has an

164 approximate spatial resolution of 1° by 1° and 75 vertical levels; iii) Global Sea Ice (GSI) version 8.0

165 (GSI8.0); iv) Global Land (GL) version 7.0, comprising the Joint UK Land Environment Simulator

166 (JULES); and v) the OASIS3 MCT coupler. All of the above individual components are summarised

167 by Williams *et al*. (2017) and detailed individually by a suite of companion papers (see Walters *et al*.

168 2017 for GA7 and GL7, Storkey *et al*. 2017 for GO6 and Ridley *et al*. 2017 for GSI8). A summary of

169 the major changes in HadGEM3 and their impacts on the climate, relative to its most recent

170 predecessor (HadGEM2), are given in Williams *et al*. (2020). Here, the *mPWP* simulation was run on

171 NEXCS, which is a component of the Cray XC40 located at the UK Met Office. NEXCS is a

172 partition of the UK Met Office's platform, Monsoon, on which the *piControl* simulation was run,

173 thereby avoiding the potential caveat discussed in Williams *et al*. (2020) concerning different

174 computing platforms.

175

176 Details of the other models discussed here, namely previous generations of the same UK Hadley

177 Centre model and all of those included in PlioMIP2, are included in the Supplementary Material

178 (Section 1).

179

180 **2.3. Full Pliocene experiment design**

181 For the most part, the *mPWP* simulation presented here follows the protocol given in H16, discussed

182 below. The main difference is that we do not modify the land-sea mask (LSM), due to technical

183 challenges of modifying the ocean LSM and coupling it to the atmosphere in this model.

184

185 *2.3.1. Greenhouse gas atmospheric concentrations, aerosol emissions and ozone*

186 Following H16, atmospheric CO_2 concentration was modified in the *mPWP* simulation, from 280 to

187 400 ppmv. All other greenhouse gases, such as CH_4, N_2O and O_2, were kept as in the *piControl*

188 simulation. Likewise, aerosol emissions (e.g. organic- and black-carbon fossil fuels) and their

189 resulting oxidants were kept as in the *piControl* simulation, consistent with previous paleoclimate

190 simulations with this model (Williams *et al*. 2020).

191

192 Under strong surface warming, the thermal tropopause rises. In simulations with prescribed ozone

193 concentration it is important that the thermal tropopause remains below the ozone tropopause, in order

194 to avoid unphysical feedbacks associated with increasing cold point temperature (see, for example,

195 Hardiman *et al*. 2019). For this reason, ozone from the *1pctCO₂* simulation of the UK Earth System

196 Model (UKESM1, see Sellar *et al*. 2019), in which CO_2 concentrations are increased relative to 1850

197 levels at 1% per year, was prescribed here. UKESM1 uses the same physical climate configuration as

198 HadGEM3, but interactively simulates ozone chemistry. The ozone was taken from a 10-year period

199 of this UKESM1 simulation (years 51-60), during which the mean surface temperature was

200 approximately 2°C warmer than the *piControl* simulation. The value of 2°C was chosen as a

201 compromise between raising the ozone tropopause enough to avoid inconsistency with the thermal

202 tropopause, without introducing significant changes in ozone forcing relative to the *piControl*. The

203 impact of the ozone modification could be explored in future work, for example by using an ozone

204 profile from a UKESM1 simulation with a higher mean surface temperature (more consistent with the

205 HadGEM3 Pliocene warming, see Section 3), or by using the methodology outlined in Hardiman *et*

206 *al*. (2019), which was used for the CMIP6 Shared Socioeconomic Pathway (SSP) scenario simulations

207 with HadGEM3.

208

209 *2.3.2. Changes to boundary and initial conditions*

210 *2.3.2.1. Palaeogeography (including land-sea mask, orography and bathymetry)*

211 The *mPWP* simulation used an identical LSM to the *piControl* simulation which, if necessary, is

212 allowed under the experimental design laid out in H16. This differs from both the standard and

213 enhanced LSMs provided by H16 (accessible, with all other required boundary conditions, from the

214 US Geological Survey's PlioMIP2 website, http://geology.er.usgs.gov/egpsc/prism/7_pliomip2.html),

215 in that in both of these the gateways in the Bering Sea, the Canadian Archipelago and Hudson Bay are

216 closed, whereas in the HadGEM3 simulations only the Canadian Archipelago/Hudson Bay gateway is

217 closed; the Bering Strait is open (see Supplementary Material, Fig. S1). Likewise, the bathymetry

218 used here is also identical to the *piControl* simulation, for the same reasons.

219

220 The orography used in the *mPWP* simulation, however, does follow the protocol of H16. Here, an

221 anomaly is firstly created by subtracting the PRISM4 modern orography from the PRISM4 Pliocene

222 orography and then, after having been re-gridded to the model's own resolution, adding this to the

223 model's existing orography (see Section 2.3.2 in H16). The results are shown in Figure 1, where the

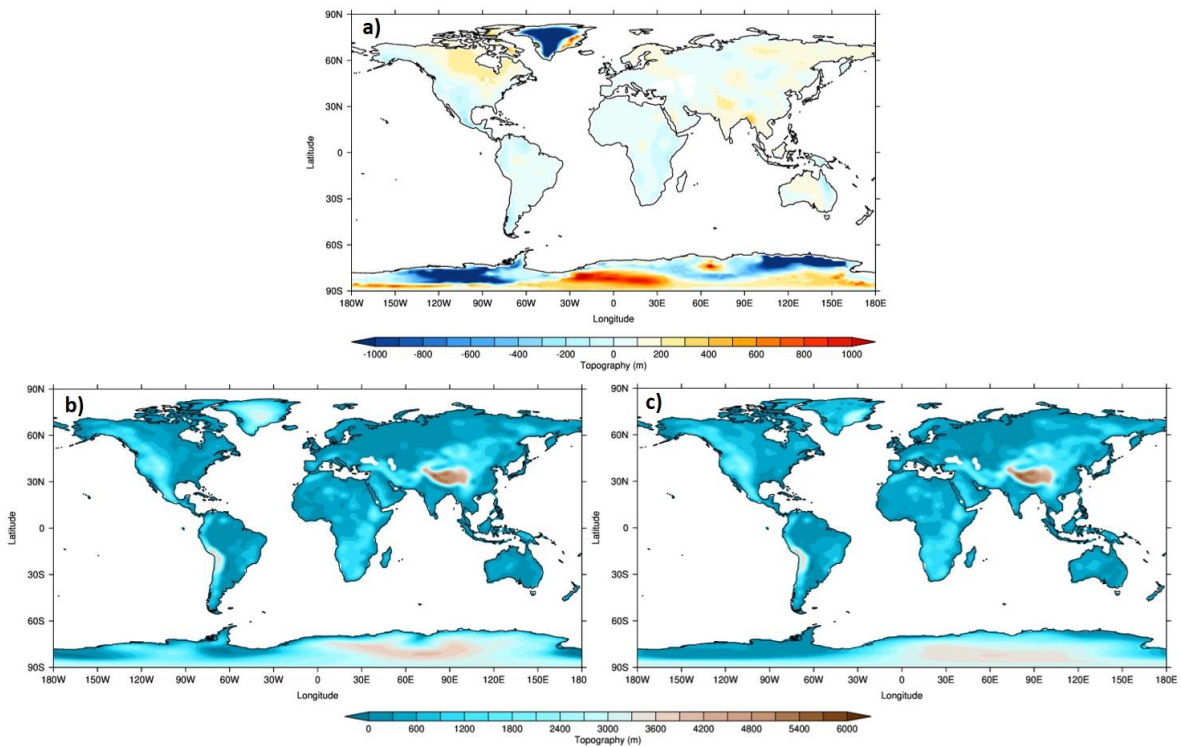
224 PRISM4 anomaly shows the largest changes are occurring over Greenland and Antarctica, with

225 smaller changes over the Himalayas, North America and Africa (Fig. 1a). When added to

226 HadGEM3's existing orography (Fig. 1b), the changes result most obviously in a lowering of

227 orography over Greenland, western and eastern Antarctica, and a raising of orography over central

228 Antarctica (Fig. 1c). Due to an early model instability relating to the steep orographic gradients in

229 western Antarctica, this region was smoothed in the final simulation (Fig. 1c).

230

231 Figure 1 - Changes to topography in HadGEM3 *mPWP* simulation. a) PRISM4 anomaly; b) Original field used in

232 HadGEM3 *piControl*; c) New field used in HadGEM3 *mPWP*, with smoothed topography over western Antarctica (final

233 version, used in simulation)

234

2.3.2.2. *Vegetation fractions (including urban, lakes and ice)*

236 As part of its GL configuration, both the *piControl* and *mPWP* simulations used the community land

237 surface model (JULES; see Best *et al*. 2011, Clark *et al*. 2011, Walters *et al*. 2019). In this land

238 surface model, sub-gridscale heterogeneity is represented by a tile approach (Essery *et al*. 2003), in

239   which each grid box over land is divided into five vegetated plant functional types (PFTs): broadleaf

240   trees (BLT), needle-leaved trees (NLT), temperate C3 grass, tropical C4 grass and shrubs. In addition

241   to these, there are four non-vegetated PFTs: urban areas, inland water (or lakes), bare soil and land

242   ice. This division of grid box into PFTs is consistent with both of the model's predecessors (see

243   Supplementary Material). With the exception of the urban tile, which was kept as PI to be consistent

244   with previous paleoclimate simulations with this model (Williams *et al*. 2020), all of these PFTs were

245   modified in the *mPWP* simulation.

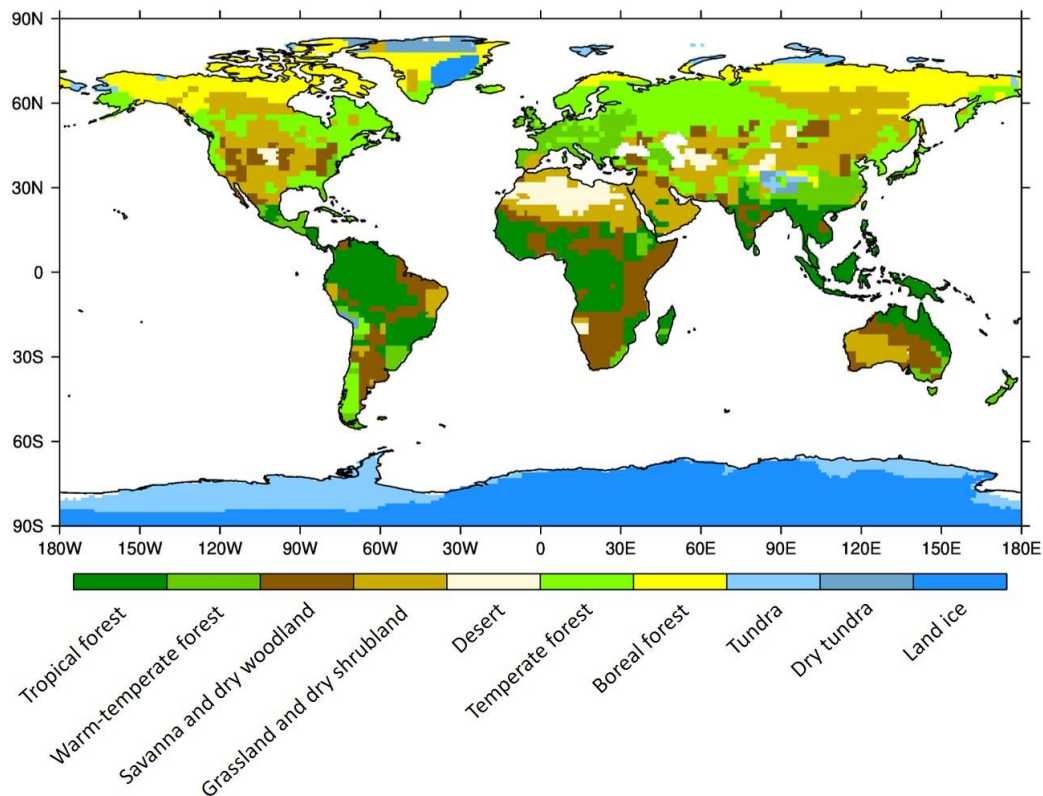
246

247   The US Geological Survey's PRISM4 (Dowsett *et al*. 2016) vegetation reconstruction from Salzmann

248   *et al*. (2008) was used, provided as a megabiome reconstruction in PlioMIP2 (H16). This can be seen

249   in Figure 2, where there are ten listed megabiomes corresponding to those used in Harrison and

250   Prentice 2003: tropical forest, warm-temperate forest, savanna and dry woodland, grassland and dry

251   shrubland, desert, temperate forest, boreal forest, tundra, dry tundra and land ice.

252

253   Figure 2 - Ten megabiomes from PlioMIP2 used create the nine PFTs used in HadGEM3 *mPWP* simulation

254

255

256

257

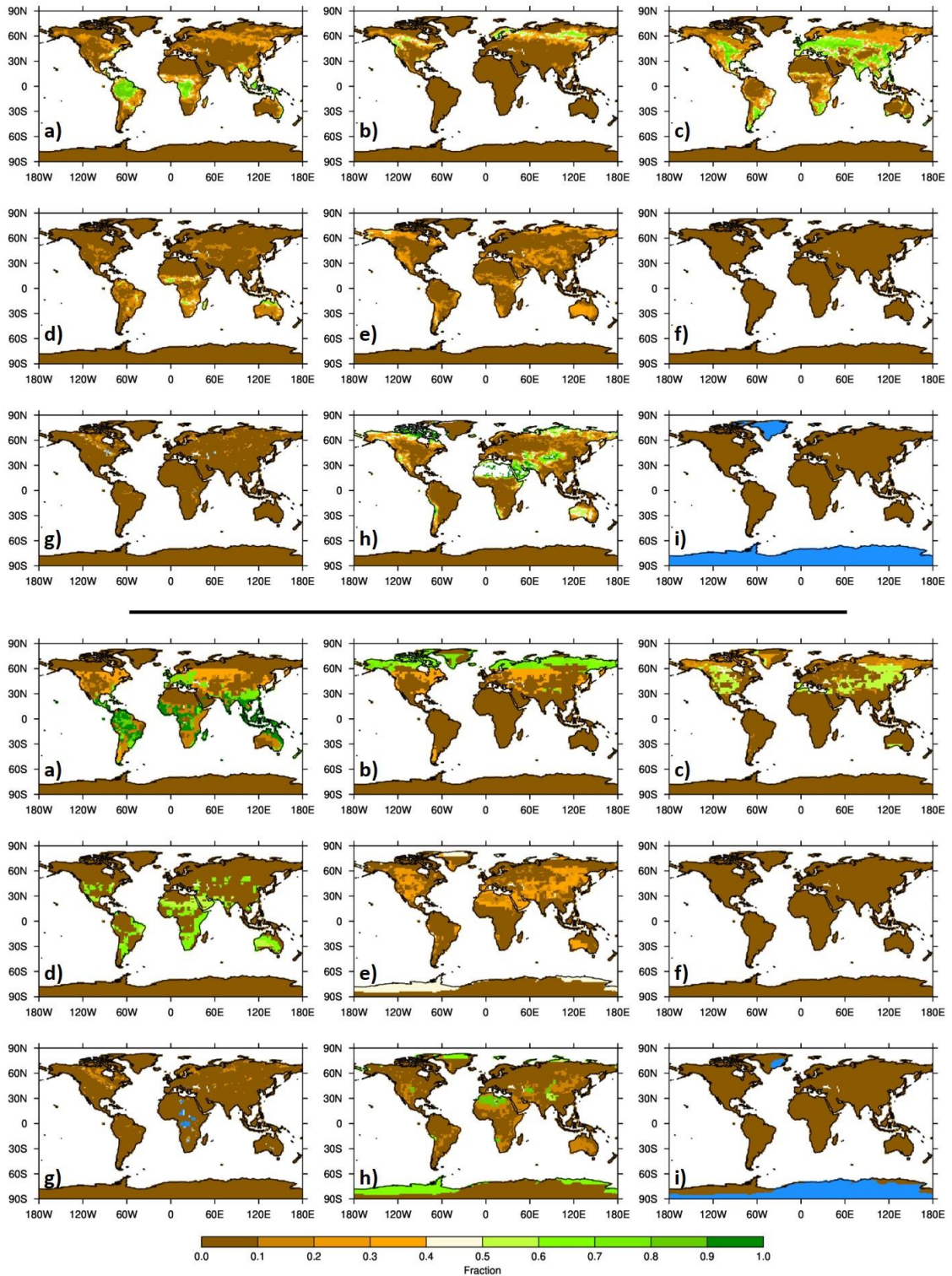
In order to translate the megabiomes from PRISM into the PFTs used by the model, a lookup table was therefore required. Minimum and maximum bounds for each megabiome were firstly obtained, based on values from Crucifix *et al*. 2005, and then estimates were made for each PFT within these bounds by mapping the preindustrial megabiomes onto the preindustrial PFT in HadGEM3; the resulting lookup table is shown in the Supplementary Material (Table S1). In this table, for example, each land grid point with the megabiome "Tropical forest" is divided amongst the model PFTs as 92% BLT, 5% bare soil, 2% tropical C4 grasses and 1% shrubs. The resulting 9 PFTs used in the *mPWP* simulation, as well as those from the original *piControl*, are shown in Figure 3. The largest fractional increases, relative to the *piControl*, occur for broadleaf trees and needleleaf trees (18% and 5%, respectively; Fig. 3a and b) and the largest decreases occur for temperate C3 grass and land ice (15% and 5%, respectively; Fig. 3c and i). In regions where there is no obvious match between the model's PFTs and the megabiomes, such as over western Antarctica (specified as tundra in the PRISM data), a closest match was provided; in this case, a mix of bare soil and shrubs.

271

Figure 3 - Nine PFTs used in HadGEM3. Top half: *piControl* simulation, bottom half: *mPWP* simulation. Values in

brackets show global mean differences (*mPWP - piControl*), expressed as a percentage. a) broadleaf trees (18%); b) needle-

leaved trees (5%); c) temperate C3 grass (-15%); d) tropical C4 grass (6%); e) shrubs (3%); f) urban areas (no change); g)

inland water (1%); h) bare soil (-12%); i) land ice (-5%)

276

277

2.3.2.3. Vegetation functional types

Alongside the vegetation fractions, both the *piControl* and *mPWP* simulations included two monthly-varying vegetation functional types, namely leaf area index (LAI) and canopy height, both of which are associated with each of the five vegetated PFTs. Given that no information was available from the PRISM vegetation reconstruction concerning these fields, two methods were used to create Pliocene LAI and canopy height. For LAI, a seasonally and latitudinally varying function was created from the zonal means of the *piControl* (Figure 4), and used to build a new field for the Pliocene, for each month and each PFT (see Fig. 4b and c for an example of the original *piControl* and the Pliocene newly-created field, respectively, both showing LAI for BLT during January). This is because, in the *piControl*, LAI varies both in time (i.e. seasonally) and space. Note that although LAI does go to zero in the *piControl*, this was not allowed in the *mPWP* simulation because the Pliocene does have some vegetation at high latitudes (see Figure 3); these functions were therefore increased by x (where x = the mean of the ten grid points containing the lowest LAI), such that there is never zero LAI. In contrast, canopy height in the *piControl* does not vary monthly, and has little variation spatially, therefore canopy height in the *mPWP* simulation is set to the global mean of the *piControl* (see Supplementary Material Fig. S2).

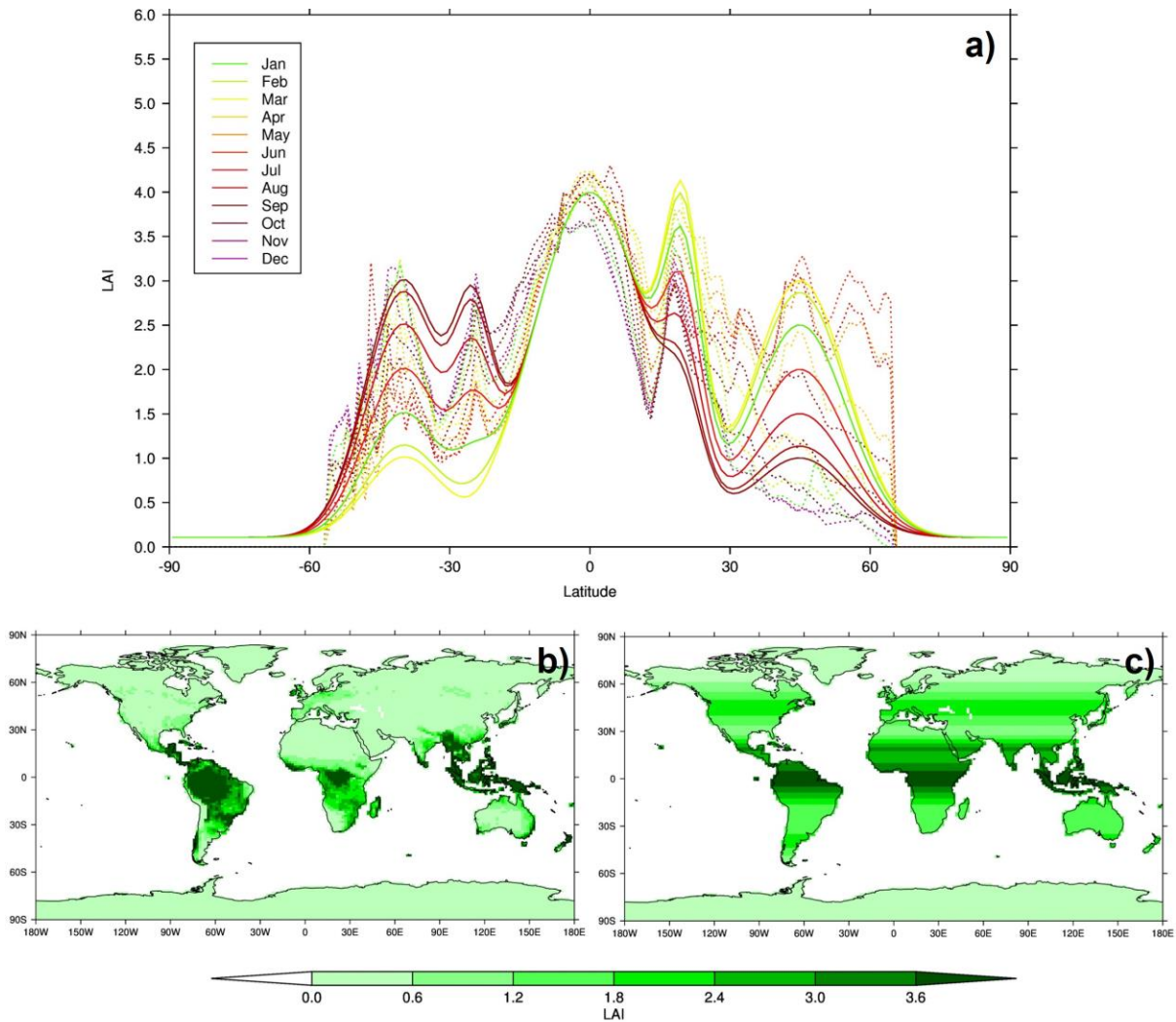



Figure 4 – LAI used in HadGEM3, for an example PFT (broadleaf trees). a) Function used to create LAI, where dashed lines show zonal mean from *piControl* simulation and solid lines show seasonally and latitudinally varying function used in the *mPWP* simulation; b) Example of functional types (broadleaf trees, January) used in *piControl* simulation; c) same as b) but for the *mPWP* simulation/

2.3.2.3. Soil properties and snow depth

Under newly-created land ice based on the new Pliocene ice mask (i.e. in regions where there is no ice in the *piControl* but ice the *mPWP* simulation), soil parameters, soil dust properties and snow depth were set to be appropriate values for existing ice regions i.e. whatever these values are under ice in the *piControl* simulation are applied to the newly-created ice regions in the *mPWP* simulation.

Conversely, and more importantly in this context (as the Pliocene represents an overall removal of ice), under newly-exposed land based on the new Pliocene ice mask (i.e. in regions where there is ice in the *piControl* but no ice in the *mPWP* simulation, primarily over Greenland and western Antarctica), the dominant vegetation fractions in these regions were firstly identified from the newly-created Pliocene vegetation. In this case, the dominant fractions were 40% shrubs and 60% bare soil.

311   Then, grid points containing this vegetation balance in the *piControl* were identified, and the soil

312   parameters, soil dust properties and snow depth values at these points were averaged. This average

313   value, for each of the above fields, was lastly inserted back into the *mPWP* simulation's newly-

314   exposed grid points; it is acknowledged that this introduces new dust emissions source regions, which

315   may well impact the resulting Pliocene climate state.

316

317   *2.3.2.4. Initial conditions*

318   Oceanic initial conditions, such as ocean temperature and salinity, were derived from the mean

319   equilibrium state of the *piControl* simulation. Some atmospheric initial conditions, such as those

320   relating to the land surface (e.g. soil moisture and soil temperature at four levels of depth), used the

321   same method as that applied to soil properties. These fields contain monthly varying values, therefore

322   appropriate timings were considered e.g. if the majority of grid points with the above balance were in

323   the Northern Hemisphere, then initial conditions during Northern Hemisphere summer were used for

324   newly-exposed regions in Greenland (and likewise during Southern Hemisphere summer for newly-

325   exposed regions in Antarctica). For the soil temperature field and particularly at upper levels, this

326   process resulted in sharp temperature gradients across western Antarctica, therefore the field was

327   spatially smoothed so that the gradients were more consistent with those in the *piControl*. Examples

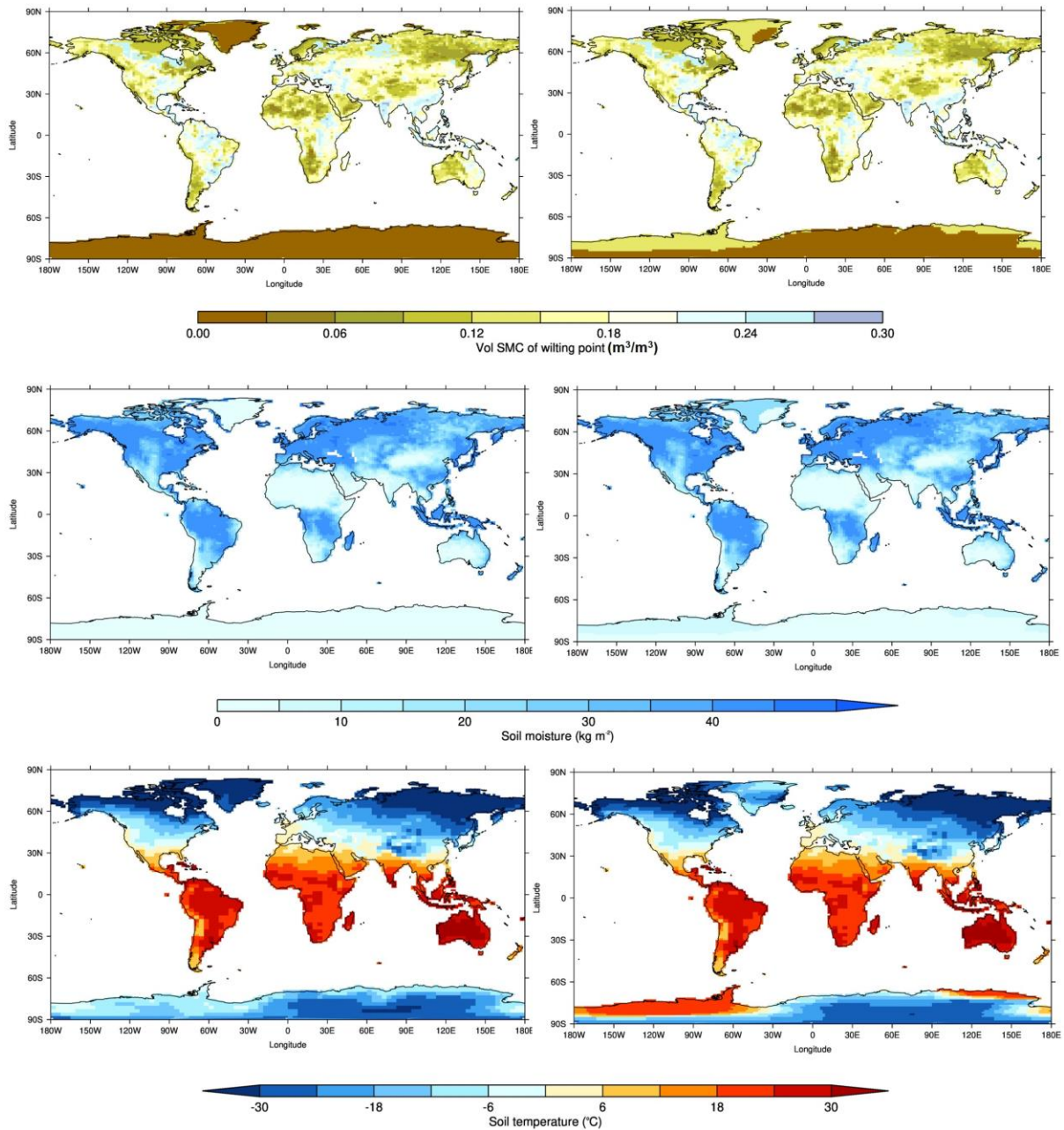
328   of the above soil-related fields are shown in Figure 5 for an example month and vertical level. A

329   complete list of the soil parameters and soil dust properties, and how each were changed relative to

330   the *piControl*, are shown in the Supplementary Material (Fig. S3 and Fig. S4, respectively).

331

332   Outside of the ice regions (i.e. outside Greenland and Antarctica), in the *mPWP* simulation the above

333   soil-related fields were kept identical as in the *piControl*.

334

335 Figure 5 – Example of soil-related fields used in HadGEM3. Left-hand column: *piControl* simulation, right-hand column:

336 *mPWP* simulation. First row: Soil parameters (example shows Volumetric soil moisture content at wilting point); Second

337 row: Soil moisture (example shows January, top-level); Third row: Soil temperature (example shows January, top-level).

338 Complete list of fields shown in Supplementary Material Fig. S3 and S4

339

2.3.3. *Changes to input parameters*

341 A small number of model input parameters were changed in the *mPWP* simulation, to make the model

342 more stable under the Pliocene boundary conditions. Firstly a parameter governing the implicit solver

343 for unstable atmospheric boundary layers was increased, and secondly three parameters for the

344 treatment of canopy snow were made consistent between BLT and NLT. The same parameter

345 changes will be included in the subsequent version of the physical model (GC4), in order to address

346 occasional model failures which were seen following the release of GC3.1. They will be described in

347   more detail in a GC4 model documentation paper, however testing of those changes for GC4 has

348   found that they have no detectable impact on model climatology.

349

350   **2.4. Modified *piControl* simulation**

351   Given that the official CMIP6 *piControl* simulation did not use the aforementioned model input

352   parameter changes, a slightly modified version of this simulation was re-run (simulation ID: u-bq637),

353   identical to the *piControl* other than including the parameter changes outlined in Section 2.3.3

354   (hereafter referred to as the *piControl_mod* simulation). This was run for 200 years, and the last 50-

355   year climatology is considered here in Sections 3 and 4.

356

357   **3. LARGE-SCALE FEATURES OF HADGEM3**

358   **3.1. Spin-up phase**

359   Consistent with other paleoclimate model experiments, the simulation should be run for as long as

360   possible to allow the model to reach a state of equilibrium, before the climatology is calculated over

361   the last 30, 50 or 100 years (Lunt *et al*. 2017). With this model, however, running for thousands of

362   years (especially important in obtaining oceanic equilibrium) was unfeasible given time and resource

363   constraints. By the end of the simulation, therefore, there was a total of 576 years for the *mPWP*

364   simulation, 526 of which are considered spin-up and 50 of which form the final climatologies; this is

365   approximately consistent with the 652 years of spin-up used by Menary *et al*. (2018).

366

367   *3.1.1. Evolution of mPWP simulation*

368   The HadGEM3 *mPWP* simulation was run in multiple parts, each starting from the endpoint of the

369   last, and each introducing additional boundary conditions so as to gradually move from PI conditions

370   to full Pliocene conditions. The *mPWP* simulation was started from the endpoint of the CMIP6

371   *piControl* simulation, specifically the last part of its spin-up phase (u-aq853), consistent with other

372   CMIP6 HadGEM3 paleoclimate simulations such as those of the mid-Holocene and Last Interglacial

373   periods (see Williams *et al*. 2020). The evolution of the *mPWP* simulation is shown in Figure 6,

374   where each stage is labelled and the resulting impact on the global mean 1.5 m air temperature is

375   shown. The first part of the *mPWP* simulation (u-bq448) is a straight copy of the CMIP6 *piControl*

376   production run (u-ar766), with no modifications other than increasing the atmospheric CO_2 to 400

377   ppmv; identical, therefore, to an E^{400} experiment following the naming convention of H16. This ran

378   for ~20 model years, before branching off to a new suite (u-br005) and introducing atmospheric ozone

379   appropriate for Pliocene conditions and Pliocene orography (see Section 2.3.1 and 2.3.2,

380   respectively). This ran for ~60 model years, before branching off to a new suite (u-br871) and

381   introducing a Pliocene-appropriate ice mask along with appropriate values for soil parameters, soil

382   dust, soil moisture, soil temperatures and snow depth over these newly created ice regions (see

383   Section 2.3.2); this, therefore would be the Eoi^{400} experiment following the naming convention of

384 H16. It should be noted, however, that at this stage this naming convention is not strictly consistent

385 with that used by H16, because they specify that orography, lakes and soils should be modified in

386 unison, and therefore *o* signifies changes to orography, bathymetry, land-sea mask, lakes and soils

387 together. In contrast, at this stage of the simulation, most boundary conditions are consistent with the

388 experimental design of H16, except vegetation, soils in non-ice regions and lakes. This ran for ~280

389 model years (during which time the task of creating appropriate Pliocene vegetation was completed),

390 before branching off to a new suite (u-bv241) and introducing a minor parameter change to allow

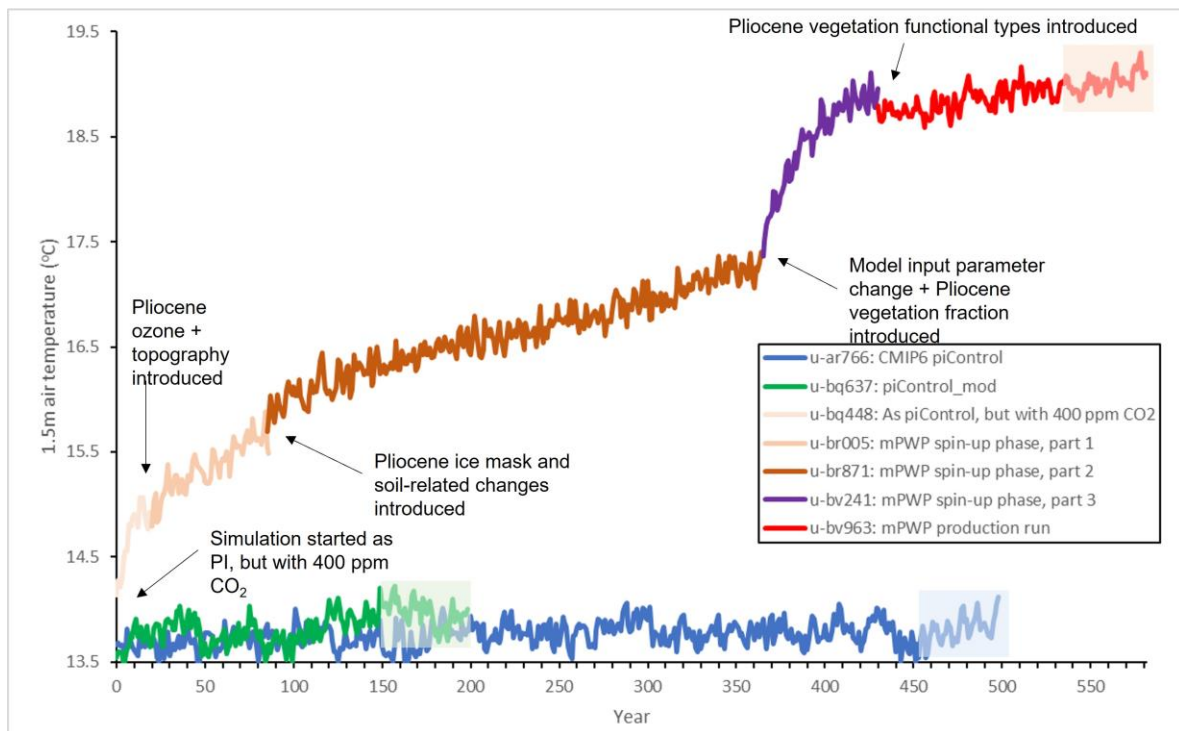
391 inclusion of the Pliocene vegetation (see Section 2.3.3), as well as the full Pliocene vegetation

392 fractions. This ran for a further ~60 years, to check the stability of the model in response to the

393 vegetation change, before branching off to a new and final suite (u-bv963), in which the full Pliocene

394 vegetation functional types were introduced. This ran for ~150 years, with the final climatology

395 (presented here in Section 3 and 4) being taken from the last 50 years i.e. allowing a 100-year buffer

396 between the final update to the model and the actual results.

397

398 Figure 6 – Annual global mean 1.5 m air temperature from the HadGEM3 *mPWP* spin-up phase and production run, as well

399 as the CMIP6 *piControl* and the *piControl_mod*. Labels show introduction of each new Pliocene element. Climatologies

400 discussed here are taken from final 50 years of each simulation (shown by shaded boxes). See Williams *et al.* (2020) for the

401 *piControl* spin-up phase that preceded these simulations.

402

403 As well as the various stages of the *mPWP* simulation, Figure 6 also shows timeseries from the

404 official ~500 year CMIP6 *piControl* simulation (Kuhlbrodt *et al.* 2018 and Menary *et al.* 2018) and

405 the 200 year *piControl_mod* conducted here, and Figure S7 shows climatologies of 1.5 m temperature

406 and surface precipitation calculated over the last 50 years of each simulation. As the figures show,

407 there is little or no difference between the two PI simulations (also suggested above in Section 2.3.3);

408   using temperature as an example, over the last 50 years of the simulations there is a mean of 13.79°C

409   and 13.97°C for the *piControl* and *piControl_mod* respectively, and a standard deviation of 0.13°C for

410   both, further confirming the negligible impact of the model parameter change in the model

411   climatology.

412

413   *3.1.2. Atmospheric and oceanic equilibrium of the mPWP simulation*

414   Concerning atmospheric equilibrium, Table 1 shows summary statistics for annual global mean 1.5 m

415   air temperature and net top of atmosphere (TOA) radiation from the last 50 years of the *mPWP*

416   simulation, compared to both the *piControl* and *piControl_mod* simulations; see Figure 6 for the

417   entire timeseries of Pliocene 1.5 m air temperature, and Figure S5 in the Supplementary Material for

418   the TOA radiation equivalent.

Variable	*piControl*	*piControl_mod*	*mPWP*
1.5m air temperature trends ($°C$ century^{-1})	0.51	-0.47	0.34
TOA radiation trends (W m^2 century^{-1})	0.02	-0.2	-0.17
Mean TOA radiation (W m^2)	0.18	0.21	0.88
Global ocean volume-mean temperature trends (°C century^{-1})	0.03	0.04	0.21
Global ocean volume-mean salinity trends (psu century^{-1})	0.0004	-0.0002	-0.004

419

420   Table 1 - Centennial trends (calculated via a linear regression) and climatology over the last 50 years of the simulations. A

421   positive TOA imbalance indicates a net loss of energy from the Earth System

422

423   Although the *mPWP* simulation is clearly warming considerably during the ~500 year run (and

424   especially when the Pliocene vegetation fraction is introduced), with trends of 0.77°C century^{-1}, it

425   levels off over the final 50 years, with trends of 0.34°C century^{-1} (Table 1). These values are higher

426   than those considered by some (e.g. Menary *et al*. 2018) to be acceptable for equilibrium, however

427   given time and resource constraints it was not possible to run the simulation further. The spatial

428   patterns of these trends, shown in Figure S6 in the Supplementary Material, shows the majority of the

429   warming occurring over high latitude regions in both Hemispheres, related to the removal of the ice

430   sheets and sea ice loss. By the end of the *mPWP* simulation, the mean TOA radiation balance is 0.88

431   W m^{-2}, significantly higher than either of the PI simulations suggesting that the *mPWP* simulation is

432   not yet in full atmospheric equilibrium. This TOA inbalance is reducing at a rate of 0.17 W m^{-2}

433   century^{-1} at the end of the simulation. A brief discussion of how the HadGEM3 *mPWP* simulation's

434   atmospheric equilibrium compares to that of the other Hadley Centre models presented here

435   (introduced in Section 4) is given in the Supplementary Material (see Section 2 and Table S2).

436

437   When the *mPWP* simulation was stopped, the global annual mean 1.5 m temperature was

438   approximately 19°C (Fig. 6). A Gregory plot (Gregory *et al*. 2004) of the evolution of TOA energy

439   inbalance and surface temperature can indicate how much more warming the model may have

440   experienced if it had been run to full equilibrium. The results of this analysis suggest the model

441   would come to equilibrium ~1.5°C higher (see Supplementary Material, Fig. S8), at 20.5°C i.e. an

442   anomaly relative to preindustrial of 6.6°C. This is the case when the extrapolation is carried out on

443   either of the final two parts of the simulation (in red and in purple in Fig. S8), suggesting that the

444   introduction of the Pliocene vegetation functional types is not having a great impact on the final

445   global mean temperature. However, this analysis is associated with some uncertainty, related to the

446   interannual variability in temperature and TOA energy inbalance, and to the fact that the linear

447   extrapolation may not be appropriate if the feedbacks vary non-linearly (e.g. Knutti *et al*. 2015).

448

449   As an example of oceanic equilibrium, Table 1 also shows summary statistics for volume integral

450   annual global mean ocean temperature and salinity from the end of the *mPWP* simulation, compared

451   to both the *piControl* and *piControl_mod* simulations; see Figure S9 in the Supplementary Material

452   for the Pliocene timeseries. Ocean temperature is steadily increasing throughout the *mPWP*

453   simulation, and likewise ocean salinity is steadily decreasing (Fig. S9). Freshwater fluxes to the

454   ocean representing iceberg calving and ice sheet basal melt are calibrated for the *piControl*, as

455   described in Sellar *et al*. (2020). These fluxes are calibrated to match the ice sheet surface mass

456   balance (SMB) expected in the *piControl*, so that salinity drift is minimised. The Pliocene SMB is

457   smaller than that in the *piControl*, and hence net flux of water to the ocean is positive, leading to the

458   salinity drift. If compute resources allowed for a much longer Pliocene simulation, this ocean flux

459   could be calibrated to Pliocene SMB once the temperature and SMB had stabilised, or calculated

460   iteratively. The long-term trends, Table 1, provide similar conclusions to those from the atmospheric

461   trends, with for example centennial temperature trends of 0.21°C century^{-1} being much higher than the

462   PI simulations (0.03°C century^{-1} and 0.04°C century^{-1} for the *piControl* and *piControl_mod*,

463   respectively). Although these values again do not meet the criteria of Menary *et al*. (2018) for

464   oceanic equilibrium, given the aforementioned computational cost of this model it was not possible to

465   run the simulations further; this is even more true in the ocean, which would require many thousands

466   of years of model simulation to reach equilibrium. This compromise has been equally necessary for

467   other computationally expensive paleoclimate simulations (e.g. Williams *et al*. 2020).

468

469   **3.2. Simulation comparison: *mPWP* versus *piControl_mod* climatologies**

470   Here the focus is on mean differences between the HadGEM3 *mPWP* simulation and its

471   corresponding modified PI simulation, *piControl_mod* (Section 2.4).   All of the following discussion

472   and figures relate to climatologies calculated over the last 50 years of the simulations, and all are

473   anomalies i.e. Pliocene - PI. Annual and seasonal mean summer/winter 1.5 m air temperature

474   (hereafter referred to as near-surface air temperature, SAT) anomalies are shown in Figure 7. The

475   annual global mean SAT anomaly for this 50-year climatology is 5.1°C. Warming relative to the PI is

476   evident throughout the year and globally, but more so over: i) landmasses (6.8°C and 4.5°C for the

477   annual mean SAT over land and ocean, respectively); ii) the Northern Hemisphere (8.5°C and 6.3°C

478   for annual mean SAT in the Northern and Southern Hemisphere extratropics (>45°), respectively).

479   Warming is also evident over high latitudes (>60°) of both hemispheres (10.9°C and 8.5°C for the

480   Northern and Southern Hemisphere, respectively, and exceeding 12°C in some places). These

481   particular metrics were chosen to be consistent with those used by H20 (see Section 4.2). Over the

482   tropics (20°N-20°S) the amount of warming is less than at higher latitudes, but the Pliocene is still

483   much warmer than the PI with annual mean SAT anomalies of 4.6°C and 3.7°C when averaged over

484   tropical land and ocean, respectively. This global and regional warming is consistent with, albeit

485   slightly warmer than, other work, namely the results from PlioMIP1 (Haywood *et al*. 2013) and

486   PlioMIP2 (see Section 4.2). The majority of the annual mean warming (Fig. 7a) in Northern

487   Hemisphere high latitudes is accounted for during that hemisphere's winter (December-February,

488   DJF) with a mean warming of 15°C (Fig. 7b), and likewise the majority of the annual mean warming

489   in Southern Hemisphere high latitudes is accounted for during that hemisphere's winter (June-August,

490   JJA) with a mean warming of 10.6°C (Fig. 7c). If the entire hemisphere, rather than >60°, is

491   considered, then this greater winter contribution to the annual mean is still true, although less so (e.g.

492   5.6°C, 6.1°C and 5.4°C for the annual, DJF and JJA means respectively in the Northern Hemisphere).

493

494   The regions of polar SAT increases, and seasonal variation, are likely explained by the changes in sea

495   ice, shown in Figure 8 (for the absolute values in sea ice fraction, see Fig. S10 in the Supplementary

496   Material). Reductions in sea ice are shown throughout the year in both hemispheres, consistent with

497   previous work (e.g. Cronin *et al*. 1993, Howell *et al*. 2016, Moran *et al*. 2006, Polyak *et al*. 2010).

498   Here, although a reduction in sea ice (of up to 70%) is evident throughout the year in either

499   hemisphere, at the seasonal timescale the largest loss (exceeding 70% in some places, such as the

500   polar Arctic and Antarctic) is seen during each hemisphere's winter (Fig. 8a and 8d). The

501   regions/timings of maximum warming (Fig. 7b-c) correspond well to the regions/timings of maximum

502   sea ice loss, implying a role for the sea ice-albedo feedback. When sea ice area is averaged over each

503   hemisphere (Fig. 8e), the Northern Hemisphere is clearly losing more sea ice in the *mPWP* simulation

504   (relative to the *piControl_mod*) than the Southern Hemisphere. However the amount of loss in the

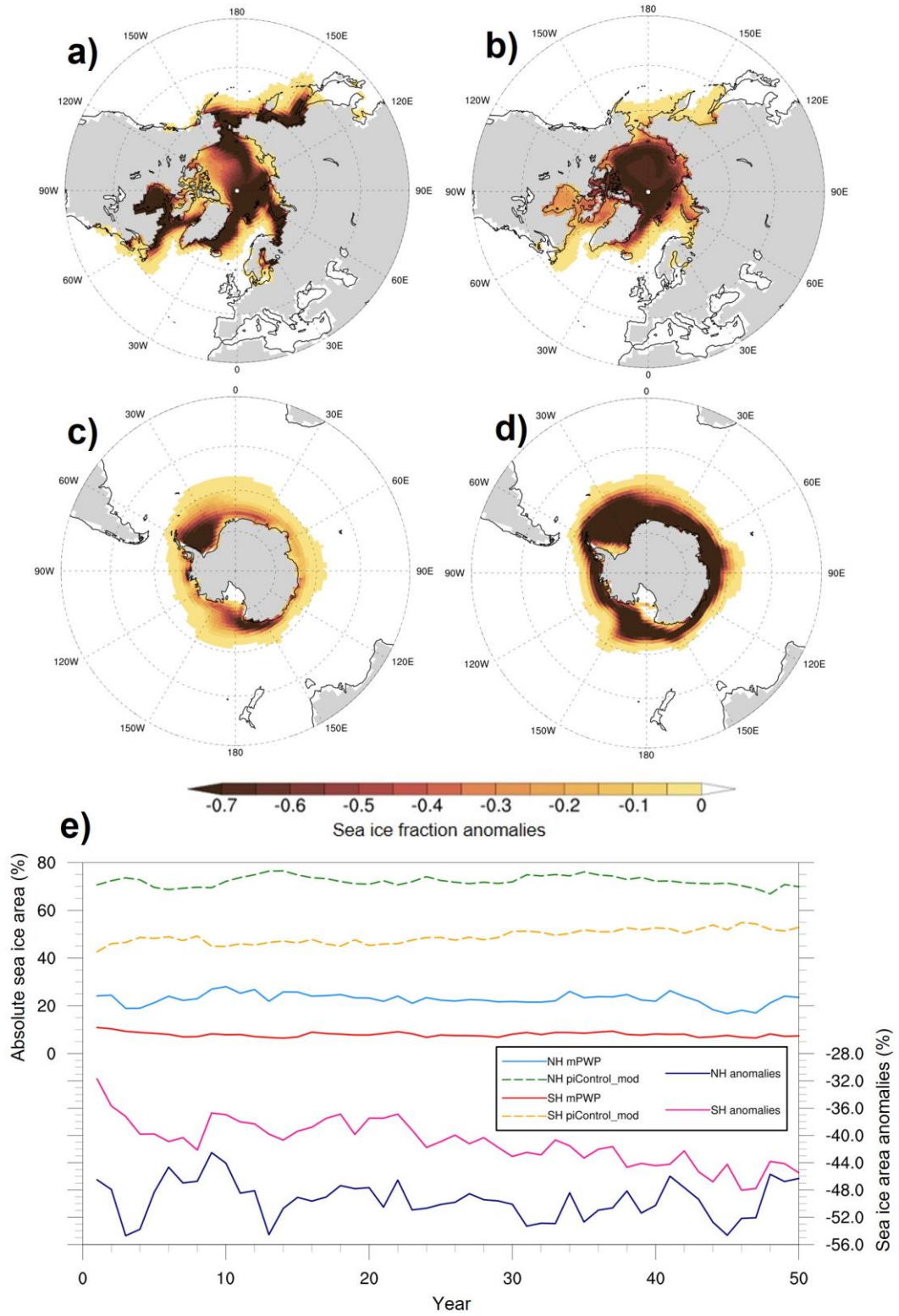
505   Southern Hemisphere is steadily increasing during the last 50 years of the *mPWP* simulation,

506   suggesting that had the model been allowed to run to full equilibrium, the difference between the

507   hemispheres would be reduced.

508

509   Figure 7 – 1.5 m air temperature climatology differences (*mPWP – piControl_mod*) from HadGEM3. a) Annual; b) DJF; c)

510   JJA

511

512 Figure 8 – Sea ice fraction climatology differences (*mPWP – piControl_mod*) from HadGEM3: a) Northern Hemisphere

513 DJF, b) Northern Hemisphere JJA, c) Southern Hemisphere DJF, d) Southern Hemisphere JJA, e) Mean sea ice area (both

514 absolute values and differences) averaged over either hemisphere

515

516 Annual and seasonal mean surface daily precipitation anomalies are shown in Figure 9. The annual

517 global mean precipitation anomaly for this 50-year climatology is 0.34 mm day^{-1}. In addition to the

518 precipitation increases at high latitudes at the annual timescale (Fig. 9a), which are again mostly

519 accounted for by changes during the Northern and Southern Hemisphere's winter (Fig. 9b and c,

520 respectively), the largest change relative to the PI is a northward displacement of the ITCZ. All

521 timescales are showing wetter conditions over oceans to the North of the equator and drier conditions

522 over oceans to the South of the equator. This is similar to work by Li *et al*. (2018), who suggested a

523 poleward movement of Northern Hemisphere monsoon precipitation in PlioMIP1. There is also a

524 noticeable enhancement of monsoon systems such as the East Asian and West African monsoon,

525 consistent with previous work (e.g. Zhang *et al*. 2013, 2016). In some places, these changes exceed

526 ~2 mm day^{-1}, geographically consistent with (albeit again much higher than) other work, such as the

527 multi-model ensemble mean (MME) from PlioMIP2 models where increases rarely exceed ~1.2 mm

528 day^{-1} (see Section 4.2). These changes, and indeed the temperature changes over Northern

529 Hemisphere landmasses, may be associated with changes to the total cloud cover, shown in Figure 10.

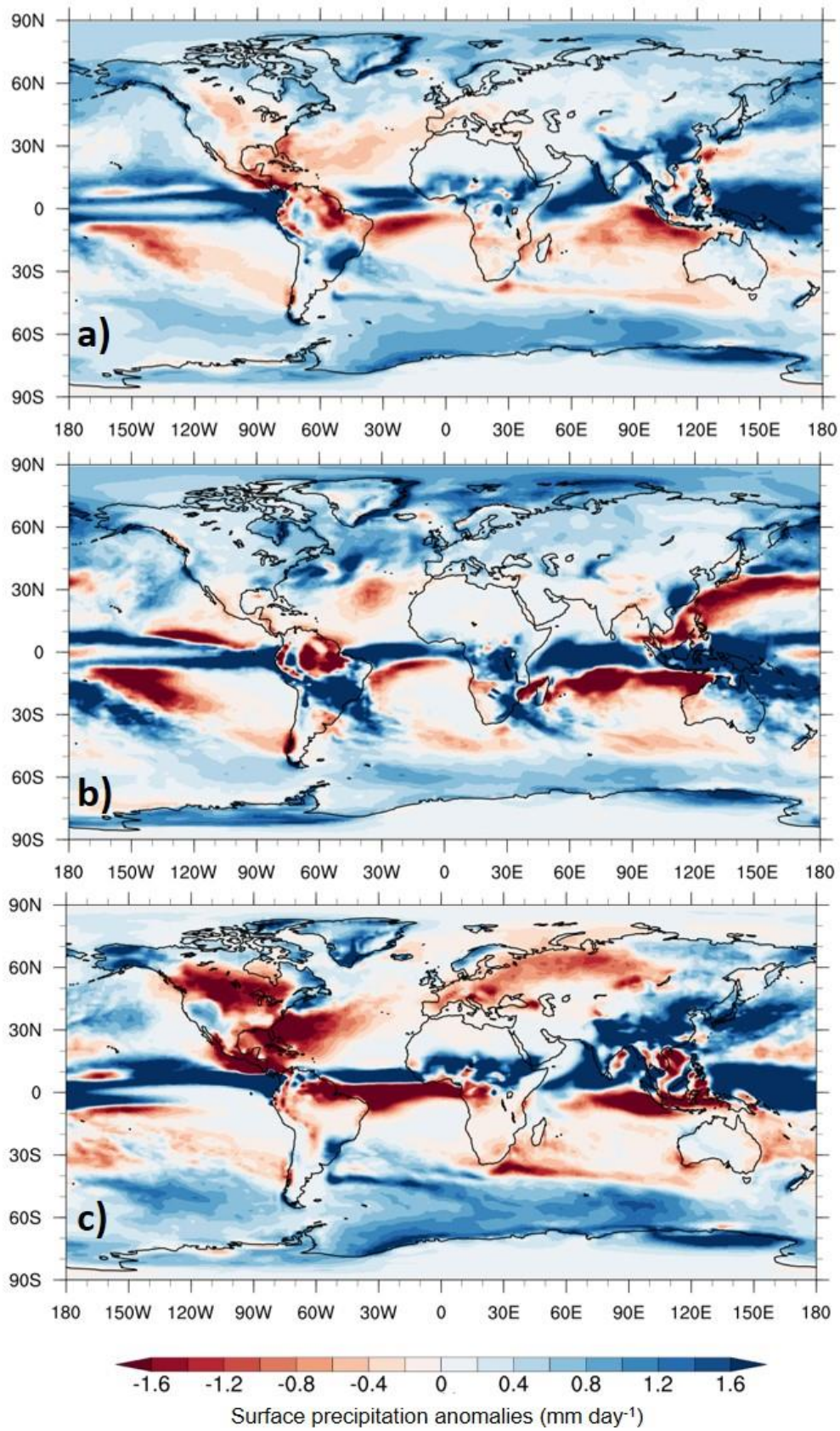
530 Although the changes are small at the annual timescale (Fig. 10a), during Northern Hemisphere

531 winter (Fig. 10b) there is a noticeable increase in cloud cover (of ~10%) over high latitude regions,

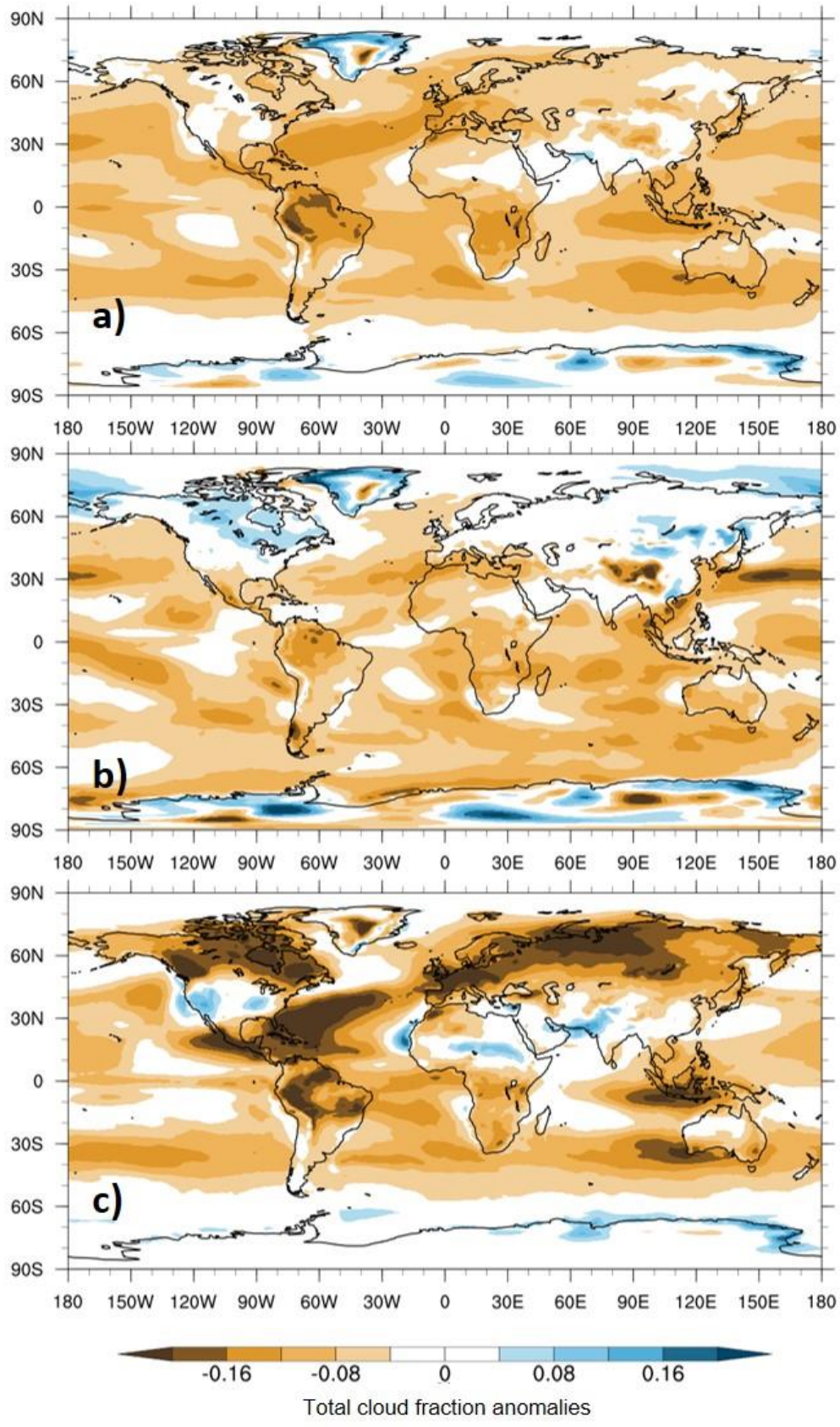
532 corresponding to the increases in precipitation. Likewise, during Northern Hemisphere summer (Fig.

533 10c) there is a large reduction (over 20% in places) in cloud cover, especially over Northern

534 Hemisphere landmasses; these regions, such as Europe and Northern Asia, correspond well to the

535 areas of decreased precipitation and increased temperature.

536

537   Figure 9 – Surface precipitation climatology differences (*mPWP – piControl_mod*) from HadGEM3. a) Annual; b) DJF; c)

538   JJA

539

540   Figure 10 – Total cloud fraction climatology differences (*mPWP – piControl_mod*) from HadGEM3. a) Annual; b) DJF; c)

541   JJA

542

543 **4. COMPARISON OF HADGEM3 WITH OTHER MODELS AND PROXY DATA**

544 **4.1. Model-model and model-data comparison: Different generations of UK model versus proxy**

545 **data**

546 Here the focus is on mean SST differences between different generations of the UK's physical climate

547 model, starting with three Pliocene simulations using the original fully-coupled climate model

548 HadCM3, then a simulation from the more recent HadGEM2 and finally the *mPWP* simulation from

549 HadGEM3. See Supplementary Material for details of these older models. For HadCM3, three

550 separate Pliocene simulations (and corresponding PIs) are used; the first two were conducted by Lunt

551 *et al*. (2011) and Bragg *et al*. (2012), and are referred to as HadCM3-PRISM2 and HadCM3-

552 PlioMIP1, respectively (see Table 2). This is to distinguish them from a third version of the same

553 model included in PlioMIP2, referred to here as HadCM3-PlioMIP2.

Model	Model name here	MIP	Boundary conditions	Reference
HadCM3	HadCM3-PRISM2	-	PRISM2	Lunt *et al*. 2011
HadCM3	HadCM3-PlioMIP1	PlioMIP1	PRISM3	Bragg *et al*. 2012
HadCM3	HadCM3-PlioMIP2	PlioMIP2	PRISM4	Hunter *et al*. 2019
HadGEM2-AO	HadGEM2	PlioMIP1	PRISM3	Tindall and Haywood 2020
HadGEM3-GC31-LL	HadGEM3	PlioMIP2	PRISM4	Presented here

554

555 Table 2 - Different generations of the UK physical climate model used here, and their involvement with PlioMIP

556

557 Multi-proxy SST data from the KM5c interglacial, compiled by McClymont *et al*. (2020), were used

558 for comparative purposes. Here, they focus on a narrow time-slice from 3.195 to 3.215 Ma, and

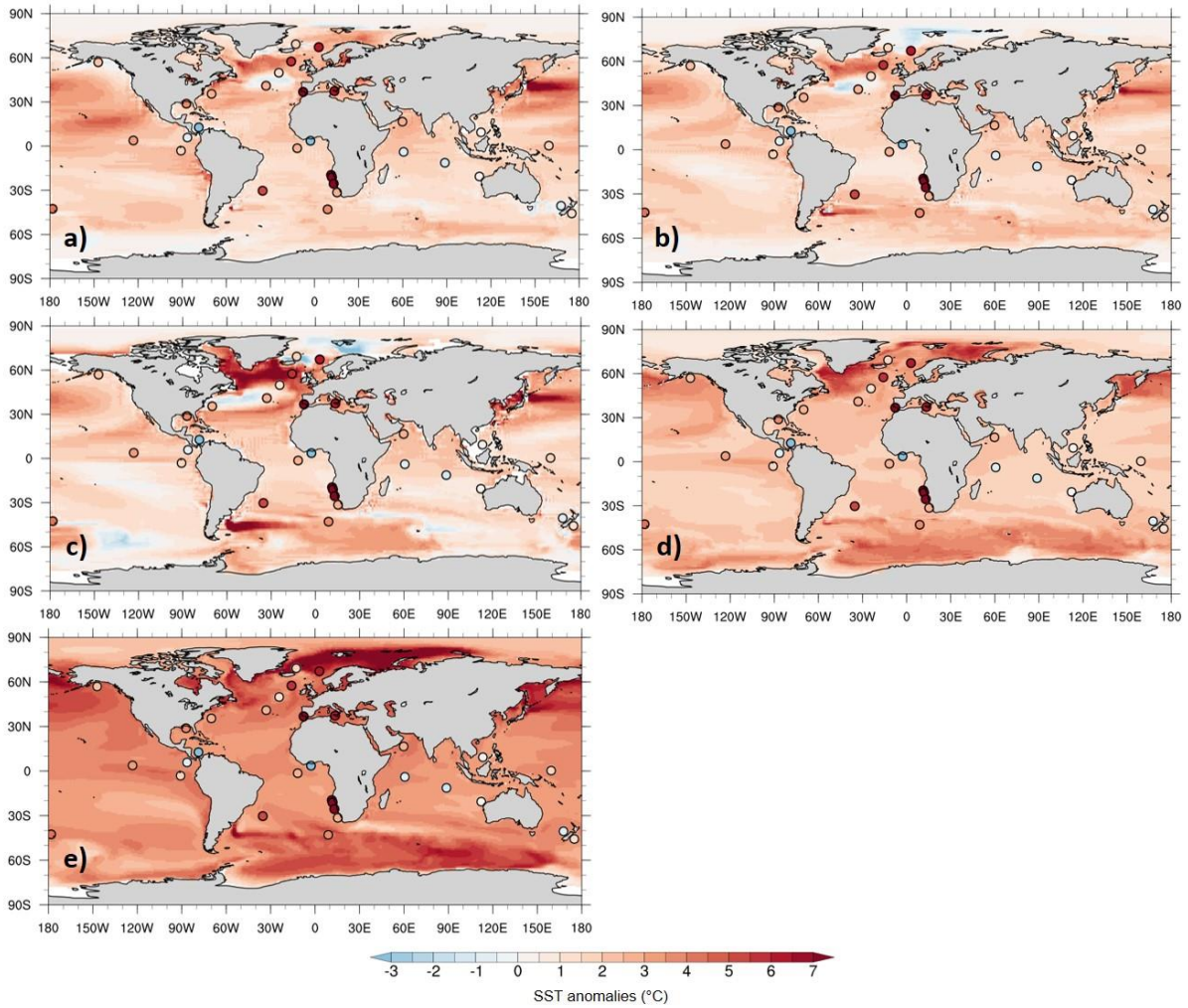
559 compile the SST data from two proxies: an alkenone-derived $U^{K'}_{37}$ index (Prahl and Wakeham, 1987)

560 and foraminifera calcite Mg/Ca (Delaney *et al*. 1985), with the resulting data comprising the PlioVAR

561 synthesis and covering 32 locations between 46°S-69°N (McClymont *et al*. 2020).

562

563 Maps of annual mean SST anomalies from the simulations, overlaid with the proxy data, are shown in

564 Figure 11 and summary statistics are shown in Table 3.

565

566 Figure 11 – Annual mean SST differences (Pliocene – PI) from different generations of the UK's physical climate model. a)
567 HadCM3-PRISM2; b) HadCM3-PlioMIP1; c) HadCM3-PlioMIP2; d) HadGEM2; e) HadGEM3. Background gridded data
568 shows model simulations, filled circles show SST proxy data from McClymont *et al.* (2020)

569

570

	HadCM3-PRISM2	HadCM3-PlioMIP1	HadCM3-PlioMIP2	HadGEM2	HadGEM3
Global mean (°C)	1.63	1.53	1.67	2.29	3.80
RMSE	3.55	3.62	3.59	3.23	3.36

571

572 Table 3 - Global annual mean SST anomalies from Pliocene simulations using different generations of the UK's physical
573 climate model, and RMSE values between simulations and SST proxy data from McClymont *et al.* (2020)

574

575

576

577   The global annual SST anomaly for HadGEM3 is 3.8°C, followed by HadGEM2 at 2.3°C, and then

578   1.7°C, 1.5°C and 1.6°C for the three HadCM3 simulations (starting with the most recent, HadCM3-

579   PlioMIP2; see Table 3). Comparing the newest model (HadGEM3) with the oldest model (HadCM3-

580   PRISM2), which have an anomaly of 3.8°C and 1.6°C respectively, clearly the most recent generation

581   is showing a much warmer Pliocene.

582

583   Comparing an earlier generation of the model with a later generation, but with identical boundary

584   conditions (HadCM3-PlioMIP1 and HadGEM2, respectively; Fig. 11b and Fig. 11d), aside from the

585   greater overall warming (2.3°C in HadGEM2 versus 1.5°C in HadCM3-PlioMIP1) already discussed

586   above, the main spatial patterns of warming are similar, with both showing the greatest warming over

587   the Labrador Sea and the north-west Pacific and HadGEM2 showing greater polar amplification

588   overall. In part thanks to this high latitude warming, root mean squared error (RMSE) values are

589   3.2°C and 3.6°C for HadGEM2 and HadCM3-PlioMIP1, respectively, showing a greater agreement

590   between the proxy data and HadGEM2 (Table 3).

591

592   Likewise, comparing the other older model with the most recent (HadCM3-PlioMIP2 and HadGEM3,

593   respectively; Fig. 11c and Fig. 11e), the spatial patterns of warming differ more widely, with

594   HadGEM3 showing widespread Northern Hemisphere high latitude warming that is not shown by

595   HadCM3-PlioMIP2 at all, other than in the Labrador Sea. HadGEM3, and indeed HadGEM2, are

596   displaying a greater extent of polar amplification in both hemispheres (Fig. 11d-e). As the warmest

597   model, HadGEM3 (RMSE = 3.4°C) is showing less agreement with the proxy data than HadGEM2

598   (RMSE = 3.2°C), likely because it is so warm that the discrepancy with the colder proxy data

599   locations (such as in the Indian Ocean, near New Zealand or off equatorial Africa) is greater (Fig.

600   11e). This is in spite of the fact that, in the warmer proxy data locations (such as in the North Atlantic

601   and Arctic) HadGEM3 is closer to the proxy data. In these regions, the earlier versions of the model

602   (Fig. 11a-c) are not even capturing the sign of change and are showing a weak cooling, in stark

603   contrast to the proxy data, that neither HadGEM2 nor HadGEM3 display (Fig. 11d-e). Where proxy

604   data suggest colder conditions, again none of the models are capturing the sign of change and all show

605   widespread warming, and this is most evident in HadGEM3 because of its particularly strong

606   warming. The fact that all of the HadCM3 simulations are showing several regions of cooling and

607   have a higher RMSE than the most recent versions suggests that this early model might be too cold.

608   In contrast, the fact that HadGEM3 has a higher RMSE than HadGEM2 suggests that, despite

609   involving significant model development (see Williams *et al*. 2020 for a summary), concerning

610   Pliocene climate HadGEM3 may actually be too warm. Therefore, whilst model development appears

611   to have improved the model's agreement with proxy data since earlier versions of the model, this only

612   appears to be true up to a certain point; the "sweet spot" appears to be HadGEM2. Moreover, given

613   the aforementioned point about the *mPWP* simulation not being in full equilibrium and being ~1.5°C

614 warmer if it had been (see Section 3.1.2), it is likely that both the SST anomaly and the RMSE values

615 would be higher when in equilibrium and therefore the performance against proxy data may be lower

616 than indicated here.

617

618 **4.2. Model-model comparison: HadGEM3 versus PlioMIP2 models**

619 Finally, the focus here is on mean differences, again considering SAT and precipitation anomalies,

620 between the *mPWP* simulation from HadGEM3 and the Pliocene simulations from all other available

621 models included in PlioMIP2 (Table 4).

Model, and modelling centre responsible for simulation	Spatial resolution (lon x lat)		ECS (°C)
	Atmosphere	**Ocean**	
CCSM4, National Centre for Atmospheric Research, US	1° x 1°	1° x 1°	3.2
CCSM4_Utr, Utrecht University, the Netherlands	2.5° x 1.9°	1° x 1°	3.2
CCSM4_UoT, University of Toronto, Canada	1° x 1°	1° x 1°	3.2
CESM1.2, National Centre for Atmospheric Research, US	1° x 1°	1° x 1°	4.1
CESM2, National Centre for Atmospheric Research, US	1° x 1°	1° x 1°	5.3
COSMOS, Alfred Wagner Institute, Germany	3.75° x 3.75°	3.0° x 1.8°	4.7
EC-Earth3.3, Stockholm University, Sweden this	1.125° x 1.125°	1° x 1°	4.3
AGISS-E2-1-G, Goddard Institute for Space Studies, US	2.0° x 2.5°	1.0° x 1.25°	3.3
HadCM3, University of Leeds, UK	2.5° x 3.75°	1.25° x 1.25°	3.5
IPSLCM5A, Laboratoire des Sciences du Climat et de l'Environnement, France	3.75° x 1.9°	2.0° x 2.0°	4.1
IPSLCM5A2, Laboratoire des Sciences du Climat et de l'Environnement, France	3.75° x 1.9°	2.0° x 2.0°	3.6
IPSL-CM6A-LR, Laboratoire des Sciences du Climat et de l'Environnement, France	2.5° x 1.26°	1.0° x 1.0°	4.8
MIROC4m, University of Tokyo, Japan	2.8° x 2.8°	1.4° x 1.4°	3.9
MRI-CGCM2.3, University of Tsukuba, Japan	2.8° x 2.8°	2.0° x 2.0°	2.8

NorESM-L, Bjerknes Centre for Climate Research, Norway	3.75° x 3.75°	3.0° x 3.0°	3.1
NorESM-F, Bjerknes Centre for Climate Research, Norway	1.9° x 2.5°	1.0° x 1.0°	2.3

Table 4 - Climate models included here from PlioMIP2 (see Haywood *et al*. 2020 for each model's reference)

A number of different metrics of SAT are shown in Figure 12 for each of the models, as well as the MME; the panels shown here are updated versions of those shown in H20, but now including HadGEM3. It should be noted that, consistent with H20, the models are listed according to their published ECS, with the highest ECS listed first (see Table 4). HadGEM3 has an ECS of 5.5 K (Andrews *et al*. 2019), compared to the 2nd highest model (CESM2) with an ECS of 5.3 K (H20). If, however, all available models within CMIP6 (i.e. not just those having conducted Pliocene simulations) are considered, then HadGEM3 has the 2nd highest ECS, just below that of CanESM5 with an ECS of 5.6 K (Zelinka *et al*. 2020).

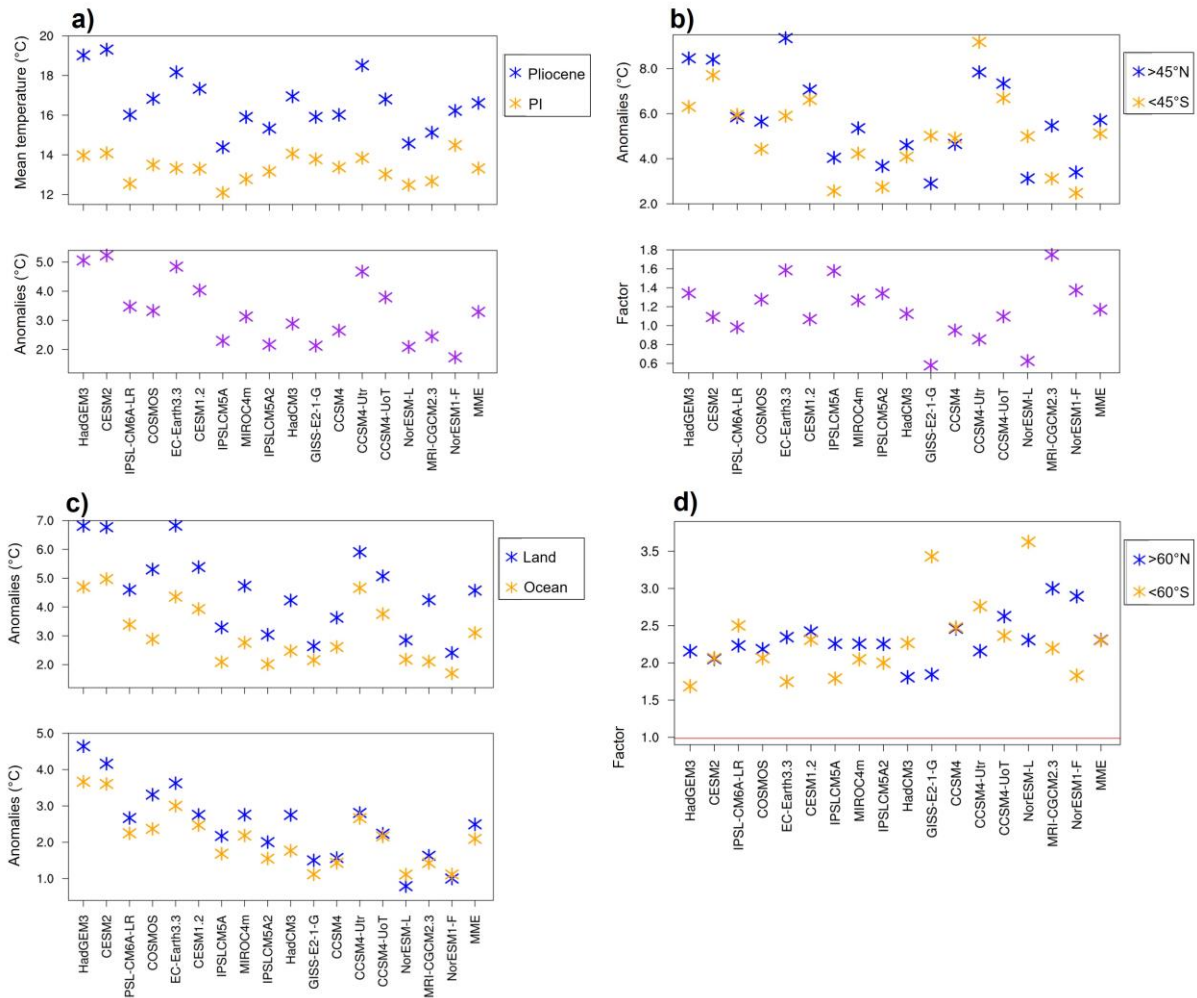



Figure 12 - SAT from Pliocene simulations from HadGEM3 and all other models in PlioMIP2. a) Global annual mean SAT (top panel) and anomalies (bottom panel); b) Extratropical (+/- 45°) annual mean SAT anomalies (top panel) and ratio (i.e.

636 >45°N divided by <45°S) between them (bottom panel); c) Land and ocean annual mean SAT anomalies, averaged globally

637 (top panel) and between 20°N-20°S (bottom panel); d) Annual mean SAT polar amplification i.e. SAT poleward of 60°

638 divided by global mean, for each hemisphere, where red line = ratio of 1 (i.e. no polar amplification). Figures reproduced

639 and adapted from Haywood *et al*. (2020)

640

641 As mentioned above (Section 3.2), the global annual SAT anomaly by the end of the *mPWP*

642 simulation is 5.1°C, making HadGEM3 one of the warmest models in PlioMIP2 and second only to

643 CESM2 (H20). This is true both in terms of its anomaly and its mean Pliocene SAT (19°C); this is

644 only lagging behind the warmest model by 0.2°C and 0.3°C for the anomalous and mean SAT,

645 respectively (Fig. 12a). HadGEM3 is much warmer than earlier global annual mean temperature

646 estimates (e.g. Haywood and Valdes 2004), and the range given by models included in PlioMIP1

647 (1.8°C to 3.6°C, see Haywood *et al*. (2013) and PlioMIP2 (1.7°C to 5.2°C, see H20). The impact of

648 including HadGEM3 amongst the models is to increase the MME anomaly by 0.1°C, from 3.2° to

649 3.3°C. Interestingly the HadGEM3 *piControl_mod* simulation is not presenting the warmest absolute

650 PI compared to the other models, coming 4th in the list, suggesting that HadGEM3 is more sensitive to

651 the Pliocene boundary conditions rather than being a generally warmer model overall.

652

653 Concerning annual global mean precipitation (Fig. 13, top panel), as mentioned above the

654 precipitation anomaly by the end of the simulation is 0.34 mm day^{-1}, making HadGEM3 not only one

655 of the warmest models in PlioMIP2 but also one of the wettest (consistent with current understanding,

656 as global precipitation is generally a function of global temperature). The range of anomalies across

657 all models during PlioMIP1 was 0.09 to 0.18 mm day^{-1} (Haywood *et al*. 2013), during PlioMIP2 it

658 was 0.07 to 0.37 mm day^{-1} (with the higher values being attributed to the models being more sensitive

659 to the updated PRISM4 boundary conditions) and the PlioMIP2 ensemble mean was 0.19 mm day^{-1}

660 (H20). Concerning the mean, it is the wettest model in terms of both its *mPWP* (3.49 mm day^{-1}) and

661 *piControl_mod* (3.15 mm day^{-1}) simulations, and both of these are much higher than the MME (3.06

662 mm day^{-1} and 2.86 mm day^{-1} for the Pliocene and PI simulations, respectively). The fact that both the

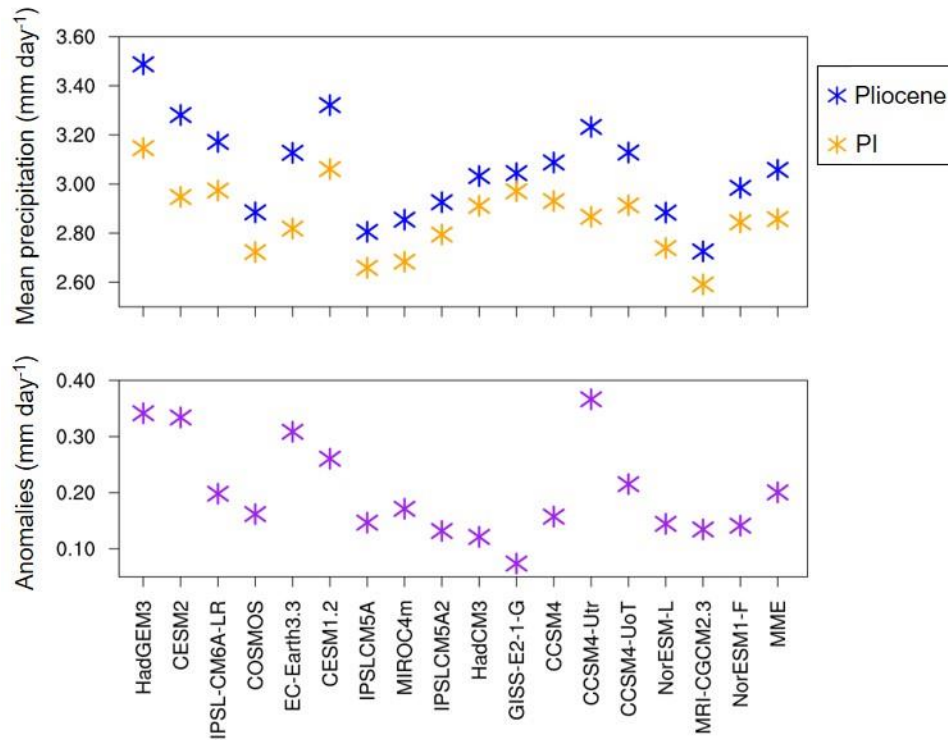
663 HadGEM3 *mPWP* and *piControl_mod* simulations are not only the wettest, but are also closer

664 together in terms of mean precipitation, means that if the anomaly is considered (Fig. 13, bottom

665 panel) HadGEM3 is not quite showing the greatest change relative to the PI; an anomaly of 0.34 mm

666 day^{-1} makes it 2nd only to CCSM4-Utr (at 0.37 mm day^{-1}). The impact of including HadGEM3

667 amongst the other PlioMIP2 models is to again slightly increase the MME anomaly, from 0.19 mm

668 day^{-1} as reported by H20 to 0.2 mm day^{-1} here.

669

670 Figure 13 - Global annual mean surface precipitation (top panel) and anomalies (bottom panel) from HadGEM3 *mPWP*

671 simulation and all other models in PlioMIP2, as well as multi-model ensemble mean (MME). Figure reproduced and

672 adapted from Haywood *et al*. (2020)

673

674 If the hydrological sensitivity (i.e. the relationship between global annual mean precipitation

675 anomalies and SAT anomalies) of the models is considered, then in line with current understanding

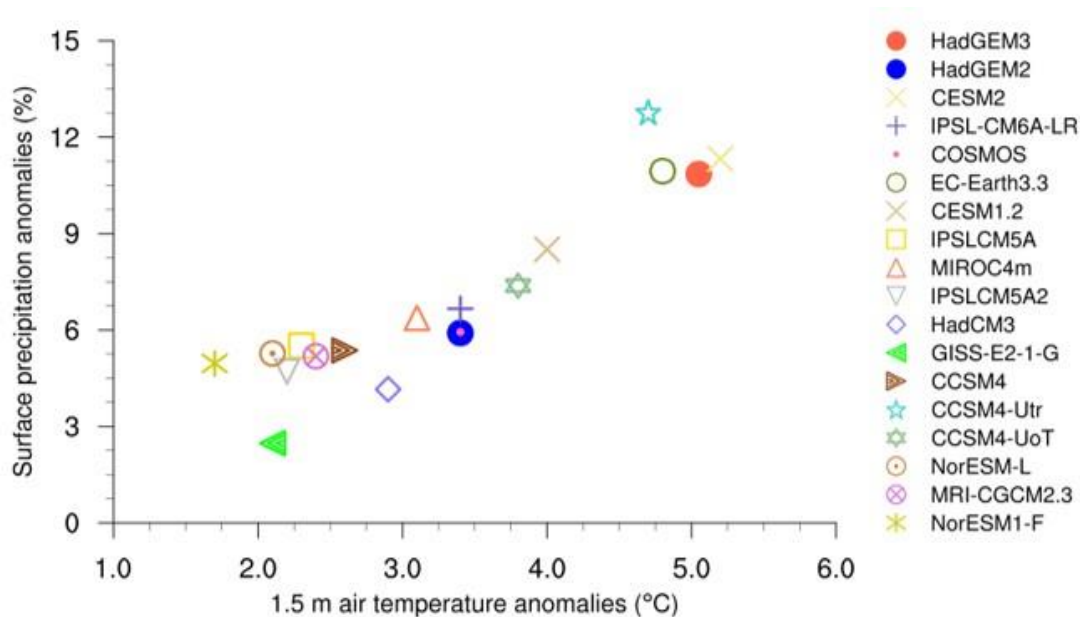
676 (e.g. Pendergrass and Hartmann 2014). there is a clear linear relationship shown by most of the

677 models, with Pliocene increases in precipitation increasing in line with SAT increases (Fig. 14). This

678 relationship is not entirely linear, however, with the aforementioned result being shown again here i.e.

679 although the HadGEM3 *mPWP* simulation is the 2nd warmest of all models in PlioMIP2, it is not the

680 wettest, suggesting that although the model is highly sensitive to the Pliocene forcings in terms of its

681 temperature response, it may be less sensitive in terms of its hydrological response.

682

683 Figure 14 - Global annual mean surface precipitation anomalies (expressed as a percentage) versus global annual mean SAT

684 from HadGEM3 *mPWP* simulation, HadGEM2 and all other models in PlioMIP2

685

686 Returning to SAT and if only extratropical warming (separated by hemisphere, above or below 45°N

687 or S) is considered, then HadGEM3 agrees with the other 11 models (out of 16) that H20 identified as

688 showing enhanced Northern Hemisphere warming, relative to the Southern Hemisphere (Fig. 12b, top

689 panel). In the Northern Hemisphere, HadGEM3 is again one of the warmest models and, at 8.46°C, is

690 considerably warmer than most other models and the MME; this, with the inclusion of HadGEM3, has

691 now increased from the 5.5°C reported in H20 to 5.7°C here. However, in the Southern Hemisphere

692 HadGEM3 is closer to many of the other models, albeit still in the top 33% of them, and with a

693 warming of 6.3°C is much closer to the MME of 5.1°C (Fig. 12b, top panel). This is further

694 demonstrated by Fig. 12b (bottom panel), showing the ratio of warming between the hemispheres

695 (calculated by dividing the Northern Hemisphere warming by the Southern Hemisphere warming),

696 where HadGEM3 is giving a ratio of 1.34 which is again close to many of the other models and the

697 MME (1.17). Considering land-sea temperature contrasts (Fig. 12c), as H20 state all of the PlioMIP2

698 models show more warming over land, both globally and across the tropics (defined as 20°N-20°S),

699 and HadGEM3 is no exception. Indeed, over either land or sea, HadGEM3 is the 2nd warmest

700 globally and warmest across the tropics, and the inclusion of this model increases the MME by 0.1-

701 0.14°C depending on whether land or sea warming is considered.

702

703 HadGEM3 is one of the largest outliers regardless of metric, however concerning polar amplification

704 this is not the case. Here, as in H20, polar amplification is defined as the ratio of SAT increases

705 poleward of 60° divided by the global mean SAT increases (Smith *et al.* 2019), calculated

706 independently for each hemisphere. Despite the HadGEM3 *mPWP* simulation qualitatively showing

707   considerable amplification at both annual and seasonal timescales (Figure 7), when quantitatively

708   compared with all other PlioMIP2 models HadGEM3 is, whilst still having amplification >1 (i.e. that

709   there is some amplification of warming around the poles), nevertheless showing considerably less

710   amplification in both hemispheres, and is also lower than the MME in both hemispheres (Fig. 12d).

711   Of all the models, HadGEM3 comes 4th-to-last for Northern Hemisphere amplification and last for

712   Southern Hemisphere amplification, and its inclusion with the other models reduces the MME ratio by

713   approximately 0.01 and 0.04 for the Northern and Southern Hemisphere, respectively. This is

714   consistent with the conclusions of H20, who note a weak relationship between ECS and amplification;

715   they observe that models with a lower ECS tend to display higher PA, whereas the opposite appears to

716   be shown here i.e. HadGEM3, with one of the highest ECS, is displaying one of the lowest amounts

717   of amplification. The amplification for all the models, as well as the MME, can be seen graphically in

718   Figure S11 in the Supplementary Material where, at first glance, HadGEM3 would appear to be

719   showing one of the largest amounts of amplification. However, and consistent with the observation

720   by H20, this is because the model is showing more warming in the tropics (relative to the other

721   models) rather than less warming at high latitudes.

722

723   Lastly, concerning SST anomalies the HadGEM3 *mPWP* simulation is warmer than most other

724   models in PlioMIP2 (Figure 15). When simulated SST is compared to the proxy data from

725   McClymont *et al*. (2020), if the models are ranked according to RMSE then the HadGEM3 *mPWP*

726   simulation (RMSE = 3.4°C; see Table 3) ranks approximately halfway amongst them. There appears

727   to be a weak relationship between the warmth of the model and agreement with proxy data, with some

728   of the other warm models (e.g. CESM2, the warmest model) showing less agreement (RMSE =

729   3.5°C) with the proxy data than HadGEM3; however, this is not always true, such as the case of the

730   CCSM4-Utr which is also comparatively warm but is showing a slightly better agreement (RMSE =

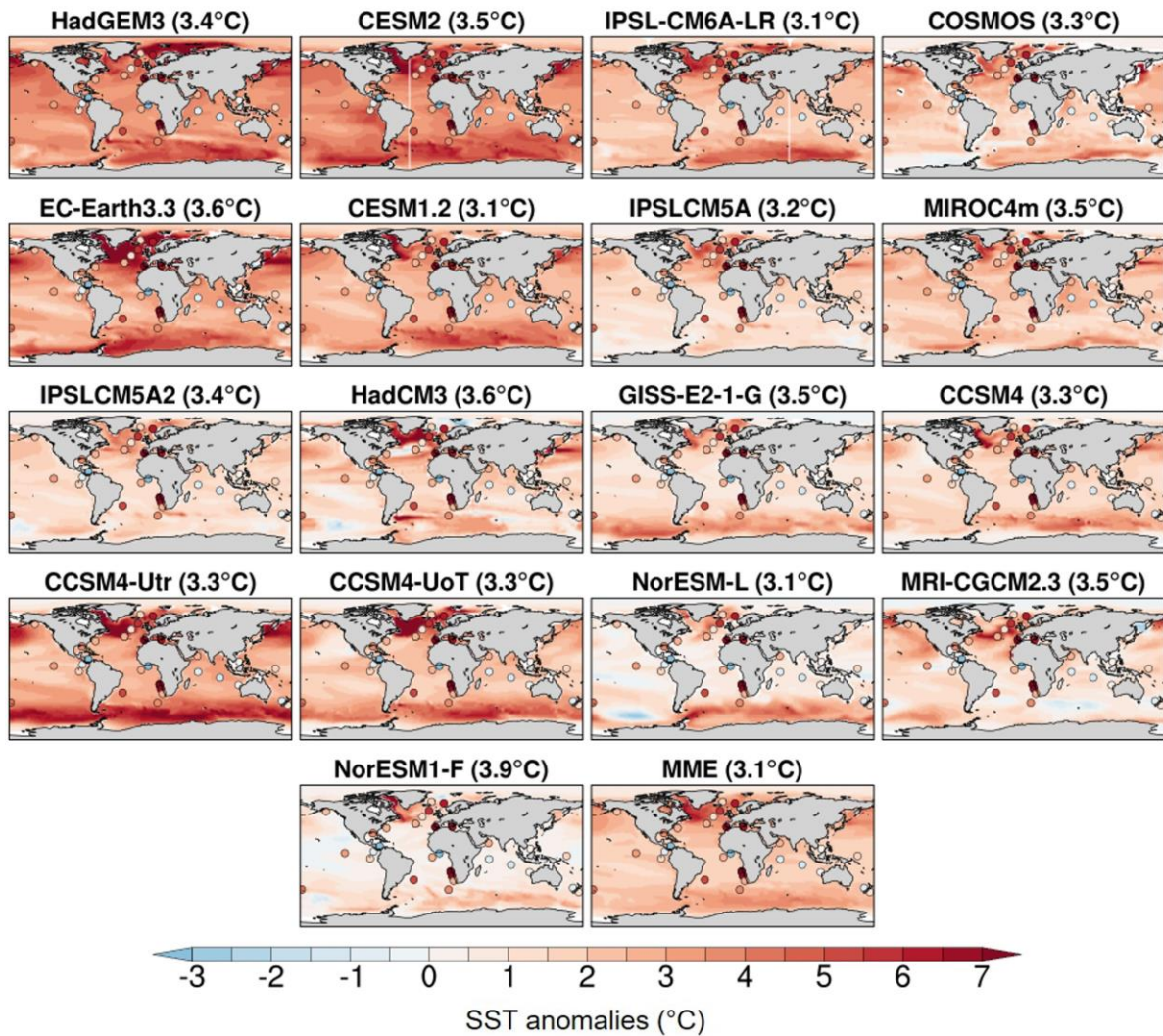
731   3.3°C) with the proxy data. It is likely that the location of the proxy data is important, as the best

732   agreement comes from the MME (RMSE = 3.1°C) which is showing warm SST anomalies over the

733   North Atlantic and Arctic (better in agreement with the proxy data there) but less warming relative to

734   HadGEM3 and CESM2 in the Southern Hemisphere (better in agreement with the proxy data in e.g.

735   the Indian Ocean).

736

737

738   Figure 15 – SST climatology differences (Pliocene – PI) from HadGEM3 *mPWP* simulation and all other models in

739   PlioMIP2, as well as multi-model ensemble mean (MME). Numbers in brackets show RMSE scores when compared proxy

740   data from McClymont *et al.* (2020)

741

742   It is likely that much of the greater warming in the HadGEM3 *mPWP* simulation, relative to the other

743   models, can be attributed to the relatively high ECS of this model. Figure 16 shows model ECS

744   against simulated Pliocene warming for all available models (see Table 4 for individual ECS values).

745   Also shown on this figure is the Earth System Sensitivity (ESS) which, for the Pliocene, can be taken

746   as the global mean temperature scaled by the CO_2 forcing for 560 ppmv compared with 400 ppmv.

747   This is because the temperature change due to the modified orography is small, and so the Pliocene

748   warming relative to preindustrial is due to the CO_2 forcing and associated feedbacks due to vegetation

749   and ice sheets, which can be interpreted as ESS (Lunt *et al.* 2010). Therefore, a plot of Pliocene

750   global mean warming against ECS will be identical to a plot of ESS against ECS, but with different

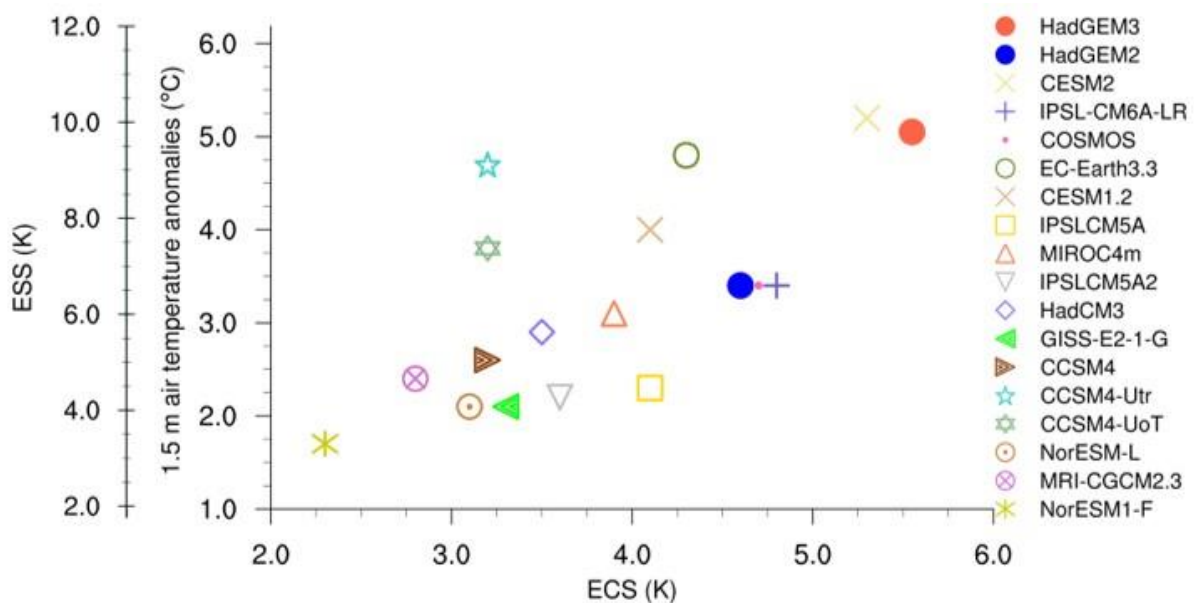
751   values on the y axis. There is a clear linear relationship between ECS and global mean warming (or

752   ESS), with the two models showing the highest ECS also having the highest Pliocene warming or ESS

753   (HadGEM3 and CESM2). Despite some outliers, such as CCSM4-Utr with a relatively high global

754 mean temperature anomaly but a relatively low ECS, this would suggest that for most models

755 Pliocene temperature anomalies (and ESS) are increasing in line with ECS.

756

757

758 Figure 16 - Global annual mean SAT anomalies versus both ESS (first y-axis) and ECS from HadGEM3 *mPWP* simulation,
759 HadGEM2 and all other models in PlioMIP2. The ESS axis is calculated by multiplying the global annual mean SAT
760 anomaly by log(560/280)/log(400/280) i.e. by 1.94, meaning the axis here goes from 1.94-11.64 K; for simplicity, this has
761 been rounded up to 2-12 K

762

763 **5. SUMMARY AND CONCLUSIONS**

764 This study has introduced the mid-Pliocene simulation using the latest version of the UK's physical

765 climate model, HadGEM3-GC31-LL, presented the experimental design and conducted a model-

766 model and model-data comparison. This study is novel, being the first time this version of the UK

767 model has been run this far back in time; only two other paleoclimate simulations using this model

768 have thus far been conducted, comprising the UK's contribution to CMIP6/PMIP4, and both of these

769 were more recent, Quaternary simulations (Williams *et al*. 2020).

770

771 The *mPWP* simulation mostly followed the experimental design defined in H16, with the exception

772 being the exclusion of a Pliocene LSM and Pliocene soils. Both of these were kept the same as PI.

773 All other boundary conditions, including CO_2, orography, ice mask, lakes, vegetation fractions and

774 vegetation functional types followed the protocol of H16, and were incrementally implemented to be

775 Pliocene, based on the PRISM4 dataset. A minor model parameter change was included to increase

776 the model's stability in light of the strong Pliocene forcing, and thus a corresponding PI simulation

777 was also run for comparison purposes   The *mPWP* simulation was run for 567 years in total, during

778 which atmospheric and oceanic equilibrium were assessed. Although not meeting the criteria used to

779 determine equilibrium in other paleoclimate simulations, especially concerning oceanic equilibrium,

780 due to computational restrictions it was not possible to run this model for the thousands of years

781 required to achieve this.

782

783 The results presented here are divided into three sections: i) a simulation comparison, in which the

784 *mPWP* simulation is compared to its corresponding *piControl_mod* simulation (Section 3.2); ii) a

785 model-model and model-data comparison, in which the most recent *mPWP* simulation is compared to

786 Pliocene simulations from previous versions of the same model, all assessed against proxy data

787 (Section 4.1); and iii) a model-model comparison, in which the most recent *mPWP* simulation is

788 compared to other models (Section 4.2).

789

790 For the first comparison, the *mPWP* simulation is behaving in line with current understanding and

791 previous work (e.g. Haywood *et al.* 2013, H20), showing a warmer and wetter world relative to the PI,

792 with the greatest warming occurring over the poles. This polar warming, which can be attributed to a

793 loss in sea ice and changes in clouds, and the changes to precipitation (such as an enhancement of

794 monsoon systems) all agree with the expected response and previous work (e.g. Cronin *et al.* 1993,

795 Howell *et al.* 2016, Li *et al.* 2018, Moran *et al.* 2006, Polyak *et al.* 2010, Zhang *et al.* 2013, 2016).

796 For the second comparison, there is a clear increase in global temperatures (as measured by SST) as

797 the model develops through time, beginning with the early Pliocene simulations using HadCM3 (Lunt

798 *et al.* 2011 and Bragg *et al.* 2012), through HadGEM2 (Tindall and Haywood 2020) and up to the

799 most recent *mPWP* simulation from HadGEM3, presented here. Up to a point, this warming results in

800 a better agreement with available proxy data. However, just as the earlier HadCM3 simulations

801 appear to be too cold relative to some proxy data, the most recent *mPWP* simulation from HadGEM3

802 appears to be too warm; the "sweet spot" appears to be the previous generation of the model,

803 HadGEM2. This would be even more the case had the *mPWP* simulation been allowed to run to full

804 equilibrium, and it is suggested that the final global mean surface temperature could have been

805 approximately 1.5°C higher if so. For the third comparison, the above conclusion that HadGEM3 is

806 too warm is further suggested by the fact that it is one of the warmest and wettest models (even at its

807 current state of equilibrium) in all of PlioMIP2 (H20), and this is true over either land or sea and

808 especially in the Northern Hemisphere. When compared to proxy SST data, HadGEM2 ranks

809 approximately halfway amongst the models, and is much too warm in certain locations, such as the

810 Indian Ocean. However, the conclusion that the model is too warm overall is argued by the fact that

811 the anomalies coming from the HadGEM3 *piControl_mod* simulation are not the warmest, suggesting

812 that rather than the model being too warm in general, the warming may be coming from the model's

813 sensitivity to the Pliocene forcing. This is consistent with the model's high ECS, which is among the

814 highest of all the most recent state-of-the-art CMIP6 models (Andrews *et al.* 2019, H20, Zelinka *et al.*

815 2020).

816

817   A number of caveats should be mentioned in this study. The question over the relatively short (but

818   unavoidable due to computational cost) run length has already been discussed, with the results

819   suggesting that the *mPWP* simulation would have been even warmer if it had been allowed to run

820   until true equilibrium. Besides this, firstly any differences to the PlioMIP2 models may be in part

821   related to the fact that the LSM used here is identical to the *piControl*, rather than using the enhanced

822   LSM following the experimental design of H16. This, as discussed above, was necessary, due to

823   technical difficulties in coupling a new LSM to the atmosphere. One of the impacts of this is

824   discussed in Zhang *et al*. (2021), who investigated Atlantic Meridional Overturning Circulation

825   (AMOC) changes during the Pliocene using the PlioMIP2 models. It was found that in contrast to

826   most other PlioMIP2 models, which stimulate a stronger AMOC in the Pliocene relative to the PI,

827   HadGEM3 shows a weaker AMOC, with a maximum of 14.3 Sv and 16.1 Sv for the *mPWP* and

828   *piControl_mod* simulations, respectively (Zhang *et al*. 2021). Secondly, using PI soil parameters and

829   soil dust properties (away from ice regions) may also have an impact on the observed warming;

830   although H16 does provide a set of palaeosol data from Pound *et al*. (2016), this was not used here

831   because of the difficulties in matching the reconstructions to the model's soil-related fields. Thirdly,

832   concerning greenhouse gas forcings, in all of the Pliocene simulations discussed here only CO_2 was

833   modified, with other gases such as methane being left as PI. Given that these trace gases will likely

834   amplify warming, especially in the extratopics (Hopcroft *et al*. 2020), leaving these as PI may be

835   resulting in a cooler climate in all of the simulations. Lastly, the large warming in the *mPWP*

836   simulation may be because certain processes, in particular vegetation, were fixed rather than being

837   interactive (although this is also the case in the majority of the other PlioMIP2 models). In particular,

838   the fact that the introduction of Pliocene vegetation in the *mPWP* simulation results in such a dramatic

839   rise in global SAT (Figure 6) deserves much further exploration. This may be highly important

840   regarding any possible impact on the climate under a Pliocene-style forcing, and therefore current

841   work is underway to investigate the role of vegetation in contributing to the model's simulated

842   warming.

843

844   **DATA AVAILABILITY**

845   Selected fields (SAT, precipitation and SST) from the HadGEM3 *mPWP* simulation are currently

846   available from the Earth System Grid Federation (ESGF) WCRP Coupled Model Intercomparison

847   Project (Phase 6), located at https://esgf-node.llnl.gov/projects/cmip6/ (last access: 18 March 2021).

848   If other fields are required, they can be made available to the public by directly contacting the lead

849   author. Likewise, access to the other model simulations considered here can be gained by contacting

850   the lead author, or the authors of the appropriate publication (see Haywood *et al*. 2020 for a list of the

851   appropriate publications). For the SST reconstructions, the data can be found within the

852   Supplementary Online Material of

853   McClymont *et al*. (2020), available online at: https://doi.org/10.5194/cp-16-1599-2020-supplement.

854

AUTHOR CONTRIBUTIONS

855

856 CJRW conducted the *mPWP* simulation, carried out the analysis, produced some of the figures, wrote

857 the majority of the manuscript, and led the paper. XR produced some of the figures. AAS, WHGR,

858 RSS, PH and EJS provided technical assistance in running HadGEM3. JCT, SJH and AMH also

859 provided technical assistance, and contributed the HadGEM2 and HadCM3 simulations. DJL

860 contributed to some of the writing. All authors proofread the paper and provided comments.

861

862 **COMPETING INTERESTS**

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

864

865 **ACKNOWLEDGEMENTS**

866 CJRW and DJL acknowledge the financial support of the UK Natural Environment Research Council

867 (NERC)-funded SWEET project, research grant NE/P01903X/1, as well as funding from the

868 European Research Council under the European Union's Seventh Framework Programme (FP/2007-

869 2013)/ERC Grant Agreement no. 340923 (T-GRES). AAS was supported by the Met Office Hadley

870 Centre Climate Programme, funded by BEIS and Defra. XR was supported by the 4D-REEF project,

871 receiving funding from the European Union's Horizon 2020 research and innovation programme

872 under the Marie Skłodowska-Curie research grant, no. 813360. PH was supported by a University of

873 Birmingham Fellowship. RSS was funded by the NERC national capability grant for the UK Earth

874 System Modelling (UKESM) project, research grant NE/N017951/1. AMH and JCT acknowledge

875 receipt of funding from the European Research Council under the European Union's Seventh

876 Framework Programme (FP7/2007-2013)/ERC, grant agreement no. 278636.

877

878 **FINANCIAL SUPPORT**

879 This research has been supported by the NERC-funded SWEET project (grant no. NE/P01903X/1),

880 the Met Office Hadley Centre Climate Programme (funded by BEIS and Defra), the European Union's

881 Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie research grant

882 (no. 813360), a University of Birmingham Fellowship, the NERC UKESM project (grant no.

883 NE/N017951/1) and the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC

884 (grant no. 278636).

885

886 **LIST OF TABLES**

LIST OF FIGURES

Figure 1 - Changes to orography in HadGEM3 *mPWP* simulation: a) PRISM4 anomaly; b) Original field used in HadGEM3 *piControl*; c) New field used in HadGEM3 *mPWP*, with smoothed orography over western Antarctica (final version, used in simulation)

Figure 2 - Ten megabiomes from PlioMIP Phase 2 used create the nine PFTs used in HadGEM3 *mPWP* simulation

Figure 3 - Nine PFTs used in HadGEM3. Top half: *piControl* simulation, bottom half: *mPWP* simulation. Values in brackets show global mean differences (*mPWP - piControl*), expressed as a percentage. a) broadleaf trees (18%); b) needle-leaved trees (5%); c) temperate C3 grass (-15%); d) tropical C4 grass (6%); e) shrubs (3%); f) urban areas (no change); g) inland water (1%); h) bare soil (-12%); i) land ice (-5%)

Figure 4 - LAI used in HadGEM3, for an example PFT (broadleaf trees, January). a) Function used to create LAI, where dashed lines show zonal mean from *piControl* simulation and solid lines show seasonally and latitudinally varying function of this zonal mean; b) example of functional types used in *piControl* simulation; c) same as b) but for the *mPWP* simulation

Figure 5 – Example of soil-related fields used in HadGEM3. Left-hand column: *piControl* simulation, right-hand column: *mPWP* simulation. First row: Soil parameters (example shows Volumetric soil moisture content at wilting point); Second row: Soil moisture (example shows January, top-level); Third row: Soil temperature (example shows January, top-level). Complete list of fields shown in Supplementary Material Fig. S3 and S4

927   Figure 6 – Annual global mean 1.5 m air temperature from the HadGEM3 *mPWP* spin-up phase and

928   production run, as well as the CMIP6 *piControl* and the *piControl_mod*. Labels show introduction of

929   each new Pliocene element. Climatologies discussed here are taken from final 50 years of each

930   simulation (shown by shaded boxes). See Williams *et al*. (2020) for the *piControl* spin-up phase that

931   preceded these simulations.

932

933   Figure 7 – 1.5 m air temperature climatology differences (*mPWP – piControl_mod*) from HadGEM3.

934   a) Annual; b) DJF; c) JJA

935

936   Figure 8 – Sea ice fraction climatology differences (*mPWP – piControl_mod*) from HadGEM3: a)

937   Northern Hemisphere DJF, b) Northern Hemisphere JJA, c) Southern Hemisphere DJF, d) Southern

938   Hemisphere JJA, e) Mean sea ice area (both absolute values and differences) averaged over either

939   hemisphere

940

941   Figure 9 – Surface precipitation climatology differences (*mPWP – piControl_mod*) from HadGEM3.

942   a) Annual; b) DJF; c) JJA

943

944   Figure 10 – Total cloud fraction climatology differences (*mPWP – piControl_mod*) from HadGEM3.

945   a) Annual; b) DJF; c) JJA

946

947   Figure 11 – Annual mean SST differences (Pliocene – PI) from different generations of the UK's

948   physical climate model. a) HadCM3-PRISM2; b) HadCM3-PlioMIP2; c) HadCM3-PlioMIP2; d)

949   HadGEM2; e) HadGEM3. Background gridded data shows model simulations, filled circles show

950   SST proxy data from McClymont *et al*. (2020)

951

952   Figure 12 - SAT from Pliocene simulations from HadGEM3 and all other models in PlioMIP2. a)

953   Global annual mean SAT (top panel) and anomalies (bottom panel); b) Extratropical (+/- 45°) annual

954   mean SAT anomalies (top panel) and ratio (i.e. >45°N divided by <45°S) between them (bottom

955   panel); c) Land and ocean annual mean SAT anomalies, averaged globally (top panel) and between

956   20°N-20°S (bottom panel); d) Annual mean SAT polar amplification i.e. SAT poleward of 60°

957   divided by global mean, for each hemisphere, where red line = ratio of 1 (i.e. no polar amplification).

958   Figures reproduced and adapted from Haywood *et al*. (2020)

959

960   Figure 13 - Global annual mean surface precipitation (top panel) and anomalies (bottom panel) from

961   HadGEM3 *mPWP* simulation and all other models in PlioMIP2, as well as multi-model ensemble

962   mean (MME). Figure reproduced and adapted from Haywood *et al*. (2020)

963

964   Figure 14 - Global annual mean surface precipitation anomalies (expressed as a percentage) versus

965   global annual mean SAT from HadGEM3 *mPWP* simulation, HadGEM2 and all other models in

966   PlioMIP2

967

968   Figure 15 – SST climatology differences (Pliocene – PI) from HadGEM3 *mPWP* simulation and all

969   other models in PlioMIP2, as well as multi-model ensemble mean (MME). Numbers in brackets

970   show RMSE scores when compared proxy data from McClymont *et al*. (2020)

971

972   Figure 16 - Global annual mean SAT anomalies versus both ESS (first y-axis) and ECS from

973   HadGEM3 *mPWP* simulation, HadGEM2 and all other models in PlioMIP2. The ESS axis is

974   calculated by multiplying the global annual mean SAT anomaly by log(560/280)/log(400/280) i.e. by

975   1.94, meaning the axis here goes from 1.94-11.64 K; for simplicity, this has been rounded up to 2-12

976   K

977

978

979

980

REFERENCES

Andrews, T., Andrews, M. B., Bodas-Salcedo, A., Jones, G. S., Kuhlbrodt, T., Manners, J., Menary, M. B., Ridley, J., Ringer, M. A., Sellar, A. A., Senior, C. A. and Tang, Y.: Forcings, feedbacks, and climate sensitivity in HadGEM3-GC3.1 and UKESM1, JAMES, https://doi.org/10.1029/2019MS001866, 2019.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B. And Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes, Geosci. Model Dev., 4, 677–699, https://doi.org/10.5194/gmd-4-677-2011, 2011.

Berntell, E., Zhang, Q., Li, Q., Haywood, A. M., Tindall, J. C., Hunter, S. J., Zhang, Z., Li, X., Guo, C., Nisancioglu, K. H., Stepanek, C., Lohmann, G., Sohl, L. E., Chandler, M. A., Tan, N., Contoux, C., Ramstein, G., Baatsen, M. L. J., von der Heydt, A. S., Chandan, D., Peltier, W. R., Abe-Ouchi, A., Chan, W.-L., Kamae, Y., Williams, C. J. R. and Lunt, D. J.: Mid-Pliocene West African Monsoon Rainfall as simulated in the PlioMIP2 ensemble, Clim. Past Discuss. [preprint], https://doi.org/10.5194/cp-2021-16, in review, 2021.

Bragg, F. J., Lunt, D. J. and Haywood, A. M.: Mid-Pliocene climate modelled using the UK Hadley Centre Model: PlioMIP Experiments 1 and 2, Geosci. Model Dev., 5, 1109-1125, doi:10.5194/gmd-5-1109-2012, 2012.

Burke, K. D., Williams, J. W., Chandler, M. A., Haywood, A. M., Lunt, D. J. and Otto-Bliesner, B. L.: Pliocene and Eocene provide best analogs for near-future climates, PNAS, 115, 13288-13293, 10.1073/pnas.1809600115, 2018

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Cox, P. M., and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation, Geosci. Model Dev., 4, 701–722, https://doi.org/10.5194/gmd-4-701-2011, 2011.

Cronin, T. M., Whatley, R. C.,Wood, A., Tsukagoshi, A., Ikeya, N., Brouwers, E. M., and Briggs, W. M.: Microfaunal evidence for elevated mid-Pliocene temperatures in the Arctic Ocean, Paleoceanography, 8, 161-173, https://doi.org/10.1029/93PA00060, 1993.

1017 Crucifix, M., Betts R. A. and Hewitt, C. D.: Pre-industrial-potential and Last Glacial Maximum global

1018 vegetation simulated with a coupled climate-biosphere model: Diagnosis of bioclimatic relationships,

1019 Glob. Planet. Chang, 45 (4), 295-312, DOI: 10.1016/j.gloplacha.2004.10.001, 2005.

1020

1021 Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J.,

1022 Jones, C. D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S.,

1023 Totterdell, I., Wiltshire, A., Woodward, S.: Development and evaluation of an Earth-System model-

1024 HadGEM2. Geosci. Model Dev. 4, 1051–1075. https://doi.org/10.5194/gmd-4-1051-2011, 2011.

1025

1026 Cox, P. M.: Description of the TRIFFID Dynamic Global Vegetation Model - Hadley Centre

1027 Technical Note 24, Met Office, Bracknell, https://jules.jchmr.org/sites/default/files/HCTN_24.pdf,

1028 2001.

1029

1030 Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, PR. and Smith, J.: The impact of

1031 new land surface physics on the GCM simulation of climate and climate sensitivity, Clim. Dyn., 15,

1032 183-203, https://doi.org/10.1007/s003820050276, 1999.

1033

1034 Delaney, M. L., Be, A. W. H., and Boyle, E. A.: Li, Sr, Mg, and Na in foraminiferal calcite shells

1035 from laboratory culture, sediment traps, and sediment cores, Geochim. Cosmochim. Ac., 49, 1327-

1036 1341, 1985.

1037

1038 Dowsett, H. J.: The PRISM palaeoclimate reconstruction and Pliocene sea-surface temperature, in:

1039 Deep-Time Perspectives on Climate Change: Marrying the Signal from Computer Models and

1040 Biological Proxies, edited by: Williams, M., Haywood, A. M., Gregory, F. J., and Schmidt, D. N.,

1041 Bath, UK, Geological Soc Publishing House, 459-480, 2007.

1042

1043 Dowsett, H. J., Robinson, M., Haywood, A., Salzmann, U., Hill, D., Sohl, L., Chandler, M., Williams,

1044 M., Foley, K., and Stoll, D.: The PRISM3D paleoenvironmental reconstruction, Stratigraphy, 7, 123-

1045 139, 2010.

1046

1047 Dowsett, H., Dolan, A., Rowley, D., Moucha, R., Forte, A. M., Mitrovica, J. X., Pound, M.,

1048 Salzmann, U., Robinson, M., Chandler, M., Foley, K., and Haywood, A.: The PRISM4 (mid-

1049 Piacenzian) paleoenvironmental reconstruction, Clim. Past, 12, 1519-1538,

1050 https://doi.org/10.5194/cp-12-1519-2016, 2016.

1051

1052   Essery, R. L. H., Best, M. J., Betts, R. A., Cox, P. M., and Taylor, C. M.: Explicit representation of

1053   subgrid heterogeneity in a GCM land surface scheme, J. Hydrometeor., 4, 530–543,

1054   https://doi.org/10.1175/1525-7541(2003)004<0530:EROSHI>2.0.CO;2, 2003.

1055

1056   Essery, R. L. H., Best, M. J. and Cox, P. M.: MOSES 2.2 Technical Documentation - Hadley Centre

1057   Technical Note 30, Met Office, Bracknell. https://jules.jchmr.org/sites/default/files/HCTN_30.pdf,

1058   2001.

1059

1060   Foley, K. M. and Dowsett, H. J.: Community sourced mid-Piacenzian sea surface temperature (SST)

1061   data, US Geological Survey data release, https://doi.org/10.5066/P9YP3DTV, 2019.

1062

1063   Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., Lowe, J. A.,

1064   Johns, T. C. and Williams, K. D.: A new method for diagnosing radiative forcing and climate

1065   sensitivity, Geophys. Res. Lett., 31, L03205, DOI:10.1029/2003GL018747, 2004.

1066

1067   Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., Mitchell, J. F. B., and

1068   Wood, R. A.: The simulation of SST, sea ice extents and ocean heat transports in a version of the

1069   Hadley Centre coupled model without flux adjustments, Clim. Dynam., 16, 147-168, 2000.

1070

1071   Hardiman, S. C., Andrews, M. B., Andrews, T., Bushell, A. C., Dunstone, N. J., Dyson, H., Jones, G.

1072   S., Knight, J. R., Neininger, E., O'Connor, F. M., Ridley, J. K., Ringer, M. A., Scaife, A. A., Senior,

1073   C. A. and Wood, R. A.: The impact of prescribed ozone in climate projections run with HadGEM3-

1074   GC3.1, JAMES, 11, https://doi.org/10.1029/2019MS001714, 2019.

1075

1076   Harrison, S. P. and Prentice, I. C.: Climate and CO_2 controls on global vegetation distribution at the

1077   last glacial maximum: analysis based on palaeovegetation data, biome modelling and palaeoclimate

1078   simulations, Glob. Change Bio., 9 (7), 983-1004, https://doi.org/10.1046/j.1365-2486.2003.00640.x,

1079   2003.

1080

1081   Haywood, A. M., Dowsett, H. J., Dolan, A. M., Rowley, D., Abe-Ouchi, A., Otto-Bliesner, B.,

1082   Chandler, M. A., Hunter, S. J., Lunt, D. J., Pound, M., and Salzmann, U.: The Pliocene Model

1083   Intercomparison Project (PlioMIP) Phase 2: scientific objectives and experimental design, Clim. Past,

1084   12, 663–675, https://doi.org/10.5194/cp-12-663-2016, 2016

1085

1086   Haywood, A. M., Hill, D. J., Dolan, A. M., Otto-Bliesner, B. L., Bragg, F., Chan, W.-L., Chandler, M.

1087   A., Contoux, C., Dowsett, H. J., Jost, A., Kamae, Y., Lohmann, G., Lunt, D. J., Abe-Ouchi, A.,

1088   Pickering, S. J., Ramstein, G., Rosenbloom, N. A., Salzmann, U., Sohl, L., Stepanek, C., Ueda, H.,

1089 Yan, Q., and Zhang, Z.: Large-scale features of Pliocene climate: results from the Pliocene Model

1090 Intercomparison Project, Clim. Past, 9, 191-209, https://doi.org/10.5194/cp-9-191-2013, 2013

1091

1092 Haywood, A. M., Tindall, J. C., Dowsett, H. J., Dolan, A. M., Foley, K. M., Hunter, S. J., Hill, D. J.,

1093 Chan, W.-L., Abe-Ouchi, A., Stepanek, C., Lohmann, G., Chandan, D., Peltier, W. R., Tan, N.,

1094 Contoux, C., Ramstein, G., Li, X., Zhang, Z., Guo, C., Nisancioglu, K. H., Zhang, Q., Li, Q., Kamae,

1095 Y., Chandler, M. A., Sohl, L. E., Otto-Bliesner, B. L., Feng, R., Brady, E. C., von der Heydt, A. S.,

1096 Baatsen, M. L. J. and Lunt, D. J.: The Pliocene Model Intercomparison Project Phase 2: large-scale

1097 climate features and climate sensitivity, Clim. Past, 16, 2095-2123, https://doi.org/10.5194/cp-16-

1098 2095-2020, 2020.

1099

1100 Haywood, A. M. and Valdes, P. J.: Modelling Middle Pliocene warmth: contribution of atmosphere,

1101 oceans and cryosphere, Earth Planet. Sc. Lett., 218, 363–377, doi:10.1016/S0012-821X(03)00685-X,

1102 2004.

1103

1104 Hunter, S. J., Haywood, A. M., Dolan, A. M. and Tindall, J. C.: The HadCM3 contribution to PlioMIP

1105 phase 2, Clim. Past, 15, 1691-1713, https://doi.org/10.5194/cp-15-1691-2019, 2019.

1106

1107 Hopcroft, P. O., Ramstein, G., Pugh, T.A.M., Hunter, S. J., Murguia-Flores, F., Quiquet, A., Sun, Y.,

1108 Tan, N. and Valdes, P. J.: Polar amplification of Pliocene climate by elevated trace gas radiative

1109 forcing, Proceedings of the National Academy of Sciences, DOI: 10.1073/pnas.2002320117, 2020

1110

1111 Howell, F. W., Haywood, A. M., Otto-Bliesner, B. L., Bragg, F., Chan, W.-L., Chandler, M. A.,

1112 Contoux, C., Kamae, Y., Abe-Ouchi, A., Rosenbloom, N. A., Stepanek, C., and Zhang, Z.: Arctic sea

1113 ice simulation in the PlioMIP ensemble, Clim. Past, 12, 749-767, https://doi.org/10.5194/cp-12-749-

1114 2016, 2016.

1115

1116 Knutti, R. and Rugenstein, M. A. A.: Feedbacks, climate sensitivity and the limits of linear models,

1117 Phil. Trans. R. Soc. A, 373, 20150146, http://dx.doi.org/10.1098/rsta.2015.0146, 2015.

1118

1119 Kuhlbrodt, T., Jones, C. G., Sellar, A. et al.: The low resolution version of HadGEM3 GC3.1:

1120 Development and evaluation for global climate, JAMES, 10: 2865-2888,

1121 https://doi.org/10.1029/2018MS001370, 2018.

1122

1123 Li, X. Y., Jiang, D. B., Tian, Z. P., and Yang, Y. B.: Mid-Pliocene

1124 global land monsoon from PlioMIP1 simulations, Palaeogeogr. Palaeocl., 512, 56-70,

1125 https://doi.org/10.1016/j.palaeo.2018.06.027, 2018.

1126

1127 Lunt, D. J., Bragg, F., Chan, W.-L., Hutchinson, D. K., Ladant, J.-B., Morozova, P., Niezgodzki, I.,

1128 Steinig, S., Zhang, Z., Zhu, J., Abe-Ouchi, A., Anagnostou, E., de Boer, A. M., Coxall, H. K.,

1129 Donnadieu, Y., Foster, G., Inglis, G. N., Knorr, G., Langebroek, P. M., Lear, C. H., Lohmann, G.,

1130 Poulsen, C. J., Sepulchre, P., Tierney, J. E., Valdes, P. J., Volodin, E. M., Dunkley Jones, T., Hollis,

1131 C. J., Huber, M. and Otto-Bliesner, B. L.: DeepMIP: model intercomparison of early Eocene climatic

1132 optimum (EECO) large-scale climate features and comparison with proxy data, Clim. Past, 17, 203-

1133 227, https://doi.org/10.5194/cp-17-203-2021, 2021.

1134

1135 Lunt, D. J., Haywood, A. M., Schmidt, G. A., Salzmann, U., Valdes, P. J. and Dowsett, H. J.: Earth

1136 system sensitivity inferred from Pliocene modelling and data, Nature Geoscience, 3, 60-64,

1137 DOI:10.1038/ngeo706, 2010.

1138

1139 Lunt, D. J., Haywood, A. M., Schmidt, G. A., Salzmann, U., Valdes, P. J., Dowsett, H. J. and

1140 Loptson, C. A.: On the causes of mid-Pliocene warmth and polar amplification, EPSL, 321-322, 128-

1141 138, doi:10.1016/j.epsl.2011.12.042, 2012

1142

1143 Lunt, D. J., Huber, M., Anagnostou, E., Baatsen, M. L. J., et al.: The DeepMIP contribution to

1144 PMIP4: experimental design for model simulations of the EECO, PETM, and pre-PETM (version

1145 1.0), Geosci. Model Dev., 10, 889-901, doi:10.5194/gmd-10-889-2017, 2017.

1146

1147 Martin, G. M. et al.: The HadGEM2 family of Met Office Unified Model climate configurations.

1148 Geosci. Model Dev. 4, 723-757. https://doi.org/10.5194/gmd-4-723-2011, 2011.

1149

1150 McClymont, E. L., Ford, H. L., Ho, S. L., Tindall, J. C., Haywood, A. M., Alonso-Garcia, M., Bailey,

1151 I., Berke, M. A., Littler, K., Patterson, M. O., Petrick, B., Peterse, F., Ravelo, A. C., Risebrobakken,

1152 B., De Schepper, S., Swann, G. E. A., Thirumalai, K., Tierney, J. E., van der Weijst, C., White, S.,

1153 Abe-Ouchi, A., Baatsen, M. L. J., Brady, E. C., Chan, W.-L., Chandan, D., Feng, R., Guo, C., von der

1154 Heydt, A. S., Hunter, S., Li, X., Lohmann, G., Nisancioglu, K. H., Otto-Bliesner, B. L., Peltier, W. R.,

1155 Stepanek, C. and Zhang, Z.: Lessons from a high-CO_2 world: an ocean view from $\sim$ 3 million years

1156 ago, Clim. Past, 16, 1599–1615, https://doi.org/10.5194/cp-16-1599-2020, 2020.

1157

1158 Menary, M. B., Kuhlbrodt, T., Ridley, J. et al.: Pre-industrial control simulations with HadGEM3-

1159 GC3.1 for CMIP6, JAMES, 10: 3049–3075, https://doi.org/10.1029/2018MS001495, 2018.

1160

1161 Moran, K., Backman, J., Brinkhuis, H., et al.: The Cenozoic palaeoenvironment of the Arctic Ocean,

1162 Nature, 441, 601-605, https://doi.org/10.1038/nature04800, 2006.

1163

1164   Pendergrass, A. and Hartmann, D. L.: Changes in the distribution of rain frequency and intensity in

1165   response to global warming, J. Clim., 27, 8372-8383, DOI:10.1175/JCLI-D-14-00183.1, 2014.

1166

1167   Polyak, L., Alley, R. B., Andrews, J. T., Brigham-Grette, J., Cronin, T. M., Darby, D. A., Dyke, A. S.,

1168   Fitzpatrick, J. J., Funder, S., Holland, M., Jennings, A. E., Miller, G. H., O'Regan, M., Savelle, J.,

1169   Serreze, M., St. John, K., White, J. W. C., and Wolff, E.: History of sea-ice in the Arctic, Quaternary

1170   Sci. Rev., 29, 1757-1778, doi:10.1016/j.quascirev.2010.02.010, 2010.

1171

1172   Pound, M. J., Tindall, J., Pickering, S. J., Haywood, A. M., Dowsett, H. J., and Salzmann, U.: Late

1173   Pliocene lakes and soils: a global data set for the analysis of climate feedbacks in a warmer world,

1174   Clim. Past, 10, 167-180, https://doi.org/10.5194/cp-10-167-2014, 2014.

1175

1176   Prahl, F. G. and S. G. Wakeham: Calibration of unsaturation patterns in long-chain ketone

1177   compositions for palaeotemperature assessment, Nature, 320, 367-369, 1987.

1178

1179   Ridley, J., Blockley, E., Keen, A. B. et al.: The sea ice model component of HadGEM3-GC3.1, GMD,

1180   11: 713-723, https://doi.org/10.5194/gmd-11-713-2018, 2017

1181

1182   Siahaan, A., and Walton, J.: The low resolution version of HadGEM3 GC3.1: Development and

1183   evaluation for global climate, JAMES, 10, 2865-2888, https://doi.org/10.1029/2018MS001370, 2018.

1184

1185   Salzmann, U., Dolan, A. M., Haywood, A. M., Chan, W.-L., Voss, J., Hill, D., Abe-Ouchi, A., Otto-

1186   Bliesner, B., Bragg, F. J., Chandler, M. A., Contoux, C., Dowsett, H. J., Jost, A., Kamae, Y.,

1187   Lohmann, G., Lunt, D. J., Pickering, S. J., Pound, M. J., Ramstein, G., Rosenbloom, N. A., Sohl, L.,

1188   Stepamek, C., Ueda, H. and Zhang, Z.: Challenges in quantifying Pliocene terrestrial warming

1189   revealed by data-model discord, Nat. Clim. Chang., 3, 969-974, https://doi.org/10.1038/nclimate2008,

1190   2013.

1191

1192   Salzmann, U., Haywood, A. M., Lunt, D. J., Valdes, P. J. and Hill, D. J.: A new global biome

1193   reconstruction and data-model comparison for the Middle Pliocene, Global Ecol. Biogeogr, 17, 432–

1194   447, https://doi.org/10.1111/j.1466-8238.2008.00381.x, 2008.

1195

1196   Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., et al.: UKESM1:

1197   Description and evaluation of the U.K. Earth System Model, JAMES, 11,

1198   https://doi.org/10.1029/2019MS001739, 2019.

1199

1200   Sellar, A. A., Walton, J., Jones, C. G., Wood, R., et al.: Implementation of U.K. Earth System Models

1201   for CMIP6, JAMES, 12, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019MS001946,

1202   2020.

1203

1204   Smith, D. M., Screen, J. A., Deser, C., Cohen, J., Fyfe, J. C., García-Serrano, J., Jung, T., Kattsov, V.,

1205   Matei, D., Msadek, R., Peings, Y., Sigmond, M., Ukita, J., Yoon, J.-H., and Zhang, X.: The Polar

1206   Amplification Model Intercomparison Project (PAMIP) contribution to CMIP6: investigating the

1207   causes and consequences of polar amplification, GMD, 12, 1139-1164, https://doi.org/10.5194/gmd-

1208   12-1139-2019, 2019.

1209

1210   Storkey, D., Megann, A., Mathiot, P. et al.: UK Global Ocean GO6 and GO7: A traceable hierarchy

1211   of model resolutions, GMD, 11: 3187-3213, https://doi.org/10.5194/gmd-11-3187-2018, 2017

1212

1213   Tierney, J. E., Poulsen, C. J., Montañez, I. P., Bhattacharya, T., Feng, R., Ford, H. L., Hönisch, B.,

1214   Inglis, G. N., Petersen, S. V., Sagoo, N., Tabor, C. R., Thirumalai, K., Zhu, J., Burls, N. J., Foster, G.

1215   L., Goddéris, Y., Huber, B. T., Ivany, L. C., Turner, S. K., Lunt, D. J., McElwain, J. C., Mills, B. J.

1216   W., Otto-Bliesner, B. L., Ridgwell, A. and Zhang, Y. G.: Past climates inform our future, Science,

1217   370, 6517, eaay3701, DOI:10.1126/science.aay3701, 2020.

1218

1219   Tindall, J. C. and Haywood, A. M.: Modelling the mid-Pliocene warm period using HadGEM2, Glob.

1220   Planet. Chang, 186, https://doi.org/10.1016/j.gloplacha.2019.103110, 2020.

1221

1222   Valdes, P. J., Armstrong, E., Badger, M. P. S., Bradshaw, C. D., Bragg, F., Crucifix, M., Davies-

1223   Barnard, T., Day, J. J., Farnsworth, A., Gordon, C., Hopcroft, P. O., Kennedy, A. T., Lord, N. S.,

1224   Lunt, D. J., Marzocchi, A., Parry, L. M., Pope, V., Roberts, W. H. G., Stone, E. J., Tourte, G. J. L.,

1225   and Williams, J. H. T.: The BRIDGE HadCM3 family of climate models: HadCM3@Bristol v1.0,

1226   Geosci. Model Dev., 10, 3715-3743, https://doi.org/10.5194/gmd-10-3715-2017, 2017.

1227

1228   Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., Furtado, K., Hill, P.,

1229   Lock, A., Manners, J., Morcrette, C., Mulcahy, J., Sanchez, C., Smith, C., Stratton, R., Tennant, W.,

1230   Tomassini, L., Van Weverberg, K., Vosper, S., Willett, M., Browse, J., Bushell, A., Carslaw, K.,

1231   Dalvi, M., Essery, R., Gedney, N., Hardiman, S., Johnson, B., Johnson, C., Jones, A., Jones, C.,

1232   Mann, G., Milton, S., Rumbold, H., Sellar, A., Ujiie, M., Whitall, M., Williams, K., and Zerroukat,

1233   M.: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0

1234   configurations, Geosci. Model Dev., 12, 1909-1963, https://doi.org/10.5194/gmd-12-1909-2019,

1235   2019.

1236

Williams, C. J. R., Guarino, M.-V., Capron, E., Malmierca-Vallet, I., Singarayer, J. S., Sime, L. C., Lunt, D. J., and Valdes, P. J.: CMIP6/PMIP4 simulations of the mid-Holocene and Last Interglacial using HadGEM3: comparison to the pre-industrial era, previous model versions and proxy data, Clim. Past, 16, 1429-1450, https://doi.org/10.5194/cp-16-1429-2020, 2020.

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H. T., Hill, R., Hyder, P., Ineson, S., Johns, T. C., Keen, A. B, Lee, R. W., Megann, A., Milton, S. F., Rae, J. G. L., Roberts, M. J., Scaife, A. A., Schiemann, R., Storkey, D., Thorpe, L., Watterson, I. G., Walters, D. N., West, A., Wood, R. A., Woollings, T., and Xavier, P. K.: The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations, JAMES, 10, 357-380, https://doi.org/10.1002/2017MS001115, 2017.

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., Klein, S. A. and Taylor, K. E.: Causes of higher climate sensitivity in CMIP6 models, Geophys. Res. Lett., 47, e2019GL085782. https://doi.org/10.1029/2019GL085782, 2020.

Zhang, R., Yan, Q., Zhang, Z. S., Jiang, D., Otto-Bliesner, B. L., Haywood, A. M., Hill, D. J., Dolan, A. M., Stepanek, C., Lohmann, G., Contoux, C., Bragg, F., Chan,W.-L., Chandler, M. A., Jost, A., Kamae, Y., Abe-Ouchi, A., Ramstein, G., Rosenbloom, N. A., Sohl, L., and Ueda, H.: Mid-Pliocene East Asian monsoon climate simulated in the PlioMIP, Clim. Past, 9, 2085-2099, https://doi.org/10.5194/cp-9-2085-2013, 2013.

Zhang, R., Zhang, Z. S., Jiang, D. B., Yan, Q., Zhou, X., and Cheng, Z. G.: Strengthened African summer monsoon in the mid-Piacenzian, Adv. Atmos. Sci., 33, 1061-1070, https://doi.org/10.1007/s00376-016-5215-y, 2016.

Zhang, Z., Li, X., Guo, C., Otterå, O. H., Nisancioglu, K. H., Tan, N., Contoux, C., Ramstein, G., Feng, R., Otto-Bliesner, B. L., Brady, E., Chandan, D., Peltier, W. R., Baatsen, M. L. J., von der Heydt, A. S., Weiffenbach, J. E., Stepanek, C., Lohmann, G., Zhang, Q., Li, Q., Chandler, M. A., Sohl, L. E., Haywood, A. M., Hunter, S. J., Tindall, J. C., Williams, C. J. R., Lunt, D. J., Chan, W.-L. and Abe-Ouchi, A.: Mid-Pliocene Atlantic Meridional Overturning Circulation simulated in PlioMIP2, Clim. Past, 17, 529-543, https://doi.org/10.5194/cp-17-529-2021, 2021.

Zhu, J., Poulsen, C. J. and Otto-Bliesner, B. L.: High climate sensitivity in CMIP6 model not supported by paleoclimate, Nat. Clim. Chang, 10, 378-379, https://doi.org/10.1038/s41558-020-0764-6, 2020.

1274

1275

1276