Quantifying Uncertainty in Model Predictions for the Pliocene (Plio-QUMP): Initial results


1. Introduction

Evidence that humankind is affecting the climate system is now overwhelming (IPCC, 2007). However, predictions of the magnitude of future climate change are limited by an incomplete knowledge of the uncertainty in climate model predictions. Uncertainty in climate models comes from four sources, firstly the ability of the model to simulate the present global climate system (known as ‘model skill’), secondly the ‘Charney Sensitivity’ (the global annual mean temperature response to a doubling of carbon dioxide (CO2) (Charney, 1979 — the ability of the model to simulate the transient changes in climate), thirdly the initial conditions for the simulation, and fourthly the boundary conditions used within the model. Boundary conditions in a fully coupled atmosphere–ocean climate model include the concentrations of trace gases, the land–sea and ice masks, vegetation cover, orography and any changes to orbital parameters. To run a climate model simulation, the model must be initially spun up into equilibrium with the specified boundary conditions. In predictive climate modelling, this is achieved by running the model using observational data sets of trace gases to drive the model which has modern settings for orography, ice and land–sea masks. The model is then run for the length of the observational period bringing it up to the point where the model becomes predictive (Johns et al., 2003). The model skill at reconstructing the climate for the observational period is tested, and if the model is accurate, it can continue to run the predictive simulations (IPCC, 2007). The limitation of this method is that the model has been tested on its skill of replicating the gradually warming climate from 1750 to the present day, (~0.75 °C over this time period (IPCC, 2007)). Climate change is predicted to be most likely at least 2 to 3 °C warmer by 2100 (IPCC, 2007), a rate and
magnitude of change far greater than anything experienced in the past 260 years. Coupled with uncertainty over the future trace gas emissions (a boundary condition uncertainty), the models show that the world will get warmer over the coming century, with predictions for global temperature change ranging from +1.6 to +6.4 °C (IPCC, 2007), and with even greater uncertainty in the regional effects of these levels of climate change.

Palaeoclimate modelling offers a method for quantifying and reducing some of these uncertainties. By selecting an appropriate time period to study, it is possible to test the model skill at replicating radically different climate states. The mid-Pliocene Warm Period (mPWP, ~3.3 to 3.0 Ma BP) provides an excellent opportunity to test model skill at reconstructing a warmer world as it was globally 2 to 3 °C warmer than the pre-industrial era (e.g. Haywood et al. 2000; Haywood and Valdes, 2004; Haywood et al. 2009a; Dowsett et al., 2010b). Elevated concentrations of CO2 in the atmosphere (estimates ranging from 360 to 425 ppm) are seen as at least a contributing factor in the warming (Raymo et al., 1996; Pagani et al., 2010, Seki et al., 2010), with another potential cause of warming being the palaeogeography, which in the mPWP was very similar to the present day. Most importantly a detailed and comprehensive dataset of palaeoenvironmental conditions is available to initially constrain or evaluate model predictions (Dowsett, 2007; Haywood et al., 2009b). Whilst there are other intervals in geological time with warmer conditions (e.g. the Eocene) or a greater abundance of high resolution data (the Last Glacial Maximum), the mPWP represents an excellent balance between higher temperatures and supply of robust proxy data for use in a test of climate model skill (Dowsett et al., 1996, Raymo et al., 2009; Dowsett et al., 2010a, 2010b).

In the past, the exploration of uncertainty in climate model predictions has been tackled by the creation of model ensembles, initially through Multi-Model Ensembles (MMEs), where a series of structurally different climate models, with different climate sensitivities (ranging from 1.5 °C to 4.5 °C; Hegerl et al., 2006) are run from the same set of initial conditions and prescribed emissions scenarios (e.g. Stott & Forest, 2007; Tebaldi & Knutti, 2007). These ensembles have dominated the recent IPCC Assessment Reports (IPCC, 2001; 2007) and major projects such as the Climate Model Intercomparison Project (CMIP; Meehl et al., 2000), and the Palaeoclimate Model Intercomparison Project (PMIP; Braconnot et al., 2007). The Quantifying Uncertainty in Model Predictions (QUMP) (Murphy et al., 2004; Collins et al., 2006; Murphy et al., 2007; Collins et al., 2010) and climateprediction.net projects (Stainforth et al., 2005; Sanderson et al., 2008a,b) have developed an alternative method to MMEs for creating ensemble predictions of future climate change, achieved through a Perturbed Physics Ensemble (PPE). PPEs use only one model structure, but by perturbing physical parameters in the model generate a large ensemble of different representations of the climate system with different climate sensitivities (Piani et al., 2005; Collins et al., 2010). Most climate models have a fixed resolution, measured in degrees of longitude and latitude that define a grid box for the atmospheric or oceanic component. Inside this grid box will be elements of the physical properties of the climate system that have to be parameterised because they are sub-grid scale. Some of these parameters have a range of plausible values and during the development of the model a value for each of these parameters is selected based on model performance and physical understanding (Murphy et al., 2004; Piani et al., 2005; Stainforth et al., 2005; Collins et al., 2006; 2010). The PPE is created by changing the values of a selection of these parameters in the settings of the climate model (Murphy et al., 2004).

QUMP has focussed on both idealised scenarios and climate projections for the 21st century (e.g. Murphy et al., 2004) with some work looking at the mid-Holocene (Brown et al., 2008) and the Last Glacial Maximum (the PalaeoQUMP project). This paper introduces the initial phase of the Quantifying Uncertainty in Model Predictions for the Pliocene (Plio-QUMP) project, a first experiment in applying PPE to a past warmer world with higher CO2. The paper presents the results from the first three members of the coupled model ensemble using the UK Met Office coupled climate model (HadCM3) including key model diagnostics (those that can be compared to proxy datasets) and the impacts on data/model comparisons.

2. Methods

2.1. Experimental design

This paper presents the initial three PPE experiments of the Plio-QUMP Project. For each ensemble member, a spin up phase of several hundred years is performed to compute the stabilising flux adjustments. After spin up, the simulations are used to initialise the three member HadCM3 coupled PPE (see Sections 2.2.1 and 2.3). The three member ensembles are then integrated for further 300 simulated years. The ensemble consists of a standard-parameter simulation, a high sensitivity simulation, and a low sensitivity simulation (see Section 2.3). The standard simulation (which includes flux corrections) was verified against previous non-flux-adjusted standard simulations used in mPWP studies and was found to display an acceptable simulation of key metrics. Therefore, the members of this ensemble are valid for comparison against previous models of the mPWP. However in this paper only the standard for this simulation was used and referred to in the comparisons with the high and low sensitivity simulations. The reasons for the use of flux corrections in these experiments compared to other mPWP modelling experiments are outlined in Section 2.3.

The simulations were tested using different forms of data/model comparison. Model surface temperatures were tested against proxy SST estimates derived from the US Geological Survey Pliocene Research Interpretation and Synoptic Mapping (PRISM) PRISM3D Mean Annual Sea Surface Temperature (MASST) dataset (Dowsett et al., 2010b; see Section 2.4). The climatological outputs were also used to force the BIOME4 vegetation model (see Section 2.2.2) with the simulated biomes being compared to the Pliocene vegetation dataset of Salzmann et al. (2008) (Section 2.4). The use of the climate data in both a direct data/model comparison with the SST data and then to drive the vegetation model makes this a two step approach, as errors in the comparison with the SST data will be inherited by the vegetation model inputs.

MPWP conditions are replicated through the creation of a series of boundary conditions within the model. The boundary conditions used in this study are the same as in a majority of mPWP modelling studies including Haywood & Valdes (2004), Haywood et al. (2007), Lunt et al. (2008a, 2008b), Haywood et al. (2009b), and Lunt et al. (2009, 2010). We used a modern land–sea mask (including a fully closed Central American Seaway), a mid-Pliocene land–ice mask and an atmospheric CO2 concentration of 400 ppm which is within the range of values presented by the data (Raymo et al., 1996, Pagani et al., 2010, Seki et al., 2010). The simulation was initialised with prescribed vegetation from the PRISM2 reconstruction, a mega-biome reconstruction with seven biomes (Matthews, 1985; Dowsett et al., 1999). The land ice mask is adjusted from the modern through reducing the land ice mask and an oceanic component. Inside this grid box will be elements of the physical properties of the climate system that have to be parameterised because they are sub-grid scale. Some of these parameters have a range of plausible values and during the development of the model a value for each of these parameters is selected based on model performance and physical understanding (Murphy et al., 2004; Piani et al., 2005; Stainforth et al., 2005; Collins et al., 2006; 2010). The PPE is created by changing the values of a selection of these parameters in the settings of the climate model (Murphy et al., 2004).
2.2. Model descriptions

2.2.1. HadCM3

This study used the UK Met Office fully coupled atmosphere–ocean general circulation model (AOGCM) HadCM3, which contains atmosphere, ocean, fixed vegetation and sea ice components (Gordon et al., 2000). The atmosphere is composed of 19 vertical levels with a horizontal grid resolution of 2.5° latitude by 3.75° longitude (Gordon et al., 2000), which equates to a grid box at the equator of 278 km latitude by 417 km longitude. The model contains a number of features that are developments from the predecessor HadCM2 (see Johns et al., 1997). This includes a radiation scheme covering six additional spectral bands in the shortwave and eight additional bands in the long wave and explicitly representing the radiative effects of all greenhouse gases not just CO₂, O₃ and H₂O (Edwards & Slingo, 1996; Gordon et al., 2000; Johns et al., 2003). Background aerosols in the model are determined using prescribed pre-industrial emission input files. The penetrative convection scheme of Gregory & Rowntree (1990) has been developed to include a parameterisation of the impacts of convection on momentum and the downdraft of convection (Gordon et al., 2000; Johns et al., 2003). HadCM3 employs the use of MOSES (Met Office Surface Exchange Scheme) including soil moisture response to temperature and on the effect of CO₂ and stomatal resistance on evapo-transpiration (Williams et al., 2001). A number of other parameterisations in the model are linked to features in cloud representation, and cloud development details of these parameterisations are found in Gordon et al. (2000). Details of the atmospheric component are in Pope et al. (2000).

The ocean component comprises 20 levels with a rigid lid on a 1.25° × 1.25° latitude–longitude grid (Gordon et al., 2000; Brierley et al., 2010) which represents a grid box of 139 km by 139 km at the equator. There are 6 ocean grid boxes for every atmospheric grid box in the coupling of the model. A key component of the ocean model is the interaction with sea ice, and every high latitude ocean grid box in HadCM3 can have sea ice cover (Gordon et al., 2000). A number of ocean basin topographies had to be edited due to grid scale and this was found to have an especially sensitive response in the North Atlantic around the Iceland–Faroe–Scotland ridge and in the Denmark Strait. Topographies were smoothed in places and channels set at certain depths in the model (Roberts & Wood., 1997; Gordon et al., 2000). A number of parameterisations exist in the ocean component to represent fluxes and mixing processes in the ocean, details of which are found in Gordon et al. (2000). The final key parameterisation in the ocean component is Mediterranean outflow to the Atlantic. In the ocean this is a crucial flow and has wide ranging impacts on Atlantic waters by venting warm, salty water into the cooler North Atlantic off Spain. However, in the HadCM3 land/sea mask, the Strait of Gibraltar is closed, so a parameterisation represents outflow of waters through the strait (Johns et al., 2003).

2.2.2. BIOME4

BIOME4 is an equilibrium vegetation model driven offline (i.e. unconnected) with climate model outputs and is used here to interpret the effects of the climate of the initial ensemble members upon likely biomes of the mPWP. A climatologically averaged year composed of 1.5 m temperature, precipitation and cloudiness are upon likely biomes of the mPWP. A climatologically averaged year interpretation of the effects of the climate of the initial ensemble members (unconnected) with climate model outputs and is used here to create a BIOME4 year of 12 months January–December. The absolute annual minimum temperature field has no time dimension in the model and was created by taking the average February (northern hemisphere cold month) and the average August (southern hemisphere cold month) temperature and merging these to create the annual coldest temperature for the BIOME4 year.

2.3. The Perturbed Physics Ensemble

HadCM3 contains over 100 parameters in the atmospheric component, of which 31 have been identified as potentially having a noticeable effect on climate when they are perturbed (Murphy et al., 2004; Collins et al., 2006; 2010). Initially, these parameters were perturbed individually (Murphy et al., 2004). The approach was developed by perturbing multi-parameter sets to create more ensemble members (Collins et al., 2006, 2010). With two or three potential settings per parameter this offers the opportunity to create an ensemble with millions of members. However, restrictions in computing power reduced this to the most skilful 129 members for the UK Met Office coupled atmosphere–slab ocean model (HadSM3) and the most skilful 17 for the coupled model (Collins et al., 2006; Webb et al., 2006; Collins et al., 2010). The rationale methodology for reducing the massive million-member ensemble down to a computationally manageable size is discussed in Collins et al. (2010).

The QUMP project has used a flux corrected version of the HadCM3 model to correct for a top of the atmosphere (TOA) radiation imbalance created by the perturbed physics simulations. During standard model construction the model parameters are adjusted to ensure that the incoming and outgoing radiations are equal. Since the parameter perturbations cause a TOA imbalance, a flux adjustment is required. The flux adjustments were applied through performing a multi-decade simulation for each member and relaxing values for the seasonal and spatial distribution of sea surface temperature and sea surface salinity values, called a Haney forcing (using a relaxation coefficient of 30 days for temperature and 120 days for salinity as in Tziperman et al. (1994)). This was applied until the model reached a minimal forcing effect from this change (TOA approximately less than 0.2 Wm⁻²). Once this happened the seasonally-varying flux adjustment input file was created and this allows the model to run with a stable climate in both a normal run and PPE experiments. In Collins et al. (2006) the use of flux adjustments was shown to have an impact on the model, causing a slowing of Atlantic Meridional Overturning Circulation (MOC), and a cooling of North Atlantic SST’s. However, the use of a different relaxation time constant for temperature and salinity (i.e. less vigorous forcing) improves the performance of the flux adjustment through reducing significantly the biases to northern hemisphere SST and sea ice that occurred in previous PPE studies. This has been shown in comparison work completed in Collins et al. (2010) between the standard (no flux adjustments) HadCM3 model, the flux adjusted HadCM3 used in Collins et al. (2006; shorter relaxation constant) and the flux adjusted HadCM3 used in Collins et al. (2010; longer relaxation constant). The Plio-QUMP project including this experiment initiated the model simulations with the improved flux corrections of Murphy et al. (2007) and Collins et al. (2010).
We created a three member PPE, which consisted of an unperturbed standard simulation and two perturbed simulations named high and low sensitivity. The ‘high sensitivity’ simulation had the parameters perturbed into a combination that created the highest Charney sensitivity of the PPE settings used in the HadCM3 QUMP experiments, with the ‘low sensitivity’ having the lowest Charney sensitivity whilst maintaining an acceptable simulation of modern climate. Table 1 shows the parameters changed, their area of influence in the model construction and the values used in each simulation in the ensemble. The final Charney sensitivities of the simulations were 7.1 °C for the high sensitivity, 2.1 °C for the low sensitivity and 3.3 °C for the standard sensitivity (values obtained from equivalent HadSM3 experiments). The Charney Sensitivity values are used to distinguish the differences between the simulations and are not seen as the defining factor for the Plio-QUMP experiments. Other members of the full ensemble have similar (and more moderate) Charney Sensitivities to each other and display different climatic effects as the parameters interact with the boundary conditions (Collins et al., 2010). The extreme outlier nature of the high and low sensitivity simulation enabled the best chance of obtaining a wide range of results.

In addition to the PPE simulations that use Pliocene boundary conditions, we also make use of experiments forced with both anthropogenic and natural forcing factors from 1860 to 2000 (Collins et al., 2010). This allows us to make comparisons and detect anomalies between mPWP simulations and simulations relevant to the historical period where modern-day observations have been made.

Analysis of the experiments from the QUMP and climateprediction.net PPEs in Rougier et al. (2009) identified which parameters drove the largest changes in HadSM3 which was mainly large scale cloud process parameters, with parameters 1–4, 7 and 18 in Table 1 showing the greatest influence. Parameter 7 (the entrainment rate co-efficient) exerted the strongest influence on the model when it was perturbed (Rougier et al., 2009).

2.4. Data

The palaeoclimate modelling results were evaluated against two different types of dataset, a multi-proxy SST dataset and a terrestrial vegetation dataset. Both come from the PRISM3D palaeoenvironmental reconstruction of the mPWP (Dowsett et al., 2010a). PRISM3D, the 4th version of the dataset (PRISMO through to 3D (Dowsett et al., 1994; 1996; 1999, 2005, 2010a, 2010b)) comprises 86 marine and 202 terrestrial data points, reconstructed sea surface temperatures, deep water temperatures, sea ice, land ice and vegetation. The PRISM reconstruction is a comprehensive dataset created with palaeoclimate modelling in mind, so the reconstruction does not create a time series for a proxy record (or stack of proxy records) as in Lisiecki & Raymo (2005) but a slab record covering an approximately ~300 kyr interval. The PRISM3D Mean Annual Sea Surface Temperature data (PRISM3D MASSST) is the latest release from the PRISM group, specifically aimed at assisting in the data/model comparison of projects such as Plio-QUMP (Dowsett et al., 2010b).

The vegetation data comprises 202 terrestrial sites originally presented in Salzmann et al. (2008), which were incorporated into the PRISM3D dataset. The biome of an area is a reflection of the general climate, and as such a vegetation data comparison with the results from a vegetation model forced with the climate model outputs adds

### Table 1

<table>
<thead>
<tr>
<th>Label</th>
<th>Sub-group</th>
<th>Identifier/properties</th>
<th>Climate sensitivity (°C)</th>
<th>Low sensitivity</th>
<th>Control</th>
<th>High sensitivity</th>
<th>Area of influence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>VFI</td>
<td>Icefall speed (ms⁻¹)</td>
<td></td>
<td>0.93976</td>
<td>1</td>
<td>0.54306</td>
<td>Large scale cloud</td>
</tr>
<tr>
<td>2</td>
<td>CT</td>
<td>Cloud droplet to rain conversion rate (×10⁻⁴ s⁻¹)</td>
<td></td>
<td>1.63E-04</td>
<td>1.00E-04</td>
<td>3.50E-04</td>
<td>Large scale cloud</td>
</tr>
<tr>
<td>3</td>
<td>RHCrit</td>
<td>Threshold of relative humidity for cloud formation</td>
<td></td>
<td>0.84077</td>
<td>0.7</td>
<td>0.7</td>
<td>Large scale cloud</td>
</tr>
<tr>
<td>4</td>
<td>CW</td>
<td>Cloud droplet to rain conversion rate (×10⁻⁴ kg m⁻³)</td>
<td></td>
<td>1.75E-04</td>
<td>2.00E-04</td>
<td>1.41E-04</td>
<td>Large scale cloud</td>
</tr>
<tr>
<td>5</td>
<td>MinSIA</td>
<td>Minimum sea ice albedo — dependence of SIA on temperature</td>
<td></td>
<td>4.25E-05</td>
<td>5.00E-05</td>
<td>3.23E-05</td>
<td>Large scale cloud</td>
</tr>
<tr>
<td>6</td>
<td>ice_tr</td>
<td>Sea ice minimum temperature</td>
<td></td>
<td>0.50666</td>
<td>0.5</td>
<td>0.60104</td>
<td>Sea ice</td>
</tr>
<tr>
<td>7</td>
<td>Ent</td>
<td>Entrainment rate coefficient</td>
<td></td>
<td>9.5243</td>
<td>10</td>
<td>3.836</td>
<td>Sea ice</td>
</tr>
<tr>
<td>8</td>
<td>lesizes</td>
<td>Ice particle size (µm)</td>
<td></td>
<td>2.89934</td>
<td>3</td>
<td>2.37726</td>
<td>Convection</td>
</tr>
<tr>
<td>9</td>
<td>Cape</td>
<td>CAPE closure — intensity of convective mass flux</td>
<td></td>
<td>30.761</td>
<td>30</td>
<td>28.099</td>
<td>Radiation</td>
</tr>
<tr>
<td>10</td>
<td>flux_g0</td>
<td>Boundary layer flux profile</td>
<td></td>
<td>16.9231</td>
<td>10</td>
<td>6.9039</td>
<td>Boundary layer</td>
</tr>
<tr>
<td>11</td>
<td>Charnoc</td>
<td>Charnock constant — roughness lengths and surface fluxes over sea</td>
<td></td>
<td>0.0128</td>
<td>1.20E-02</td>
<td>0.0134</td>
<td>Boundary layer</td>
</tr>
<tr>
<td>12</td>
<td>soillev</td>
<td>Number of soil levels accessed for evaportranspiration</td>
<td></td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>Land surface</td>
</tr>
<tr>
<td>13</td>
<td>lambda_d</td>
<td>Asymptotic neutral mixing length parameter</td>
<td></td>
<td>0.3356</td>
<td>0.15</td>
<td>0.135</td>
<td>Boundary layer</td>
</tr>
<tr>
<td>14</td>
<td>cnv_diffl</td>
<td>Free convective roughness over sea</td>
<td></td>
<td>4.14E-03</td>
<td>1.30E-03</td>
<td>1.67E-03</td>
<td>Boundary layer</td>
</tr>
<tr>
<td>15</td>
<td>oil_diffl</td>
<td>Ocean ice diffusion</td>
<td></td>
<td>3.67E-04</td>
<td>3.75E-04</td>
<td>3.73E-04</td>
<td>Sea ice</td>
</tr>
<tr>
<td>16</td>
<td>dyndel</td>
<td>Order of diffusion operator</td>
<td></td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>Dynamics</td>
</tr>
<tr>
<td>17</td>
<td>dyndiff</td>
<td>Diffusion e-folding time</td>
<td></td>
<td>8.513</td>
<td>12</td>
<td>8.5</td>
<td>Dynamics</td>
</tr>
<tr>
<td>18</td>
<td>eactfl</td>
<td>Cloud fraction at saturation boundary layer</td>
<td></td>
<td>0.79598</td>
<td>0.5</td>
<td>0.51262</td>
<td>Large scale cloud</td>
</tr>
<tr>
<td>19</td>
<td>eactfrp</td>
<td>Cloud fraction at saturation free troposphere value</td>
<td></td>
<td>0.64799</td>
<td>0.5</td>
<td>0.50631</td>
<td>Large scale cloud</td>
</tr>
<tr>
<td>20</td>
<td>k_gw</td>
<td>Surface gravity wavelength</td>
<td></td>
<td>1.98E+04</td>
<td>2.00E+04</td>
<td>1.67E+04</td>
<td>Dynamics</td>
</tr>
<tr>
<td>21</td>
<td>k_lee</td>
<td>Trapped lee wave constant</td>
<td></td>
<td>2.97E+05</td>
<td>3.00E+05</td>
<td>2.50E+05</td>
<td>Dynamics</td>
</tr>
<tr>
<td>22</td>
<td>gw_gw</td>
<td>Starting level for gravity wave drag</td>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>Dynamics</td>
</tr>
<tr>
<td>23</td>
<td>s_sphe</td>
<td>Non-spherical ice particles shortwave radiation properties</td>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>Radiation</td>
</tr>
<tr>
<td>24</td>
<td>c_sphe</td>
<td>Non-spherical ice particles longwave radiation properties</td>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>Radiation</td>
</tr>
<tr>
<td>25</td>
<td>rhparam</td>
<td>Flow dependant RHcrit</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Radiation</td>
</tr>
<tr>
<td>26</td>
<td>vertfl</td>
<td>Vertical gradient of cloud water</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>Radiation</td>
</tr>
<tr>
<td>27</td>
<td>canopy</td>
<td>Surface canopy energy exchange</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Land surface</td>
</tr>
<tr>
<td>28</td>
<td>cnv_upd</td>
<td>Convective anvil updraft factor</td>
<td></td>
<td>0.51016</td>
<td>Off</td>
<td>Off</td>
<td>Convection</td>
</tr>
<tr>
<td>29</td>
<td>anvil</td>
<td>Convective anvil shape factor: radiative properties of convective cloud</td>
<td></td>
<td>2.59138</td>
<td>Off</td>
<td>–</td>
<td>Convection</td>
</tr>
<tr>
<td>30</td>
<td>sto_res</td>
<td>Dependence of stomatal conductance on CO₂</td>
<td></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>Land surface</td>
</tr>
<tr>
<td>31</td>
<td>f_rough</td>
<td>Forest roughness lengths</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Land surface</td>
</tr>
<tr>
<td>32</td>
<td>sw_absn</td>
<td>Shortwave water vapour continuum absorption</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Radiation</td>
</tr>
</tbody>
</table>
another dimension to testing the model results. The data/model comparison is achieved through the use of a Kappa statistic test (see Section 2.5.2.) to look for areas of agreement between the data set and the BIOME4 model outputs.

2.5. Statistics

2.5.1. Student’s t-test

Student’s t-test is a method of testing the null hypothesis between two datasets. Using a null hypothesis that any variation between the model data (i.e. ‘high minus standard’) is due to natural variability between model simulations, significant changes in the model results can be identified. The t-test was applied to the 30 year average climatologies with all the results being tested to a confidence level of 95%. Any areas where the anomaly between the model simulations was significant indicated that the change was due to more than just the model variability in the two simulations.

2.5.2. Cohen’s Kappa Statistic

Cohen’s Kappa Statistic (Cohen, 1960) quantitatively assesses the agreement between two sets of categorisations, whilst taking into account chance agreements. The Kappa Statistic (k) is calculated by subtracting the proportion of expected agreement (Pe) from the proportion of observed agreement (Po) and this result is normalised through dividing it by the maximum possible difference (1 − Pe) to generate the Kappa Statistic:

\[ k = \frac{(P_o - P_e)}{(1 - P_e)}. \]

The Pe value includes the expectation that an element of agreement by chance exists and this is allowed for in the Kappa Statistic (Cohen, 1960; Prentice et al., 1992).

The values for the test range from 0 (agreement by chance) to 1 (perfect fit) (Cohen, 1960; Jenness & Wynne, 2005). By allowing for agreement by chance in the statistical test, even the smallest difference in the results between the different simulations is indicating a statistically significant difference between the data/model comparisons for each simulation. Weaknesses of the Kappa Statistic include that the cause of the difference is not specified; it could be one site or a number of sites causing the change in the result. It is only possible to state that the closer to 1 (perfect fit) the Kappa Statistic is, the better the data/model comparison. The Kappa Statistic does not make any allowance for how wrong a site mismatch is, so being a very close biome reconstruction or a complete opposite biome reconstruction does not alter the end value, the test simply views two sites which do not match the data. Two forms of the statistic have been applied to the Plio-QUMP simulations, the full 28 biome classification and an 8 megabiome classification following Harrison and Prentice (2003) and Salzmann et al. (2009). The reason for using the megabiome classification, as well as a full biome classification, is the increased confidence in the statistical test applied since having a large number of categories with a low sample in each is less robust than having fewer categories with more samples in each. Ideally a minimum of 50 samples per category should be used, and 75–100 samples for more than 12 categories (Congalton and Green, 1999; Jenness and Wynne, 2005). This is difficult for palaeobotanical studies where sample sizes are restricted by many factors such as deposition, taphonomy, preservation and limited exposure. This makes the megabiome classification’s Kappa scores more statistically robust than that for the full biome classification. The Kappa Statistic test has been used successfully in previous palaeoclimate vegetation data/model studies (Salzmann et al., 2008; 2009; Salzmann et al., 2009; Haywood et al., 2009b; Pound et al., 2011).

### Table 2

<table>
<thead>
<tr>
<th>Simulation</th>
<th>Global mean annual temperature anomaly</th>
<th>Global mean annual precipitation anomaly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pliocene minus modern (high sensitivity)</td>
<td>3.3 °C</td>
<td>0.15 mm/day</td>
</tr>
<tr>
<td>Pliocene minus modern (standard simulation)</td>
<td>2.5 °C</td>
<td>0.18 mm/day</td>
</tr>
<tr>
<td>Pliocene minus modern (low sensitivity)</td>
<td>2.4 °C</td>
<td>0.12 mm/day</td>
</tr>
<tr>
<td>Pliocene high sensitivity minus standard</td>
<td>1.5 °C</td>
<td>−0.18 mm/day</td>
</tr>
<tr>
<td>Pliocene low sensitivity minus standard</td>
<td>−0.1 °C</td>
<td>−0.06 mm/day</td>
</tr>
</tbody>
</table>

3. Results

3.1. Climate metrics

3.1.1. Pliocene minus modern anomalies

The standard Pliocene simulation was compared to a ‘modern’ simulation covering the period 1950–1985, the time period of Reynolds & Smith (1995), whose data was used to produce the core top estimates of the PRISM3D MA SST dataset by the PRISM group. Two comparisons were plotted, with the standard simulations for the...
temperature and precipitation fields. Global mean values were calculated for all three ensemble members and the results are shown in Table 2.

The Pliocene is modelled as being warmer than modern. Fig. 1a shows this is predominantly through warming in the high latitudes and polar regions (the area of intense warming on East Antarctica is caused by differences in the Pliocene and the modern land ice mask). Temperatures in the tropics are only marginally warmer than the modern values, which fits with the pattern in the PRISM data of an enhanced equator to pole temperature gradient in the Pliocene compared to the modern (Dowsett et al., 2010a). Fig. 1b shows that changes in precipitation between the Pliocene and the modern simulations are subtle, with a few areas exhibiting large increases in precipitation (e.g. eastern Pacific), and some areas of large decreases in the precipitation anomaly (e.g. Amazonia and Indonesia).

Table 2 reinforces the global plots shown in Fig. 1 with the global mean averages for Pliocene temperature shown to be 2.5 to 3.3 °C warmer than the modern simulations and marginally wetter by 0.12 to 0.18 mm/day. These differences will be driven by the difference in boundary conditions — i.e. orography (Rocky Mountains 50% lower than modern) and CO₂ values (Pliocene 400 ppmv and modern ~320 ppmv) as the modern simulations are the corresponding PPE member. The interaction between the difference in the Pliocene and modern boundary will have caused the variation in global means (Table 2) and regional differences (Fig. 1) displayed between the three simulations.

3.1.2. Perturbed physics simulations

As shown in Table 1, the high sensitivity simulation has a Charney Sensitivity of 7.1 °C compared to the 3.3 °C of the standard simulation and the 2.1 °C of the low sensitivity simulation. In terms of the effect on model temperatures for the simulations, this leads to the expected conclusions that for the vast majority of the Earth’s surface the anomaly plot for ‘high minus standard’ (Fig. 2a) shows a warm anomaly. The largest warm anomaly is seen in the higher latitudes and over continental North America and Asian subtropics. The ‘low minus standard’ anomaly plot (Fig. 2b) shows a cooling anomaly over oceanic regions and high latitudes with some areas of warm anomaly on continental areas.

The Student’s t-test shows that the changes in ‘high minus standard’ anomaly are significant (at a 95% confidence level) except for a region in the mid-latitude North Pacific and the east coasts of China and Japan (Fig. 2c), which are the two large cool anomalies on the plot (Fig. 2a). For the ‘low minus standard’ anomaly there are slightly more areas where the anomaly was not significant (Fig. 2d), but not in areas containing sites for data/model comparison (both SST and vegetation).

The ‘high minus standard’ plot shows up to 14 °C warming anomaly over the high latitude oceans whilst only showing minimal warming in the tropics (Fig. 2a). This is important as the tropics have been identified as showing no clear data/model mismatch in previous Pliocene studies, with the weakness in the models being at the higher latitudes in the data/model comparisons (Fig. 5). In continental areas, the model predicts a warm anomaly in Australia of 2 to 4 °C and in

---

Figure 2. Mean annual temperature anomalies for 'high sensitivity minus standard' (2a) and 'low sensitivity minus standard' (2b) in °C. Student’s t-test was applied to the anomaly plots and the insignificant regions, which are plotted in grey and overlain over the 'high minus standard' anomaly (2c) and the 'low minus standard' anomaly (2d).
continental USA, of 3 to 4 °C. The tropical rainforest belt through South America, central Africa and Southeast Asia displays very little change with a warm anomaly in the range of 0.5 to 1 °C, which is the smallest significant change in continental temperatures.

The precipitation patterns in the two anomaly plots display a greater range in the ‘high minus standard’ anomaly (Fig. 3a) than in the ‘low minus standard’ anomaly (Fig. 3b). In both experiments there is little change in global average precipitation, although there may be a significant change regionally. The overall pattern in the ‘high minus standard’ anomaly plot is that higher latitudes increase their precipitation anomaly whilst tropical and extra-tropical areas see a decrease in the precipitation anomaly. In the ‘low minus standard’ anomaly plot the general trend is for a slight reduction in the precipitation anomaly, but there are fewer and weaker areas of major reduction.

The ‘high minus standard’ precipitation anomaly shows an important reduction in the daily average rainfall for the continental USA and Australia. The reduction in Australia is approximately 0.25 mm/day, but the reduction in continental USA is approximately 1 mm/day. Likewise, the temperature plots for this anomaly show a much reduced level of warming in the tropical rainforest belt, and this is mirrored in the precipitation anomaly plots with increases of approximately 2 mm/day through northern South America, central Africa and Indonesia.

For the most part, the Evaporation–Precipitation (E–P) plots show a more positive moisture budget which could be due to a net increase in the precipitation anomaly over most continental areas and high latitudes oceans with a net increase in the evaporation anomaly from tropical oceans. The patterns for both the ‘high minus standard’ anomaly (Fig. 4a) and ‘low minus standard’ anomaly (Fig. 4b) are very similar. The ‘high minus standard’ anomaly has greater areas of net evaporation over land than the ‘low minus standard’ anomaly, but at the same time has more intense areas of net precipitation in tropical forest regions. The patterns displayed highlight the net drying anomaly for continental areas in the ‘high minus standard’ simulation, which have the potential to change the vegetation types between the high and standard simulations, in turn causing a change in the skill of the data/model comparison between the BIOME4 model and the palaeobotanical data (Section 3.3).

The range in the temperature difference between the ‘high minus standard’ anomaly compared to the ‘low minus standard’ anomaly is associated with the slightly greater statistical significance of the ‘high minus standard’ anomaly compared to the ‘low minus standard’ anomaly temperature plots. The variation in the climate sensitivity values for the three simulations ensured that the effects of the high simulation in comparison with the standard were greater and far more likely to override any anomalies due to natural variability in the models. The Student’s t-test failed to find any model grid points where the precipitation data or the E–P data are insignificant for either the ‘high minus standard’ or the ‘low minus standard’ anomalies. This is probably due to a weakness in the choice of test, but as these data
Fig. 5. Data/model comparison using the PRISM3D Mean Annual Sea Surface Temperature (MA SST) dataset. PRISM3D MASST minus standard (5a), PRISM3D MASST minus high sensitivity simulation (5b) and PRISM3D MASST minus low sensitivity simulation (5c) in °C. Root mean square errors (RMSE) were calculated for each comparison as (a) 4.37, (b) 3.25 and (c) 4.38.

### 3.2. Data/model comparison

**SSTs**

Fig. 5a displays the present areas of reduced skill in the data/model comparison with the PRISM3D MASST dataset. These areas are focussed in the higher latitudes and range from 4 to 14 °C in the North Atlantic and Arctic Ocean in areas where the anomalies were shown to be statistically significant. The reduced skill is best characterised in the North Atlantic where the high concentration of data points (DSDP/IODP sites 410, 552, 606, 607, 608, 609, 610, 907, 909, 911 — Dowsett, 2007 for more details) highlights the progressive reduction in skill of the data/model comparison moving northwards. This has been a significant area of data/model difference for mPWP climate HadCM3 simulations (Robinson, 2009). Away from the high concentration of data points in the North Atlantic, other key areas of ocean through-flow to be noted are oceanic gateways and upwelling regions and the tropical Pacific.

The high sensitivity simulation (Fig. 5b) provides the closest fit to the PRISM data of the three simulations. This simulation decreases the discrepancy between the data and the model by 3 to 6 °C, with the largest increases in the highest latitude data sites. This result is to be expected with the temperature data produced by the model and the warming at the surface being driven by the perturbed parameters creating the high climate sensitivity of 7.1 °C. The general pattern observed in Fig. 2a is repeated in this data/model comparison, with the warm anomaly of the high sensitivity simulation mainly at higher latitudes compared to the tropical regions.

The low sensitivity simulation fails to improve the data/model comparison, but despite the cooler ocean surface temperatures shown in Fig. 2b (for “low minus standard” anomaly), the data/model comparison is not significantly poorer between the low and the standard simulations. There is one small area of improvement north of Iceland where the model/data comparison at one site is improved by about 1 to 2 °C, to within the uncertainty in the data values.

In areas away from the North Atlantic, all three simulations perform very well in the data/model comparison. The standard simulation data/model comparison shows that the model overestimates warming in the tropical regions, causing a negative data/model anomaly, but one that is within the analytical error of the data (approximately +/− 1.5 °C; Dowsett et al., 2010b). There is a slight increase in this negative anomaly in the high sensitivity simulation, of about a further 1 °C, with no change in the low sensitivity simulation. This causes a unilateral cooling anomaly across the equatorial Pacific region where there are data points for comparison. There is no noticeable change in the data/model comparison between the high and standard simulations in the areas of upwelling with data points (off western Africa and South America), and whilst the low sensitivity simulation is marginally weaker than the standard simulation in this area, the change is negligible. A similar lack of change in the comparison for all three simulations occurs near the Indonesian gateway.

In terms of the comparison between the PRISM3D MASST data and the ensemble members, the high sensitivity simulation is the most skillful of the three shown in this ensemble for recreating the conditions of the mPWP based on the RMS errors (3.25 for the high sensitivity compared to 4.37 (Standard) and 4.38 (low sensitivity)).

### 3.3. Data/model comparison — Biomes

The outputs from the BIOME4 model were compared with the Piacenzian Stage palaeobotanical database of Salzmann et al. (2008) using Kappa Statistics (Section 2.5.2). The results of this data/model comparison show that the standard version of HadCM3 in this ensemble produces the best agreement between the data and the model. The standard simulation produced a Kappa score of 0.201 for the full BIOME classification and 0.229 for the Megabiome classification (Fig. 6a), with scores of 0.186 (full) and 0.172 (mega) for the high sensitivity simulation (Fig. 6b) and scores of 0.120 (full) and 0.162 (mega) for the low sensitivity simulation (Fig. 6c). These results indicate that the high sensitivity simulation was better than the low sensitivity ensemble member in comparison to the palaeobotanical
data. The vast majority of the regions of poor data/model comparability reflect the model simulating less precipitation than is required for the reconstruction of the palaeo-data. In tropical regions this leads to a loss of forest and its replacement with savanna, probably as a result of poor representation of the total rainfall. In higher latitudes the weaker vegetation reconstructions are probably related to the
seasonal cycles in the rainfall in the model, compared to the palaeo-data. This is reflected in the reduced amount of extreme errors in the higher latitudes, with patterns being similar but slightly different to tropical latitudes which see large changes in the vegetation type.

Regional comparison between the model-predicted biomes and the palaeobotanical data shows that all three simulations have areas of real difficulty. There is no agreement over Australia in any of the comparisons of model output with the data. Over Australia the model output is too dry leading to the prediction of desert and xerophytic tropical shrubland biomes, whereas the palaeobotanical data show that Australia was dominated by tropical forest biomes in the northeast, tropical savanna across the northwest and centre and temperate to warm-temperate forest and woodland biomes in the southwest and southeast.

In northern South America, the model simulations correctly predict tropical forest except in the northeast. Data from Chile show the presence of tropical savanna, whereas the model predicts warm-temperate to temperate forest. In southern South America the data again disagree with model output. Palaeobotanical data from Argentina show the presence of temperate deciduous broadleaved savanna and temperate grassland during the mid-Pliocene. The standard and low simulations predict temperate needle-leaf forests and temperate sclerophyll woodland for this region, whereas the high simulation predicts tropical xerophytic shrubland and desert. South America is one of the areas that is unavoidable weak for climatological simulation predicts tropical xerophytic shrubland and desert. South America is one of the areas that is unavoidably weak for climatological data from palaeobotany with some information coming from vertebrate palaeontology (Salzmann et al., 2008). However, it is unlikely this is the cause of the poor relationship between the model and the data in this region.

Data for the Arabian Peninsula suggests xerophytic tropical shrubland with temperate grassland towards the Mediterranean coast. The standard and high sensitivity simulations predict extensive desert coverage for this region. The low sensitivity simulation produced an expanse of xerophytic tropical shrubland, as shown in the data.

In Asia the standard simulation agrees with the aridity of the Tibetan plateau and the warm temperate evergreen mixed forest around Southeast Asia. The rest of Asia is poorly modelled in comparison with the data. The high and low sensitivity simulations show less agreement in Asia than the standard.

All three simulations produce a very similar Antarctica, with a prediction of tundra on the coast of the continent and through the West Antarctic Peninsula (the ice mask was fixed in BIOME4, so vegetation can only occur where there was no ice sheet). This prediction matches the palaeobotanical evidence from the Dry Valleys region of the Transantarctic Mountains. However, the dating of this site is controversial and recent work suggests it may be Miocene in age, not mid-Pliocene (Ackert & Kurz, 2004; Ashworth et al., 2007).

At the highest latitudes of North America, the high sensitivity simulation shows better agreement with the data than the standard and low sensitivity simulations around Ellesmere and Meighen Islands. However, it is less skilful than the standard in its predictions for Alaska. All three simulations produce good data/model agreement for Greenland. The west and Gulf coasts of America are poor in all the simulations, being too warm, with the high sensitivity simulation also becoming too dry. The east coast of North America is consistently comparable to data points below 40°N. Above this latitude the single datum and model predictions have no agreement.

Central America is predicted in all three simulations to have a vegetation of xerophytic tropical shrubland whereas the palaeo-data shows it to have been warm-temperate evergreen broadleaf and mixed forest to tropical evergreen broadleaf forest.

For the Iberian Peninsula, the BIOME4 outputs all suggest a tropical dry to temperate dry climate whereas the data indicate only a temperate dry climate dominant during the mid-Pliocene, with minor areas of warm-temperate evergreen mixed forest. This is most likely caused by the difference between model and palaeoclimatic total annual rainfall and seasonality. There may also be an issue here and elsewhere surrounding the resolution of the model, as the Iberian Peninsula is covered by only six model grid boxes.

Scandinavia is well modelled by the standard and high sensitivity simulations, with agreement with the taiga forest shown by the palaeo-data. The low sensitivity simulation is poorer here as it predicts tundra. Western Europe shows good agreement between the model and the data in all the simulations, with the predicted warm temperate evergreen broadleaf and mixed forest matching well with the majority of the data points. Around the eastern Mediterranean coast, the model simulations (especially the high sensitivity simulation) predict temperate sclerophyll woodland and shrubland, whereas the data show warm temperate mixed forest. This is probably due to differences in data and model predictions for annual precipitation and seasonality. In Eastern Europe, the low sensitivity simulation is less skilful than the high sensitivity simulation which performs well in this area, matching some of the data with its prediction of warm temperate mixed forest. However, the area it predicts this for extends farther, than the data which shows a change to cooler, drier, more open biomes to the east of the Black Sea.

In Africa, the most noticeable result in any of the simulations is the lack of a Sahara Desert in North Africa, which has been replaced by a xerophytic tropical shrubland in the low sensitivity simulation. The standard and high sensitivity simulations do predict the Sahara Desert, and a more extensive tropical rainforest than the low sensitivity simulation. Beyond that, all three simulations show a common error when compared with the fossil data. This is the transition from desert/shrubland to savanna and tropical grassland. This occurs a grid square further south than the most northern data occurrence. Southern Africa is particularly poorly modelled, with tropical deciduous broadleaf woodland modelled instead of the tropical savanna, and xerophytic tropical shrubland instead of temperate sclerophyll woodland.

The regional comparison highlights that the reduced skill in the model relates to a number of areas where seasonality is strong and where precipitation is high. A number of areas have been highlighted as being too dry and as a result, generate a drier biome than that which the data indicate existed at the time.

4. Discussion

Palaeoclimate data affords a means of testing climate model experiments in a way that is not possible with future climate projections. The SST data produced by the PRISM group was initially used to drive atmosphere only climate models (i.e. Haywood et al., 2000) and later for data/model comparisons with fully coupled AOGCMs (Dowsett et al., 2011-this issue). Whilst the dataset has developed into an ever more detailed palaeoenvironmental reconstruction for the mpPWP, with a detailed vegetation reconstruction added (Salzmann et al., 2008), the SST dataset is the only quantitative temperature reconstruction for the period. This has led to a situation where the aim of improving mpPWP modelling studies is to generate model simulations which increase the warmth in the higher latitude oceans, so as to tackle the weakness in the model when compared to the data whilst not weakening the areas of good data/model agreement. The results from the ‘PRISM3D MASST minus high sensitivity simulation’ data/model comparison (Fig. 5b), show that a simulation with a higher Charney sensitivity could increase the model skill in comparison with this dataset. However, when the high sensitivity simulation was used to force the BIOME4 model, it produced a vegetation prediction that had less agreement with the palaeobotanical dataset than with the standard simulation. Primarily, this was over land areas such as North America where the high sensitivity simulation was shown in anomaly plots with the standard simulation (Figs. 2a, 3a and 4a) to produce a warmer, drier climate.
The warm anomaly which improved the skill of the model in comparison with the PRISM3D MASST dataset reduced its skill in comparisons with the vegetation dataset. There is an insufficient amount of rainfall for the vegetation patterns in many regions to match the palaeo-data, leading to a prediction of drier climate vegetation compared to the palaeo-data.

The improvement of mPWP model skill is not going to be found through increased temperatures alone. Despite its all-round reduced skill compared to the standard and high sensitivity simulations, the low sensitivity simulation may indicate a way forward. The low sensitivity simulation showed an interesting contrast between the land and the ocean (Fig. 2b). Ocean areas tended to be less sensitive, yet land areas showed general greater sensitivity to the changes in the parameterisations (except at the highest latitudes), causing a warm anomaly. A reversal of this pattern with warm anomaly oceans and little change to the terrestrial areas (in comparison with a mPWP standard simulation) could improve the skill of mPWP models.

Whether or not this combination is possible is not known at present. The full ensemble of 17 members (the three shown here, plus 14 further variations as described in Collins et al., 2010), is the next stage for the Plio-QUMP Project. The usefulness of using two different proxies (SST and vegetation) to test the skill of the modelling simulations has been displayed in these initial results. Although the 17 member ensemble is not likely to include a perfect mPWP model, it will generate a range of models that enables us to quantify the uncertainty in the model predictions and to illustrate (with constraints for skill areas) where the climate model is performing skilfully in comparison with the available proxy data and where it is failing to perform with good skill.

These initial experiments have produced interesting results generating new ideas about both the types of parameters that have been perturbed and the importance of boundary condition uncertainty. The initial ensemble, as with the full ensemble to follow, is only perturbing parameters in the atmospheric component of the HadCM3 model. QUMP projects aim to look at predictive climate change over the coming century, and atmospheric parameters are the only ones that act on such a timescale — Collins et al. (2007), undertook an oceanic perturbation and failed to find significant results over the timescale of the next century. However, in a data/model comparison with SSTs, and over the longer timescales of palaeo-experiments, these oceanic perturbations could become a stronger influence on the climate model simulation.

The data/model comparison of Fig. 5b illustrates that the high sensitivity simulation was unable to achieve all the necessary warmth in the higher latitudes to align data and model results, even as it modelled continental areas that were too warm for the palaeo-vegetation to have existed. Also, all three members of the ensemble showed little variation around ocean gateways, in the tropical Pacific and in areas of upwelling. All these areas showed an over-estimation of warmth by the model in comparison with the proxy data that are available in these regions. These are issues relating to uncertainty in the boundary conditions of the model. Boundary condition uncertainty in the mPWP modelling studies could be an explanation for why the high sensitivity simulation was still unable to generate enough warmth in the North Atlantic to match that indicated by the data. There are two key boundary conditions that could have affected this. The height of the Rocky Mountains was set at 50% of their modern height in the PRISM2 reconstruction, but this value may not be realistic. If they were high the Rocky Mountains would have affected atmospheric circulation around the North Atlantic, which could exert a higher latitude influence (Hill et al., 2011-this issue). The other boundary condition is the ocean bathymetry. There has been detailed research into the bathymetry of the mid-Pliocene ocean, focussing on key tropical gateways such as Indonesia and the Central American Seaway. Recent work has shown that the Central American Seaway was closed during the mPWP (Lunt et al., 2008a) and that the Indonesian gateway was in a modern configuration by this period in the Pliocene (Karas et al., 2009, 2010). One recently investigated region where bathymetry could affect the modelling results is the Greenland–Scotland ridge (Robinson et al., 2011-this issue). The recent work has shown that it is reasonable, on geological grounds, to adjust the height of the ridge for modelling purposes, and that when these changes are included in models there is an increase in high latitude North Atlantic SSTs. Combining this work with the work of the Plio-QUMP ensemble could further reduce errors in the data/model comparison involving the MASST dataset, without causing too much warming on land for the vegetation reconstruction to be degraded.

All published work on QUMP projects to date has been on predictive climate change over the next century, so this study represents the first data/model comparison for a PPE in a warmer than modern palaeoclimate. The two end members used in this simulation have been shown to be statistically valid versions of the HadCM3 model (Collins et al., 2010). They both perform within the range of validation tests that were undertaken by Collins et al. (2010) and for that reason they were considered acceptable simulations for use in the Plio-QUMP Project. It will not be until the full ensemble is completed and analysed that a full understanding of the performance of these end members will be realised. It is important for both predictive QUMP experiments and for the quantifying of mPWP uncertainty that the full ensemble is produced. Only once it has been completed will the full range of potential model shall the results be known. Whilst these end members are the extremes of Charney sensitivity for the HadCM3 QUMP ensemble members, there is no certainty that these represent the end members in range of changes to vegetation and SSTs in the data/model comparisons. It must be noted though that these experiments lack the interaction of earth system feedbacks, such as climatically driven changes in vegetation, which could play a prominent role in the changes in climate between ensemble members.

5. Conclusions

The Quantifying Uncertainty in Model Predictions for the Pliocene (Plio-QUMP) Project offers a tremendous opportunity to address the differences in the skill of different model simulations of the mid-Pliocene Warm Period (mPWP; 3.3 to 3.0 Ma BP) through data/model comparisons and to use this work to quantify uncertainty in model predictions.

The initial results were important indicators as to the direction of the project as they highlighted that it will require a measured and balanced approach to dealing with the weaknesses in previous Pliocene HadCM3 simulations. The high sensitivity simulation produced an improved data/model comparison with the PRISM3D Mean Annual Sea Surface Temperature (MASST) dataset, especially in the North Atlantic and Arctic Oceans. However, it was less skilful in comparison with the palaeobotanical reconstruction than the mPWP standard simulation. The low sensitivity simulation was less skilful in both data/model comparisons than the high sensitivity simulation and the standard simulations.

The high sensitivity simulation performed less skilfully in the vegetation data/model comparison because it warmed both areas over ocean and land by large amounts. Whilst this improved the Pliocene Research Interpretation and Synoptic Mapping (PRISM) 3D MASST comparison, it caused a drying out of areas such as Australia and North America in BIOME4, which reduced the agreement with the palaeobotanical data in some continental regions. Both the high and low sensitivity simulations yielded positives and negatives in the data/model comparisons. It will be a combination of elements from both runs that will form an ensemble member (or members) giving the best results. It is important to consider the possibility that a couple of simulations will produce improved data/
model comparisons bracketing the palaeo-data and that these will be used to quantify the uncertainty. Next, the Plio-QUMP Project will initiate the full 17 member perturbed physics ensemble (PPE) to create the full ensemble of simulations to be compared with data from the PRISM3D MASST dataset and the palaeoecological vegetation dataset of Salzmann et al. (2008).

It is evident that whilst the atmospheric parameter PPEs are a good starting point for the Plio-QUMP Project, the project will not be complete without analysis of other causes of uncertainty, both from perturbing ocean parameters and from analysing the uncertainty created by key boundary conditions on land and in the oceans.

The 'low minus standard' simulation (Fig. 2b) displayed a contrasting temperature pattern between the ocean and land. Unlike our final full PPE, our initial three-experiment ensemble is not sufficiently large to investigate the parameters that cause this contrast. The Plio-QUMP Project will also investigate which parameters are exerting the strongest climatological effects on the model. This is important for understanding the impacts that are made when we perturb the model physics and is vital for understanding and quantifying the uncertainty.

Acknowledgements


