Modelling Pliocene Climate with Perturbed Physics Ensembles

by

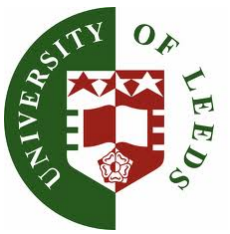
James Owen Pope

BSc (Hons) Environmental Geoscience (University of Edinburgh)

MSc by Research Global Environmental Change (University of Edinburgh)

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy



The University of Leeds

School of Earth and Environment

May 2015

In Memoriam

The Reverend Canon Colin Pope

6th July 1951 - 16th September 2009

"There be of them, that have left a name behind them, that their praises might be reported. And some there be, which have no memorial; who are perished, as though they had never been; and are become as though they had never been born; and their children after them. But these were merciful men, whose righteousness hath not been forgotten. With their seed shall continually remain a good inheritance, and their children are within the covenant. Their seed standeth fast, and their children for their sakes. Their seed shall remain for ever, and their glory shall not be blotted out. Their bodies are buried in peace; but their name liveth for evermore. The people will tell of their wisdom, and the congregation will shew forth their praise."

Ecclesiasticus 44: 8-15

Declaration of Authorship

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 3 is based on the paper submitted in 2011 and published as: Pope, J.O., Collins, M., Haywood, A.M., Dowsett, H.J., Hunter, S.J., Lunt, D.J., Pickering, S.J. & Pound, M.J. 2011. Quantifying Uncertainty in Model Predictions for the Pliocene (Plio-QUMP): Initial Results. Palaeogeography, Palaeoclimatology, Palaeoecology. 309. 128-140. The author produced the text and figures in the paper which are reproduced in this thesis with the exception of Figure 3.10, which was produced by Dr Matthew Pound.

Signed:

_____

Date:

_____

Acknowledgements

Undertaking a PhD has been an immense personal and educational challenge, however I am deeply grateful to have had the opportunity to experience the challenge. First and foremost, I have to thank Professor Alan Haywood, his passion for research is infectious and his guidance and support have ensured I have produced this thesis, I most certainly could not have done it without him. Similarly the support of the rest of the Sellwood Group for Palaeo-Climate has been invaluable. I thank Dr Aisling Dolan for her energy, encouragement and her continuing willingness to answer questions ranging from the scientific to the obvious, Dr Stephen Hunter for his invaluable assistance in teaching and guiding me into using Matlab and his excellent statistics script, Dr Daniel Hill for providing his sharp insights into my work and resulting useful discussion, Dr Ulrich Salzmann and Dr Matthew Pound for their assistance with terrestrial biomes reconstructions, SAT data and the Kappa statistics used as part of my DMC rankings. Thanks also to Dr Julia Tindall for her advice and proof reading of chapters within this thesis. I should apologise to Caroline Prescott, originally a Masters student, whom I introduced to climate modelling and dragged her into this world, therefore I forgive her frequent questions, it has been a privilege to be involved in your development. My thanks also to Fergus Howell for providing useful discussions on the sea ice parameterisations in HadCM3, which led to me delving into that area in this thesis. Finally, a huge thank you to Dr Stephen Pickering, for his patience and perseverance when dealing with my computing related questions, his assistance makes a huge difference to the whole group.

My thanks also to Professor Matthew Collins for introducing me to the world of QUMP and perturbed physics ensembles. His patience and speed of answering my e-mailed questions is greatly appreciated. Additionally, working within the Met Office has been an aspiration of mine since my mid-teens, so thank you for making that happen for a couple of weeks! Similarly, thanks to Dr Daniel Lunt for interesting discussions on the Pliocene, the application of the PPEs within the epoch and his feedback on this thesis. To the PRISM team at the US Geological Survey, I acknowledge the CASE Award that accompanied this project. Additionally the discussions we had with respect to the palaeo-data and its application to the work within this thesis were incredibly useful. On a personal level, Dr Harry Dowsett, Dr Marci Robinson and Danielle Stoll always added enjoyment to any conference we attended together. I am also indebted to you all for your generosity and kindness, when I was stranded at the AGU Fall Meeting in December 2010. Your kindness in that moment of need meant so much to me then, even if Danielle hoodwinked me into cooking dinner and then drank all the wine.

A PhD is a rollercoaster of an experience, one that is made far more enjoyable by the friendship and support of those around you, both within and outside the University. In no particular order, thanks to: Dr Ben Parkes, Dr Amber Leeson, Dr Ryan Hossaini, Dr Mike Hollaway, Dr Eimear Dunne, Dr Sarah Wallace, Dr Jenn Brooke, Dr Cat Scott, Dr David Tupman, Dr Bradley Jemmett-Smith, Dr Chris Wilson, James Witts, Dr Lyndsey Fox, Chris Poole, Tom Fletcher, Palaeo@Leeds, Dr Aidan Farrow, Dr Ruza Ivanovic, Dr Lauren Gregoire, Dr Richard Morris, Dave & Jen Fotheringham, Pip Thompson, Bexley Kemp, Tom Henderson, Magdalena Sarna, Jordan Dias, Dr Tamsin Edwards, New Rover CC, Tobias Dalton, the Rocsoc, Kevin, Myer & Laurrie Butler, Matthew Boyce, Richard Rigby, The ARC1 Support Team, LICS and all my fellow Friends of the Pliocene. Whether it was a pint, a paper I needed, the chance to unwind, sport, ICT support, a refreshing change, a touch of reality or entertainment at conferences, you have all played a role in helping me through the duration of my PhD life. To my mum, sister and grandma, I guess I have finally finished university. I thank you for being the home support and encouragement I have always known was there, despite my consistent attempts to be anywhere but home! You have always encouraged me to push my boundaries, and for this I am eternally grateful.

Finally, I have dedicated this thesis to my dad, who died two weeks before I started. I fondly remember the night I drew a sketch of glacial/inter-glacial cycles on a napkin in a bar in Dublin, and I remember your enthusiasm and pride when we talked about what I was going to undertake. I know you would have been proud whatever had happened, but this thesis is for you and I hope it is fitting of your memory.

Abstract

Uncertainty in model simulations arises due to the construction of the model (structural uncertainty), the representation of sub-grid scale processes (parameter uncertainty) or the input of model boundary conditions. Perturbed physics ensembles (PPEs) produce an ensemble of simulations using a single climate model. A PPE produces different representations of climate by altering the tuning of parameterisations representing processes occurring on sub-grid scales, such as clouds and radiation. A PPE has been produced to investigate model parameter and boundary condition uncertainty for the mid-Pliocene Warm Period (3.264 to 3.025 Ma BP). Through the use of a PPE, 14 versions (13 perturbed members and the Standard version) of the UK Met Office atmosphere-ocean general circulation model HadCM3 were created. The full ensemble was re-run to assess the impact of simultaneously changing physical boundary conditions for orography, ice sheets and vegetation in combination with perturbed physics. Finally the effect of the potential range in reconstructed mid-Pliocene CO_2 was investigated through a sub-ensemble of the PPE.

Using data-model comparisons (DMCs), the ensemble members with higher than the Standard values of Charney sensitivity were better able to simulate the magnitude of high latitude mid-Pliocene warming. The strongest performing ensemble members for the DMCs displayed Charney sensitivities of 4.54°C, 4.62°C and 5.40°C, above the upper bound of the IPCC likely range (1.5 to 4.5°C). However, these warmer members with higher Charney sensitivities weakened the data-model comparison in the tropics. Ensemble members with lower than Standard values of Charney sensitivity, close to the lower bound of the IPCC likely range, better resolved temperature reconstructions in the tropics, but were unable to resolve high latitude warming. It is evident that the PPE is able to achieve the magnitude of mPWP warming but not the spatial distribution of the warming.

The investigation into boundary condition uncertainty using the PPE reveals that the PRISM3D physical boundary conditions lead to improved simulations of the mPWP climate than the PRISM2 boundary conditions. For the range of atmospheric CO_2 concentrations, the results from the sub-ensemble indicate that lower values of CO_2 lead to reduced performance of the PPE members compared to the palaeo-data. The conclusion is that concentrations of CO_2 below 350 ppmv for the mPWP would make simulating high latitude climates very difficult for climate models.

Contents

List of Figures

List of Tables

Abbreviations

ACC	Antarctic Circumpolar Current	ACP	Atlantic Coastal Plain
AGCM	Atmosphere Only General Circulation Model	AMSL	Above Mean Sea Level
anvil	Convective Anvils Shape Factor (PPE Parameter)	AOGCM	Atmosphere-Ocean General Circulation Model
AR4	4th Assessment Report	B	An ensemble member with Charney sensitivity = 2.42°C
BASISM	British Antarctic Survey Ice Sheet Model	BIOME4	Mechanistic Equilibrium Vegetation Model Version 4
BP	Before Present	c_sphlw	Non-spherical Ice Particles Longwave Radiation Properties (PPE Parameter)
c_sphsw	Non-spherical Ice Particles Shortwave Radiation Properties (PPE Parameter)	CAM	NCAR Community Atmosphere Model
canopy	Surface Canopy Energy Exchange (PPE Parameter)	Cape	CAPE Closure (PPE Parameter)
CAS	Central American Seaway	CCSM3	NCAR Community Climate System Model Version 3
CCSM4	NCAR Community Climate System Model Version 4	CH_4	Methane
Charney Sensitivity	Climate response to a doubling of CO_2 over 100 years	Charnoc	Charnock Constant - Roughness lengths and surface fluxes over sea (PPE Parameter)
CLAMP	Climate Leaf Analysis Multivariate Program	CLIMAP	Climate Long-Range Investigation and Mapping
Climate Sensitivity	An alternative name for Charney sensitivity	CPDN	climateprediction.net a PRDC approach to creating a perturbed physics ensemble on UKMO models
CMIP	Coupled Model Intercomparison Project	cnv_rl	Free convective roughness over sea (PPE Parameter)

cnv_upd	Convective Anvils Updraft Factor (PPE Parameter)	CO_2	Carbon Dioxide
CT	Cloud droplet to rain conversion rate (PPE Parameter)	CW_{land}	Cloud droplet to rain conversion rate (PPE Parameter)
Cw_{sea}	Cloud droplet to rain conversion rate (PPE Parameter)	D	An ensemble member with Charney sensitivity = 2.88°C
DMC	Data-Model Comparison	DOT	Deep Ocean Temperature
DSDP	Deep Sea Drilling Project	dyndel	Order of diffusion operator (PPE Parameter)
dyndiff	Diffusion *e*-folding time (PPE Parameter)	EACF	The combined name for both the 'eacfbl' and 'eacftrp' parameterisations
eacfbl	Cloud Fraction at Saturation Boundary layer (PPE Parameter)	eacftrp	Cloud Fraction at Saturation Free Troposphere Value (PPE Parameter)
EAIS	East Antarctic Ice Sheet	ECHAM	Max Plank Institute GCM
EECO	Early Eocene Climatic Optimum (52-50 Ma BP)	EGMAM	Max Plank Institute ESM
EMIC	Earth System Model of Intermediate Complexity	ENSO	El Nino Southern Oscillation
Ent	Entrainment rate coefficient (PPE Parameter)	E-O	Eocene-Oligocene
EOT	Eocene-Oligocene Transition (33.7 Ma BP)	E-P	Evaporation minus Precipitation
EPICA	European Project for Ice Coring in Antarctica	ESS	Earth System Sensitivity
ETOPO5	Earth Topography at a 5 Minute Spacing	F	An ensemble member with Charney sensitivity = 3.75°C
f_rough	Forest Roughness Lengths (PPE Parameter)	FAMOUS	Fast Met Office/UK Universities Simulator
flux_g0	Boundary layer flux Profile (PPE Parameter)	GCM	General Circulation Model
GENESIS	Global Environmental and Ecological Simulation of Interactive Systems	GENIE	Grid Enabled Integrated Earth System Model

GISP	Greenland Ice Sheet Project	GISS	Goddard Institute of Space Studies
Glimmer	GENIE Land Ice Model with Multiply Enabled Regions	GRIP	Greenland Ice Core Project
GrIS	Greenland Ice Sheet	GSR	Greenland-Scotland ridge
gw_lev	Starting Level for Gravity Wave Drag (PPE Parameter)	H	An ensemble member with Charney sensitivity = 3.44°C
HadAM3	Hadley Centre Atmospheric Model Version 3	HadCM2	Hadley Centre Coupled Model Version 2
HadCM3	Hadley Centre Coupled Model Version 3	HadGEM1	Hadley Centre Global Earth System Model Version 1
HadSM3	Hadley Centre Slab-Ocean Model Version 3	High Sensitivity	Initial results ensemble member, equivalent to Q with Charney sensitivity = 7.11°C
I	An ensemble member with Charney sensitivity = 4.40°C	ice_tr	Sea Ice Minimum Temperature
Icesize	Ice Particle Size (PPE Parameter)	ICTP	International Centre for Theoretical Physics
IMC	Intra-Model Comparison	IODP	Integrated Ocean Drilling Program
IPCC	Intergovernmental Panel on Climate Change	ITCZ	Inter-Tropical Convergence Zone
ITF	Indonesian Through Flow	J	An ensemble member with Charney sensitivity = 3.90°C
JUMP	Japan Uncertainty Modelling Project	K	An ensemble member with Charney sensitivity = 4.44°C
k_gwd	Surface Gravity Wavelength (PPE Parameter)	k_lee	Trapped Lee Wave Constant (PPE Parameter)
kg/m^3	Kilograms per Cubic Metre	Kyrs	Thousand years
L	An ensemble member with Charney sensitivity = 4.88°C	lambda	Asymptotic Neutral Mixing Length Parameter (PPE Parameter)
LGM	Last Glacial Maximum (21 Ka BP)	LIG	Last Inter-Glacial (130-114 Ka BP)
Low Sensitivity	Initial results ensemble member with Charney sensitivity = 2.19°C	LR04	Lisiecki-Raymo 2004 Stack

LSM	Land-Sea Mask	M	An ensemble member with Charney sensitivity = 4.54°C
Ma	Million Years Ago	MASST	Mean Annual Sea Surface Temperature
MAT	Modern Analogue Technique	Mg/Ca	Magnesium/Calcium
MinSIA	Minimum Sea Ice Albedo(PPE Parameter)	MIROC	Model for Interdisciplinary Research on Climate
MIS	Marine Isotope Stage	mm/day	Millimetres Per Day
MMCO	mid-Miocene Climatic Optimum (18-16 Ma BP)	MMCT	mid-Miocene Climate Transition (14.8-12.9 Ma BP)
MME	Multi-model ensemble	MOSES	Met Office Surface Exchange Scheme
MPE	Multi-physics ensemble	MPT	mid-Pleistocene Transition (1.25-0.75 Ma BP)
mPWP	mid-Pliocene Warm Period	MSLP	Mean Sea Level Pressure
Myrs	Million years	N	An ensemble member with Charney sensitivity = 4.62°C
N_2O	Nitrous Oxide	NADW	North Atlantic Deep Water
NASA	National Aeronautics and Space Administration	NCAR	National Center for Atmospheric Research
NHG	Northern Hemisphere Glaciation	NorESM	Norwegian Earth System Model
O	An ensemble member with Charney sensitivity = 4.79°C	ODP	Ocean Drilling Program
oi_diff	Ocean Ice Diffusion (PPE Parameter)	P	An ensemble member with Charney sensitivity = 5.40°C
Pa	Pascals	PETM	Palaeocene-Eocene Thermal Maximum (~55.5 Ma BP)
PFT	Plant Functional Type	Pg	Petagram
PI	Pre-Industrial	Plio	Pliocene
PlioMIP	Pliocene Modelling Intercomparison Project	Plio-QUMP	Quantifying Uncertainty in Model Predictions for the Pliocene
PlisMIP	Pliocene Ice Sheet Modelling Intercomparison Project	PMIP	Palaeoclimate Modelling Intercomparison Project
ppbv	Parts Per Billion by Volume	PPE	Perturbed Physics Ensemble
ppmv	Parts Per Million by Volume	PRDC	Public Resource Distributed Computing

PRISM	Pliocene Reconstruction, Interpretation and Synoptic Mapping	PRISM1	1st iteration of the PRISM palaeo-environmental reconstruction
PRISM2	2nd iteration of the PRISM palaeo-environmental reconstruction	PRISM3D	3rd iteration of the PRISM palaeo-environmental reconstruction
Q	An ensemble member with Charney sensitivity = 7.11°C	QUMP	Quantifying Uncertainty in Model Predictions
RHCrit	Threshold of relative humidity for cloud formation (PPE Parameter)	rhparam	Flow dependant RHcrit (PPE Parameter)
RMSE	Root Mean Square Error	s_sphlw	Non-spherical Ice Particles Longwave Radiation Properties
s_sphsw	Non-spherical Ice Particles Shortwave Radiation Properties (PPE Parameter)	SAT	Surface Air Temperature
SH	Southern Hemisphere	SMC	Soil Moisture Content
SME	Single Model Ensemble	soillev	Number of Soil Levels accessed for Evapotranspiration (PPE Parameter)
SRES	Special Report on Emissions Scenarios	SST	Sea Surface Temperature
Standard	The unperturbed version of HadCM3 with a Charney sensitivity of 3.30°C	sto_res	Dependence of Stomatal Conductance on CO_2 (PPE Parameter)
sw_absn	Shortwave Water Vapour Continuum Absorption (PPE Parameter)	THC	Thermohaline Circulation
TOA	Top of the Atmosphere	U^k_{37}	Ratio of saturated to unsaturated carbon bonds in Alkenones
UKCP09	UK Climate Projections 2009	UKMO	United Kingdom Met Office
USGS	United States Geological Survey	vertcld	Vertical Gradient of Cloud Water (PPE Parameter)
VF1	Ice Fall Speed (PPE Parameter)	Vostock	Antarctic Ice Core Site
WAIS	West Antarctic Ice Sheet	Wm^{-2}	Watts Per Square Metre
$\delta^{13}C$	Change in the ratio of stable carbon isotopes (^{13}C & ^{14}C) from a laboratory standard	$\delta^{18}O$	Change in the ratio of stable oxygen isotopes (^{16}O & ^{18}O) from a laboratory standard

Chapter 1: Introduction

Evidence that humankind is affecting the climate system is now overwhelming (IPCC, 2007). The only method for producing projections of future climate are dependant upon computer models of the climate system. These projections are limited by an incomplete knowledge of the uncertainty in climate models. Strategies for dealing with the uncertainty within the model simulations focus on creating ensembles of simulations producing a range of model results. The ensembles are developed using the observational period bringing them up to the point where the model becomes predictive (Johns *et al.*, 2003). The ensemble members skill at reproducing the climate for the observational period is tested, and if the member is accurate, it can continue as a member of the ensemble and be run for the predictive simulations (IPCC, 2007). The limitation of this method is that the ensemble 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)). Whereas 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. Palaeoclimate modelling offers an alternative for testing model simulations and model ensembles and a method for quantifying model uncertainties. By selecting an appropriate time period to study, it is possible to test the model skill at replicating radically different climate states.

The thesis aims to investigate the role of parameter uncertainty on data-model mismatches for a warmer than modern climate from the geological record. The aim will be achieved through the use of a perturbed physics ensemble (PPE). The PPE will then be utilised for assessing the role of boundary condition uncertainty on data-model mismatches through adjusting both representations of the Earth's surface and atmospheric greenhouse gas concentrations. The thesis will outline the development of these investigations and the methods for modelling and analysing results generated during the course of the study.

Chapter 1 introduces the Pliocene epoch in relation to the climate of the Cenozoic Era (the last 65 million years of the geological record) and outlines why the Pliocene is a useful epoch for palaeoclimate investigation. It will then review the work on modelling the Pliocene, focusing mainly on the recent work undertaken in the last 10 to 15 years, but also includes a brief review of early Pliocene climate modelling. The review will highlight the areas of uncertainty in Pliocene modelling to date which represents areas where this ensemble could improve Pliocene simulations. It will also review the

methods used to reconstruct the Pliocene through geology, geochemistry and micropalaeontology, focussing on the work of the Pliocene Research Interpretations and Synoptic Mapping (PRISM) group at the US Geological Survey over the last twenty years. The other aspect to this review is modelling uncertainty, which occurs in a variety of forms. The second section of Chapter 1 focusses on discussing the forms of model uncertainty and methodologies utilised to attempt to reduce or quantify the uncertainty. The section will end with a discussion of previous PPEs and the different methodologies used to create the ensembles. Finally there will be an outline for the rest of the thesis.

1.1- Key Features of the Earth's Climate System

The Earth's climate system represents the interaction of incoming energy with the atmosphere, ocean and land surface components of the Earth. The dominant form of energy into the climate system is solar forcing - the energy from the sun (Beer et al., 2000). Two descriptions of the solar forcing are commonly used: total solar irradiance (TSI) and solar insolation (Kohler et al., 2010; Kopp & Lean, 2011). TSI refers to the solar energy flux at the top of the atmosphere for a square metre orientated towards the sun at a distance of 1 AU (Beer et al., 2000). Solar insolation refers to the solar energy flux received at the Earth's surface and its magnitude is governed by the latitude of the site and the season during which it is measured (Beer et al., 2000). Earth system components can all modify the climatic response to the solar forcing, a forcing that itself is modified on various timescales by changes in the magnitude of solar output and the Earth's orbit around the Sun (Berger, 1978; Frohlich, 2006; Kohler et al., 2010). Solar insolation is strongly affected by the orbital parameters, whilst TSI is modified by changes in solar output and alterations in the eccentricity orbital component (the distance between the Earth and the Sun - Beer et al., 2000). Even small variations in solar forcing are able to cause global and regional climate changes (Lean & Rind, 2009; Kopp & Lean, 2011).

The magnitude of the forcing on the climate system is expressed as the radiative forcing, a perturbation to the energy budget of the Earth's climate system from an external source (Collins, W.D. et al., 2006). Radiative forcing is defined as the net change in TSI at the top of the atmosphere. The net change is calculated as the total incoming shortwave radiation minus reflected shortwave radiation and outgoing longwave radiation. It is expressed as a rate of change for a unit area, measured as Watts per metre squared (Wm^{-2} – IPCC, 2001; 2007). The climate system responds to a change in radiative forcing through an alteration (such as warming) of earth system

components (such as surface temperature – Soden et al., 2008). Radiative forcing is affected by variation in solar output, land surface changes (affecting planetary albedo), changes to atmospheric greenhouse gas concentrations and resulting feedbacks from within the climate system (Collins, W.D. et al., 2006; Kohler et al., 2010). The changes in land surface conditions and greenhouse gases can be caused by natural or anthropogenic changes (Foley et al., 1994; Matthews et al., 2004). In palaeoclimate studies, natural variations in greenhouse gas or land surface forcings are key drivers of modifications to radiative forcing, along with changes in the orbital forcing (Kutzbach & Liu, 1997; Petit et al., 1999).

The response of global temperatures to changes in radiative forcing is determined by the sensitivity of the climate system, known as equilibrium climate sensitivity or Charney sensitivity. Charney sensitivity is defined as the equilibrium (or steady state) temperature increase for a doubling of atmospheric CO_2 (Charney, 1979; PALAEOSENS, 2012). This has been estimated by climate models to be equivalent to a radiative forcing increase of ~3.5 Wm^{-2} (Raper et al., 2002). However, until the climate system reaches equilibrium, a smaller increase in surface temperatures is observed than that which would be expected for a given forcing (Gregory & Forster, 2008). A significant reason for this is the role of the ocean in the climate response to changes in radiative forcing. Of the warming caused by an increase in radiative forcing, 90% is stored in the oceans allowing only 10% of the warming to interact with the atmosphere and land surface components (IPCC, 2007; Loeb et al., 2012; Balmaseda et al., 2013). Since the 1960's the oceans have accumulated 20 times the heat of the atmosphere (Levitus et al., 2005; IPCC, 2007). However, the ocean response to changes in radiative forcing is slow. A $1m^2$ water column to 2500m depth in the ocean requires 300 years to be warmed by 1°C for a radiative forcing of 1 Wm^{-2}. The heat flux to the oceans is represented within climate models, and is modelled as increasing as Charney sensitivity increases (Raper et al., 2002). Whilst increased Charney sensitivity indicates higher equilibrium warming, the rate of that warming is tempered by the increased ocean heat flux.

The transient climate response (TCR) was devised to estimate the magnitude of the response of the climate system to a forcing, before the system reaches equilibrium. TCR is the modelled response for a doubling of CO_2, with CO_2 increasing at 1% year (Gregory & Forster, 2008). TCR is smaller than the Charney sensitivity because it does not allow for the effect of ocean warming on the climate (Knutti et al., 2006; IPCC, 2007). Additionally, TCR is easier to constrain at higher values than Charney sensitivity, owing to a non-linear relationship between TCR and Charney sensitivity

(Knutti et al., 2006). Within the palaeoclimate community there is also the concept of Earth System Sensitivity (ESS - Lunt et al., 2010) and its relationship to Charney sensitivity. Earth System Sensitivity includes the temperature related response of vegetation, ice sheet and oceans to changes in radiative forcing. As an example, large ice sheets respond slowly to increases in global temperature, with estimates that it may take several millennia for ice sheets to respond to anthropogenic warming (Alley et al., 2005). On this timescale, the climate system will have reached radiative equilibrium with the change in forcing. However, the longer term responses such as ice sheets will lead to further changes in the climate (Lunt et al., 2010). ESS therefore represents the extended sensitivity of the climate system to any perturbation in radiative forcing. ESS values are larger than the Charney sensitivity values by a factor of ~1.5 for models in the Pliocene Modelling Intercomparison Project (PlioMIP) multi-model ensemble (Haywood et al., 2013a).

However, fully coupled atmosphere-ocean general circulation models (AOGCMs) do not explicitly state a Charney sensitivity, ESS or TCR setting to control the warming within the simulation. The sensitivity of a simulation is the result of the parameterisations made within the model and the tuning of model parameters. Climate models represent the transfer of energy through the various components using mathematical formulae on the grid scale and utilise parameterisations to represent them on sub-grid scales (Murphy et al., 2004). These parameters, the methodology for choosing them and the tuning values selected, lead to uncertainty in climate model output. The uncertainty that arises is discussed in Section 1.4 and represents they cornerstone of work contained within this thesis.

1.2. The Pliocene Epoch of the Cenozoic Era.

The Pliocene epoch (5.3 to 2.6 Ma BP) represents the last period of sustained warmth before the emergence of a cooler climate characterised by the onset of glacial/interglacial cycles during the Pleistocene (Brierley & Federov, 2010). Based on The Geologic Time Scale 2012 (Gradstein et al., 2012), the Pliocene is the second and final epoch of the Neogene period, one of three periods of the Cenozoic era.

1.2.1. The Cenozoic (65 Ma BP to Present)

The Cenozoic is divided into three periods, the Paleogene (66.0 to 23.0 Ma BP) which contains the Paleocene, Eocene and Oligocene epochs, the Neogene (23.0 to 2.6 Ma BP)

containing the Miocene and Pliocene epochs and the Quarternary (2.6 Ma BP to present) containing the Pleistocene and Holocene epochs. The Cenozoic saw six major climatic shifts including:

- Periods of sudden warming or sustained warmth
    - Paleocene-Eocene Thermal Maximum (PETM - ~56.0 Ma BP - Zachos et al., 2001; 2005; Pagani et al., 2006a,b; Wiejers et al., 2007)
    - Early Eocene Climatic Optimum (EECO - 52 to 50 Ma BP - Lowenstein & Demicco, 2006, Sagoo et al., 2013)
    - Mid-Miocene Climatic Optimum (MMCO - 18 to 16 Ma BP - Flower & Kennett, 1994)
- Periods of sudden cooling
    - Eocene-Oligocene Climate Transition (EOT - 33.7 Ma BP – Lear et al., 2008; Liu et al., 2009)
    - Middle Miocene Climate Transition (MMCT - 14.8 to 12.9 Ma BP - Flower & Kennett, 1994; Zachos et al., 2001; Micheels et al., 2007).
- Orbitally forced Glacial/Interglacial Cycles (Pleistocene 2.6 Ma to 10 Ka BP – Clark et al., 1999)

These events are well represented in terms of magnitude and timing in the Zachos compilation curve of $\delta^{18}O$ benthic oxygen isotopes (Zachos et al., 2001; 2008).

The Paleogene world began free of permanent ice caps and global temperatures elevated by 8 to 12°C compared to the present day (Zachos et al., 2001; 2008). It was punctuated with periods of rapid warming such as the PETM and more stable warm climate periods such as the EECO. The warming at the PETM was mainly caused by the massive release over 10 Kyrs of greenhouse gases (3000 to 6800 Pg of carbon - Panchuck et al., 2008; Zeebe et al., 2009). The source of which was most likely from methane hydrates (Zachos et al., 2001), indicated by a negative carbon isotope excursion in globally distributed records (Kennett & Stott, 1991; Bowen et al., 2004; Pagani et al., 2006a,b; Sluijs et al., 2007; Dunkley-Jones et al., 2010).

At 33.7 Ma, a rapid large cooling transition took place at the EOT and the first evidence for a sustained ice sheet on East Antarctica is detected (Zachos et al., 1996; Lear et al., 2000; 2008; DeConto & Pollard, 2003). A possible driver was the opening of the Drake passage and Tasman ocean gateway at this time triggering the development of the Antarctic Circumpolar Current (ACC), which isolates Antarctica and the Southern Ocean from warmer sub-tropical waters, allowing the development of a cooler Antarctic

climate (Kennett, 1977; Katz et al., 2011). Based on modelling studies supported by the data from a boron isotope record (Pearson et al., 2009), the cooling was also driven through atmospheric CO_2 declining through a 750 ppmv threshold enabling Antarctic ice to grow (DeConto & Pollard, 2003; DeConto et al., 2008).

The climate stabilised through the Oligocene and entering the Miocene epoch a period of warming occurred at the MMCO (approximately +6°C with a weakened equator to pole gradient - Flower & Kennett, 1994, Bruch et al., 2007) with a permanent but not extensive ice sheet on Antarctica (Flower & Kennett, 1994; Zachos et al., 2001). The Miocene climate then cooled through the MMCT before stabilising in the Pliocene epoch which saw East Antarctic ice sheets at 50% of their modern size and the development of ice on Greenland (approximately 33% of present size - Dowsett et al., 1999; 2005).

The causes of warming and cooling in the Miocene could be due to several Earth system drivers. CO_2 has been shown to be a driver of many warm climate changes (Shellito et al., 2003). Kurschner et al. (2008) showed results from stomatal indices that indicated that Miocene climate and CO_2 displayed long term coupling. However it is not easily reconstructed in the Miocene with values from alkenones suggesting pre-industrial level concentrations throughout the Miocene for both the MMCO or MMCT (Pagani et al., 2005; LaRiviere et al., 2012). The research indicates that temperature and CO_2 was decoupled during the Miocene (Pearson & Palmer, 2000) and that the climate changes were driven by changes in ocean gateways or orography. Gateways such as the Central American Seaway and the Tethys Ocean affect ocean heat flow (Flower & Kennett, 1994; Bice et al., 2000; Von der Heydt & Djikstra, 2006; Herold et al., 2008; 2012) while orographic change through the dynamic uplift of the Himalayas & Tibetan Plateau (Currie et al., 2005), the central Andes (Gregory-Wodzicki, 2000) the Alps (Kuhlemann et al., 2006) or the North American cordillera (Foster et al., 2010) during the Miocene, caused changes to atmospheric circulation, heat distribution and precipitation (Quade et al., 1989; Spicer et al., 1990; Harrison & Yin, 2004; Harris, 2006; Jimenez-Moreno et al., 2008; Foster et al., 2010).

These changes would have led to impacts on Miocene climate and vegetation (Dutton & Barron, 1997; Pound et al., 2011). However, other CO_2 reconstructions such as data collected from stomatal indices or pedogenic carbonate display much greater variation in Miocene CO_2 (Kurschner et al., 1996; 2008; Ekart et al., 1999). Modelling by You et al. (2009) and van de Wal et al. (2011) suggests that CO_2 was 450 to 580 ppmv during the Miocene. It is most likely that the Miocene climate was the response to a combination of changes: CO_2, orography and ocean circulation along with feedbacks

from vegetation that in some regions may have been greater than the CO_2 forcing (Henrot et al., 2010). These variables mean that it is hard to ascertain the exact causes of climate change during the Miocene making it hard to investigate the climate system due to uncertainty in the drivers of the climate change.

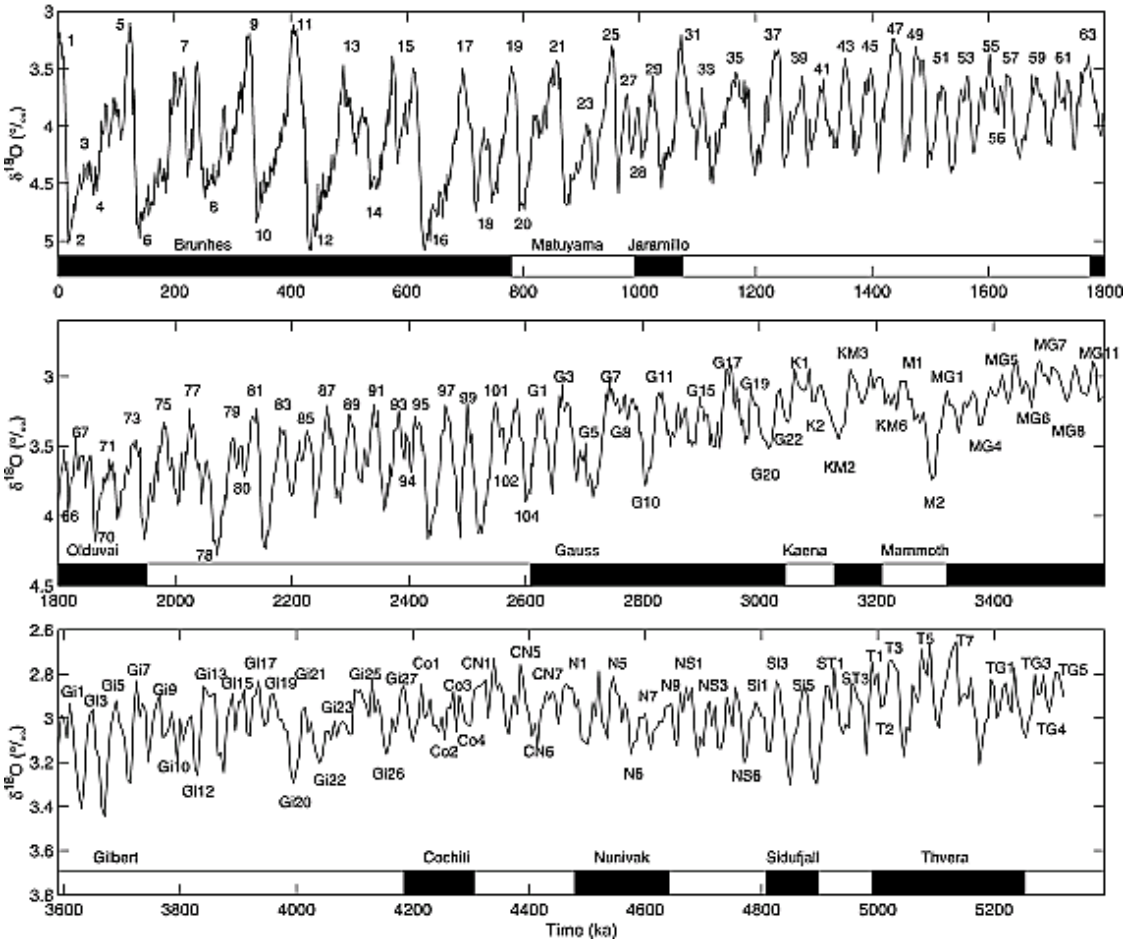


Figure 1.1. The Lisiecki-Raymo (LR04) stack of 57 benthic $\delta^{18}O$ records over the last 5.2 million years (Lisiecki & Raymo, 2005). Benthic $\delta^{18}O$ records represent changes in bottom water temperatures and global ice volume with larger values representing cooler temperatures and/or increased ice volume. The transition from Pliocene warm climates to orbitally forced glacial/interglacial cycles is clearly displayed between 2.8 and 1.8 Ma BP.

The Pliocene epoch is the period of geological time from 5.3 to 2.6 Ma BP (Gradstein & Ogg, 2009), prior to the Pleistocene epoch (Dowsett et al., 2005; Lawrence et al., 2006; Lunt et al., 2009). A wide range of data and modelling studies have shown that the Pliocene had a climate approximately 3°C warmer than the present day (i.e. Dowsett et al., 1996; Sloan et al., 1996; Haywood et al., 2000a; 2009a; 2013a; Haywood & Valdes 2004; Lunt et al., 2008a,b). The warming was concentrated at higher latitudes with palaeo-data suggesting that Pliocene tropical oceans were a similar temperature to the present day. Enhanced poleward heat transport was the proposed method for the distribution of heat to higher latitudes which were up to 14°C warmer than the present

day (Dowsett & Poore, 1991; Dowsett et al., 1992; 1994; Dowsett, 2007; Dowsett & Robinson, 2006). The cause of this climate change was addressed through a novel modelling strategy to produce factorisations of the components of mPWP warming, which attributed most warming to CO_2 (Lunt et al., 2012a) with elevated concentrations similar to the present day levels (i.e. Pagani et al., 2010; Seki et al., 2010).

The climate of the Pleistocene (2.6 Ma BP to 11,500 years BP) is dominated by orbital forcing. The Pliocene-Pleistocene boundary was redrawn in 2009, as the emergence of climate dominated by these orbitally forced glacial cycles were detected beginning in late Pliocene climate reconstructions (Figure 1.1). As a result the last 800 Kyrs (2.6 to 1.8 Ma BP) were redefined from the Pliocene to the Pleistocene (Gradstein & Ogg, 2009). Pleistocene orbital forcing was initially dominated by the 41 Kyr cycle, but after the mid-Pleistocene Transition (MPT- 1.25 to 0.75 Ma BP), the 100 Kyr cycle began to force climate (Clark et al., 2006). The last 800,000 years of Pleistocene climate are recorded in the ice cores such as Vostok & EPICA on Antarctica (Wolff et al., 2010). These cores hold information on the atmospheric composition, temperature, precipitation, aerosol deposition and the identity of the aerosol sources (volcanic, cosmogenic or anthropogenic – Petit et al., 1999). Research into ice cores has shown a correlation between greenhouse gas concentration and temperature changes (Petit et al., 1999), with the glacial/inter-glacial events being triggered by changes in orbital forcing and amplified by the resulting climate feedbacks from greenhouse gases (Ruddiman, 2003).

An important consideration for the changes at the Plio-Pleistocene boundary is the cause of the change in the influence of orbital forcing, which dominated the climate during the Pleistocene, which it had not done during the Pliocene despite evidence that the forcing was apparent (Lisiecki & Raymo, 2005 – Figure 1.1). The suggested driver is a decrease in CO_2 from ~400 ppmv to 280 ppmv allowed the formation of large Northern Hemisphere ice sheets and their resulting ability to amplify climate changes, causing the glacial/inter-glacial cycles (Lunt et al., 2008a,b). It is suggested by Ruddiman (2003) that ice sheets should have begun to reform approximately 10,000 years ago, but that early human CO_2 and CH_4 emissions stalled the growth and avoided a glacial climate. The theory fits with the hypothesis that greenhouse gases act as threshold for ice sheet formation and orbital forcing driving of climate.

Whether reconstructing palaeoclimate environments from modelling or data, there are several important considerations when choosing a time period to reconstruct. Firstly,

changes to continental configurations can have a significant impact on the flow of atmospheric and oceanic currents which can radically affect the climate system (Foster et al., 2010; Hill et al., 2011). As a result there are no analogues between the Palaeocene or Eocene and the anthropogenic warming of the present day (Haywood et al., 2011b). Secondly, when reconstructing past environments, a vital tool for validating coupled atmosphere-ocean model simulations or for forcing atmosphere only simulations is palaeo-data. The data is often produced from sediment cores, drilled from the ocean floor. However, the older record is sparser, and the reduced amount of available data for periods such as the EECO, requires solutions such as producing *"background"* records from before the event to increase available testing of the model (Lunt et al., 2012b). With a young Atlantic ocean and most of the Pacific below the calcium compensation depth (Palike et al., 2012), there are fewer Eocene sediments which contain calcite shelled organisms available for producing reconstructions of climate, leaving records with hiatuses (Coxall et al., 2005). Where proxy data is available, there are considerations regarding any post-depositional processes with diagenesis potentially altering the signals produced from analysis of the proxies. If unknown or not accounted for this could lead to a misleading reconstruction (Pearson et al., 2001; Williams et al., 2005; Lunt et al., 2008c). Finally, any proxies which start from the early geological principle of uniformitarianism (Hutton, 1788; Lyell, 1830) using techniques to develop temperature data from species such as the modern analogue technique (MAT – Hutson, 1979; Dowsett et al., 1999), become harder in deeper time periods. The older the sediments, the more likely a species is extinct or evolution has altered the habits of the species being analysed and therefore the inferred data on the climate preferences of the palaeo-species (Murray, 2001; Des Marias & Juenger, 2010).

Ideally, a palaeoclimate modelling study requires a prolonged warmer than present climate with a near modern palaeogeography with access to extensive palaeo-data and where the cause of the warming is identified and well understood. The only period of time which fits this requirement, is the Pliocene, and specifically a sub-section, the mid-Pliocene Warm Period.

1.2.2. mid-Pliocene Warm Period (mPWP)

The mid-Pliocene Warm Period (mPWP) was the last climatically stable period directly prior to the cyclical Pleistocene glacial/interglacial climate and as such is of significant interest to palaeo-climatologists. Lasting from 3.264 to 3.025 Ma BP, it is a period of sustained global mean annual surface temperatures approximately 3°C above the pre-

industrial levels with a very similar palaeo-geography to the present day (Haywood et al., 2000a; Haywood & Valdes, 2004; Haywood et al., 2009a; Dowsett et al., 2010b). The mPWP is the most parsimonious warm climate epoch for investigating the possible effects of anthropogenically forced climate change in the 21st century (Haywood et al, 2002b; 2009a,b; Haywood & Valdes, 2006; Dowsett, 2007; IPCC, 2007; Dowsett et al., 2009b; Lunt et al., 2009).

The Pliocene has other advantages that make it particularly suitable for study in this way. Firstly there is an extensive palaeo-environmental dataset available for use in producing boundary conditions and datasets for model testing, made available by the United States Geological Survey (USGS) Pliocene Research Interpretation and Synoptic Mapping (PRISM) group. Recently released in its 4th iteration (PRISM3D – Dowsett et al., 2005; 2009a; 2010a; Dowsett, 2007; Hill et al., 2007; Chandler et al., 2008; Salzmann et al., 2008), this dataset compiles the most comprehensive data reconstruction for sea surface temperatures (SSTs), deep ocean temperatures, sea level, vegetation, orography and ice sheets in a climate system as warm as that projected for the late 21st century. Secondly the Pliocene's relevance for study comes from it is level of atmospheric CO_2, which was elevated compared to the pre-industrial and was the most likely reason for the temperature of the Pliocene. Estimates of Pliocene atmospheric CO_2 range from 300 to 425 ppmv (Raymo et al., 1996; Pagani et al., 2010; Seki et al., 2010), despite the uncertainty in the different proxy methods we can be confident that CO_2 was higher in the Pliocene than during the pre-industrial (see Section 1.4.4).

The combination of warmer climate and availability of palaeo-data, makes the mPWP a fascinating test bed for palaeoclimate modelling experiments with the potential for developing both models and palaeo-environmental reconstructions.

1.3 – Modelling the Pliocene

1.3.1. Why Model the Pliocene?

The mPWP is also often described as an equilibrium climate (i.e. IPCC, 2007; Pagani et al., 2010), a time when the climate system has fully responded to the perturbations in radiative forcing, which within the mPWP reflects both the response to Charney and Earth System Sensitivities (see Section 1.1). Therefore, the mPWP offers an opportunity to understand the response of the Earth's climate system to a perturbation in its radiative forcing, akin to the impacts of present climate change (IPCC, 2007).

Within the mPWP the change in radiative forcing (compared to the pre-industrial) is largely the response to an increase in concentrations of atmospheric CO_2, with some influences from changes in the land surface & orography (Haywood & Valdes, 2004; Lunt et al., 2012a). Haywood & Valdes (2004), determined that for a Pliocene minus pre-industrial mean annual temperature anomaly of 3.1°C, that 1.9 Wm^{-2} radiative forcing change was caused by CO_2 , with surface albedo changes (2.3 Wm^{-2}) and cloud cover (1.8 Wm^{-2}) the other causes of the radiative forcing change. Lunt et al. (2012a), utilising increased computer power to run sensitivity studies, applied a factorisation approach, determined that Pliocene mean annual warming was dominated by was CO_2 (36 to 61%). There were also strong contributions from orography (0 to 26%), vegetation (21 to 27%) and ice sheets (9 to 13%).

Micropalaeontological studies (Dowsett & Poore, 1991) identified that in the North Atlantic, the Pliocene was significantly warmer than the modern day and the last interglacial period. In conjunction with data produced by the PRISM groups early Northern Hemisphere reconstructions (Dowsett et al., 1994), Chandler et al. (1994) were able to run the NASA Goddard Institute for Space Studies (GISS) Atmosphere only General Circulation Model (AGCM) for the Pliocene. They modelled anomalies to the present day control simulation of 8 to 12°C warmer in the Pliocene Arctic, while maintaining close to modern temperatures at the equator. The work developed into the first global reconstruction from the PRISM group (see Section 1.3.3 for iterations and developments of this work), which was used to run the US National Center for Atmospheric Research (NCAR) GENESIS AGCM to produce the first global simulation for the Pliocene (Sloan et al., 1996). The work reinforced the findings of early modelling and the PRISM groups early reconstructions, reconstructing the Pliocene world as approximately 3.6°C warmer than the present day control simulations.

However, as shown in Section 1.2, the Cenozoic era provides many examples of periods of time that could be studied and modelled for understanding the climate system, with either warmer sustained stable climates or transient climate states which are closer in form to the effects of anthropogenic warming (Haywood et al., 2011b). The Pliocene presents the *'best case'* for modelling climates with the intention of improving our understanding of the future climate and climate models. The other sustained stable warm climates such as the MMCO or EECO or transient climates such as the PETM are sufficiently further back in geological time for the effects of plate tectonics to have made noticeable changes to the orography and bathymetry. The closure of the Central American Seaway through the Miocene into the early Pliocene caused the reduction in and the eventual ceasing of two way water flow between the Pacific and the Atlantic,

which aided the formation of North Atlantic Deep Water (NADW) and thermohaline circulation (THC – Lutz, 2011). Therefore, periods of transient palaeoclimate change are lacking a representation of a key modern climatological feature. Additionally, further back in time the spatial and temporal availability of data becomes increasingly sparse, which impedes both the creation of boundary conditions for models and data for testing the model against (Huber & Caballero, 2011; Lunt et al., 2012b). While the glacial/interglacial cycles of the past 2.6 Myrs, and especially the last 800 Kyrs, have greater data coverage than the Pliocene and a transient climate, the climate change is not as extreme as the present anthropogenic events (Haywood et al., 2009a).

The Pliocene is recent enough to have a near modern orography and land-sea mask, as well as having a climate warmer than the modern, similar to that which is predicted by the Intergovernmental Panel on Climate Change (IPCC) as being "*very likely"* by 2100 (IPCC, 2007). It also has the most extensive global data coverage through the PRISM palaeo-environmental reconstructions of any warm period in the geological record.

1.3.2. Recent Pliocene Climate Modelling

The development in the latest versions of non-flux adjusted fully coupled atmosphere ocean general circulation models (AOGCMs i.e. HadCM3, CCSM3) around the year 2000, coupled with increased certainty about the effects of greenhouse gases on the climate, saw an increase in palaeoclimate modelling and modelling of the mPWP. The PRISM2 boundary conditions, first in the AGCM HadAM3, then in the slab ocean version HadSM3 and the AOGCM HadCM3 were used to model the Pliocene climate. Haywood et al. (2000a,b; 2001; 2002a,b,c) and Haywood & Valdes (2004; 2006) investigated global and regional climate changes during the mPWP and used AOGCM simulations to force an offline vegetation model (BIOME4).

Developments followed this work such as Lunt et al. (2008a,b) coupling mPWP simulations in the HadCM3 AOGCM to an offline ice sheet model. Glimmer v1.04, was used to investigate Pliocene ice sheets and how different boundary condition changes (CO_2, ocean gateways, orography) could affect the onset of Northern Hemisphere glaciation (NHG) at the Plio-Pleistocene boundary (Lunt et al., 2008a,b). Climate models have also been used in conjunction with offline models such as BIOME4, or the British Antarctic Survey Ice Sheet Model (BASISM) to develop the PRISM dataset for ice sheets (Hill et al., 2007) and vegetation (Salzmann et al., 2008; 2009). Work has also been undertaken studying known annual variability in the climate system such as the El Nino Southern Oscillation (ENSO – Haywood et al., 2007a; Bonham et al., 2009;

Scroxton et al., 2011), to investigate how such systems respond in a warmer world. Most recently, Lunt et al. (2010) investigated the relationship between Charney sensitivity and Earth System Sensitivity (ESS) during the Pliocene (Section 1.1). Lunt et al. (2012), investigating the causes of Pliocene warmth produced factorisation analysis of Pliocene warming (Section 1.3.1) and also using an energy balance analysis, they determined the relative regional influence of greenhouse gas, orographic, vegetation and ice sheet changes.

For example, Lunt et al., (2012) determined ice sheet changes were the dominant form of warming in the high southern latitudes, whilst orography and CO_2 dominated Northern Hemisphere polar amplification. There were also Pliocene changes in radiative forcing from the effect of changes in the orbital forcing. Performing similar analyses on the PlioMIP MME, Hill et al. (2014) determined that the trends across the ensemble of models were consistent, but with varying degrees of influence. Tropical warming in the mPWP was dominated by greenhouse gas forcings. However, warming was enhanced by cloud influences on planetary albedo with varying degrees of impact across the ensemble members (Hill et al., 2014). However, at higher latitudes, Pliocene warming, whilst still influenced by the range of energy balance components, the dominant forcing was from the clear sky albedo, although there was a slight offset of this warming from cloud albedo. These forcings are strongly linked to the distribution of ice sheets and vegetation, boundary condition forcings, highlighting the importance of the specified boundary conditions.

The majority of this work has been based on simulations conducted on a single model and the results from this one model simulation or set of simulations used to look at the climate impacts being investigated. The Pliocene Model Intercomparison Project (PlioMIP), brings together 14 modelling groups worldwide to address this, running Pliocene simulations based on the same initial and boundary conditions (Haywood et al., 2010, 2011a). The project allows the study of several aspects of the mPWP climate using a range of structurally different models in a multi-model ensemble (MME), similar to projects such as the Coupled Model Intercomparison Project (CMIP - Meehl et al., 2000) and the Palaeoclimate Model Intercomparison Project (PMIP - Braconnot et al., 2007). A sub-study from the PlioMIP project, the Pliocene Ice Sheet Model Intercomparison Project (PlisMIP - Dolan et al. 2012) aims to investigate the responses of ice sheet models to reconstruct Pliocene ice sheets (see Section 1.3.3 *iv*).

Finally, increases in computing power have allowed the early development of transient simulations through the Pliocene using Earth system models of intermediate

complexity (EMIC). The CLIMBER-2 EMIC (Petoukhov et al., 2000; Ganopolski et al., 2001) was used to investigate the change in Pliocene climate due to CO_2 forcing and response changes in vegetation and ice sheets through transient simulations (Ganopolski et al., 2011; Willeit et al., 2013). The results indicated that global mean annual Pliocene warming (compared to pre-industrial) varied between 1.9 and 2.8°C due to changes in orbital forcing and resulting feedbacks involving ice sheets and vegetation. The 41 kyr (obliquity) Milankovitch orbital cycle is the dominant cycle within the mPWP palaeo-archives (i.e. the Lisiecki-Raymo stack, Figure 1.1), but the effects on solar insolation and resulting feedbacks on mPWP climate are non-trivial (Willeit et al., 2013). The developing ability to reconstruct mPWP climate with orbital resolution and to also transiently model across orbital cycles has led to the development of the first Pliocene time slice (Haywood et al., 2013b). The time slice represents a paradigm shift in the methodology for modelling the Pliocene in the future (see Section 1.3.3 *vii*). By selecting a time slice, environmental factors that change on orbital timescales, such as ice sheets and vegetation, affects on the palaeo-data are reduced, allowing a better comparison between the palaeodata and the model simulations (Haywood et al., 2013b – see Section 1.3.3 *vii*).

1.3.3. The Evolution of the PRISM Palaeo-Environmental Reconstruction

i) PRISM Overview

The PRISM project was created with two primary goals. First to identify the causes and variability in the mPWP climate and the second was to create a dataset that could be used by climate modelling groups to investigate the mechanisms of the climate change (Dowsett et al., 1999).

Following the modern day, Last Interglaciation (LIG) and Pliocene reconstruction in Dowsett & Poore (1991) the PRISM group was created and released a Northern Hemisphere only reconstruction on an 8° x 10° grid using marine and terrestrial proxies (PRISM0 - Dowsett et al., 1994). A global reconstruction developed from this (PRISM1 - Dowsett et al., 1996) and was further developed in the PRISM2 dataset (Dowsett et al., 1999). Both PRISM1 & PRISM2 were produced on 2° x 2° grids for use in climate models as boundary conditions. The continuing improvement in computing power allowed climate models to include fully coupled atmosphere and oceans as standard. Combined with developments in palaeo-proxy techniques, the PRISM3D reconstruction was developed. PRISM3D has four major developments compared to previous PRISM reconstructions (Dowsett et al., 2010a): comprehensive vegetation

(Salzmann et al., 2008) and ice sheet (Hill et al., 2007) reconstructions, both created with the aid of a data-model hybrid methodology and a deep ocean temperature reconstruction (Dowsett et al., 2009a; 2010a). PRISM3D remained on the 2° x 2° grid (with deep ocean temperatures on a 4° x 5° grid) with a sea surface temperature (SST) reconstruction using the same methodology as PRISM2 (Dowsett et al., 2009a). It also included a Mean Annual Sea Surface Temperature (MASST) dataset for point by point comparison with outputs from climate models (Dowsett et al., 2010b). A summary of the PRISM datasets and developments in number and type of localities is shown in Table 1.1.

Version	Resolution	Coverage	Sea Surface Temperatures	Deep Ocean Temperature	Vegetation	Ice Sheets	Key References
PRISM 0	8° x 10°	Northern Hemisphere	24 sites	--	45 Sites	Sea Level Estimates	Dowsett et al., 1994
PRISM 1	2° x 2°	Global	64 Sites	--	74 Sites	Sea Level Estimates	Dowsett et al., 1996; Thompson & Fleming, 1996
PRISM 2	2° x 2°	Global	77 Sites	--	74 Sites	Sea Level Estimates & Modelling	Dowsett et al., 1999
PRISM3D	2° x 2°	Global	115 sites	27 Sites 4° x 5° Resolution	202 Sites	Data-Model Hybrid Approach.	Dowsett, 2007; Dowsett et al., 2010a,b; 2012; Salzmann et al., 2008; Hill et al., 2007

Table 1.1. The evolution of the key components of the PRISM palaeo-environmental reconstructions over the course of the project to date. The references indicate the key papers for each reconstruction. Deep ocean temperatures represent water depths of 1000 to 4500m.

ii) Sea Surface Temperature Reconstruction

The PRISM project originated out of a North Atlantic SST reconstruction and SSTs have remained the flagship data of the project ever since. Initially used to act as the SST boundary condition for use in AGCMs, the SST component of the PRISM reconstruction has continued to play a prominent role with the development of AOGCMs, becoming a key data-model comparisons (Dowsett et al., 2010b). Primarily, SST data is generated from microfossils extracted from sediment cores obtained from the ocean floor by the Integrated Ocean Drilling Program (IODP) and its predecessors the Deep Sea Drilling Project (DSDP) and the Ocean Drilling Program (ODP). These international projects drill cores throughout the global oceans and the data generated from the cruises is widely used in palaeoclimate reconstructions such as the Zachos compilation curve (Zachos et al., 2001) or in individual research projects using a single site such as Reed-Sterrett et al. (2010 – ODP site 1022) as well as the PRISM project. The PRISM team take a sediment core and identify the PRISM section, defined as 3.264 to 3.025 Ma BP using the transition of oxygen isotope stage boundaries M2/M1 to G21/20 (Lisiecki & Raymo, 2005; Dowsett et al., 2010a). The period is also identified by other biostratigraphic and magnetostratigraphic markers (Dowsett et al., 2010a). The length of the period and its boundaries allow for easy correlation between sections that originate from all over the globe (Dowsett & Poore, 1991).

Once the PRISM section has been identified in a core, the first task is to identify the species in the section. Faunal identification could be a source of error due to investigators having different opinions on a species, however this is negated in the PRISM dataset by having a principle investigator who re-identifies and cross checks all samples to ensure taxonomic consistency across every analysis (Dowsett, 2007; Dowsett et al., 2010a). The PRISM team creates taxonomic grouping schemes, assumed to have similar environments across taxonomic categories (Dowsett, 2007). The decisions made in the PRISM reconstruction have meant that if this assumption creates an error it will be an error that underestimates warming (Dowsett, 2007), as the warmer tolerance microfossils are more fragile and under go dissolution preferentially resulting in a cooler estimate of SSTs (Dowsett et al., 2012)

Once the taxonomy is completed, the estimation of SSTs is undertaken using a variety of methods. The transfer function technique is a series of equations that relates microfossil abundances in modern (or core-top) samples to the oceanographic factors such as SST and salinity (Dowsett, 2007). Created by Imbrie and Kipp (1971), this technique has been extensively used in palaeoclimate research in projects such as

CLIMAP (1981; 1984), LIG studies (i.e. Briskin & Berggren, 1975; Ruddiman et al., 1986) and initial work on Pliocene-Miocene samples (Poore, 1981). The PRISM group developed these initial studies into 18 counting categories of modern and Pliocene taxa, to apply the technique to the mPWP (Dowsett & Poore, 1990; Dowsett, 1991; 2007; Dowsett et al., 1996; 1999). The equations are then used to generate a temperature reconstruction and the equations can be tested to produce an estimate of the standard error and seasonal sensitivity of the transfer functions (Dowsett & Poore, 1990; Dowsett, 2007). Another method is the modern analogue technique (MAT – Hutson, 1979; Dowsett & Robinson, 1998), where a relationship can be shown statistically between a Pliocene species and its modern relation enabling Pliocene SSTs to be reconstructed (Dowsett et al., 1999; Dowsett, 2007). For the mPWP some species are now extinct, requiring modern species to be grouped to allow comparison with Pliocene taxa (Dowsett & Poore, 1990; Dowsett, 1991; Dowsett & Robinson, 1998). Despite this potential weakness, MAT has been shown to compare well to PRISM and modelling estimates of Pliocene SSTs (Marques De Silva et al., 2010).

Once the samples have been calibrated and used to produce estimates of SST, then the data that is representative of the 300kyr PRISM time slab must be chosen (Dowsett et al., 1999; Dowsett, 2007). The PRISM team decided to focus on a reconstruction of mPWP mean interglacial conditions to minimise potential errors from correlating peaks and troughs between different sample sites widely separated geographically (Dowsett et al., 1994; 1996; 1999; Haywood et al., 2002a; Dowsett, 2007). Termed the *"Peak Averaging Method"* (Dowsett & Poore, 1990), it targets the warm peaks in an SST time series through a core (where a warm peak is a sample warmer than those directly above or below it in the core - Dowsett & Poore, 1991; Dowsett et al., 2005; Dowsett & Robinson, 2006; Dowsett, 2007). All the values that are considered to be valid estimates are averaged to produce the warm peak average for that core site (Dowsett, 2007). The validity comes from a test of the communality of the data. Communality is a function of the explanation of the variance in the data by the factor model and ranges from 0 to 1 (Dowsett et al., 2005; Dowsett, 2007). An acceptable communality threshold is used by the PRISM team of 0.7, equating to a minimum of 84% of the variance being explained by the factor model (Dowsett, 2007), values failing to achieve that level are not included in the reconstruction (Dowsett & Robinson, 1998; Dowsett et al., 2005; Dowsett, 2007). From the warm peaks the 'August' (summer) and 'February' (winter) temperatures are produced (Dowsett, 2007).

These two peak averages are combined with other proxy techniques for estimating SST such as magnesium/calcium palaeothemometry (Mg/Ca) and alkenones (Dowsett,

2007; Dowsett et al., 2010a,b). Mg/Ca is a method which uses the ratio of the observed increase in biogenic calcite as temperature increases, a process observed in laboratory culture experiments and in the analysis of Quaternary foraminifera (Lear et al., 2000; Lear, 2007). Magnesium substitutes for calcium in the carbonate lattice of a calcite shelled organism such as foraminifera as the temperatures increase (Lear et al., 2000; Lear, 2007). Mg/Ca is biased by the ratio of Mg/Ca in seawater, which can not be completely resolved for palaeo-oceans. The uncertainty from this is is accounted for in the Mg/Ca calibration curves, but is also reduced because of the long residence time of both ions in the ocean (Lear, 2007).

Alkenone unsaturation index palaethemometry is based on long chained organic compounds (lipids) which are created by some species of haptophyte algae (Lawrence et al., 2007; Pagani et al., 2010). These lipids come in different lengths of carbon chains (C_{37}, C_{38} & C_{39}) and a different number of carbon double bonds. While C_{38} and C_{39} lipids are also temperature sensitive, the C_{37} lipid is analytically easiest to calculate a temperature relationship ratio from. Using C_{37} (which has between two and four double bonds), a ratio (U^k_{37}) is calculated by the number of two double bond lipids to the sum of two and three double bond lipids. As the U^k_{37} ratio increases, the alkenone unsaturation decreases (there are fewer double bonds) and this is shown to relate to an increase in near surface ocean temperatures (Lawrence et al., 2007). The alkenones are considered 'near surface' temperature proxies as the algae live in the photic zone, but at a depth which is hard to determine as it is based upon the nitrate supply to the surface waters from deep waters (Oukouchi et al., 1999). At high and low latitudes, alkenones are formed near the surface, but at mid-latitudes they are closer to the thermocline (Oukouchi et al., 1999). Both these proxies lack seasonality and where faunal data exists, are considered supplementary information to strengthen confidence in the reconstruction (Dowsett et al., 2010b). However, for some sites in the dataset, there is insufficient faunal data, resulting in these proxies being used. Where Mg/Ca or alkenones have been used, the MASST reconstruction accepts a potential error of +/- 1.2°C for Mg/Ca and +/- 1.3°C for alkenones (Dowsett et al., 2010b).

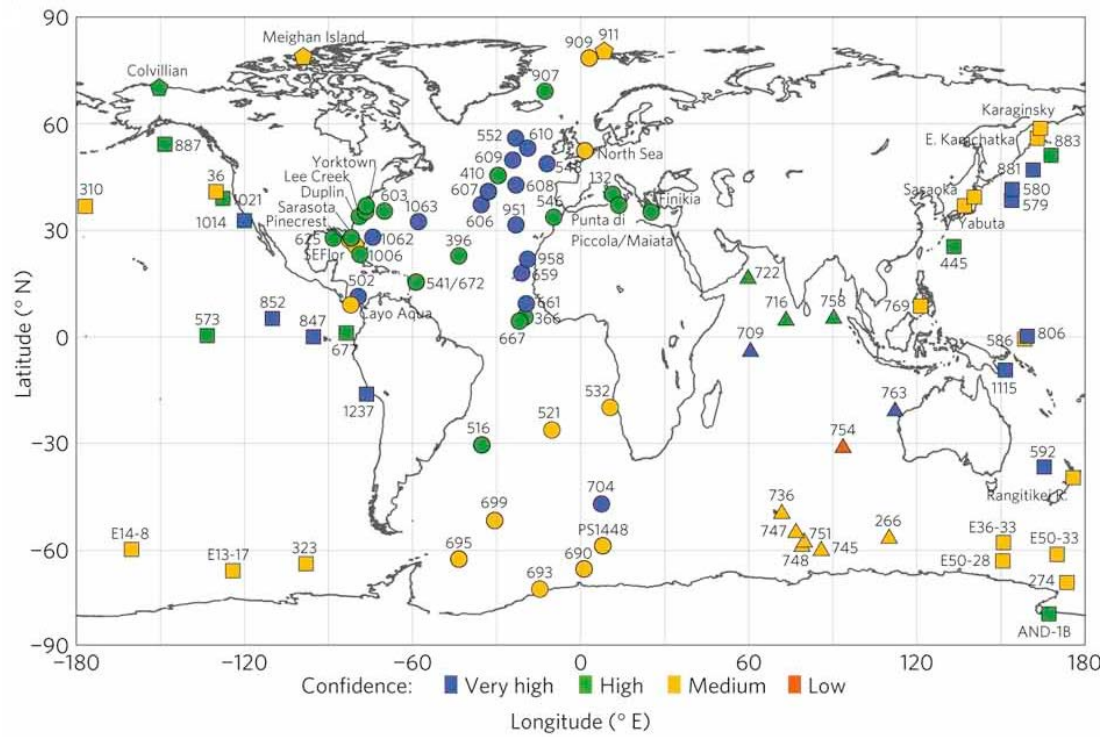


Figure 1.2. The geographical location of the PRISM3D MASST sites reproduced from Dowsett et al. (2012). The shape of each site indicates which ocean basin each site originates from (Circles = Atlantic; Squares = Pacific; Triangles = Indian). The colour of the shape reflects the confidence rating assigned to each site by the PRISM team.

The data for each site is collated and although no one proxy is considered to be better than another, the PRISM3D MASST dataset ideally uses the February and August values and a derived faunal mean SST value, which is supplemented by Mg/Ca and/or alkenone proxies that fall between the February and August values (Dowsett et al., 2010b). At a few sites, the alkenone or Mg/Ca proxies do not fall between the February and August values. In those cases, each sites mean annual value was determined using expert judgement on a case by case basis assessing the constituent proxies and the possible factors that could have affected their value in the water column or during post processing. Details on these sites and how the final analysis was derived can be found in Dowsett et al. (2010b).

The strength of the PRISM3D MASST reconstruction is the variety and flexibility of techniques and species used while adhering to methods to minimise potential errors. Different analytical techniques perform better in certain oceans, i.e. MAT in the Pacific Ocean (Dowsett & Robinson, 1998; Dowsett, 2007) with transfer functions used extensively in the Atlantic Ocean (Dowsett & Poore, 1990; Dowsett et al., 1994; 1996; Dowsett, 2007). The inclusion of additional proxies ensures that no single environmental variable can affect the mean annual estimate for the site (Dowsett et al., 2010b). The confidence of the PRISM team in the MASST data points has been assessed based upon the quality of the dating, the quantity and preservation of the micro-fossils

and the techniques used (Dowsett et al., 2012). The MASST data, a mean annual value, means that sites excluded from the initial PRISM3D global climate reconstruction (Dowsett et al., 2010a) as they do not provide seasonal information are included in the MASST dataset as they do include mean annual data. The MASST dataset therefore has increased geographical coverage shown in Figure 1.2 (Dowsett et al., 2010b).

The PRISM3D MASST dataset is required for data-model comparisons with AOGCMs, as climate model output is generally assessed through creating mean annual averages for variables over an averaging period (usually the last 30 years of simulation). The traditional seasonal PRISM reconstruction, important for understanding the Pliocene oceans and as a boundary condition for AGCMs, is not used as frequently for comparison with AOGCMs. As well as in this thesis, the MASST dataset is being used for data-model comparisons in other Pliocene modelling studies (Dowsett et al., 2012; Haywood et al., 2013a).

iii) Deep Ocean Temperatures

The deep ocean temperature reconstruction is presently being developed by the PRISM team as part of work to develop our understanding of palaeo-ocean circulation, salinity and temperatures (Dowsett et al., 2010a). Initial work on 27 DSDP and ODP sites covering water depths from 1000m to 4500m was done using Mg/Ca analysis of a bottom water ostracode genus *Krithe* (Cronin et al., 1996; 2005). Early work using deep sea temperatures in the model initialisation (Dowsett et al., 2006) has reproduced conditions of increased North Atlantic Deep Water formation and SSTs seen in the PRISM data reconstructions for deep water masses and SSTs (Dowsett, 2007).

iv) Ice Volume & Sea Level

A range of geological evidence indicates that mPWP sea level was different to the present day, with reduced ice sheet volume contributing the majority of the sea level rise. mPWP sea level estimates (Table 1.2) fall into a bracket of 20 to 30m above mean sea level (AMSL) for the mPWP, with PRISM reconstructions since PRISM2 using an mPWP sea level of +25m (+/- 17m) AMSL (Dowsett et al., 1999; 2010a; Dowsett, 2007; Haywood et al., 2010; 2011a).

Sea Level Reconstruction (Above Mean Sea Level - AMSL)	Locality Used	Reference
35m +/- 18m	Orangeburg Scarp	Dowsett & Cronin, 1990
20m to 25m	Eneetak Atoll	Wardlaw & Quinn, 1991
25m	Atlantic Coastal Plain & Benthic $\delta^{18}O$	Krantz, 1991
30m	LR04 Benthic $\delta^{18}O$ Stack	Raymo et al., 2009
15m to 30m	IODP Sites 925 & 926	Dwyer & Chandler, 2009
30m	Sediments in Wangauni Basin, New Zealand	Collated in Dowsett et al., 2010a
22m	Statistical Reanalysis of Previous Estimates	Miller et al., 2012

Table 1.2. Sea level estimate for the mid-Pliocene Warm Period and the locality the estimate is based upon.

Using the estimate for sea level, the next phase of the reconstruction focuses on global ice volumes. A variety of proxies are used to determine the ice volume and extent involving coastal geomorphology, climatological reconstruction with ice sheet modelling, oxygen isotope analysis and analogue comparison (Dowsett, 2007; Hill et al., 2007). In PRISM2, the Greenland Ice Sheet (GrIS) was reduced by 50% (compared to modern) giving 4m of sea level rise (Sohl et al., 2009; Haywood et al., 2010). The West Antarctic Ice Sheet (WAIS) is removed completely giving a 6m sea level rise and the East Antarctic Ice Sheet (EAIS) is reduced and redistributed to account for the remaining 15m of sea level rise (Pollard & DeConto; 2009; Sohl et al., 2009; Haywood et al., 2010; 2011a). Some subtle changes have been made between PRISM2 and PRISM3D. Hill et al. (2007) modelled the Pliocene climate and then coupled this to the British Antarctic Survey Ice Sheet Model (BASISM) to generate reconstructions of the ice sheet sizes and topographies. The result was that GrIS contributes an extra 1.6m of sea level rise (now 5.6m – Hill, 2009) in PRISM3D. The EAIS is flattened and extended covering a wider area (in comparison with PRISM2), however the ice sheet has a smaller volume than in PRISM2. The result with the maintained total absence of the

WAIS is 22m of sea level rise from ice sheet loss in PRISM3D (Dowsett et al., 2010a). While considered fixed in the PRISM boundary conditions, it has been shown that the ice sheets would have been susceptible to orbital variability and would have waxed and waned during the mPWP as the orbital conditions changed (Dolan et al., 2011). The PRISM ice sheet reconstruction as with all elements of the dataset produces an average representation of interglacial conditions during the mPWP (Haywood et al., 2010, 2011a).

v) Land-Sea Mask & Orographic-Bathymetric Reconstruction

The land-sea mask in the PRISM3D reconstruction has been developed from the PRISM2 reconstruction in which grid boxes were either land or sea. The areas that are 100% continental or ocean remain labelled in this way as in PRISM2, but the marginal coastal regions have been changed (Haywood et al., 2010). In PRISM2, the ETOPO5 *'five minute topographic grid'* (Edwards, 1992; Dowsett et al., 1999) was used. If the majority of points in a PRISM grid box (2° x 2°) were above the PRISM +25m AMSL estimate then the grid square was determined as 'land' and vice versa (Dowsett et al., 1999). The shape of coastlines and the appearance of the continents is altered, in a similar way to the land-sea mask in a climate model, representing the best estimate for a land-sea mask at the time.

The PRISM team developed this for PRISM3D with the marginal grid boxes becoming fractional. The 100% land or ocean boxes remain the same, but now instead of marginal boxes being determined as ocean or land based on analysis of the grids main constituent, they are designated as marginal and a fraction of land and sea in the box is determined (Haywood et al., 2010, 2011a). In PRISM3D, continental areas are given vegetation data, ice data (where applicable) and elevation with oceanic areas given SST, bottom water temperatures and (where applicable) sea ice and the marginal areas receive all the relevant data required (both ocean and land data).

The orography for the land-sea mask is another important element of the reconstruction. The orography of the PRISM3D land-sea mask was changed with the Rockies and the Andes raised towards their present day height an increase of 1,500 to 2,000m (Markwick, 2007; Sohl et al., 2009; Haywood et al., 2010). The PRISM2 orography had reconstructed the Rockies as being 50% of their present day height. The adjustment was made as a result of research showing that there had been minimal change since before the Pliocene (McMillan et al., 2006; Moucha et al., 2008). The surface orography plays an important role in enhancing the warmer climate of the

Pliocene caused by the elevated levels of CO_2 in the atmosphere, but a lower height of the Rockies affects northward heat flow in climate models (Hill et al., 2011). Modelling studies with reduced Rocky Mountains compared to modern have displayed warmer and wetter American winters (Kutzbach et al., 1989; Seager et al., 2002; Foster et al., 2010). The higher mountains prevent zonal air flow deflecting the jet stream south which cools the north east of North America compared to if there was a reduced height for the Rocky mountains (Foster et al., 2010). The contribution of changes in orography between the PRISM2 and PRISM3D reconstructions can be tested by altering the orographic boundary condition in the model simulations for the Pliocene and a Pre-industrial control. For example, Hill et al. (2011) showed that changes in the orography of the Rocky Mountains affected the location and the intensity of the North Atlantic Oscillation in the model simulations.

Oceanic bathymetry is also an important boundary condition, controlling the development of key water masses in the deep oceans. The key boundaries in the Pliocene ocean are the Central American Seaway (CAS), the Greenland-Scotland ridge, the Indonesian Passages and the Bering Strait. The CAS closure through the gradual tectonic uplift of the Isthmus of Panama prevented the mixing of Pacific and Atlantic waters driving the formation of the thermohaline circulation and the transport of heat to higher latitudes (Lutz, 2011). It had been proposed that this occurred later in the Pliocene (~2.6 Ma) leading to the formation of Northern Hemisphere glaciation (Saranthein et al., 2009). Lunt et al. (2008a,b) shows that this was not likely based on modelling results, similarly most data supports the early Pliocene (~5 Ma) effective full closure of the CAS.

Work by Karas et al. (2009; 2010) used DSDP and IODP sites: 214, 709, 757, 758, 763 & 806 to investigate the timing of the changes in Indonesian Through Flow (ITF) of the Indonesian Passages during the Pliocene using analysis of Mg/Ca ratios in planktonic foraminifera. They found that the timing of this change was during the later stages of the mPWP and into the late Pliocene. The shoaling of the Indonesian Passages changed the composition of the water forming the ITF from warm salty South Pacific waters to cooler fresher North Pacific waters into the Indian Ocean. Krebs et al. (2011) found that the shoaling of the Indonesian Passages from the Pliocene into the modern led to a reduced ITF causing the aridification of the Australian continent as SSTs cooled in the passages and continental regions warmed up. SSTs warmed in the Equatorial Pacific warm pool and this led to a modelled increase in precipitation over the warm pool region (Krebs et al., 2011). The change is consistent with the PRISM3D vegetation data, which has regions of tropical forest and tropical grassland in Pliocene Australia

compared to modern arid outback vegetation (Salzmann et al., 2008). However with models using a modern land-sea mask this makes it hard to reproduce the required precipitation over Australia.

The Greenland-Scotland ridge (GSR) has been investigated as a potential barrier to heat from the North Atlantic penetrating into the Arctic in Pliocene modelling, an area of key data-model discrepancy. The PRISM data has always indicated SSTs in the Arctic that AOGCMs are unable to replicate (Haywood & Valdes, 2004; Robinson, 2009; Robinson et al., 2011; Dowsett et al., 2012). The height of the GSR is controlled by the mantle hotspot beneath Iceland and its activity affects the bathymetry of the ridge system, with less vigorous (than modern) activity in the Pliocene resulting in the ridge being ~300m deeper (Jones et al., 2002; Robinson et al., 2011). Robinson et al. (2011) investigated the effects of altering different ridge section heights and the response of the model looking at both SSTs and deepwater production. They found that the deepening of the ridges increased deepwater production and Arctic SSTs. Although it did not fully remove the data-model mismatch in the high latitudes, it could be a way of tackling the issue in future modelling studies (Robinson et al., 2011).

The Bering Strait has acted as an occasional gateway for flow between the Pacific and Arctic oceans, with the opening of the gateway controlled by tectonic activity in the region and global sea level changes (Matthiesson et al., 2009). The opening of the gateway allows the flow of water between the North Pacific and North Atlantic, which reduces the development of NADW reducing the intensity of THC (Shaffer & Bendtsen, 1994). In glacial periods, increased ice cover reduces sea level closing the gateway (Shaffer & Bendtsen, 1994; Goosse et al., 1997; Hasumi, 2002), but during the Pliocene, sea level was ~25m AMSL, making the gateway tectonically controlled. The age of an open Bering Strait comes from analysis of shallow marine fauna. When the strait is closed, North Pacific and Arctic fauna have different compositions compared to when the Strait is opened, with the Arctic record becoming dominated by North Pacific fauna (Marincovich & Gladenkov, 1999). However, the age for the opening event ranges from the late Miocene (~6 Ma) to the Late Pliocene (~3 Ma - Marincovich & Gladenkov, 1999; Gladenkov et al., 2002). The uncertainty has an important effect on the reconstruction of Pliocene oceanography and modelling results (Robinson et al., 2011).

Topography and bathymetry are currently evolving sections of the PRISM dataset with ongoing work from modelling, geophysics and geology to create stronger reconstructions for the mPWP boundary conditions.

vi) Vegetation, Lakes & Soils

PRISM1 & PRISM2 used the data from fossil pollens and plant macro-fossils from the 75 terrestrial sites available in the dataset created by Thompson & Fleming (1996), these sites where dated using a mixture of radiometric dating, tephrochronology and biostratigraphy (Dowsett et al., 1999). The data was arranged into seven land cover categories (desert, tundra, grassland, deciduous forest, coniferous forest, rainforest and land ice - Dowsett et al., 1999) based on the 22 categories used by Matthews (1985) and pollen from the sites was assigned based on qualitative estimates of the change in the temperature and precipitation for that area compared to the modern and how this would affect the likely biome (Dowsett et al., 1999).

PRISM3D uses a 202 site terrestrial and marine reconstruction with 28 biome classes (Salzmann et al., 2008). A Tertiary Environmental & Vegetation Information System (TEVIS) database of fossil pollens, leaves, wood and palaeosol carbonates was created from the PRISM2 database locations and from literature of other studies on Pliocene vegetation (Salzmann et al., 2008). The interpretation by the original authors of the research included in the database was maintained and the data was collated into a consistent form using the 28 biomes which are found in the BIOME4 offline equilibrium vegetation model (See Chapter 2, Section 2.2.2 for BIOME4 description).

If the information available for a site suggested two types of biome, the climatically warmer biome type was selected (Salzmann et al., 2008). The TEVIS database also stored information on the dating technique used, type of sample and preservation (Salzmann et al., 2008). As many of the terrestrial sites come from short sequences with few dating points (Dowsett, 2007), many of these sites have poor temporal resolution and chronological dating. The 202 site PRISM3D vegetation reconstruction is a likely reconstruction for the Piacenzian age (3.6 to 2.6 Ma BP) of which the mPWP is a sub-section of time used by the PRISM group (Salzmann et al., 2008, Dowsett et al., 2010a).

The 202 site reconstruction was used to create a global reconstruction of the mPWP vegetation by coupling the HadAM3 model (with SSTs from the PRISM3D reconstruction) with the BIOME4 model. The vegetation model was then compared to the database of results from the 202 sites using ArcGIS (Salzmann et al., 2008). The final vegetation reconstruction for the Pliocene for the PRISM3D dataset was produced using a synthesis of information from the data and models (Salzmann et al., 2008; Dowsett et al., 2010a).

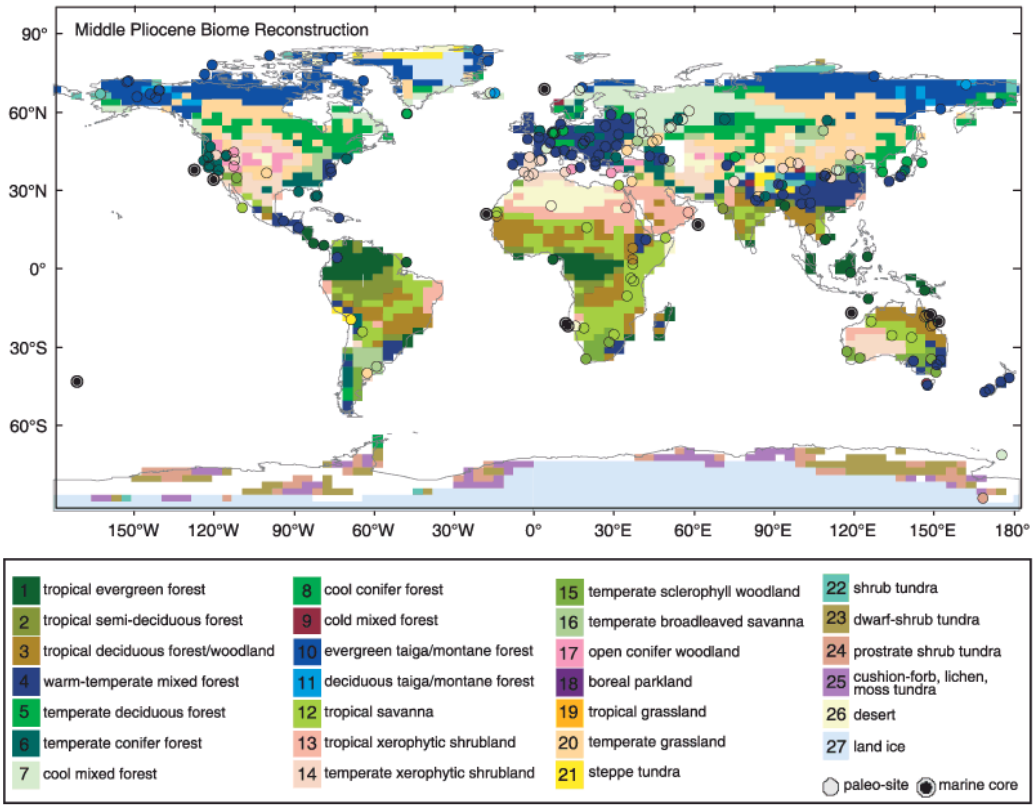


Figure 1.3. The PRISM3D vegetation biome data-model hybrid reconstruction from Salzmann et al. (2008) which serves two purposes as a mid-Pliocene dataset for use in data-model comparisons and as a model boundary condition. Each circle represents one of the 202 data sites which are categorised into biomes using the biomes programmed into the BIOME4 model. The base colours are produced using the strongest performing versions of the HadAM3/Biome4 models in comparison with the data. Biome number 28 (Barren) is not listed in the figure key above.

The TEVIS database has been refined to produce a vegetation dataset that can be utilised for quantitative terrestrial SAT data-model comparisons (Salzmann et al., 2013). The refining of the TEVIS dataset saw sites from Russia, Canada, Mexico, France, Turkey, Germany and Portugal added. Sites were removed from Iceland and Antarctica (due to dating uncertainty) along with sites above 1000m AMSL or from sediment cores further than 250 Km offshore (Salzmann et al., 2013). The changes reduced the size of the dataset to 45 sites which could be used to produce a quantitative SAT value. Where published the temperature value assigned by the authors was applied, these SATs were calculated using a range of methods including CLAMP (Climate Leaf Analysis Multivariate Program – Wolfe, 1993), Coexistence Approach (Mosbrugger & Utescher, 1997), modern analogue vegetation distribution (semi-quantitative) & multi-proxy approaches using these and other proxy methods (Salzmann et al., 2013). Where the authors had not published a temperature estimate, the Coexistence Approach was applied (Salzmann et al., 2013). A qualitative estimate of confidence in the

reconstructed SAT value was applied to each site based on the dating, preservation and temperature estimate methods used

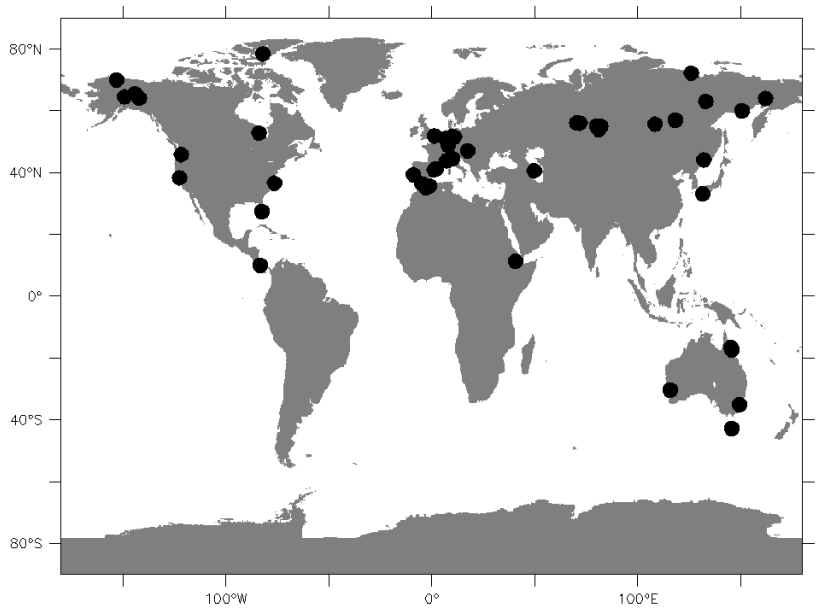


Figure 1.4. Geographical coverage and location of the 45 sites in the vegetation SAT reconstruction of Salzmann et al. (2013).

Data for the vegetation SATs is spatially limited to the Northern Hemisphere, with the exception of four Australian sites. There are two equatorial sites, including the only African site, with most of the sites being in the mid-high latitudes of North America and Eurasia.

To date, all simulations (including this projects) undertaken for the mPWP have used modern reconstructions and values for soils and lakes. Studies of Holocene climate utilising Holocene specific data for soils and lakes have displayed significant regional impacts on precipitation (Krinner et al., 2012; Pound et al., 2013). A number of mPWP studies have shown weaknesses in reproducing precipitation in tropical and sub-tropical regions and this could be a reflection of the use of modern soils and lakes (Pound et al., 2013).

To create a mPWP soils and lakes database, available data was collected into a database (similar to the TEVIS database used for the vegetation reconstruction – Salzmann et al., 2008; Pound et al., 2013). Age, geographical distribution and dating methods used for the site, for the soils a soil type was also documented (Pound et al., 2013). The coverage of the soils data was not global (54 sites in total), for inclusion as a climate model boundary condition, it was combined with the biome reconstruction and from this global soils for the mPWP were generated (Pound et al., 2013). The rationale for this methodology is based on the relationship between soils and biomes produced

(Pound et al., 2013). For the lake reconstructions, the percentage of model grid cell (3.75° x 2.5°) that was lake was estimated for two scenarios designed to bracket the uncertainty in the size of mPWP lakes, creating a wet and dry reconstruction (Pound et al., 2013). Initial modelling studies were undertaken to investigate the impact of these new boundary conditions which suggest that while these changes do not exert a large global impact on climate, however generated some important regional feedbacks, with increased precipitation through North Africa and North America (Pound et al., 2013). Contoux et al. (2013) modelled the forcing of Megalake Chad in the mPWP, finding regional seasonal impacts on winds and precipitation distribution, but no significant impact upon biome distribution. These developments will be included in the next iteration of the PRISM dataset, PRISM 4.

vii) The PRISM Time Slice

The first PRISM time slice is centred on 3.205 Ma BP (with a range from 3.207 to 3.204 Ma BP) a warm interval in the LR04 Benthic stack (Lisiecki & Raymo, 2005) indicated by a negative oxygen isotope excursion of 0.21 to 0.23‰ (Haywood et al., 2013b). The interval is already part of the PRISM time slab, a logical decision as this period already has the most extensive data available for it and the closest to modern topographic reconstruction (Haywood et al., 2013b). The size of the excursion is important as it improves the signal to uncertainty ratio, reducing the impact of natural variability in the climate reconstruction (Haywood et al., 2013b). Reconstructions of orbital forcings using calculations of the Milankovitch cycles from Laskar et al. (2004) can be used to determine the top of the atmosphere insolation forcing for the entire length of the PRISM interval. The selected time interval has an insolation forcing which is very similar to modern determined by seasonal and regional analysis. The advantage of this choice is that it reduces the impact of a potential cause of the Pliocene warmth (Willeit et al., 2013) and an attributing elements which cause the warming simulated by AOGCMs (Lunt et al., 2012a). Finally allowing easier interpretation of mPWP results in the context of Charney and Earth system sensitivity studies (Haywood et al., 2013b).

Sea level is estimated for the period as being 22m +/- 10m higher than present day (Haywood et al., 2013b), equivalent to the near complete loss of the Greenland and West Antarctic Ice Sheets (Naish et al., 2009; Pollard & DeConto, 2009; Dolan et al., 2011). The interval is now suitable to be used for palaeo-environmental reconstructions by the PRISM team (and other groups) to generate palaeo-proxy reconstructions for the time slice. The advantage of this time slice is that it is easily found in core sections due to its location relative to major magnetostratigraphic

boundaries and isotopic marker boundaries (Haywood et al., 2013b). Additionally the climate of the slice allows for a degree of error (a *'zone of tolerance'* - Haywood et al., 2013b) in finding the slice when analysing cores. The main weakness of moving to a time slice is that the geographical coverage of the proxy data, and the number of available sites will be reduced (Haywood et al., 2013b) as not all sites will contain data that is robust enough to make it to the final data compilation. The long term aim is to create a series of time slices which investigate the variability in Pliocene climate due to various changing conditions such as orbital forcing, greenhouse gases or in glacial climates within the Pliocene (Haywood et al., 2013b). Along with lakes and soils data, the PRISM Time Slice will from the basis of the PRISM4 reconstruction, the next iteration of the PRISM palaeo-environmental reconstruction for the mPWP.

viii) Data Uncertainty

As Sections *ii* & *vi* indicate, there is an inherent uncertainty on the palaeo-data produced for the mPWP. The uncertainty will affect the size of the mismatch between models and data and therefore quantification of the data uncertainty is an important part of the reconstruction. Recent work has been undertaken to provide a quantified uncertainty on the palaeo-data for SSTs (Dowsett et al., 2013) and the SATs (Salzmann et al., 2013). For the SSTs uncertainty was calculated based on the standard deviation of each sites warm peak estimate throughout the PRISM time slab. The warm peak averaging method is outlined in Section *ii*.

SAT uncertainty arises from the temporal and bioclimatic range on the data (Salzmann et al., 2013). Bioclimatic uncertainty represents the range of potential climates that the identified species could survive in, while temporal range reflects the climate change in the region over the period of time dated and uncertainty in the dating of the fossil record at each site (Salzmann et al., 2013). Bioclimatic uncertainty came from analysis of the temperature estimates used at each site (e.g. CLAMP, Coexistence Approach & modern analogues – Section *vi*). The reconstructed fossil assemblages at each site represent a group of flora that would have grown in a window of climatic tolerance, outside of which the assemblage could not have grown. The bioclimatic range represents the uncertainty that this window produces on the mean temperature for that site (Salzmann et al., 2013). The final SAT estimates (Section *vi*) included the full climatic range for each site including (where available) bioclimatic and temporal uncertainty combined into one uncertainty value for each site (Salzmann et al., 2013).

The uncertainty estimates produced and discussed for both the SSTs and SATs are a quantitative number and unrelated to the confidence in a site, which is calculated based on the analysis of dating, preservation and the temperature estimate methodology (Salzmann et al., 2013). For the data-model comparisons the uncertainty range can be applied as a window to see whether models can achieve the range in the data for a site. For sites where the model is still unable to achieve the window, the nearest boundary of the window to the data can be used to calculate the mismatch.

1.3.4. mPWP Data-Model Mismatches

Initial mPWP modelling studies on AGCMs used PRISM SSTs as a boundary condition in the simulation to create a Pliocene model (i.e. Sloan et al., 1996; Haywood et al., 2000a). Simulations using AOGCMs, do not require prescribed SSTs, which allows a fuller investigation of the climate, however the simulation lacks the ground truthing from prescribed SSTs. Therefore to determine the validity of a mPWP AOGCM simulation, data-model comparisons (DMCs) to palaeo-environmental datasets are undertaken (Dowsett et al., 2012). For the mPWP these datasets include: PRISM3D MASST dataset (Dowsett et al., 2010b), SATs based on vegetation (Salzmann et al., 2013) and a vegetation biome dataset (Salzmann et al., 2008). These methods have been used extensively throughout the mPWP to test model simulations (Haywood et al., 2004; 2009b; 2013a; Pope et al., 2011; Dowsett et al., 2012).

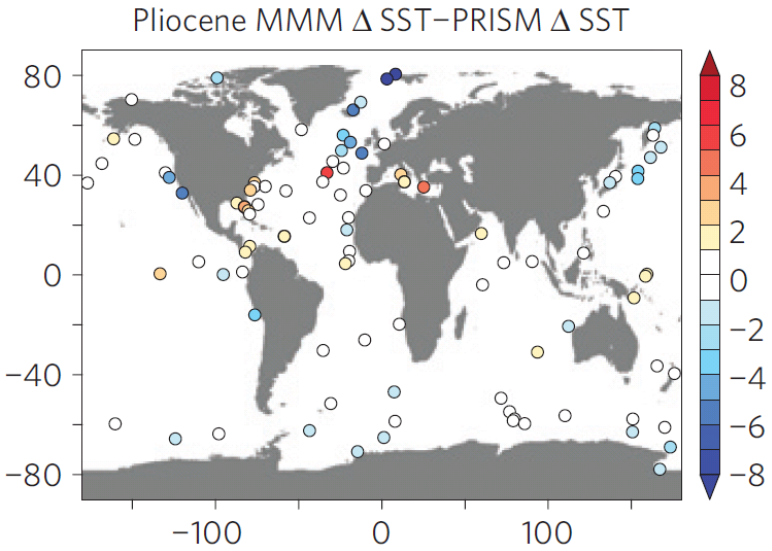


Figure 1.5. mPWP data-model comparison using a multi-model mean from four climate models in the PlioMIP project (CCSM4, HadCM3, GISS-ER and MIROC4m) and the PRISM3D SST data. The DMC is presented as the multi-model mean (mPWP minus pre-industrial) minus the PRISM3D SST (mPWP minus pre-industrial). The figure is reproduced from Dowsett et al. (2012). The figure highlights the existing sea surface temperature mismatch between mPWP simulations and the palaeo-data.

From the variety of Pliocene work published over a range of different climate models, an inability to effectively model the magnitude of warming reconstructed at high latitudes in comparison with the data reconstructions is evident (Figure 1.5). The mismatch is especially prevalent in the North Atlantic, where a series of data sites, considered to be of high or very high confidence by the PRISM group (Dowsett et al., 2012 & Figure 1.2), are the source of the greatest mismatch with the model, increases progressing northwards. In the tropics there is little or no data-model mismatch.

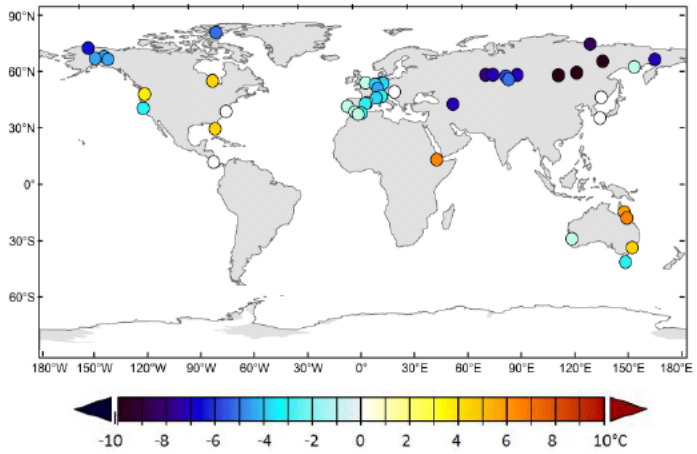


Figure 1.6. mPWP data-model comparison using a multi-model mean from the PlioMIP project (Haywood et al., 2013a) and the vegetation SAT data (Salzmann et al., 2013). The DMC is presented as the multi-model mean (mPWP minus pre-industrial) minus the SAT data (mPWP minus pre-industrial). The figure is reproduced from Salzmann et al. (2013). The figure highlights the existing terrestrial surface temperature mismatch between mPWP simulations and the palaeo-data.

The high latitude data-model mismatch observed in the SSTs is replicated in the SATs (Figure 1.6), however a paucity of suitable data through the tropics means that the small observed mismatch in tropical SSTs can not be analysed in the vegetation data. Mid-latitude sites in Europe and the USA indicate that the mismatch decreases moving towards the equator, which is inline with the SST DMC.

The challenge for future mPWP modelling is to understand the existing mismatch between data and models. To investigate this mismatch there are potential experimental designs that can be applied. These experimental designs can be used to understand the causes of the mismatch: due the uncertainty in the models through ensembles (i.e. PlioMIP – Haywood et al., 2010; 2011a and this PPE), uncertainty in the boundary conditions (Hill et al., 2011; Robinson et al., 2011), uncertainty due to variability in orbital forcings over the PRISM time-slab (time slice – Haywood et al., 2013b) or through uncertainty in the palaeo-data used (i.e. Salzmann et al., 2013).

By designing modelling experiments which can address the model based causes of the data-model mismatch, it may be possible to reduce the mismatch and isolate the uncertainty. Through tackling the uncertainty in mPWP simulations, the understanding of the climate system and its response to changes in forcings can be better understood.

1.4 – Uncertainty in Modelling Climate

Understanding the form, size and impact of uncertainty in climate modelling is vital to understanding both past changes in the Earths climate system and changes due to increases in greenhouse gas concentrations over the coming centuries (Hawkins & Sutton, 2009; 2011). There are four main types of uncertainty in models:

- Natural variability
- Model uncertainty (parameter and structural)
- Scenario uncertainty
- Boundary condition uncertainty

Natural variability, scenario uncertainty and boundary condition uncertainty affect climate model simulations in different proportions. In palaeoclimates, the scenario is not a cause of uncertainty, but boundary conditions are, whereas for a future climate projection, natural variability and scenario uncertainty are important uncertainties (see Section 1.4.1). Model uncertainty affects all modelling simulations and is an important uncertainty to quantify.

Rougier & Crucifix (2012) in an essay on uncertainty in climate science and climate policy described the uncertainty in climate change as "*epistemic uncertainty*", an uncertainty that arises from *"limitations in our knowledge and resources"*, and an uncertainty that can be reduced with further investigation.

The purpose of their essay was to question: 1) what is the uncertainty, 2) what is its form and 3) are the climate science community tackling this uncertainty in the correct way? (Rougier & Crucifix, 2012). Climate modelling is affected by a variety of causes of uncertainty. It is important to identify these uncertainties, ways of tackling them and also identifying the uncertainties that can not be quantified or reduced. Increasing our understanding of the climate system through quantifying uncertainty leads to increased confidence in projections of climate change due to anthropogenic warming. Equally important is communicating the uncertainty to policy makers as it is only through communication that appropriate decisions can be made on future climate change policy (Murphy et al., 2009a; Yokohata et al., 2012). To achieve this, it is

important to understand the main types of uncertainty in climate models and the impact of these uncertainties.

1.4.1. Origins of Key Model Uncertainties

The climate system is a highly complex series of interactions (Tebaldi & Knutti, 2007). Uncertainty arises in model simulations because we do not fully understand the system we are modelling (Sexton et al., 2012a). Quantification of uncertainty is important for people working on adaptation and mitigation to climate change, as it impacts upon adaptation and mitigation strategies. Adaptation planners often use offline models forced by climate model outputs i.e. New et al. (2007) who assessed water resources under changing climates. However, the volume of model output data produced by an ensemble of simulations covering a range of future scenarios and Charney sensitivities is too great for use in offline models. To fully assess the possible impacts of climate change and determine the appropriate adaptation strategy requires too much data to easily disseminate (Sexton et al., 2012a). Therefore, probabilistic estimates for the projection from the ensemble and its uncertainty allows these groups to produce an adaptation strategy for a *'most likely'* climate change scenario with a confidence range that can be quantified (Sexton et al., 2012a).

The first form of uncertainty is natural (also called internal) variability, which is inherent within the climate system and also in climate models (Murphy et al., 2009a) as the climate system is an example of a chaotic system. It was hypothesised in the early 20th century that with the correct equations and the ability to calculate them, weather forecasts could be made. These equations were calculated and with computer power it has been shown that it was possible to make accurate weather forecasts a short time (1 or 2 days) ahead (Shukla, 1998). It was also discovered that any small uncertainties in the initial conditions quickly led to ever increasing errors and degrading the accuracy of the forecast (Shukla, 1998). These errors, the response of the mathematical equations representing a nonlinear dynamic system, led to the definition of the climate system as a chaotic system. A chaotic system meaning a system that was sensitive to interior changes in the initial conditions of the system (Lorenz, 1963; Shukla, 1998). Averaging periods are used to minimise the effect of natural variability in model simulations (Murphy et al., 2009a). The length of the averaging period plays an important role in mitigating the size of the internal variability in a simulations climate metrics (Rowell, 2012). For transient simulations (such as investigating future climate change), an averaging period is not possible, however Hawkins & Sutton (2011) determined that internal variability was constant over the length of a simulation. When

averaging periods can be used, typically the last 30 years of a simulation are averaged, producing a climatological mean reduces the impact of the internal variability. However, it is not possible to fully remove this uncertainty from the simulations, or to effectively quantify it, so it has to be left as a '*known unknown'* in modelling projects (Murphy et al., 2009a).

The second form of uncertainty is model uncertainty, which breaks down into two key uncertainties: structural and parameter uncertainty. Structural uncertainty is related to the choices made during the construction of the model (such as grid resolution) and the way that climate processes are represented within the model components (Collins, 2007). The way that climate processes are represented can have extreme results, such as the example of clouds and radiation. Clear sky shortwave feedbacks result in a warming effect in some models and a cooling in others (Murphy et al., 2009a). The differences in the creation of the model are the structural uncertainties that are sampled by multi-model ensembles (MMEs). Parameter uncertainty is the uncertainty that arises from the choices regarding the representation of sub-grid scale processes. Once the resolution of a climate model has been set, there are then a series of processes that will occur on a scale smaller than the size of a grid box, called sub-grid scale processes. These are parameterised in the model as a numerical value (Murphy et al., 2004), assigned during the *'tuning'* of the climate model. Tuning refers to the initial creation of the model, when parameters are adjusted to a potential value (the tuning) and the resulting model version tested against climate metrics (Randall & Wielicki, 1997; Mauritsen et al., 2012). The tuning process starts with short runs, leading to longer runs altering different parameters using the standard baseline simulation usually a pre-industrial simulation (Mauritsen et al., 2012). The tuning process focuses on the decisions made by the model development team often focused on key sub-grid scale processes that have inherent uncertainty within them, such as cloud processes (Mauritsen et al., 2012).

The parameterisation of cloud processes has proven to be exceedingly difficult in GCMs, with complexities arising from the spatial scale and range of processes involved in forming clouds (Xie et al., 2002). Climate models calculate the amount of cloud at each layer of the atmosphere in each grid box column (i.e. 19 layers in HadCM3) using the values for temperature and water vapour to calculate the cloud type and size. The types and sizes of clouds are based on information collected by research flights (e.g. Abel & Shipway, 2007) and other measurements of clouds (such as satellite images - Oreopoulos & Khairoutdinov, 2003). Observations are combined with modelling results from single column models (SCMs) and cloud resolving models (CRMs) to create

the parameterisations in the grid box to calculate cloud types for that climatology (Xu et al., 2000; Xie et al., 2002). For example, warm rain convection in HadCM3 was parameterised by driving a high resolution cloud resolving model (CRM) with data collected from research flights in the 'Rain in Cumulous over Ocean' (RICO) project (Abel & Shipway, 2007). The results from these studies can be used to determine the style of and values for parameterisations in climate models (Abel & Shipway, 2007). However, data does not cover the entire globe leading to some parameterisations being globally uniform when they are not in reality (Curry et al., 1996). At high latitudes there are noticeably different boundary layer clouds compared to the rest of the globe (Curry et al., 1996). Combined with the unique effects of high latitude radiation, sea ice/ice sheet and cloud interactions, the high latitudes are characterised by a significantly different cloud regime to the rest of the world (Curry et al., 1996). Additionally there are not exact values for the data collected and so the parameters have a range that creates the uncertainty (Twohy et al., 1997).

The example above is replicated throughout the climate model. As a result a range of parameters can be identified which display an uncertainty, and from these the parameters are chosen for perturbing in the 'Quantifying Uncertainty in Model Predictions' (QUMP) ensembles (Murphy et al., 2004; Collins et al., 2006). Within climate models, a range of sensible potential values for each parameterisation exists. The final combination which produces the model are tested to ensure they produce an acceptable version of the model (tested against datasets for key metrics such as temperature and precipitation). However, there is the possibility that the final combination is right for the wrong reasons and this creates the uncertainty in model projections (Collins et al., 2006). Knowledge of the performance of models relies on a good understanding of parameter and structural uncertainty. To quantify the effect of structural and parameter uncertainty on climate projections, ensembles of models are created.

Climate model design trends towards increasing the resolution of the model, which can reduce the required number of parameterisations within the model (as smaller grid boxes leave fewer unresolved sub-grid scale processes to be estimated). However, climate models are also being developed to include more components, moving from AOGCMs towards Earth System Models (ESMs). These include the carbon cycle, aerosol transport, dynamic vegetation and ice sheets. The inclusion of more components increases the uncertainty from parameterisations as more unresolved parameters are added to the model, which interact with other already uncertain parameters which could reduce model performance (Rougier & Crucifix, 2012).

The third type of uncertainty is scenario uncertainty. In work looking at future climate projections, uncertainty originates from the changes to greenhouse gas concentrations in the future. The actual values can not be predicted as a number of variables exist which will affect global emissions (IPCC, 2000). As a result, detailed emission scenarios were created to generate a range of predicted concentrations to aid the creation of the climate projections (SRES scenarios - IPCC, 2000; RCPs - Moss et al., 2010). The different scenarios allow simulations to be run investigating future climate change, but by design, represents a source of uncertainty as it is impossible to be certain, which (if any) of the scenarios will be most representative of the future. Hawkins & Sutton (2009; 2011) determined the contribution of scenario uncertainty to future climate change projections. For temperature, scenario uncertainty is the dominant form of uncertainty by 2100, but for precipitation the dominant form of uncertainty is model uncertainty. The challenges of ascertaining future greenhouse gas emissions do not apply to work in palaeoclimate, but boundary condition uncertainty replaces scenario uncertainty. In addition to scenario uncertainty is initial condition uncertainty. Primarily a feature of meteorological forecasting model simulations, initial condition uncertainty refers to the point a model simulation is started from, which can influence the final results (Tebaldi & Knutti, 2007), even up to decadal timescales. Decadal cycle internal variability features such as the AMOC influence the simulated climate, based on when in a cycle the simulation starts (Bryan et al., 2006). In shorter length simulations, a suite of different starting points are used to analyse the initial condition uncertainty in an ensemble (Tebaldi & Knutti, 2007).

The final type of uncertainty is boundary condition uncertainty. Boundary conditions are the representation of non-calculated processes that are important to the representation of the climate system in the model. Boundary conditions are prescribed within a climate model and remain static throughout the duration of the run. These include: continental configuration, orography, bathymetry, ice sheets, vegetation and atmospheric composition (Haywood et al., 2011a). Boundary condition uncertainty also occurs due to the representation of known boundary conditions in the model. An example of this uncertainty is the Atlantic and the Mediterranean, which in reality are joined by a narrow seaway, but grid box representation in HadCM3 joins Africa and Spain closing the seaway, which is represented as a diffusive pipe (Johns et al., 2003; Ivanovic et al., 2013). Changing the representation from a pipe to an open seaway causes a shoaling of the Mediterranean outflow (from 1500m to 1000m), leading to a deeper penetration of warmer Mediterranean waters further into the North Atlantic (Ivanovic et al., 2013). The increased outflow (seaway compared to pipe) leads to

reduced temperatures in the Labrador Sea (~3°C cooling), and hotspots of ~3°C warming in the North Atlantic and in the Barents Sea (Ivanovic et al., 2013). The pipe and how it represents Mediterranean outflow has been demonstrated to have an effect on Northern Hemisphere climates. The uncertainty surrounding its most accurate representation is a boundary condition uncertainty in the model. The main forms of uncertainty for the ensemble, as with any palaeoclimate modelling study are model (structural and parameter) and boundary condition uncertainty. Investigations have been undertaken in both of these forms of uncertainty and this research is now outlined.

1.4.2. Tackling Structural Uncertainty: Multi Model Ensembles (MME)

Structural uncertainty is tackled through the creation of a multi-model ensemble (MME – i.e. Tebaldi & Knutti, 2007), such as Coupled Model Intercomparison Project (CMIP - Meehl et al., 2000) and Palaeoclimate Model Intercomparison Project (PMIP - Braconnot et al., 2007). MMEs are generated by running a series of model simulations using structurally different models with the same forcings (Stott & Forest, 2007; Tebaldi & Knutti, 2007). MMEs are not stronger compared to a single diagnostic than any single ensemble member. It is when the MME is tested against a range of model diagnostics and metrics that the MME out performs its individual members, which gives the best result for the climate simulations (Lambert & Boer, 2001; IPCC, 2007; Tebaldi & Knutti, 2007).

The advantage of a MME is that it is an ensemble of different models. MMEs cover a wide range of models with different structural configurations (how the climate system is represented and interpreted), a model '*gene pool*' (Collins et al., 2011). The sampling of structurally different models when averaged reduces biases and errors in individual simulations. The result is a climatological mean that has less uncertainty than an individual model simulation, due to the impact of structural choices upon feedbacks and forcings within the models being reduced. MMEs also ensure that each ensemble member has been rigorously tested to make sure it performs well as a climate model and generates a plausible and stable control simulation when it is run (IPCC, 2007). The weaknesses of the MME are around its conception. A MME is by requirement an *'ensemble of opportunity'* (Tebaldi & Knutti, 2007), an *'ad-hoc'* ensemble (Rougier & Crucifix, 2012), where the sampled models are not chosen but are volunteered by institutes that have run the required simulations. The result is a limit on the potential size of the ensemble (the gene pool) with the number of samples at most in the order of a dozen simulations. Even then, the gene pool is not guaranteed to contain that number

of fully independent models, with many similarities (shared components) between models, that cause replication of similar systematic errors (Knutti, 2010; Collins et al., 2011). Additionally, within an MME, there are models that may contain some common components which reduces the probing of the structural uncertainty because the replication bias' the analysis towards the replicated components (Masson & Knutti, 2011; Knutti et al., 2013). For example, the NorESM model shares atmosphere, sea ice and land surface components with the CESM model and utilises the same model coupler (Zhang et al., 2012).

1.4.3. Tackling Parameter Uncertainty: Perturbed Physics Ensembles

A perturbed physics ensemble (PPE) explores the impact of potential alternative parameter values in the parameter space on the uncertainty in a climate projection. Investigating the uncertainty in parameterisations in a climate model assumes that the standard version of the model is not the best possible representation of the climate system for that models design. Therefore, altering the parameters could produce a model with a better representation of the climate (Murphy et al., 2004; Collins et al., 2006, 2011; Sexton et al., 2012a) assuming that the basis for the physics is solid (Ingram, 2012). The largest PPEs have been conducted by the UK Met Office Quantifying Uncertainty in Model Predictions (QUMP) Project and the climate*prediction*.net (CPDN) project, using different members of the third generation of UK climate models, HadAM3 (AGCM), HadSM3 (50m slab ocean) and HadCM3 (AOGCM). These PPEs have been created and run on a range of different modelling platforms.

1.4.3.1 – The QUMP and climateprediction.net Ensembles

The QUMP and climate*prediction.net* projects have been undertaken in the last decade using the HadCM3 family of models. They have focussed on perturbing parameters in the atmospheric component of the model, using expert judgement (Murphy et al., 2004; Collins et al., 2006) to select the parameters to be perturbed and have over time created more detailed methods for generating large ensembles. More recent studies have also started perturbing parameters in other components in the model such as the oceans, carbon or sulphur cycles (Ackerley et al., 2009; Brierley et al., 2010; Booth et al., 2012) and decadal prediction (Sexton et al., 2012b). Ideally, the selection of parameters to be perturbed would be independent of the expert judgement presently used, however at present, it is not possible to avoid the subjective decision making (Shiogama et al., 2012).

Model	Model Version	Perturbed Component	PPE Methodology	No. of Members	Reference
UK Met Office	HadAM3	Atmosphere	QUMP	53	Murphy et al., 2004
	HadSM3	Atmosphere	QUMP	128	Collins et al., 2006; Webb et al., 2006
	HadSM3	Atmosphere	UKCP09	280	Muprhy et al., 2009a
	HadSM3	Atmosphere	climateprediction.net	2,578	Piani et al., 2005; Stainforth et al., 2005
	HadSM3	Atmosphere	QUMP	224 (128 + 96 new)	Clark et al., 2010
	HadCM3	Atmosphere	QUMP	17	Collins et al., 2006; 2011
	HadCM3	Ocean	QUMP	7	Collins et al., 2007; Brierley et al., 2010
	HadCM3	Atmosphere	UKCP09	17	Murphy et al., 2009a; Sexton et al., 2012a
	HadCM3	Atmosphere	THC-QUMP/Latin Hypercube	22	Vellinga & Wu, 2008; Hodson et al., 2013
	HadSM3	Sulphur Cycle	QUMP	243	Ackerley et al., 2009
	HadCM3C	Carbon Cycle	QUMP	17	Booth et al., 2012
	HadCM3	Atmosphere & Sulphur	climateprediction.net	~9,000	Rowlands et al. 2012
	HadCM3	Atmosphere & Ocean	Latin Hypercube	200	Irvine et al., 2013
	HadCM3L	Atmosphere	climateprediction.net	160	Frame et al., 2009
	FAMOUS	Atmosphere & Ocean	Latin Hypercube	100	Gregoire et al., 2011
NCAR	CAM3.1	Atmosphere	Bayesian Framework	518	Jackson et al., 2008
	CAM4.0	Atmosphere	Bayesian Framework	3000	Covey et al., 2011
	CAM3.5 (Slab)	Atmosphere	CAMCUBE	81	Sanderson, 2011
	CCSM3.5	Land Surface	CAMCUBE	108	Fischer et al., 2010
MPI	ECHAM5	Cloud/Radiation	Expert Solicitation	50	Klocke et al., 2011
	EGMAM	Cloud Parameters	Expert Solicitation	32	Nierhorster & Collins, 2009
ICTP	AGCM	Atmosphere	Precipitation Tuning	10	Neelin et al., 2010
JUMP	MIROC3.2	Atmosphere	Ensemble Kalman Filter	32	Yokohata et al., 2010
	MIROC5	Atmosphere	Ensemble Kalman Filter	42	Shiogama et al., 2012
	MIROC5	Atmosphere	Multi-Physics Ensemble	8	Watanabe et al., 2012

Table 1.3. Summary table of key perturbed physics ensembles. The model and methodology for creating the PPE is outlined along with the number of ensemble members and the key reference for each ensemble.

The QUMP project started at the UK Met Office working on HadAM3 and was developed to include HadSM3 and HadCM3 (Table 1.3). One key expert assumption has been used in the QUMP ensemble creation, which is the decision to not perturb some parameters. It has been assumed that this would not exert a significant change to the climate

produced in the ensemble (Murphy et al., 2009a). The early experiments were followed up by a series of ensembles and elements of the climate system looking at changes to extreme event frequency (Barnett et al., 2006) and global and local feedback mechanisms (Webb et al., 2006).

The UK Climate Projections 2009 project (UKCP09 – Murphy et al., 2009a) utilised HadSM3 and HadCM3 ensembles within the QUMP framework and a regional climate model to look at the impacts of both forcing and model physics uncertainty on projections of future anthropogenic climate change on the British Isles. UKCP09 is publically available in the UK for developing adaptation plans for the potential impacts of climate change. It forms a resource that explains how the projections are created and how to use them for creating adaptation plans for climate change. The data available covers climatological parameters (temperature, precipitation, air pressure, humidity and cloud cover), observed climatological trends and also projections for sea level rise, storm surge frequency, currents and salinity (Murphy et al., 2009a). UKCP09 dealt with three of the four main types of modelling uncertainty with an averaging period used to reduce the impact of natural variability. Three emissions scenarios were used to look at the impact of scenario uncertainty and the use of probabilistic projections using the QUMP ensemble to account for modelling uncertainty. By creating the '*grand ensemble'* of parameter and scenario condition uncertainties, UKCP09 was able to increase the scope of the modelling uncertainty investigated. The results produced which they considered the strongest model projections for the 21st century, were packaged for use in adaptation planning. UKCP09 included a '*discrepancy*' term (Sexton et al., 2012a) a measure of the extent of parameterisations, the way they had been approximated in the model and how this affected the ability of a model to reproduce the real world climate system. A model output was represented as a function of its parameters and settings plus the model discrepancy (Sexton et al., 2012a). The discrepancy term was applied to the UKCP09 probabilistic projections resulting in a narrowed probability distribution of Charney sensitivity for global reconstructions and UK climates (Sexton et al., 2012b). Data from UKCP09 has been used by studies looking at climate impacts on building design (Eames et al., 2011; Watkins et al., 2011), species distribution (Buckley et al., 2011) and extreme event hazard management (Blenkinsop et al., 2012) amongst many potential impacts from climate change.

Paper	Model	Charney Minimum Value	Charney Maximum Value
Stainforth et al., 2005	climate*prediction*.net HadSM3	1.7	9.9
Sanderson, 2011	CAM3.5	2.2	3.2
Shiogama et al., 2012	MIROC5	2.2	3.2
Piani et al., 2005	CPDN HadSM3	2.2	6.8
QUMP HadCM3 Ensemble	HadCM3 QUMP	2.2	7.1
QUMP HadSM3 Ensemble	HadSM3 QUMP	2.2	6.9
Watanabe et al., 2012	MIROC MPE	2.3	5.9
Jackson et al., 2008	CAM3.1	2.4	3
QUMP HadAM3 Ensemble	HadAM3 QUMP	2.4	5.4
Yokohata et al., 2010	MIROC3.2	4.5	9.6
	Average of PPEs	**2.43**	**6.1**
	CMIP3	*2.1*	*4.4*
	CMIP5	*2.1*	*4.7*

Table 1.4. The maximum and minimum Charney sensitivities for perturbed physics ensembles were available. For comparison of the range in each PPE, the minimum and maximum Charney sensitivities for the CMIP3 and CMIP5 multi-model ensembles is displayed as well as the average for these available PPEs. The HadAM3, HadSM3 and HadCM3 QUMP estimates represent the Charney sensitivities declared for the ensemble members.

The climate*prediction*.net project (Allen, 1999; Allen & Stainforth, 2002) outsourced its computing requirements by getting members of the public to download versions of the HadAM3 and HadSM3 models with 6 parameters for large scale cloud processes perturbed simultaneously. By using home computing power (a Public Resource Distributed Computing (PRDC) approach - Rougier et al., 2009) CPDN were able to run many more ensemble members than would be computationally feasible on a single super computer (Piani et al., 2005; Stainforth et al., 2005). Initial work investigated quantifying uncertainty in Charney sensitivity from the ensembles (Piani et al., 2005; Stainforth et al., 2005) with later work looking at the response of the climate in these ensemble members to rising greenhouse gases and also looking at the feedback effects (Sanderson et al., 2008a,b). climate*prediction*.net found a wide range for Charney sensitivity (1.7 to 9.9°C – Table 1.4), which was attributed to clear-sky longwave feedback in the entrainment rate parameter, compared to other models (Sanderson et al., 2008a,b; Sanderson 2011).

Output from the QUMP and climate*prediction*.net projects were combined for comparing the two PPE experiments for HadSM3 (Rougier et al., 2009). The aim was to analyse the different results for Charney sensitivity from both experiments and to also see if statistical tools could identify the most important parameters that were being perturbed. Firstly, owing to simplifications in the climate*prediction*.net study required to get the project onto home computers, it was not possible to combine the two ensembles for direct comparison (Rougier et al., 2009). An emulator for QUMP sensitivity was created across the range of parameters perturbed in the QUMP methodology, enabling a detailed sensitivity analysis to be developed on the ensemble members. The results showed that QUMP has a marginally greater Charney sensitivity than the climate*prediction*.net methodology, but also a greater uncertainty in the Charney sensitivity (at a confidence interval of 95%) than climate*prediction*.net (Sanderson et al., 2008a; Rougier et al., 2009). Hodson et al. (2013) assessed structural and parameter uncertainty in the Arctic using the CMIP3 MME and a HadCM3 PPE perturbing the QUMP parameters. However, the parameterisations were chosen using Latin Hypercube Sampling. As with Vellinga & Wu (2008), no flux adjustments were applied to the ensemble, named 'THC-QUMP' (Table 1.3). Hodson et al. (2013) noted that a lack of observations especially involving the ocean heat transport and sea ice extent were the main source of uncertainty in the models. Between the ensembles, ocean heat transport was found to be a structural uncertainty (as it played little role in the THC-QUMP ensemble) while the sea ice albedo parameterisations were a significant parameter uncertainty (Hodson et al., 2013).

The projects described above have focussed on the perturbation of atmospheric parameterisations, where the bulk of the QUMP PPE research has been undertaken. More recently work has begun to investigate parameters outside the atmospheric component of HadCM3. Collins et al. (2007), undertook an oceanic perturbation and did not discover significant results over the timescale of the 21st century (compared to atmospheric components – Murphy et al., 2009a). Brierley et al. (2010) detected vertical heat flows that when perturbed had a significant affect in comparison with natural variation, although the changes are smaller than those related to atmospheric perturbations. However, in a data-model comparison with SSTs over the longer timescales of palaeo-experiments, oceanic perturbations could become a stronger influence on the simulated climate. The sulphur cycle was investigated (Ackerley et al., 2009) showing that the uncertainty within the model, due to parameterisations of the aerosol processes was similar in magnitude to scenario uncertainty. Booth et al. (2012, see also Booth & Jones, 2011) investigated the sensitivity of the climate response to

global warming from perturbations in terrestrial carbon cycle processes in HadCM3. By perturbing parameters linked to leaf processes affecting temperature, nutrient (including CO_2) fixation and water on photosynthesis and two soil process parameters affecting soil respiration and water content, they discovered a greater plausible range in CO_2 emissions from carbon-climate feedbacks in one scenario (SRES A1B) than the emissions range covering the full range of SRES scenarios. The largest uncertainty came from the metabolism of photosynthesis as it responds to temperature. The main implications of this work are that a poor understanding of carbon cycle feedbacks is weakening our understanding of the impacts from future climate change (Booth et al., 2012).

Jackson et al. (2012) used the HadSM3 and THC-QUMP HadCM3 ensembles for future climate change experiments under constant, doubling and quadrupling CO_2 simulations. The key difference in this study compared to most HadCM3 PPEs (including this ensemble) is that no flux adjustments were applied to the PPE. The study was able to investigate the sensitivity of meridional overturning circulation (MOC) to the uncertainty in climate models under different greenhouse forcing scenarios. Although no modelled shutdown of MOC occurs, there was a reduction in the strength of MOC showing a negative relationship to global mean temperature that is not usually seen in a MME (Jackson et al., 2012). Clark et al. (2010) used an updated set of simulations for the HadSM3 PPE (Table 1.3) to investigate whether the heat wave risk was limited by global warming targets (i.e. warming not in excess of 2°C by 2100 – May et al., 2008). They found that some ensemble members still showed a chance of extreme heatwaves despite having achieved the global warming targets, indicating that global averages may not fully prevent some extreme events on regional scales, an important consideration for setting targets for mitigation, frequently discussed as global means.

The extensive database of simulations produced by the QUMP project has led to many global and regional experiments. Work has been conducted in recent years investigating effects on rainfall patterns and water management in the UK. Changes in UK extreme precipitation response to climate changes were dependent on season. With changes to winter extremes in an A1B scenario (compared to a 1920s control simulation) detectable by 2010, but summer extreme precipitation changes were not detectable until the 2080s (Fowler et al., 2010). Using the PPE data combined with water resource models, Lopez et al. (2009) evaluated the ensemble against data and also tested the climate data under changing demand scenarios. Repeating the results with CMIP3 data, Lopez et al. (2009) determined that the PPE investigated a broader range of modelling uncertainty than the MME and was therefore of greater use in

investigating adaptation strategies. Although using PPEs requires more expertise, it is rewarded by more robust planning decisions (Lopez et al., In Prep).

Frame et al. (2009) combined some of these different component ensembles to create a PPE where multiple components have been perturbed to investigate the relationship between different component uncertainties. 50,000 simulations were produced, with analysis of this data still underway. Rowlands et al. (2012) took ~9,000 simulations with atmospheric and sulphur cycle perturbations and analysed the response of the climate in these ensemble members for climate change under an SRES A1B scenario to 2050. They found their results were consistent with the IPCC projections (IPCC, 2007) for this time period, but that they were to the higher end of the results expected from a MME.

Rowell (2012) investigated four ensembles, HadSM3, HadCM3 and HadCM3C QUMP ensembles along with the CMIP3 MME (Meehl et al., 2000; 2007) with the aim of understanding sources of uncertainty for regional precipitation under future climate change. Uncertainty in atmospheric parameters was found to play a large role in precipitation patterns over tropical land and mid-latitude continental regions during the summer months. The same is true from uncertain sea ice processes in polar regions and it is concluded that over these regions, better understanding of the representation of these processes could improve models. However in arid regions, internal variability is the dominant form of uncertainty and there is no methodology for model improvement to be found for it (Rowell, 2012).

The projects that have been run using the QUMP methodology have shown that the impacts of the changes to the version of the model in each PPE member has had a *'coherent'* response on the climate system (Sexton et al., 2012a). The results give confidence for people to test using a PPE experiment based on the QUMP framework to get sensible modelling results for the climate they are testing.

1.4.3.2. Other Perturbed Physics Ensembles

Some studies have been undertaken using versions of the HadCM3 model, based upon but not strictly adhering to the QUMP methodology. Joshi et al. (2008) returned to single parameter perturbations for 7 selected parameters to create a PPE to investigate land/sea contrasts in the simulation of future climate change based on the Murphy et al. (2004) HadAM3 ensemble, generating a sensitivity test using the end member values for these parameters. Large land/sea contrasts were noted in ensemble members where a perturbed parameter increased the land cloud cover compared to the ocean

(Joshi et al., 2008). Outside of future climate change projections, there has been some palaeoclimate work, using PPEs based on the QUMP framework, looking at the LGM with the PalaeoQUMP project and some work in the mid-Holocene (Brown et al., 2008). Two PPEs have been produced using the low resolution version of HadCM3 known as FAMOUS (Jones et al., 2005) for the LGM (Gregoire et al., 2011) and the early Eocene (Sagoo et al., 2013).

Gregoire et al. (2011) produced a tuned version of FAMOUS using eight parameters in the atmospheric component and two parameters from the ocean component of the model. A Latin Hypercube Sampling approach was used to select parameter sets that explored the full range of model parameter space (Gregoire et al., 2011). A 100 member ensemble was run for the present day and the LGM, with ensemble members tested using the Arcsin Mielke Score (Watterson, 1996) against model development climate metrics and LGM SSTs from the MARGO dataset (Gregoire et al., 2011). Ensemble members were ranked from these tests and a final selection of *'good'* ensemble members were chosen. Selecting a single *'optimal'* version of the model, may lead to processes that represent key climatological processes being poorly represented in that version and a resulting weakness in the application of that version to different climates from the tuning period. For that reason, Gregoire et al. (2011) concluded that an ensemble of *'good'* model versions should be used as the final versions. Sagoo et al. (2013) produced a PPE for the early Eocene (55 to 48 Ma BP), utilising the methods of Gregoire et al. (2011). Simulations were run for 6000 years, with selected simulations extended to 10,000 years to allow the climate to reach equilibrium within the model. 17 simulations made the final ensemble, with 82 simulations failing to run, and a further 4 determined as being unstable (Sagoo et al., 2013). The final ensemble covered a range of potential temperatures and was compared to data from the Eocene. The ensemble member that was judged as being *'best'* was also the strongest member of the Gregoire et al. (2011) ensemble. A tuning of FAMOUS that can represent the warmest and coldest climates of the last 65 million years can be presumed to be a strong version of FAMOUS for recreating the impacts of future climate change. It also indicates that parameterisations influencing model climates affect the representation of warming and cooling.

The NCAR Community Atmospheric Model (CAM – AGCM with prescribed SSTs or slab ocean modes – Neale et al., 2013) and the Community Climate System Model (CCSM – AOGCM – Gent et al., 2011) have been used in PPE studies across several model versions. Jackson et al. (2008) used CAM3.1 (with prescribed SSTs) to investigate the response of optimised parameters selected using a Bayesian framework to study the

response of climate to a doubling of CO_2. The parameters were chosen from the representation of clouds and convection, and the optimisation was achieved by using Bayesian inference combined with a stochastic sampling strategy to produce sensitivity tests and generate PDFs. From the PDFs, new parameter values were selected to produce the ensemble (Jackson et al., 2008). Results indicate Charney sensitivity nearer 3°C rather than the CAM Standard value of 2.4°C (Table 1.4), but this was not an exhaustive selection. Covey et al. (2011) used CAM versions 3.6 & 4.0 (the latest iteration of the model) with prescribed SSTs in a PPE to study the response of historical measures of model performance to changes in the parameterisations. Sanderson (2011) used CAM3.5 (in slab mode) using methodology based upon climate*prediction*.net to create a PPE where the selected parameters were studied for their regional and global scale impacts on climate change. These results were then compared to climate*prediction*.net results. Charney sensitivity for the CAM ensemble ranged from 2.2 to 3.2°C, which is more tightly constrained than the climate*prediction*.net ensemble, due to differences in the clear sky cloud feedbacks in HadSM3 compared to CAM.

Perturbing land surface parameters in the Community Land Model (CLM – Lawrence et al., 2012) component of CCSM3.5, Fischer et al. (2010) investigated uncertainties in the projection of extreme events using a 108 member ensemble perturbing five parameters. The perturbed parameters investigated snow albedo, vegetation albedo, maximum rate of carboxylation, water depth and subsurface run off and roughness length. Although they found that there was a smaller range in the response to the perturbations compared to an atmospheric parameters PPE using the same model, there was a change in both the mean and seasonal responses of the model to the perturbations. Regionally this change was greater than displayed by the CMIP3 MME. The result comparing the land surface PPE to the atmospheric PPE is similar to the results from QUMP ensembles investigating other components with range in responses smaller than the atmospheric component, but still a notable effect (Murphy et al., 2011).

The ECHAM (AGCM – Roeckner et al., 2006) and the EGMAM (AOGCM – Huebener et al., 2007) models from the Max Plank Institute have also created PPEs. Using EGMAM, Niehorster & Collins (2009) investigated a sub-set of parameters focussing on entrainment rate and cloud formation (such as droplet size and ice fall speed). The parameters chosen were those that were the same between the two models (with the perturbed values based on the HadSM3 QUMP ensemble (Collins et al., 2006; Webb et al., 2006). They found that impacts on surface air temperatures were similar to

corresponding perturbations to HadSM3. HadSM3 had a greater magnitude of change due to structural differences in cloud feedback between the models (Section 1.4.3.1). Overall it was suggested that a collection of PPEs performed on different models alongside MMEs would produce the best results for evaluating and understanding climate model uncertainties (Niehorster & Collins, 2009).

Haerter et al. (2009) investigated the impacts on aerosol radiative forcing from the uncertainty in cloud parameters. Two experiments were run, one producing single perturbations of the selected parameters, across a range of ten values bracketing the expected parameter range. A second multi-perturbed parameter ensemble was created using Latin Hypercube sampling to create a 100 member ensemble. The ensembles investigated seven cloud parameters chosen for their uncertainty and relation to cloud optical properties which are strongly influenced by aerosols (Haerter et al., 2009). The largest uncertainty in IPCC published calculations of radiative forcing arises from aerosol radiative forcing (-1.8 to -0.3 Wm^2). Uncertainty in model aerosol parameters can account for 1.5 Wm^2 of forcing in the ECHAM5 model (Haerter et al., 2009). Lohmann & Ferrachat (2010) conducted a similar PPE, but with fewer perturbed parameters. They found that the parameter uncertainty was less than the structural uncertainty for the total anthropogenic aerosol effect. Klocke et al. (2011) ran the ECHAM5 model creating a 50 member PPE, focusing on the relationship between Charney sensitivity and cloud & radiation errors within the model with results compared to the CMIP3 MME. A tight range of parameter values for twelve parameterisations such as entrainment rate and cloud water to rain conversion was assigned with parameter sets determined by Latin Hypercube sampling. They found it was possible to get the range of Charney sensitivity in a MME from a PPE perturbing just a single parameter, especially the entrainment rate coefficient in agreement with QUMP ensemble analysis (Rougier et al., 2009; Joshi et al., 2010).

Neelin et al. (2010) used the International Centre for Theoretical Physics (ICTP) AGCM (Molteni, 2003) coupled to a slab ocean to investigate the disagreement between models in seasonal and regional precipitation and to guide the choices when choosing the best parameter sets for climate models. Four parameters (minimum wind speed, relative humidity, cloud albedo and a viscosity parameterisation) were perturbed, with parameter sets chosen through the application of a statistical method, the quadratic metamodel fit (Neelin et al., 2010). They found that using their method for choosing the optimal parameterisations tended towards the limits of the feasible parameter range. The results are relatively unique with parameterisations used in other PPEs coming from across the range of feasible values.

The Japan Uncertainty Modelling Project (JUMP) use versions of the MIROC AOGCM (K-1 Model Developers, 2004) to create PPEs. Yokohata et al. (2010) created a 32 member ensemble for MIROC3.2 using an ensemble Kalman filter (Annan et al., 2005a,b) to select the combinations of 13 perturbed parameters that best recreated key observed climate variables and compared it to the equivalent ensemble from the QUMP project for HadSM3. They found similar responses between the two models for the parameterisations applied, but that structural differences played an important role in the variations, concluding that both parameter and structural differences are important for understanding uncertainty in models. Yoshimori et al. (2011) used the MIROC3.2 ensemble, but ran it for the LGM. They found that the range of Charney sensitivities in the LGM experiment was reduced (compared to a $2xCO_2$ experiment), but the response of ensemble members in both experiments gave strong support to using LGM experiments to support future climate change experiments. Using MIROC5, Shiogama et al. (2012) created an ensemble which would not require flux adjustments unlike most AOGCM PPE experiments (i.e. Collins et al., 2006; 2011) using Latin Hypercube sampling of cloud parameterisations of phase and turbulence. Ocean metrics can be investigated because the ocean is not being adjusted to reduce drift from imbalances in the top of the atmosphere radiation budget. An ensemble mean Charney sensitivity of 2.2 to 3.2°C was found, similar to results from the CAM model (Sanderson, 2011 – Table 1.4). Watanabe et al. (2012) developed a multi-physics ensemble (MPE), which is a hybrid ensemble, investigating structural differences and parameter differences between models. It is most useful when comparing similar models with differences in the targeted area of model uncertainty, such as shortwave cloud feedbacks (Watanabe et al., 2012). The parameterisations for cumulus convection, large scale condensation, cloud microphysics and turbulence from MIROC3.2 were inserted into MIROC5 individually and as a group. The results indicated that no one parameter scheme controlled the differences between MIROC3.2 and MIROC5, but that the coupling of parameter schemes did affect the model results (Watanabe et al., 2012).

Across the variety of PPEs, there have been a number of methods and parameters investigated. The original QUMP ensembles investigated the widest set of atmospheric parameters, but sampled a much more tightly constrained range for these parameters, with the parameter values selected by expert opinion. The majority of other PPEs have used a smaller set of parameters (7 – 12 parameters), but used Latin Hypercube sampling to fully investigate the parameter space for these sets of parameters. Only the CCSM and HadCM3 models have had PPEs designed to investigate the uncertainty in other components of the model. Consistently across the various PPEs the response of

temperature varies across the ensemble, with some members displaying a greater range in features than MMEs where comparisons have been undertaken. However, are the PPEs that are being used actually creating realistic versions of the models used? The QUMP method fixed on the target of producing realistic versions of the model that could have been the final release version, however, ensembles generated through more thorough sampling of the parameter space, may not replicate this. Therefore, can it be assumed that these ensembles are valid for use?

The reliability of MMEs and PPEs was investigated by Yokohata et al. (2012), using two MMEs: CMIP3 AOGCM and AGCM with slab ocean ensembles and four PPEs: HadSM3, HadCM3 QUMP, MIROC3.2 AGCM and CAM 3.1 ensembles. Further development included the CMIP5 archive and added the MIROC5-AOGCM and MIROC5 MPE. The reliability of each ensemble was investigated using a rank histogram based on whether observations can be considered to have been sampled from within the ensemble members. The methodology was based upon the investigation into the CMIP3 ensemble in Annan & Hargreaves (2010) which had determined that the CMIP3 MME was a reliable ensemble to generate probabilistic projections for climate change. The histograms should be level, but rarely are, with a dome distribution for ensembles that are over-dispersed (the ensemble data fully encompasses the observational data – Figure 1.7a) and a 'U' shape for ensembles that are under-dispersed (Figure 1.7b; Yokohata et al., 2012; 2013). Under-dispersion indicates that the ensemble either underestimate (the lowest rank member of the histogram) or overestimate (the highest rank) the observational data, meaning the ensemble does not represent the data fully.

The Annan & Hargreaves (2010) MME result was upheld for both the CMIP3 and CMIP5 ensembles (Yokohata et al., 2012; 2013). The consistency of the rank histograms between the previous analyses indicates that the method and the results for CMIP MMEs are robust (Yokohata et al., 2013). The PPEs and the MPE sampled all displayed the 'U' shaped rank histogram indicating the ensemble was not reliable (Yokohata et al., 2012; 2013).

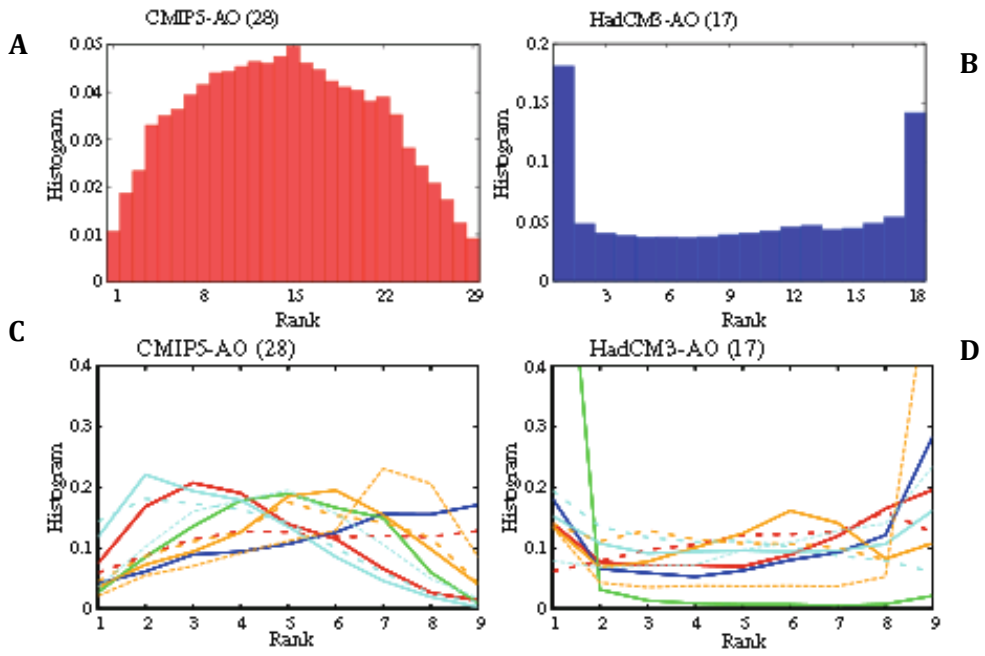


Figure 1.7. Rank histograms for reliability reproduced from Figures 1 & 2 of Yokohata et al. (2013) for A) CMIP5 all metrics, B) HadCM3 PPE all metrics, C) CMIP5 by metric and D) HadCM3 PPE by metric. Metrics represented include: SAT (red), SAT trend (Red dashed), precipitation (blue), MSLP (green), shortwave radiation (yellow) and longwave radiation (light blue) measured as net radiation (solid), cloud feedback (dotted) and clear sky (dashed).

Yokohata et al. (2013) showed that the HadCM3 QUMP ensembles are very close to the threshold of reliability. Breaking the rankings down into individual metrics determined that for variables such as surface air temperature and precipitation the climate means of the PPEs investigated were also reliable (close to dome shape Figure 1.7d; Yokohata et al., 2012; 2013). However there was not a wide enough sampling range for variables such as mean sea level pressure (MSLP) or radiation and cloud forcings in PPEs. Yokohata et al. (2012) were unable to determine if this was a feature of the models or the methodology to select the parameterisations. It was found to be more of a problem for ensemble members with higher Charney sensitivity values, generally outside of the IPCC range for expected Charney sensitivity. The result is supported by work from Joshi et al. (2010) which found for HadSM3 the highest Charney sensitivity ensemble member had a potentially unrealistic value for entrainment rate, which had been found to exert the greatest influence on HadSM3 Charney sensitivity by Rougier et al. (2009). Both ensemble methods are important for reducing uncertainty in climate projections, but careful experimental design is required to ensure that the PPEs are reliable (Rougier et al., 2009; Joshi et al., 2010; Yokohata et al., 2012).

1.4.3.3. Summary of PPEs

Perturbed Physics Ensembles (PPEs) have been run on different members of the CMIP3 MME family looking at a range of causes of parameter uncertainty in these models. The most thoroughly investigated is the HadCM3 family of models, with parameterisations involving atmospheric, ocean, carbon cycle and sulphur cycle components perturbed (Table 1.3). As with other PPEs, the key parameterisations involve the representation of radiation and cloud processes (Webb et al., 2006; Niehorster & Collins, 2009; Sanderson, 2011; Watanabe et al., 2012)). In all models that have undergone PPE investigation there has been a wide range of variation once parameters are perturbed, with some models displaying different responses to the perturbations. However it is consistent that atmospheric parameters are the most important component of GCMs for future climate projections. PPEs for the HadCM3 models have been shown to investigate a range of Charney sensitivities similar to or greater than MMEs (Table 1.4), however, this is not uniform across all models used for PPEs (Collins et al., 2011; Klocke et al., 2011; Sanderson, 2011; Shiogama et al., 2012). The widest variation in PPE Charney sensitivity in comparison amongst PPEs and to MMEs comes in the maximum estimate, with consistently similar lower bound estimates of Charney sensitivity around 2.2°C (Table 1.4).

Results from PPEs have been compared to the CMIP3 & CMIP5 MMEs to assess their reliability. PPEs have been found to perform reliably and within expected ranges for key variables such as temperature and precipitation with less reliability for mean sea level pressure or clear sky shortwave radiation (Yokohata et al., 2012; 2013). The cause of the unreliability is due to under dispersion of the model results compared to the observations, resulting in an over-estimation or under-estimation of model results by the ensemble members compared to observations. The authors suggest that this is a result of the initial PPE methodology, targeting Charney sensitivity at the cost of the representation of dynamical processes, leading to a narrow representation of the variables within the ensemble (Yokohata et al., 2012). While this could be deemed a weakness of a PPE over an MME, the PPE does perform well against key climate variables for understanding future and past changes in climate and also provide the best solution for investigating model parameter space (Yokohata et al., 2012). Of the PPEs tested in the Yokohata et al. studies, HadCM3 is close to the reliability threshold, giving confidence in its use for investigating parameter uncertainty.

1.4.4. Boundary Condition Uncertainty

In palaeoclimate reconstructions the uncertainty in the boundary conditions stems from uncertainties due to the spatial and temporal limits of data reconstructions. This leads to uncertainty arising from:

- Palaeo-environmental data reconstructions (SSTs, vegetation)
- Topography and bathymetry
- Ice sheets and sea level
- Greenhouse gas concentrations

Early modelling of past climates used AGCMs (i.e. Chandler et al., 1994; Haywood et al., 2000a) which required prescribed SSTs which were provided from reconstructions for the time period to be modelled. Uncertainty in AGCM simulations arises from uncertainties in the methods used to create the SST reconstruction.

To generate a dataset with the required global coverage to reproduce a warm period of palaeoclimate interest, often requires data covering time scales of hundreds of thousands to even millions of years (Lunt et al., 2012b; Haywood et al., 2013b; Sagoo et al., 2013). Within these timescales orbital variations affect the drivers of climate change. While these datasets produce good 'average' conditions for a time period, they are unable to aid a development of a full understanding into the drivers of the change. The low temporal or spatial availability of data sometimes required the use of data from unfavourable regions, such as areas of upwelling. Brierley et al. (2009) applied a correction to their early Pliocene data for sites that were located in regions of upwelling, based on core top data. This correction is a huge uncertainty for a reconstruction in a model, as it requires trust that conditions have remained consistent in the strength of the upwelling, which may not be a realistic assumption (Etourneau et al., 2009; Filippelli & Flores, 2009).

There is also uncertainty in the technique applied to a site to calculate the SST reconstructions, which comes from the analytical techniques used to create the reconstruction. As discussed in Section 1.3.3 *ii* Mg/Ca analysis requires assumptions about the ratio of magnesium to calcium in the palaeo-ocean (Lear, 2007). Additionally, while alkenones are assumed to be *"near surface"*, their water depth of formation is dependent upon latitude (Oukouchi et al., 1999). Reconstructions based on faunal analysis require assumptions about the environments of modern and palaeo taxa (Dowsett & Robinson, 1998). A number of palaeo-environmental reconstructions will utilise multi-proxy reconstructions across the sites. Uncertainty can be created when

two proxy techniques have a large mismatch, such as the comparison of SSTs from Mg/Ca and alkenones from ODP site 847 (Dekens et al., 2008). There is the possibility that both proxies are recording the temperature, with different results as each method reconstructs different parts of the water column or temperatures from different seasons (Dowsett & Robinson, 2006; Haywood et al., 2013b). Uncertainty can also arise during the development of a new proxy. TEX_{86} is a new proxy and the relationship to temperature is calibrated to core top samples, but this may not represent the palaeo SSTs at all sites (Schouten et al., 2002). Hollis et al. (2012) showed strong correlation between the SSTs from TEX_{86} and the inorganic proxies at some sites, but for other sites there was a proxy-proxy mismatch of ~5°C for Palaeogene SSTs. The mismatch, focussed in the Southwest Pacific, was largely due to the TEX_{86} calibration used, with TEX_{86}-$_{Low}$ being a better calibration than previous calibrations (Hollis et al., 2012). The final uncertainty comes from the method for extrapolating from a few sites to a global reconstruction. AGCMs take values from these reconstructions and then extrapolate to create a global reconstruction using modern SST maps (i.e. PRISM SSTs – Dowsett, 2007) or fitting a constructed zonal mean temperature to the global oceans by applying empirical calculations (Brierley et al., 2009).

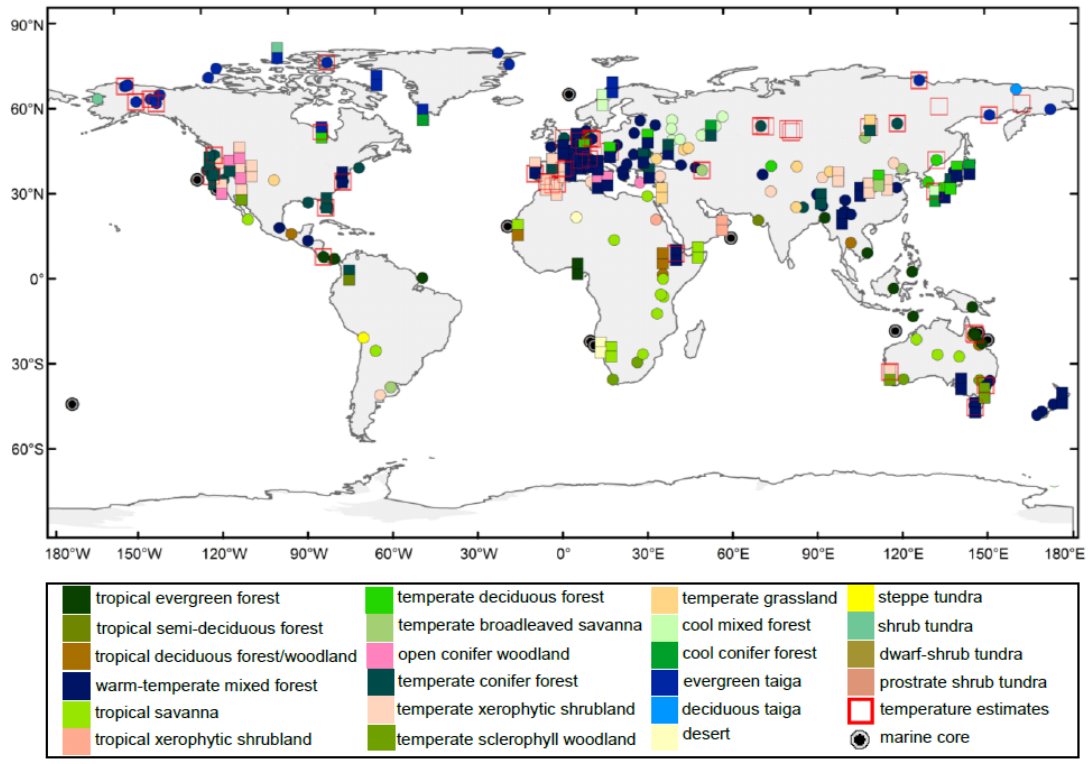


Figure 1.8. The 202 data sites for the Salzmann et al. (2008) vegetation biome reconstruction (see Figure 1.3) updated to indicate sites which have either a warm or a cold climate biome interpretation. Sites in squares identify potential changes to the biome reconstructed and the two potential biomes are displayed together.

AOGCMs, which simulate their own ocean temperatures, do not require the input SSTs removing the occurrence of this uncertainty. However, most AOGCMs still require prescribed vegetation which remains fixed throughout a simulation. Palaeo-vegetation reconstructions are often temporally limited due to poor age control in the samples (Salzmann et al., 2008), leaving subjective choices to be made by the investigator. As a result, the PRISM3D vegetation reconstruction (Salzmann et al., 2008) is a reconstruction for the Piacenzian age (3.6 to 2.6. Ma BP), compared to the 300 Kyr PRISM interval (3.264 to 3.025 Ma BP). Uncertainty in the reconstruction can affect the model results, as the land surface scheme interacts with the atmosphere component of the model. Figure 1.8 highlights where there are potential deviations from the PRISM3D biome reconstruction when allowing for variability due to changes in orbital forcing or greenhouse gas concentrations. Examples of vegetation uncertainty between the two reconstructions exists in the USA, between the reconstruction of open conifer woodland or temperate xerophytic shrubland or in central Russia where conifer forest changes from temperate to cool. The changes in vegetation affect surface albedo and regional precipitation, and changes in the type of vegetation are very important over high latitudes due to interactions with snow and winter albedo (Betts & Ball, 1997). Sensitivity studies have shown that running mPWP simulations with pre-industrial vegetation can reduce mean annual temperatures by 0.6°C (Hill et al., 2011), making vegetation uncertainty an important factor to consider in any reconstruction of the mPWP climate. Recent work (Pound et al., 2013) has shown the importance on regional scales of using a more accurate mPWP lakes and soils reconstruction.

The topography and bathymetry of a geological period is vital to palaeoclimate modelling studies forming the land-sea mask within the model. Key features for uncertainty in the reconstructions arise from the elevation of mountain ranges which affect atmospheric flows (Foster et al., 2010; Hill et al., 2011) and oceanic gateways (Krebs et al., 2011; Hill et al., 2011; Robinson et al., 2011). The reconstructions are based on a variety of evidence, from rock exposures, geophysical modelling of mantle plumes, palaeontology and geochemistry. Changes to these features are considered to have been major drivers of climates in various periods, such as the Eocene-Oligocene Transition, where the opening of the Drake Passage is proposed as being a driver of the development of an EAIS and rapid global cooling (Kennet, 1977; Katz et al., 2011). Similarly the closure of the Central American Seaway was proposed as the trigger for Northern Hemisphere glaciation (Saranthein et al., 2009), although modelling studies suggest this was not the cause (Lunt et al., 2008a). Miocene climate transitions have been proposed to have been controlled by orographic changes such as the uplift of the

Rockies or Himalayas (Currie et al., 2005; Foster et al., 2010) and not by changes in greenhouse gas concentrations. Throughout this work, the main uncertainties centre on the timing of the events, a change in the PRISM3D compared to the PRISM2 boundary conditions arose because of a change in the timing of the uplift of the Rockies from late Pliocene to Miocene (McMillan et al., 2006; Moucha et al., 2008).

Ice sheet and sea level reconstructions are important as ice sheets exert an influence upon the global climate through albedo (Clark et al., 1999; Weaver et al., 1999) and regional climate as a topographic feature (Gregory et al., 2012). Ice sheets are reconstructed through a number of approaches (e.g. Hill et al., 2007). The methods used include: geomorphological reconstruction of ice sheets (Kleman & Borgstrom, 1996), sea level reconstruction (Dowsett & Cronin, 1990), benthic $\delta^{18}O$ isotopes (Zachos et al., 1997; Naish et al., 2001), palynology (Willard, 1996) and data-model hybrid approaches (Hill et al., 2007; Hill, 2009). The hybrid approach uses climate models to force offline ice sheet models to reconstruct ice sheet development (DeConto & Pollard, 2003) or as a fixed element of a palaeo-environmental reconstruction (Hill et al., 2007). Changes in sea level affect the flow of water through key gateways, such as the Bering Strait, which can be opened or closed by changes in sea level (Shaffer & Bendtsen, 1994; Goosse et al., 1997; Hasumi, 2002).

Study	Analytical Method	CO_2 Value (ppmv)
Kurschner et al., 1996	Stomatal Indices	360 to 380
Raymo et al., 1996	$\delta^{13}C$ Marine Particulate Organic Matter	329 to 435
Pearson & Palmer, 2000	Boron Isotopes	280
Tripati et al., 2009	Boron/Calcium	280 to 300
Pagani et al., 2010	Alkenones	280 to 400
Seki et al., 2010	Alkenones	370 to 400
Seki et al., 2010	Boron Isotopes	400
Bartoli et al., 2011	Boron Isotopes	300 to 400
Stults et al., 2011	Stomatal Indices	351

Table 1.5. Estimates of mid-Pliocene atmospheric CO_2 and the analytical methodology used to calculate the estimate.

Greenhouse gas concentrations especially atmospheric carbon dioxide (CO_2), are generally a poorly constrained boundary condition for many palaeo-modelling studies. Unlike the last 800 Kyrs were the atmospheric concentration can be extracted from the ice core record producing a well constrained value for the gases (Luthi et al., 2008), deeper time reconstructions require estimates of CO_2 from a range of sources. Sources

include: alkenones, leaf stomata, pedogenic carbonates, boron isotopes and boron/calcium ratios (Table 1.5). There is no perfect proxy for CO_2 with many caveats being required for all the processes, which adds to the uncertainty generated. The effect can be seen in several studies, either applying the same method to different sites (Pagani et al., 2010) or applying the same methods as other work and getting different results (Seki et al., 2010). The variation between methods and locations is attributed to several factors. Pagani et al. (2010) analysed 6 sites spread across the globe, covering both the Atlantic and Pacific oceans. A potential bias for the alkenone CO_2 proxy is the concentration of phosphate in the near surface ocean and Pagani et al. (2010) had to make assumptions on oceanic phosphate concentrations over the last 5 million years. Although all the records display the same trend over the time period of a decreasing CO_2 level from 5 Ma to the present day, it creates a situation where there is uncertainty in the actual values for CO_2 (Table 1.5).

Seki et al. (2010) used boron isotopes and alkenones to investigate CO_2 with a multi-proxy approach at two sites. They found a good multi-proxy agreement at both sites and were able to produce an estimate for Pliocene CO_2 using both alkenones and boron isotopes. While the alkenones compared favourably with similar sites produced by Pagani et al. (2010), the boron data was very different to the data produced by Pearson & Palmer (2000). Seki et al. (2010) attribute this to analytical issues in the Pearson study (see Foster, 2008), preservation of material and the choice of species used between the two studies. Lower quality preservation for the Pearson & Palmer (2000) data would have altered the value for the boron isotopes and this would create this discrepancy. The multi-proxy nature of the *'Seki et al.'* study gives a greater confidence in their estimates of CO_2 at 370 to 400 ppmv compared to the 280 ppmv of the Pearson study.

1.4.5. Uncertainty Summary

The four types of uncertainty (natural variability, scenario, model and boundary condition uncertainty) in modelling climates arises from our level of knowledge and understanding of what is being modelled (Murphy et al., 2009a). Quantifying and understanding the model uncertainties is achieved through creating ensembles, either to investigate the structural uncertainty through MMEs or the parameter uncertainty in PPEs. Versions of these ensembles for future climate projection and for a range of palaeoclimate reconstructions are and/or have been undertaken. Boundary condition uncertainty for palaeoclimate simulations is less thoroughly investigated, due to uncertainty in the palaeo-environmental reconstructions creating a wide range for

some variables such as greenhouse gases. While crucially important for future climate projections, scenario uncertainty is not an issue for palaeoclimate modelling studies, replaced by boundary condition uncertainty. By using experimental designs to reduce or quantify these uncertainties, it is possible to improve the robustness of model reconstructions and projections of climate, as well as our understanding of the causes of uncertainty in models.

1.5 Project Rationale

Evidence that humankind is affecting the climate system is now overwhelming (IPCC, 2007). However, projections of the magnitude of possible future changes are limited by an incomplete knowledge of the skill of models for making these projections. Models are tested on their ability to reconstruct the observational period (Braconnot et al., 2012), however the magnitude of climate change in this period is relatively small (~0.75°C - IPCC, 2007). Palaeoclimate offers a solution to this with time periods that display a range of changes to global temperatures that can be reconstructed from palaeo-data enabling the dynamics of warmer world climates to be investigated (Haywood et al., 2009a; Braconnot et al., 2012). Although no perfect analogue from the geological record exists for the likely projections of 21st century climate change (Haywood et al., 2011b), the Pliocene, specifically the mid-Pliocene Warm Period (mPWP) is the most parsimonious epoch to study in this regard. It has a similar to modern continental configuration, elevated atmospheric CO_2 levels and warmer global mean temperatures compared to modern. The mPWP also has the largest global palaeo-environmental reconstruction for a warmer than modern climate enabling models to be more thoroughly tested than any other time in the geological record. As a result, the mPWP is an important geological period to study with models and palaeo-data to understand the climate system of a warmer than modern world.

The George Box axiom, *"All models are wrong, but some models are useful"* (Box & Draper, 1987) is ably demonstrated by mPWP modelling studies. The majority of mPWP climate model based studies have utilised a single climate model (Section 1.3.2) and have compared the results from the simulations to the PRISM palaeo-data. Data-model comparisons from these single models have highlighted strengths and weaknesses of individual climate models and enabled the identification of regions where a climate model is unable to reproduce the data (Section 1.3.4), but are unable to explain why. However, no account of the inherent uncertainty in climate model simulations has been accounted for in this previous work. A range of climate projection

studies investigating structural and parameter uncertainty have displayed the scale of the uncertainty on results produced by a climate models (Section 1.4.2 & 1.4.3).

Two initiatives have been undertaken to tackle this weakness in mPWP climate modelling. The first, is the Pliocene Model Intercomparison Project (PlioMIP – Haywood et al., 2010; 2011a; 2013a), which has produced a multi-model ensemble tackling the structural uncertainty in Pliocene climate modelling bringing together modelling groups from across the globe. The second is the subject of this thesis, the perturbed physics ensembles, designed to investigate parameter uncertainty in mPWP climate simulations. Although unconnected, the two are complementary, focussed on increasing our understanding in the mismatches between model and data reconstructions of the mPWP. The main component of the data-model mismatch for the mPWP focusses around the inability of fully coupled atmosphere-ocean general circulation models (AOGCMs) to simulate the high latitude warming of the mPWP climate, especially through the North Atlantic (Section 1.3.4).

Existing PPEs have focussed on future climate change projections (i.e. Collins et al., 2006; 2011), with work underway to investigate PPEs at the Last Glacial Maximum (Gregoire et al., 2011) and the Eocene (Sagoo et al., 2013). The Pliocene PPE presented here represents the first investigation of parameter and boundary condition uncertainty in a warmer than modern climate with similar continental configuration. It is the first palaeo-PPE to be tested against SST, SAT and vegetation biome data, using these three datasets to produce a combined ranking of the ensemble members.

1.5.1. Aims and Objectives

The thesis will investigate the contribution of parameter uncertainty and boundary condition uncertainty on the simulation of mPWP climate. Parameter and boundary condition uncertainty represent two of the three areas of model uncertainty relevant to palaeoclimate (see Section 1.4.1) forming vertices of the *"PMIP Triangle"* (Haywood et al., 2013a). The ensembles will be assessed in comparison with palaeoclimate data. The focus of these data-model comparisons will be to determine how the ensemble members vary in comparisons to individual palaeo-datasets and also across the combined range of palaeo-datasets available for use. The variation in ensemble member performance will enable the assessment of the ensemble and the impact of the two forms of uncertainty upon the simulation of the mPWP climate within HadCM3.

The aim of the thesis is to identify the ensemble members which reduce the existing data-model mismatch and which ensemble member performs best across the whole

range of palaeo-data utilised in this thesis. By utilising both sets of boundary conditions used in previous mPWP simulations, our PPE can also identify which boundary conditions produce the best data-model comparisons across the different palaeo-datasets used and in the combined rankings. The PRISM3D boundary conditions represent an improved understanding of the mPWP, but it is important to test whether they produce a stronger representation of the mPWP climate than PRISM2 boundary conditions. Similarly, the potential range of values of mPWP atmospheric CO_2 will be included within the assessment of ensemble member performance, allowing for uncertainty in this boundary condition.

The goals of the investigation are:

1. To investigate the mPWP global and regional climate responses to the parameter perturbations, including the effect on these responses from the potential range for mPWP atmospheric CO_2 concentrations.
2. To investigate the effect of changing the physical boundary conditions within the model, from PRISM2 to PRISM3D and the interaction between the boundary condition changes and the parameter perturbations
3. To investigate the effect parameter perturbations and boundary condition changes have on data-model comparisons to vegetation derived surface air temperature, sea surface temperature and vegetation biome data.

The perturbed physics ensemble presented here represents a thorough investigation of the uncertainty in mPWP climate simulations with AOGCMs due to the representation of sub-grid scale parameterisations and the physical boundary conditions. It is unlikely that any single ensemble member will provide a perfect reconstruction of the mPWP climate, however the range of performance across the ensembles and DMCs will provide useful information in where mPWP modelling can be improved. All models are wrong, but some members of our PPE are useful for reducing data-model mismatches.

1.6. Thesis Outline

The thesis will illustrate the progression of our PPE from conception, through the initial results to the full ensembles investigating parameter and boundary condition uncertainty and the ranking of these ensemble members with respect to palaeo-data.

i) Chapter 1 – Introduction

Chapter 1 has introduced the Pliocene as a geological epoch in the context of the Cenozoic and described the modelling and palaeo-data produced for the mid-Pliocene

Warm Period (mPWP). The existing data-model mismatch was discussed and the challenges to simulating mPWP climates that the mismatch poses. It also discussed sources of uncertainty in climate models and presents methods for tackling and quantifying uncertainty.

ii) Chapter 2 - Methodology

Chapter 2 will outline the methods to be used throughout the thesis. The HadCM3 and BIOME4 models are described along with the methods for creating Pliocene and modern simulations. The process for creating and running the perturbed physics ensemble members, including applying the flux adjustments and changes to the model ancillary files (where required) will be described. The two sets of physical boundary conditions, the PRISM2 and PRISM3D reconstructions, are described with differences between the two highlighted. The CO_2 sub-ensemble is introduced with the values for CO_2 used in the sub-ensembles outlined and the methodology for choosing the sub-ensemble members discussed. There is also a discussion about the methodological decisions made in the creation of these ensembles and why the QUMP method was followed to generate the ensemble members. Finally the palaeo-data and the methodologies for testing the performance of each ensemble member is outlined.

iii) Chapter 3 - Initial Results

Chapter 3 discusses an initial ensemble of the HadCM3 Standard and the two ensemble end members (in terms of their Charney sensitivity) that were run. The purpose of the initial ensemble was to test the feasibility of the PPE members and the methods to be used for testing the members, to determine their suitability for use in the investigation. The chapter is based on work published in Pope et al. (2011), but for this thesis additional metrics were produced to enhance the analysis. It finishes with an analysis of the developments that were made to the initial results methodology and were incorporated into the methodology in Chapter 2.

iv) Chapter 4 – Intra-Model Assessment of mid-Pliocene Climate: Can Geological Data and Climate Models be Reconciled?

Chapter 4 investigates the impact of the parameter uncertainty upon the simulation of the mPWP with HadCM3 across our ensemble members. Intra-model comparisons will highlight differences between the ensemble members and key climate features will be picked out, such as Polar Amplification within the ensemble. The importance of the potential range in mPWP atmospheric CO_2 will also be assessed through the CO_2 sub-

ensemble. Data-model comparisons will assess the performance of the ensemble members to the palaeo-data. From this, a conclusion as to the performance of the ensemble members in terms of the simulation of the mPWP will be drawn. The chapter will aim to answer three main questions:

1. *Is it possible to discriminate the model ensemble members using geological proxy data?*
2. *Are there implications for model estimates of Charney sensitivity?*
3. *Can models be reconciled with mPWP Proxy-data?*
4. *Which Methodology Produces the Best Comparison of Model and Data?*

v) Chapter 5 – On The Importance of Accurate Boundary Conditions for Modelling the mid-Pliocene Warm Period

Chapter 5 presents the results from the comparison between the PRISM3D and PRISM2 boundary condition ensembles, with the aim being to determine which set of boundary conditions represent the best basis for modelling the mPWP. The chapter will test the assumption that the improvements and refinements in the PRISM3D boundary conditions, compared to PRISM2, will be represented through improved model simulations. The same format of intra-model comparisons and data-model comparisons used in Chapter 4 will be utilised here to assess the performance of the boundary condition ensembles. The Chapter will aim to answer three research questions:

1. *Which boundary condition set produces the strongest mPWP simulations?*
2. *What are the implications of this result for Modelling the mPWP?*
3. *What is the role of boundary condition uncertainty in mPWP data-model mismatches?*

vi) Chapter 6 – Conclusions

Chapter 6 reviews and concludes the results from Chapters 4 & 5 within the context of the research aims for the thesis and outlines potential future uses of perturbed physics ensembles. A key conclusion will focus on how the simulations develop our understanding of AOGCM simulations of mPWP climate and the use of data-model comparisons to assess their performance. The chapter will finish with suggestions of future development for the work, drawing on the full range of work undertaken in the course of the thesis.

Chapter 2: Methodology

2.1 Introduction

The work undertaken for this thesis involves creating, running and testing a perturbed physics ensemble (PPE) based on the methodology of the UK Met Office 'Quantifying Uncertainty in Model Predictions' (QUMP) project. The perturbed physics ensembles utilise a fully coupled atmosphere-ocean general circulation model (AOGCM) set up for simulating the Pliocene. The investigation requires the application a variety of different models, modelling methods and analysis techniques. The climate modelling will be undertaken using the UK Met Office AOGCM, HadCM3. The PPE is created using the methodology of Collins et al. (2006; 2011) developed from the work of Murphy et al. (2004). The HadCM3 model is set up for simulating the Pliocene by inputting Pliocene specific physical boundary conditions, work which has been described for HadCM3 in several Pliocene modelling simulations (i.e. Haywood et al., 2000a; 2011a; Haywood & Valdes 2004; Lunt et al., 2008a,b). The model boundary conditions were produced by the US Geological Survey (USGS) Pliocene Research Interpretation and Synoptic Mapping (PRISM) group (Dowsett et al., 2010a; Haywood et al., 2010). An offline equilibrium vegetation model, (BIOME4 – Kaplan, 2001) will be used to model potential vegetation cover driven by the model outputs from the PPE members.

Results from the ensemble members will be analysed for changes in climate responses due to the parameterisations through intra-model comparisons, which will focus on the differences between an ensemble member and the HadCM3 Standard simulation for a variety of climate metrics. The performance of the ensemble members will be ranked using data-model comparisons to sea surface temperature (SST) data, surface air temperature (SAT) data and a vegetation biome reconstruction. Root mean square errors will be used to rank the temperature data and Cohen's Kappa statistic (Cohen, 1960) will be used for ranking the simulated biomes. These methods have previously been applied individually to assess the performance of Pliocene model simulations (SSTs - Haywood et al., 2004; Dowsett et al., 2012; SATs – Salzmann et al., 2013; vegetation biome - Haywood et al., 2009b), however Pope et al. (2011, Chapter 3) was the first use of combined (temperature and vegetation biome palaeo-data) methods to assess and rank Pliocene simulations.

Chapter 2 outlines the description of the models used, the datasets used to create the Pliocene boundary conditions for the model, and then the methods for analysing the

results (including statistical tests used) through intra-model comparisons, data-model comparisons and exploring the uncertainty. There will also be a discussion regarding the QUMP methodology, specifically the advantages and disadvantages of the decision to follow it.

2.2 Experimental Design

The aim of this thesis is to investigate parameter and boundary condition uncertainty in the HadCM3 model in a warmer world from the geological past using a perturbed physics ensemble. These aims will be achieved through three main components, the perturbed physics ensembles (PPEs) that will be created and the intra-model comparisons (IMCs) and data-model comparisons (DMCs) that will be used to assess the impact of parameter and boundary condition perturbations on climate simulations of the mPWP.

i) Perturbed Physics Ensembles

The perturbed physics ensembles have all been created using the basis of the UK Met Office QUMP project. The perturbed parameters utilised in this thesis to create the PPE was based on the AO-PPE-A ensemble from Collins et al. (2011), designed as part of the QUMP project. In total, 6 ensembles have been created and run, these were:

- Initial results ensemble – 6 members (3 PRISM2 boundary conditions, 3 Modern)
- Modern ensemble – 16 members and the Standard
- Pliocene ensemble (PRISM2 boundary conditions) – 16 members and the Standard
- Pliocene ensemble (PRISM3D boundary conditions) – 16 members and the Standard
- 300 ppmv Pliocene ensemble (PRISM3D Boundary Conditions) – 4 members and the Standard
- 350 ppmv Pliocene ensemble (PRISM3D Boundary Conditions) – 4 members and the Standard

All, but the initial results ensemble were run on the ARC1 high performance computing (HPC) resource at the University of Leeds. The initial results ensemble was run at the UK Met Office and results from this ensemble were only used in Chapter 3 (published in Pope et al., 2011). The perturbed physics ensemble used and the methodology for creating the PPE is discussed in Section 2.4.3. The impact on the perturbed physics ensembles of the change in HPC resources is discussed in Section 2.6.1.

Each ensemble was created by spinning up the Standard (unperturbed version of HadCM3) on the HPC platform being used (Met Office or ARC1). For both the Modern,

Pliocene PRISM2 and Pliocene PRISM3D Standards an existing control simulation was used to initialise the Standard simulation. These existing control simulations have been now been run for several thousand years. The Standards were run for a further 500 years in line with the PlioMIP experimental design (Haywood et al., 2011a) to establish the simulations on the HPC platform (differences between these resources such as compilers can cause minor influences on simulation climate means – Lunt et al., 2012a). Once the Standard had achieved 500 years, it was used to initialise each ensemble member, which was then run for the experimental phase, a further 300 years (Figure 2.1). Of this 300 year period, the final 30 years (years 271 to 300) were used to produce climatological averages for analysis in both intra-model and data-model comparisons.

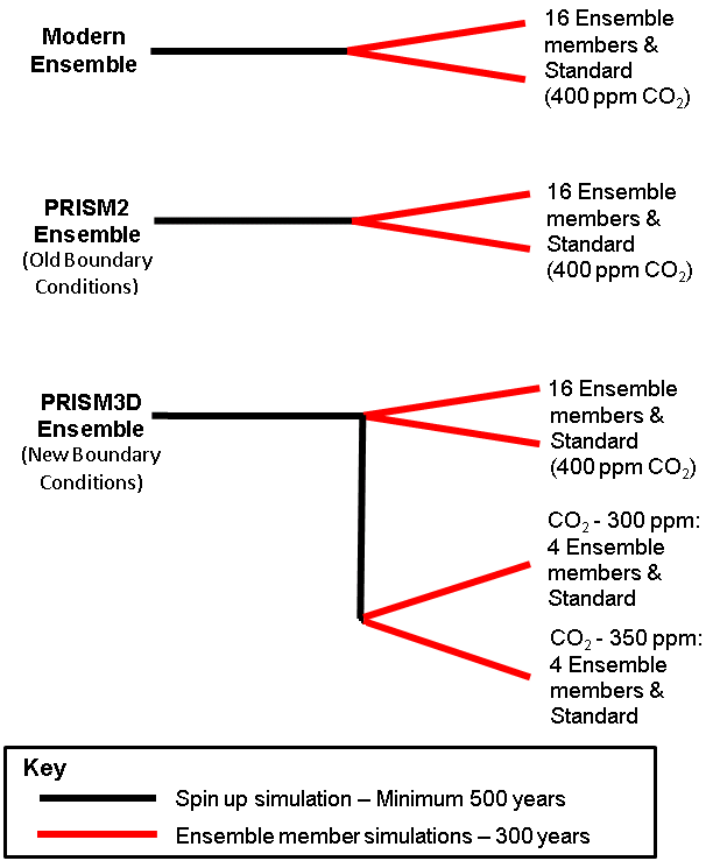


Figure 2.1. Graphical outline of the experimental design regarding the simulations produced for the investigation of Pliocene climate through perturbed physics ensembles.

The 300 year experimental phase represents a compromise between achieving computational efficiency and achieving an equilibrium climate state. It is not enough time for the simulated deep ocean to have reached equilibrium with the forcing changes imposed on the simulations. However, by running the simulations for 300 years, an approximate (or quasi) equilibrium can be reached with the atmosphere, land surface, sea surface and ocean mixed layers (Hewitt et al., 2002).

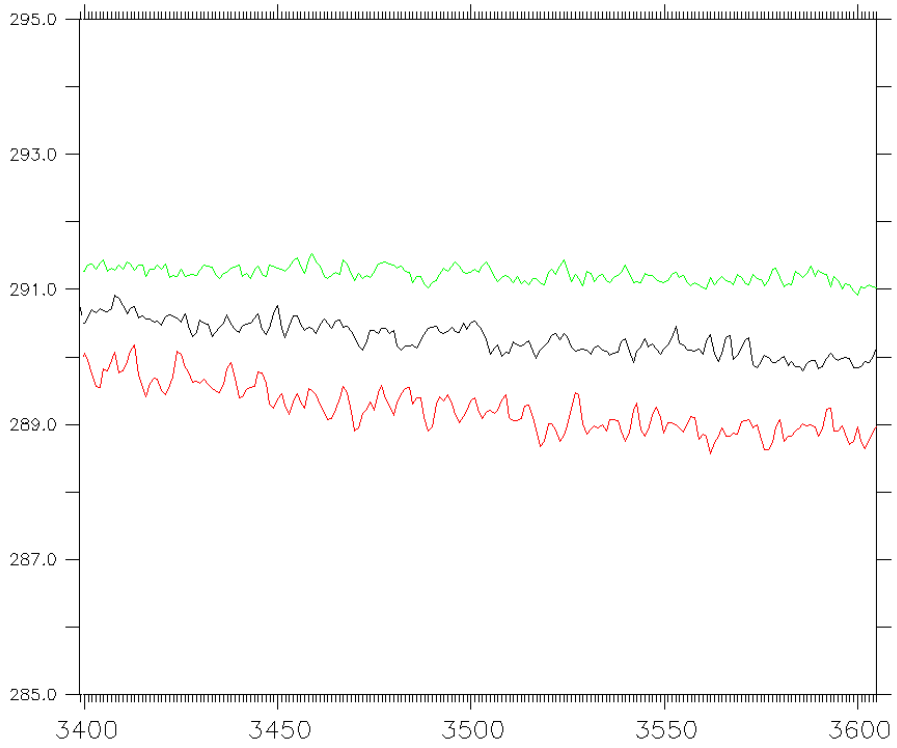


Figure 2.2. Global average time series plots for surface air temperature (K) for ensemble members: B (red), M (green) and the Standard (black) for the final 200 years of the simulations.

The ensemble members are judged as having achieved the state of approximate equilibrium if the residual change in forcing from the parameterisations in the surface air temperatures displays minimal trend. For ensemble member M, there are no perceptible trends in the temperature over the final century. For member B, there is a slight trend for a decrease in temperatures (0.2°C). However, this is not present in the final 50 years of the simulation. A similar trend occurs through the Standard. With the simulations reaching the quasi-equilibrium for the surface metrics, it is possible to take the averaging period as representative of the final climate of the simulation. All ensemble members were checked for these criteria prior to including them in the final analysis. Three ensemble members, A, C & R failed to meet these criteria and as a result these ensemble members were not included in the final analysis. The impact of this is further discussed in section 2.6.1.

Two full Pliocene ensembles have been run, labelled the PRISM2 and PRISM3D ensembles. These ensembles represent variations in the physical model boundary conditions available to be applied for modelling the mPWP. Both sets of boundary conditions were produced based on a palaeo-environmental reconstruction by the US Geological Survey PRISM team. However, the PRISM3D reconstruction (Dowsett et al., 2005; 2010a) represents an updated palaeo-environmental reconstruction compared to the PRISM2 (Dowsett et al.,

1999) reconstruction. The PRISM2 full ensemble is discussed in Section 2.4.1 and the PRISM3D full ensemble in 2.4.4. *i*.

Two sub-ensembles using different concentrations of atmospheric CO_2 were also produced. The full Pliocene ensembles were run with atmospheric CO_2 concentrations of 400 ppmv, however palaeo-data exists which suggest it could have been a lower value. The sub-ensembles use atmospheric CO_2 concentrations of 300 ppmv and 350 ppmv to investigate the potential range of atmospheric CO_2. These sub-ensembles are discussed further in Section 2.4.4. *ii*.

ii) Intra-Model Comparisons (IMCs)

Ensemble members will be compared to each other through intra-model comparisons to investigate large scale features and regional responses to the perturbed parameters compared to the Standard simulation. A range of model metrics will be used covering key climatological features such as temperature and precipitation. The methodology applied for creating and assessing the IMCs is outlined in Section 2.5.2.

iii) Data-Model Comparisons (DMCs)

The ranking of ensemble member performance will be achieved through data-model comparisons using three types of palaeo-data: a sea surface temperature (SST) dataset, surface air temperature (SAT) dataset and a terrestrial vegetation biome reconstruction. Root mean square error and Cohen's Kappa statistic will be used to determine the rankings for each ensemble member. By utilising these three palaeo-datasets, the final ranking of the ensemble members will not be biased towards an individual metric, but be based upon a member's ability to reconstruct mPWP climate across the terrestrial and oceanic realms.

The full details of the application of each DMC method used is outlined in Sections 2.5.3 (SSTs), 2.5.4 (SATs) & 2.5.5 (vegetation biomes). A methodological decision to use the modern Standard simulation to create the anomalies to the ensemble members for the DMCs could be seen as biased towards the ensemble Standard simulation. The rationale for this decision is outlined in Section 2.6.4.

2.3. Models

2.3.1. HadCM3

The PPE will be undertaken using the UK Met Office fully coupled atmosphere-ocean general circulation model (AOGCM) HadCM3, which contains atmosphere, ocean, vegetation and sea ice components (Gordon et al., 2000). HadCM3 was used to produce the perturbed physics ensembles for three principle reasons. Firstly, the most extensively tested PPEs have been produced for the HadCM3 models (see Chapter 1, Section 1.3.3.1). Secondly HadCM3 has been extensively used in palaeo-climate simulations because its ratio of model years to wall clock time is high allowing long simulations to be undertaken. Lunt et al. (2012) highlighted the weaknesses for palaeoclimate studies resulting from model simulations taking many months limiting the scope of the results. Finally, HadCM3 has been utilised in MMEs such as PlioMIP (Bragg et al., 2012; Haywood et al., 2013a), therefore utilising HadCM3 for the perturbed physics ensembles allows broad comparisons with other uncertainty work for the mPWP.

The atmosphere is comprised of 19 vertical levels with grid dimensions of 2.5° x 3.75° on a latitude-longitude grid (Gordon et al., 2000), which equates to a grid box at the equator of 278 Km latitude by 417 Km longitude. The model contains many features that are developments from its predecessor HadCM2 (see Johns et al., 1997) including a radiation scheme covering 6 additional spectral bands in the shortwave and 8 additional bands in the longwave. The model explicitly represents the radiative effects of all greenhouse gases not just CO_2, O_3 and H_2O (Edwards & Slingo, 1996; Gordon et al., 2000; Johns et al., 2003). Background aerosols in the model have been parameterised to include their effects on the climatology (Cusack et al., 1998; Johns et al., 2003). The penetrative convection scheme of Gregory & Rowntree (1990) has been developed to include a parameterisation of the impacts of momentum on convection and the downdraft of convection in the model (Gordon et al., 2000; Johns et al., 2003). HadCM3 employs the use of MOSES1 (Met Office Surface Exchange Scheme version 1– Cox et al., 1999) with developments in soil moisture responses to temperature (freezing/melting) and on the effect of CO_2 and stomatal resistance on evapotranspiration (Williams et al., 2001). Several parameterisations in the model are linked to features in cloud representation and cloud development. These are especially crucial surrounding the partitioning of mixed phase clouds and cloud formation such as the radius of cloud droplets and droplet numbers in clouds. Further details of these parameterisations are found in Gordon et al. (2000). There have also been several minor changes to the atmospheric component (compared to HadAM2 - Johns, 1996) and these are detailed in Pope et al. (2000).

The ocean is based on a Bryan-Cox type model (Roberts & Wood, 1997) comprised of 20 levels with a rigid lid on a 1.25° x 1.25° latitude-longitude grid (Johns et al., 1997; Gordon et al., 2000) which represents an equatorial grid box of 139 Km by 139 Km. The rigid lid results in no volumetric change in the ocean regardless of evaporation from the ocean or riverine inputs (Gordon et al., 2000). There are 6 ocean grid boxes for every atmospheric grid box in the coupling of the model. A key feature of the ocean component is the interaction with sea ice, and every high latitude grid box in HadCM3 has the ability to have sea ice cover (Gordon et al., 2000). Changes in sea ice albedo due to temperature change are represented using a linear response. A maximum sea ice albedo of 0.8 occurs at a temperature of -10°C and decreases linearly to 0.5 at the minimum sea ice temperature of 0°C (Crossley & Roberts, 1995; Gordon et al., 2000). The relationship represents the changes in sea ice albedo as warming occurs, reducing snow cover and increasing the number of melt ponds (Gordon et al., 2000). In places, ocean basin topographies had to be edited due to grid scale, this was found to have an especially sensitive response in the North Atlantic around Greenland-Iceland-Faeroes-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), this is an important consideration with respect to boundary condition uncertainty as both Pliocene and modern simulations use the same land sea mask.

Several parameterisations exist in the ocean fluxes and mixing ratios and the absorption of shortwave radiation in the surface waters, further details of which can be found in Gordon et al. (2000). Freshwater fluxes are calculated using the balance of precipitation and evaporation over land including a term for run off from land ice and allowing for sea ice formation (Pardaens et al., 2003). Freshwater is delivered to the coasts based on the configuration of river catchments and estuaries (Ivanovic et al., 2013). An important parameterisation in the ocean component is Mediterranean outflow to the Atlantic Ocean. In the ocean, this is a crucial flow and has an important impact on Atlantic waters masses 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, therefore a parameter (in the form of a diffusive pipe) exists to represent outflow of waters through the strait (Johns et al., 2003; Ivanovic et al., 2013).

The model runs with a time step of 30 model minutes for all processes and the atmosphere and ocean are coupled at the end of each model day (Gordon et al., 2000). The HadCM3 model has been widely utilised in climate modelling projections, palaeoclimate studies and in MMEs including the CMIP3 (Meehl et al., 2007); PMIP2, PMIP3 (Braconnot et al., 2007; 2012) and PlioMIP (Bragg et al., 2012; Haywood et al., 2013a). HadCM3 has been shown

to perform consistently strongly in tests against other AOGCMs as well as reproducing the main features of climate (Lambert & Boer, 2001; IPCC, 2007; Bragg et al., 2012).

2.3.2. BIOME4

BIOME4 (Kaplan, 2001) is a mechanistic equilibrium vegetation model which can be run off-line from the HadCM3 AOGCM and is used in these experiments to interpret the effects of the climates of ensemble members upon simulated biomes of the mPWP. BIOME4 has been previously used in palaeoclimate studies to assess the effects on vegetation of changes in climates (i.e. Haywood et al., 2002c; Krebs et al., 2011), and in producing the PRISM3D vegetation distribution through a data-model hybrid approach (Salzmann et al., 2008). BIOME4 is used in this thesis to produce predicted biomes based on the output of ensemble members which can be tested against the PRISM3D reconstruction using Cohen's Kappa statistic (Cohen, 1960).

It is driven by a year of monthly outputs from the ensemble member's climatological averaging period for surface temperature (at 1.5m), precipitation, cloudiness and absolute minimum temperature combined with inputs of soil moisture, soil depth and CO_2. BIOME4 is programmed with 28 biome classifications, which are determined based on the combination of dominant and sub-dominant plant functional types (PFT). There are 12 PFTs which cover distinct categories of flora from Arctic to tropical environments (e.g. tropical grassland or cool conifer woodland (Salzmann et al., 2009)). Using the climate inputs and the CO_2 concentration, BIOME4 calculates whether the growth of each PFT could occur within those climatic conditions through the calculation of net primary productivity (NPP). Parameters such as photosynthetic pathway and seasonal fluxes in temperature and precipitation are used along with semi-empirical rules related to balances between forest and grass taxa to determine the potential PFTs at each grid point; the combination of potential PFTs then determines the final biome (Kaplan, 2001). BIOME4 was run at the atmospheric grid resolution of HadCM3, enabling the required fields to be taken directly from the HadCM3 outputs without adjusting the resolution and interpolating inputs.

BIOME4 can be run using either absolute or anomaly climate inputs. Absolute climate forcing is the input method used for the BIOME4 plots produced and used in Chapter 3 (and Pope et al., 2011). To create the inputs for the absolute method, the 30 years of the simulation used for the climate averaging are taken and separated into constituent months. The 30 monthly means for each month are averaged for the required inputs (1.5m temperature, precipitation and cloudiness) and then a whole year of average

months is created and then used to drive the BIOME4 model (i.e. taking the 30 Januarys in the averaging period, averaging these to produce an average January, repeating for every month). The averaging method maintains the seasonal variations throughout the year compared to taking the annual means for the fields required, which would remove the seasonal variation. The absolute minimum temperature field is created by taking the average coldest monthly temperature for each hemisphere from the 30 year period of climatological averaging.

To run BIOME4 in anomaly mode, the anomaly inputs are created through producing a climate anomaly, for example 'Pliocene minus modern' for an average year. The anomaly inputs are combined with a set of standardised climate inputs taken from an observations run produced at the Bristol Institute for Global Dynamic Environments (BRIDGE) at the University of Bristol. For the 1.5m temperature, cloudiness and absolute minimum temperature, the BIOME4 anomaly is created by adding the anomaly climate inputs to the observed climate:

$$I = (P-m) + O \qquad \text{(Eq. 2.1)}$$

Where 'I' represents the input for the BIOME4 model, 'P' represents the climate metric from the Pliocene simulation, 'm' represents the modern simulation climate metric and 'O' represents the observations.

For precipitation & cloudiness the BIOME4 anomaly is created by multiplying the observed climate by the anomaly climate inputs

$$I = (P/m) * O \qquad \text{(Eq. 2.2)}$$

Soil parameters (soil water holding capacity & soil water percolation index) are also added as an input to the model along with the climatological variables. There are two methods for adding the soil parameters, either a globally constant variable or with geographic variation based on present day data (known as the 'alternative soil parameters'). The initial results (Chapter 3) used the geographically constant soil inputs while the full ensembles (Chapter 5) used the alternative soils. The changes in methodology between Chapter 3 and Chapter 5 were the result of discussions that arose as a result of the initial results paper (Pope et al., 2011) and the limitations in that methodology. The result of the discussions was the use of anomaly climate inputs and geographically varying soils. The final input variable to BIOME4 is CO_2, which is set (manually) to the value used in the primary simulation, for 'Pliocene minus modern' comparisons, CO_2 is at set at Pliocene levels of 400 ppmv.

2.4. Creating the Perturbed Physics Ensemble

2.4.1. Creating a Pliocene Model Simulation

The HadCM3 model is set up to simulate the Pliocene epoch and the mPWP specifically by adjusting the model boundary conditions. PRISM's second global palaeo-environmental reconstruction, the PRISM2 dataset (Dowsett et al., 1999) was used to create the Pliocene boundary conditions for the initial results and the PRISM2 full ensemble. These were the standard boundary conditions that have been used extensively in Pliocene climate modelling simulations such as Haywood et al. (2000a,b); Haywood & Valdes (2004); Lunt et al. (2008a,b) and in a data-model comparison study (Dowsett et al., 2011).

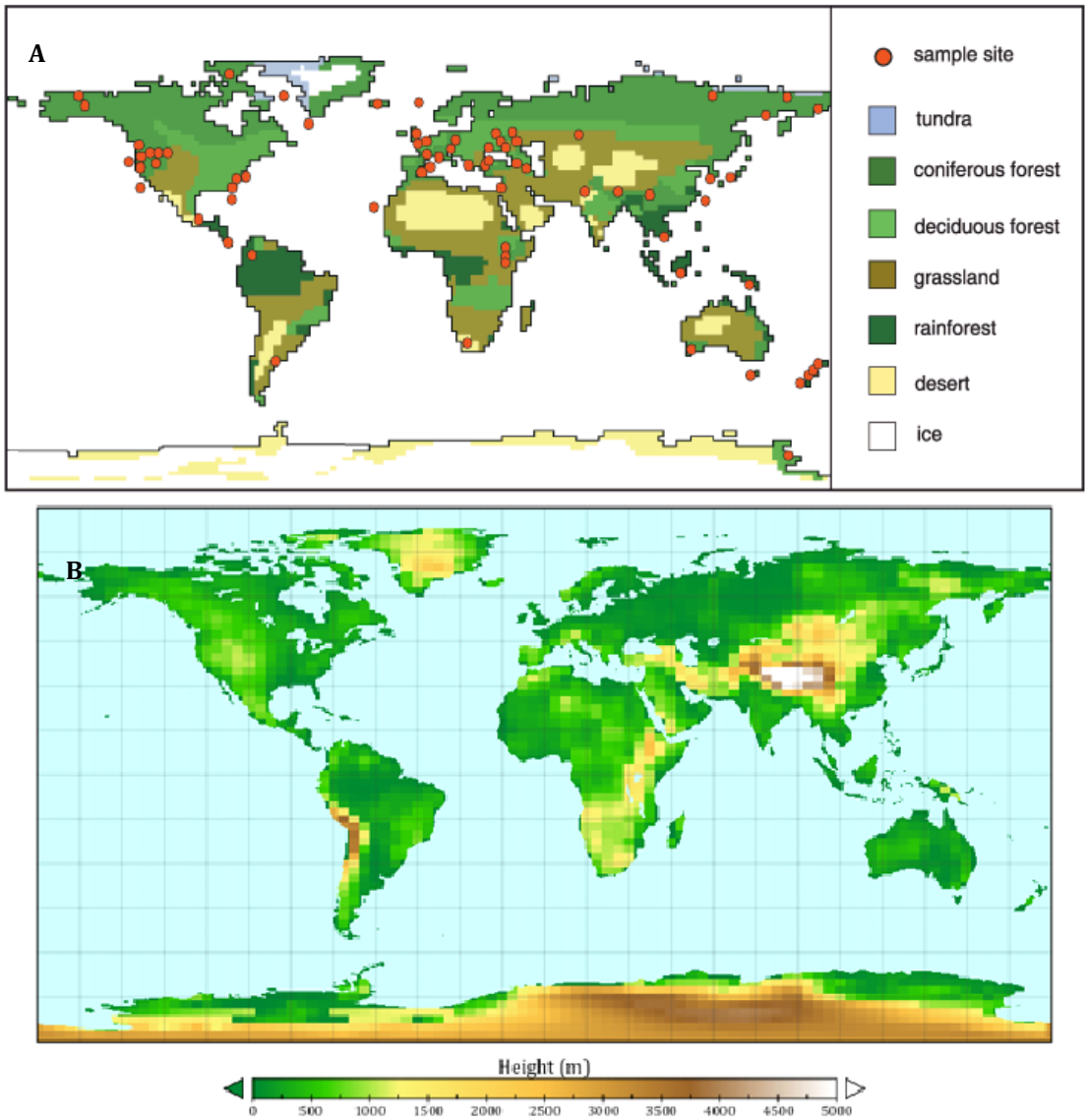


Figure 2.3. The PRISM2 physical boundary conditions used to set up the model to run a mPWP simulation: A) vegetation & ice sheets & B) topography. The figure is reproduced from Dowsett et al. (1999).

The boundary conditions consist of using a modern land-sea mask (including a fully closed Central American Seaway (CAS)), with changes to the orography: the Rocky Mountains were reduced 50% from modern height, and the East African plain was raised by 500m (Figure 2.3b). It has been suggested (Saranthein et al., 2009) that the CAS could have been open during the Pliocene. However for the mPWP, there is no evidence that it was open enough to have influenced mPWP simulations (Lunt et al., 2008a). The vegetation scheme is based on the PRISM2 megabiome reconstruction using 7 biome types (desert, rainforest, grassland, deciduous forest, evergreen forest, tundra and ice) reconstructed from 74 terrestrial data points. The land ice mask includes a 50% reduction in the volume of the Greenland Ice Sheet and 33% in the volume of the Antarctic Ice Sheet, equivalent to a rise in sea level of 25m. The PRISM2 vegetation and ice sheet reconstruction is shown in Figure 2.3a, soils, lakes and rivers are all set as modern throughout the ensemble members. With the exception of sea ice (initially set by the PRISM2 reconstruction, but a dynamic feature in HadCM3), all of these boundary conditions are fixed for the duration of the model simulation. The PRISM2 reconstruction also included a sea surface temperature reconstruction based on 77 marine locations, but this boundary condition is not required in a fully coupled AOGCM simulation. The PRISM2 boundary conditions make a total of 23 alterations to variables compared to the modern ancillary files used, changing features such as soil heat and moisture capacity (Lunt et al., 2012a).

The model is forced with an atmospheric CO_2 concentration of 400 ppmv, which is within the range of values suggested for the Pliocene from a variety of data sources and methodologies (Raymo et al., 1996; Pagani et al., 2010; Seki et al., 2010). All other trace gases were set at pre-industrial levels as per the experimental design for the PlioMIP project (Haywood et al2011a).

As discussed in Chapter 1 (Section 1.2.3 *v*), the use of the PRISM2 orography in the land-sea mask could impact upon northwards heat transfer (Hill et al., 2011) and a change in the climate of North America (Seager et al., 2002, Foster et al., 2010). Additionally changes to the vegetation patterns could impact on the results due to changes in the parameterisations to the land surface vegetation. For consistency with previous studies on the mPWP using the HadCM3 model, the PRISM2 boundary conditions were preferred to the PRISM3D boundary conditions for the first full ensemble. However to investigate the effect of this decision, a boundary condition ensemble using the PRISM3D boundary conditions will also be undertaken (Section 2.4.4).

2.4.2. Creating a Modern Model Simulation

For producing 'Pliocene minus modern' comparisons a modern simulation for each ensemble member was required. The PRISM dataset generates its core top values from the work of Reynolds & Smith (1995), which represents an average of the climate through the late 1950s to the early 1980s. Average values for greenhouse gases (CO_2 – 334 ppmv, CH_4 - 1602 ppb, N_2O – 281 ppb) were calculated for this period using data from the Mauna Loa observatory (as documented in Keeling et al., 2005) and the US National Climate Data Center (NCDC) and used to generate the forcings for this modern HadCM3 simulation. These greenhouse gas values were held constant for the duration of the spin up and experimental periods. Boundary conditions for ice sheets, orography and vegetation are also set to the modern. The modern simulation is designed to represent the averaging period used by Reynolds & Smith (1995) to create the model equivalent of the core top data in the PRISM dataset. The modern simulations in Chapter 3 came from averaging the transient QUMP simulations between 1950 to 1985 for the three simulations produced for the initial results.

2.4.3 Creating the Perturbed Physics Ensemble

All the perturbed parameters in these ensembles are specific to the atmospheric component of the HadCM3 model. These parameters and their values have been chosen through work on the QUMP project. The initial work using the atmosphere only model, HadAM3, identified 100 parameters that could be potentially perturbed and individually perturbed the 32 parameters seen as having the largest impact on uncertainty (Murphy et al., 2004). The QUMP work developed to target the effects of multi-perturbed parameter combinations, with the 32 parameters perturbed simultaneously (Collins et al., 2006; Webb et al., 2006). However the potential size of a multi-parameter PPE is beyond the bounds of computing power, running to millions of potential members. The ensemble members were selected using a Bayesian statistical framework to select the *'best'* ensemble members (Collins et al., 2006), where the likelihood of each ensemble member to achieve a set series of assessment metrics, determined whether it would be considered as a potential ensemble member.

The 53 single perturbed parameter experiments of Murphy et al. (2004) were used to infer the effect of multiple perturbed parameters based on the linear output from these 53 simulations (Collins et al., 2006; Webb et al., 2006). The parameter ranges came from a uniform distribution between the maximum and minimum of the potential ranges for the parameters based on expert judgement or experimental data (Murphy et al., 2004; Collins

et al., 2006). 3.6 million potential members of the ensemble were produced (all the combinations of the 32 perturbed values for the parameters). The members were split into 64 groupings of resulting model Charney sensitivity (Collins et al., 2006; Webb et al., 2006). The *'Climate Prediction Index'* (CPI – Murphy et al., 2004) a scoring method which utilises root mean square error for climate metrics where there are suitable observations or re-analysis data was used to rank the potential members. The assessment metrics used to create the CPI were those used in the initial testing of the Standard released version of HadCM3 and ensured that the ensemble member was a valid climate model (Murphy et al., 2004). The best 20 runs from each of the 64 groupings, were selected (the determination of the strength of the run from the Bayesian framework), creating a subset of 1280 potential model parameter sets. The best scoring simulation using the CPI (i.e. the model with greatest skill) was selected from each Charney sensitivity bin. A second simulation was also chosen from each bin, using the highest ranked ensemble member that samples a different region of parameter space (Webb et al., 2006). The method ensured that two similar simulations were not selected from each bin, allowing the maximum exploration of parameter space using skilful models within the bounds of the available computer resources. For HadSM3, this process was refined to the best 128 perturbed members (added to the Standard to create a 129 member ensemble - Collins et al., 2006; Webb et al., 2006). For HadCM3 it was iterated down to the best 16 perturbed members along with the Standard version to create the 17 member ensemble by examining the table of sensitivities and parameters from the 128 perturbed members of the HadSM3 ensemble. The ensemble members were chosen by selecting the ensemble members that covered the best range of Charney sensitivities and perturbed parameters (Collins et al., 2011).

The Bayesian framework is used because it can be applied to any complexity of model, although changes are required to the assumptions and final implementation of the framework (Collins et al., 2012).

$$c = M(p, R) \qquad \text{(Eq. 2.3)}$$

Where 'c' represents the model outputs, 'M' represents a climate model (HadCM3 for the perturbed physics ensembles) which is controlled by the parameterisations 'p' and the initial forcings 'R' (based on Collins et al., 2006; 2012).

$$P(c|o) \propto P(c)P(o|c) \qquad \text{(Eq. 2.4)}$$

Bayes theorem calculates the probability (P) of a result *'c'* given the data *'o'*. However there is uncertainty in the data, so the formula includes a term which is the *"probability of the data given c"*, essentially a measure of the fit of the model and data (the use of '|'

denotes *'conditional on'* - Rougier, 2007). A crucial problem for the application of this theorem to the climate modelling problem is that the ideal data (o) for this application in climate modelling is the same as the end result hoping to be achieved (i.e. temperature response to forcing over the 21st century). As a result, a surrogate set of data is required, that the model can be tested against allowing the validation of the model to project future climate. The surrogate data comes from utilising key model test metrics and datasets of observations over the 20th century. The data allows model *'faux'* predictions of these past events in terms of mean climate reconstruction, response to large volcanic eruptions or the variability in climate over the period. A final term assumes the probability of *'c'* being accurate before any inclusion of the data (Collins et al., 2006).

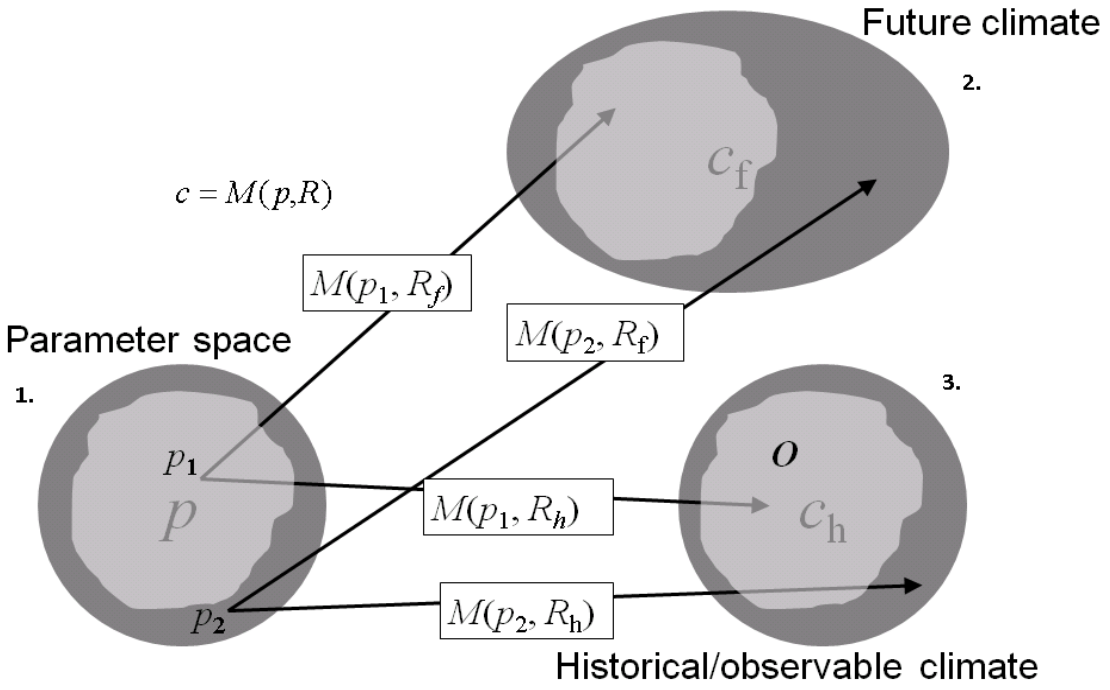


Figure 2.4. A graphical representation of the production of perturbed parameter combinations for the PPE (which originate from the QUMP project) in three variations of the model, 1) the parameter space of the model *'M'*, 2) the climate projection for model *'M'* for future climate and 3) the climate projection for model *'M'* for historical/observed climate. Reproduced from Collins et al. (2012).

In Figure 2.4, the grey circles represent (1) parameter space (p) of model (M) and the climate projections produced by that model M for future climate change scenarios (2) and the historical or observed climate (3). The target (for the QUMP project) was to select realistic parameter combinations for producing a future climate reconstruction (2), but this must be inferred based on the performance of the model for reproducing the past climate variables. Parameter combinations p_1 and p_2 represent potential ensemble

members and when they are run in the model (M) with the same forcing (R_h) to produce a result for the historical/observed climate (Collins et al., 2006; 2012).

Where these model results (c) are in a light grey box this represents a region were the parameterisations reproduce a historical/observed climate that is within observational error. The dark grey regions of the parameter space represent parameters that produce a historical/observed climate which is less likely to be consistent with the observations. A parameter combination that produces an acceptable historical/observed climate is inferred to produce an acceptable future climate projection (Collins et al., 2006; 2012). The resulting combinations can be used to produce a Bayesian weighting on the probability distribution of climate response for the future, which can be used in conjunction with other metrics (such as the CPI) to rank potential parameter combinations and therefore choose ensemble members (Collins et al., 2012).

The parameters that have been perturbed, the values used and where in the model this change is affected is shown in Table 2.1a,b. Using results from previous QUMP and climate*prediction*.net experiments Rougier et al. (2009) determined the most influential parameters were numbers 1 to 4, 7 & 18 (from Table 2.1a,b). The colours have been added to the boxes to highlight how parameters change in comparison with the Standard through the ensemble members.

	Label	Sub-group	Identifier/Properties	Area of Influence
-	CS		Charney Sensitivity	
1	VF1		Ice Fall Speed (ms^{-1})	Large Scale Cloud
2	CT		Cloud Droplet to Rain Conversion rate (S^{-1})	Large Scale Cloud
3	RHCrit		Threshold of Relative Humidity for Cloud Formation	Large Scale Cloud
4a	CW	Land	Cloud Droplet to Rain Conversion Rate ($kg\ m^{-3}$)	Large Scale Cloud
4b		Sea	Cloud Droplet to Rain Conversion Rate ($kg\ m^{-3}$)	Large Scale Cloud
5	MinSIA		Minimum Sea Ice Albedo - Dependence of SIA on Temperature	Sea Ice
6	ice_tr		Sea Ice Minimum Temperature	Sea Ice
7	Ent		Entrainment Rate Coefficient	Convection
8	Icesize		Ice Particle Size (μm)	Radiation
9	Cape		CAPE Closure - Intensity of Convective Mass Flux	Convection
10	flux_g0		Boundary Layer Flux Profile	Boundary Layer
11	Charnoc		Charnock Constant - Roughness Lengths & Surface Fluxes Over Sea	Boundary Layer
12	soillev		Number of Soil Levels Accessed for Evapotranspiration	Land Surface
13	lambda		Asymptotic Neutral Mixing Length Parameter	Boundary Layer
14	cnv_rl		Free Convective Roughness Over Sea	Boundary Layer
15	oi_diff		Ocean Ice Diffusion	Sea Ice
16	dyndel		Order of diffusion Operator	Dynamics
17	dyndiff		Diffusion *e*-folding Time	Dynamics
18	eacfbl		Cloud Fraction at Saturation Boundary Layer	Large Scale Cloud
19	eacftrp		Cloud Fraction at Saturation Free Troposphere Value	Large Scale Cloud
20	k_gwd		Surface Gravity Wavelength	Dynamics
21	k_lee		Trapped Lee Wave Constant	Dynamics
22	gw_lev		Starting Level for Gravity Wave Drag	Dynamics
23a	s_sph	sw	Non-spherical Ice Particles Shortwave Radiation Properties	Radiation
23b		lw	Non-spherical Ice Particles Longwave Radiation Properties	Radiation
24a	c_sph	sw	Non-spherical Ice Particles Shortwave Radiation Properties	Radiation
24b		lw	Non-spherical Ice Particles Longwave Radiation Properties	Radiation
25	rhparam		Flow Dependant RHcrit	Large Scale Cloud
26	vertcld		Vertical Gradient of Cloud Water	Large Scale Cloud
27	canopy		Surface Canopy Energy Exchange	Land Surface
28	cnv_upd		Convective Anvils Updraft Factor	Convection
29	anvil		Convective Anvils Shape Factor: Radiative properties of convective cloud.	Convection
30	sto_res		Dependence of Stomatal Conductance on CO_2	Land Surface
31	f_rough		Forest Roughness Lengths	Land Surface
32	sw_absn		Shortwave Water Vapour Continuum Absorption	Radiation

Table 2.1a. The name and short label for each parameter perturbed in the full ensemble. The area of influence within the model climate is also indicated. Identifiers in Table 2.1a apply to values in Table 2.1b.

	Label	Standard	B	D	F	H	I	J	K	L	M	N	O	P	Q
-	**CS**	**3.3**	**2.42**	**2.87**	**3.75**	**3.43**	**4.39**	**3.9**	**4.44**	**4.88**	**4.54**	**4.62**	**4.8**	**5.4**	**7.11**
1	VF1	1	1.0413	1.12286	1.37049	1.42575	0.65138	0.50546	0.53499	0.52716	0.87884	0.57054	0.99225	0.64807	0.54306
2	CT	1.00E-04	6.10E-05	1.10E-04	3.92E-04	3.80E-04	9.90E-05	3.29E-04	2.44E-04	1.97E-04	3.76E-04	2.50E-04	2.36E-04	3.08E-04	3.50E-04
3	RHCrit	0.7	0.71434	0.72846	0.7	0.7	0.79788	0.82018	0.67225	0.7	0.7	0.68958	0.7	0.87689	0.7
4a	CW_land	2.00E-04	2.76E-04	1.95E-04	2.72E-04	6.78E-04	2.70E-04	1.72E-03	1.04E-03	1.11E-04	1.67E-04	1.49E-04	1.91E-04	3.23E-04	1.41E-04
4b	CW_Sea	5.00E-05	6.90E-05	4.85E-05	6.80E-05	1.69E-04	6.75E-05	4.30E-04	2.60E-04	2.33E-05	4.01E-05	3.47E-05	4.73E-05	8.07E-05	3.23E-05
5	MinSIA	0.5	0.50429	0.52653	0.5095	0.52116	0.62855	0.63406	0.63533	0.53872	0.58108	0.58361	0.64619	0.64455	0.60104
6	ice_tr	10	9.69	8.105	9.3214	8.4886	2.8044	2.5977	2.5501	7.2343	4.5845	4.4896	2.1429	2.2044	3.836
7	Ent	3	3.49381	2.78167	4.29985	4.85597	2.4295	2.98215	2.16414	3.75359	3.83715	3.1626	3.61375	4.51062	2.37726
8	Icesize	30	34.675	30.023	29.606	28.07	27.276	29.069	38.293	31.774	30.326	32.906	27.214	32.485	28.099
9	Cape	Off	Off	1.97	1.54	Off	1.39	1.4	Off	Off	2.43	Off	Off	1.28	Off
10	flux_g0	10	8.2498	17.8212	8.5368	15.5547	5.083	8.6838	8.0411	6.5362	5.2381	8.5441	9.8194	12.0714	6.9039
11	Charnoc	1.20E-02	0.0185	0.0192	0.0128	0.0155	0.0141	0.0157	0.0153	0.0132	0.0123	0.0132	0.0123	0.0173	0.0134
12	soillev	4	2	2	2	4	4	2	3	4	4	4	4	4	4
13	lambda	0.15	0.2163	0.3419	0.30901	0.15263	0.15956	0.48694	0.14737	0.10945	0.17026	0.41642	0.13059	0.35284	0.135
14	cnv_rl	1.30E-03	3.44E-03	2.95E-03	3.81E-03	3.54E-03	1.81E-03	9.95E-04	4.47E-04	9.36E-04	4.18E-03	1.14E-03	2.43E-03	1.82E-03	1.67E-03
15	oi_diff	3.75E-04	3.72E-04	3.63E-04	3.71E-04	3.44E-04	3.54E-04	3.59E-04	3.74E-04	3.62E-04	3.45E-04	3.59E-04	3.53E-04	3.71E-04	3.73E-04
16	dyndel	6	4	4	4	6	6	6	6	6	6	4	6	4	6
17	dyndiff	12	9.948	16.41	8.569	11.075	9.593	13.096	11.414	8.795	15.535	7.832	19.842	12.222	8.5
18	eacfbl	0.5	0.67342	0.67577	0.73237	0.59233	0.50408	0.73756	0.54746	0.50077	0.56161	0.62522	0.51981	0.51606	0.51262
19	eacftrp	0.5	0.58671	0.58789	0.61619	0.54616	0.50204	0.61878	0.52373	0.50038	0.5308	0.56261	0.50991	0.50803	0.50631
20	k_gwd	2.00E+04	1.46E+04	1.44E+04	1.16E+04	1.65E+04	1.99E+04	1.18E+04	1.17E+04	1.98E+04	1.55E+04	1.62E+04	1.20E+04	1.95E+04	1.67E+04
21	k_lee	3.00E+05	2.19E+05	2.16E+05	1.74E+05	2.48E+05	2.98E+05	1.77E+05	1.76E+05	2.97E+05	2.32E+05	2.43E+05	1.80E+05	2.92E+05	2.50E+05
22	gw_lev	3	3	3	3	3	5	3	5	5	5	5	5	5	3
23a	s_sph_sw	2	2	2	7	2	2	2	7	2	2	2	2	2	7
23b	s_sph_lw	1	3	3	7	3	3	3	7	3	3	3	3	3	7
24a	c_sph_sw	3	1	1	7	1	1	1	7	1	1	1	1	1	7
24b	c_sph_lw	1	1	1	7	1	1	1	7	1	1	1	1	1	7
25	rhparam	0	0	0	1	1	0	0	0	1	1	0	1	0	1
26	vertcld	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	canopy	0	1	0	0	0	0	0	0	1	0	1	0	0	1
28	cnv_upd	Off	Off	Off	0.81295	0.14225	Off	Off	Off	Off	0.84813	Off	0.92694	0.80149	Off
29	anvil	Off	Off	Off	1.06076	2.51386	Off	Off	Off	Off	1.50822	Off	2.78748	1.39652	Off
30	sto_res	1	1	1	0	0	0	0	1	1	1	1	1	0	1
31	f_rough	0	0	0	0	3	3	0	3	3	3	3	0	0	3
32	sw_absn	0	0	1	0	0	0	0	0	1	0	0	1	0	0

Table 2.1b. The labels and values for the parameters perturbed for each ensemble member in the full ensemble. The colour of the box indicates whether value is: greater (yellow), smaller (blue) or same (white) as Standard. The information included in Tables 2.1a & b is obtained from in Collins et al. (2006) & Rougier et al. (2009).

The process of creating a PPE, which in essence, detunes the model to apply a new tuning, causes model drift creating an imbalance in the top of the atmosphere (TOA) radiation budget (Collins et al., 2011). The imbalance is usually removed during the construction and tuning of the model. Flux adjustments are used to reduce regional errors and increasing the reliability of QUMP experiments for use in regional climate change. They also allow the full effects of the perturbed parameters to be explored without the need to ensure that a balanced TOA radiation budget is strictly adhered to (Murphy et al., 2007). The flux adjustments were applied through performing a multi-decade simulation for each member. Values for the seasonal and spatial distribution of sea surface temperature and sea surface salinity values were relaxed through a Haney Forcing (Haney, 1971 - using a relaxation coefficient of 30 days for temperature and 120 days for salinity as in Tziperman et al., 1994). The forcing is applied until the model reached a minimal forcing effect from this change (TOA approximately less than 0.2 Wm^{-2}). Once this has happened the seasonally-varying flux adjustment input file was created and this allows the model to run with a stable climate in both standard and ensemble member simulations.

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 (AMOC), and a cooling of North Atlantic SSTs. 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. The improvement occurs through reducing significantly the biases to Northern Hemisphere SSTs and sea ice that occurred in previous PPE studies. The result was confirmed in comparison work completed in Collins et al. (2011) between the control (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., (2011; longer relaxation constant). The perturbed physics ensemble members were initiated with the improved flux adjustments (of Murphy et al., 2007; Collins et al., 2011).

HadCM3 was the first UK Met Office model to not require flux adjustments as standard in the model simulations. The removal of flux adjustments was an improvement in model performance and it is a weakness of the QUMP framework that this correction is required in the model. However, it does not create a statistically weaker version of the Standard simulation compared to non-flux adjusted versions of the HadCM3 model (Collins et al., 2011).

The flux adjustments are applied through a modification file in the ocean component of HadCM3. The flux adjustment files are linked to the perturbations made to the model, so the same file was used in the Pliocene and modern simulations for each ensemble member

and the Standard simulation. The Standard simulation was spun up with its flux adjustment file and then each ensemble member started from this spin up, with each ensemble member was started with its flux adjustment file. The majority of the individual parameter perturbations were applied using a Unix shell script which hand edited the relevant sections of the model code after the processing of the job in the Unified Model User Interface (UMUI) on the Providing Unified Model Access (PUMA) service so that the values for each ensemble member could be set to the prescribed setting (see Table 2.1b). Ensemble members H, I, K, L, M, N & Q included an adjusted parameterisation of surface roughness for forest vegetation. The change was implemented across these ensemble members by creating a new vegetation ancillary file. The surface roughness parameter perturbation was applied through using the PRISM biome numbers as a template to then change the values of surface roughness. The file could then be recompiled into a new ancillary file and applied to the simulations. A new vegetation ancillary file was created for both PRISM2 and PRISM3D boundary conditions.

2.4.4 Creating Boundary Condition Ensembles

i) PRISM3D Boundary Conditions Ensemble

As discussed in Chapter 1, the changes to the PRISM2 model boundary conditions to create the PRISM3D model boundary conditions, have the potential to alter the climate simulated by climate models in some key areas. The raising of the Rockies towards modern height from the half modern height used in the PRISM2 reconstruction affects atmospheric circulation impacting on both continental North America and the Arctic increasing northward heat flow and MSLP (Hill et al., 2011). The change from a 7 classification megabiome to a 28 classification biome vegetation boundary condition creates a more realistic land surface scheme affecting regional responses to climate forcing, however modern soils, lake and river reconstructions were still used. Recent work (Pound et al., 2013) has reconstructed mPWP soils and lakes. They found some improvements to the simulations of the mPWP and the importance of these recent developments in the context of my modelling are discussed in Chapter 5. Finally, adjustments to the ice sheets for PRISM3D could have albedo and elevation impacts on the climate. As with PRISM2 boundary conditions, the PRISM3D boundary conditions will remain fixed for the duration of the model simulations.

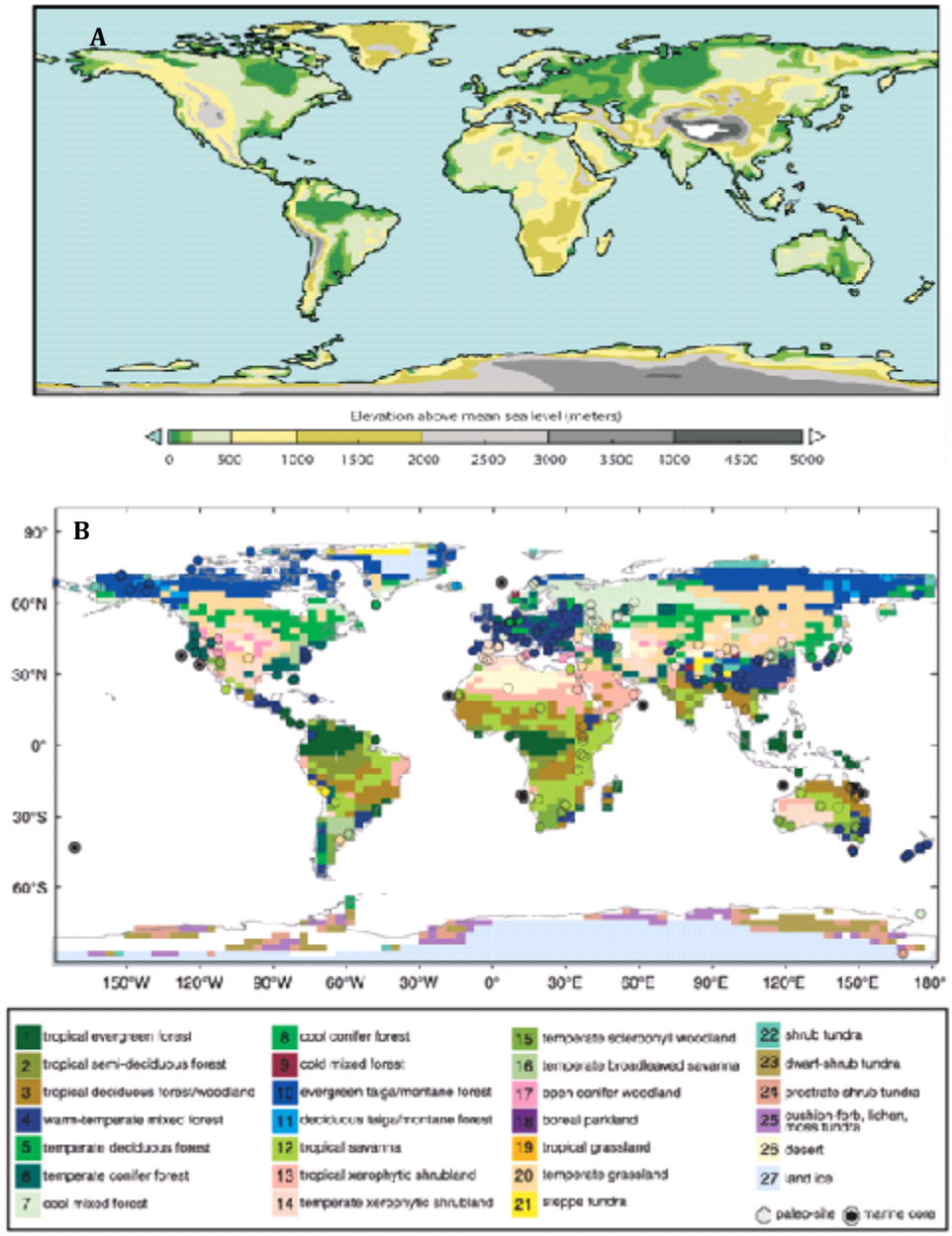


Figure 2.5. The PRISM3D physical boundary conditions used to set up the model to run a mPWP simulation: A) topography & B) vegetation & ice sheets. Figure created from figures in Dowsett et al. (2010a).

The PRISM3D boundary conditions were applied in the same way as the PRISM2 boundary conditions, with changes to the required ancillary file inputs (for applying PRISM3D boundary conditions in HadCM3 - see Bragg et al., 2012). The physical changes were a returning of the Rockies to near modern elevation, a 28 biome vegetation scheme and changes to the elevation and spatial distribution of Greenland and Antarctic ice sheets. The PRISM3D ensemble will be used to investigate the effects of changes to the physical model boundary conditions and how they interact with the perturbed parameters. The changes between PRISM2 and PRISM3D boundary conditions represents an uncertainty in

the reconstruction of the mPWP climate and it is important that both boundary conditions are used to assess the effect of physical boundary condition uncertainty.

ii) Pliocene Atmospheric CO_2 Concentration Boundary Condition Sub-Ensemble

Variation in mPWP CO_2 was discussed in Chapter 1 (Section 1.3.4), with CO_2 in the Pliocene likely to have ranged from 300 to 425 ppmv (i.e. Pearson & Palmer, 2000; Pagani et al., 2010; Seki et al., 2010). To investigate the full range of potential CO_2 in the mPWP, a sub-selection of ensemble members were selected to be run with CO_2 values of 300, 350 and 400 ppmv (the full PPE CO_2 standard).

The sub-ensembles were run for 300 and 350 ppmv CO_2 values for the PRISM3D boundary conditions. Ensemble members were selected to represent the spread of full ensemble results and include the highest ranked simulation. For the CO_2 sub-ensembles CO_2 was set to either 300 and 350 ppmv and then the Pliocene Standard simulations were spun up for 500 years as per the full ensembles. The length of this spin up will not be enough to bring the model into full equilibrium (specifically the deep ocean, which takes many centuries to achieve (Huber & Sloan, 2001)); however the atmosphere and surface oceans will have reached a quasi-equilibrium (Section 2.2 *i* discusses the spin up of Pliocene simulations). Time constraints on producing the sub-ensemble enforce the inability to fully spin up the deep ocean. The sub-ensemble members and Standards were then run for a further 300 years. The sub-ensemble will investigate the relationship between CO_2 and the PPE members, to investigate if an ensemble member performs better using an alternative yet plausible value for mPWP CO_2, as a combination of CO_2 changes and parameter changes could produce a more skillful mPWP simulation.

2.5. Testing the Ensemble Members

Testing the perturbed physics ensembles has two main objectives: intra-model comparisons (IMCs) and data-model comparisons (DMCs). The IMCs focus on the differences between the ensemble members due to differences in boundary conditions or differences due to the perturbed parameters. DMCs provide the tests that rank the ensemble members through analysis of data-model mismatches using two techniques and three datasets. Section 2.5 outlines the ensemble comparisons produced along with the IMC and DMC methods that will be used.

2.5.1. Comparing Ensemble Simulations

To compare differences between ensemble simulations and for creating the data-model comparisons, five types of '*model minus model*' comparison are created:

- *"Pliocene minus modern ensemble member"* – The modern ensemble member is subtracted from its Pliocene equivalent. There are no differences between the parameterisations themselves, only in the interaction of parameters and the different boundary conditions between the Pliocene and modern ensembles.
- *"PRISM3D minus PRISM2 ensemble member"* – The PRISM2 (old boundary conditions) ensemble member is subtracted from its corresponding PRISM3D (new boundary conditions) ensemble member to display differences between the Pliocene boundary conditions used in the perturbed physics ensembles and other mPWP modelling work. The output aids the investigation of differences caused by the interactions between the perturbed parameters and the boundary condition changes.
- *"Ensemble member minus Standard"* – The Standard version for the ensemble is subtracted from each ensemble member (both PRISM2 and PRISM3D ensembles). The results display the differences due to the parameterisations with boundary conditions maintained between the simulations.
- *"PRISM3D/PRISM2 ensemble member minus modern Standard"* – The modern Standard is subtracted from each Pliocene ensemble member (for both PRISM2 and PRISM3D ensembles). It enables the differences between the ensemble members due to the parameterisations to be incorporated alongside the changes between Pliocene and modern boundary conditions. The output will be used in the DMCs to SSTs and SATs which are released as 'Pliocene minus modern' datasets which is necessary to reduce bias' that occur due the effect of latitude on SSTs (Dowsett et al., 2011). The comparison will also be used to create the BIOME4 climate inputs. The decision to use this anomaly for the DMCs is discussed in Section 2.6.4.

Through these, the effects of boundary condition and parameter changes on ensemble members can be thoroughly explored in the IMCs and the ensemble members ranked in comparison to palaeo-data in the DMCs. Only the *"Pliocene ensemble member minus modern Standard"* comparison will be used in the DMCs, for the SST, SAT and biome DMCs.

2.5.2. Intra-Model Comparisons

Intra-model comparisons (IMCs) refer to the climate metrics tested within the ensemble members. No mPWP palaeo-data will be used for these comparisons, meaning they can not determine the relative performance of each ensemble member as a mPWP simulation. However, IMCs indicate how the ensemble members are performing and aid the interpretation of the data-model comparisons results. The perturbed parameterisations

can affect a range of features of the simulated climate, requiring metrics that investigate across this range. Some metrics (surface temperature, sea surface temperature and precipitation) are key constituents of the DMCs and the IMCs for these metrics will offer initial indication of the relative strength of an ensemble member for reducing data-model mismatches. The IMCs will be used to investigate large scale and regional features in the ensemble results that indicate that an ensemble member may improve (or weaken) the data-model mismatches.

The statistical significance of IMCs (differences between ensemble members and the Standard) will be tested using Student's T-test, a widely used statistical method of testing the significance of a null hypothesis between two datasets (Student, 1908; Chervin & Schneider, 1976). Using a null hypothesis that any variation between simulations (i.e. *"ensemble member minus Standard"*) is due to natural variability between model simulations, significant changes in the model anomalies can be identified as being the result of the perturbation made to an ensemble member. A two sample T-test was applied to the 30 year average climatologies with all the results being tested at a confidence level of 95%. The Student's T-test will be applied to model metrics produced from the climatological means for surface air temperature (SAT), and sea surface temperature (SST) metrics as these are the metrics utilised in direct comparison to palaeo-data. The Student's T-test results will be overlaid on the plot for the metric being assessed for each simulation with insignificant regions being stippled, indicating clearly which changes are significant while allowing the climate metric results to be clearly displayed.

A major advantage of the T-test is that it an easy to apply and widely recognised test that gives a useful indication of the significance of anomalies between model simulations. However, the test requires an assumed Gaussian (normal) distribution for the variable being tested, and while this is possible for a majority of climate metrics, it is not possible for precipitation (which has a Poisson distribution (Bardsley, 1984)). The result is that the metric can not test the significance of anomalies involving precipitation for the intra-model comparisons (Pope et al., 2011). The t-test is applied as a sequence of grid box by grid box comparisons which are then combined to plot the significance of the change in a simulated climate metric in comparison to the ensemble or modern Standard. The use of this is well established (i.e. Chervin & Schnieder, 1976), and each of these local grid box significances are easily tested. However, as each grid box is affected by the spatial interactions within the model outputs, the issue surrounding field significance arises (Livezey & Chen, 1983; Wigley & Santer, 1990). Based on work by Knutson et al. (1999) using AOGCMs it can be assumed that should the t-test return greater than 77% of the comparison as significant that the overall result represents a globally significant result

across the chosen field (Knutson et al., 2006). While this represents an approximation, it gives a broad indication as to whether the global differences being observed in the IMCs are representative of changes in the model parameters compared to the Standard or the result of natural/internal variability within the ensemble members.

2.5.3. Data-Model Comparisons: Sea Surface Temperature (SST)

The PRISM3D MASST dataset (Dowsett et al., 2010b) will be compared to the model outputs through a 'model minus data' site by site global comparison and root mean square error (RMSE) calculation (Hyndman & Koehler, 2006). The 30 year annual mean model SST for each MASST data point will be compared with the MASST value for that location. The RMSE for each model simulation can then be calculated. While anomaly plots of data-model comparisons (e.g. Dowsett et al., 2011; Pope et al., 2011; Robinson et al., 2011) provide a visual indication of the data-model comparison, the RMSE value will give a statistically significant ranking of which ensemble member has provided the best data-model comparison with the lowest RMSE indicating the best ensemble member.

The root mean square error (RMSE) is the statistical test used to rank the each ensemble member for the SST and SAT data-model comparisons. The RMSE is a commonly used scale dependant accuracy measure which means it has a non-normalised scale (controlled by the scale of the data – Hyndman & Koehler, 2006). These statistical tests are useful for comparing data which have the same scales. An RMSE score of 0 represents perfect agreement between model and data (Moriasi et al., 2007), with RMSE scores less than half of the standard deviation being considered as low (Singh et al., 2004; Moriasi et al., 2007). Weaknesses of the RMSE include: its sensitivity to outliers in data leading to results becoming biased to large events, that it can not be used to compare data with different scales (Hyndman & Koehler, 2006) and its susceptibility to spatial dependence of errors (Goodchild et al., 1992).

For the purposes of this investigation, these weaknesses do not cause much concern as the RMSE is only used to compare temperatures from models and data, both in the same scale. The sensitivity to outliers would normally be considered a weakness, but the treatment of DMC outliers is vital to determining which ensemble member is strongest. By squaring all errors, the RMSE is not biased by an ensemble member which reduces negative mismatches in the data-model comparison (model colder than data), but produces large positive mismatches (model warmer than data), which potentially cancel out negative mismatches. The squaring of all errors removes the negative element and so all the errors add together to make an overall error which can be calculated. The RMSE will be utilised

in ranking the performance of ensemble members in comparison to each other and the Standard. However, it is important to determine whether variations in the RMSE represent a significant change in the performance of the ensemble member compared to the Standard in relation to the palaeo-data. Part of this assessment will involve utilising the translation of RMSE scores into percentage changes and also through replicating the RMSE analysis using a different climate averaging period. The inclusion of the additional averaging period is to determine whether the RMSE is assessing the differences between ensemble members and palaeo-data due to the perturbed parameters or the impact of the internal variability within the model simulations. The ensemble members will be extended for 30 years (years 301 to 330) to create new climatological averages for assessing the influence of the change in averaging period on the RMSE score for the DMC. For Chapter 4, the difference between ensemble members RMSE scores compared to the Standard will be generated and for Chapter 5 the difference in RMSE scores between equivalent ensemble members across the two sets of boundary conditions. For simulations were the difference in RMSE due to the change in averaging period exceeds the difference due to the intra-ensemble or boundary condition comparisons will not be considered a robust result.

The MASST dataset contains 100 marine sites (Figure 2.6) drilled as part of the Deep Sea Drilling Project (DSDP) and it is successors, the Ocean Drilling Program (ODP) and the Integrated Ocean Drilling Program (IODP). The PRISM team analysed planktonic foraminifera using a variety of techniques including faunal and floral assemblage palaeothermometry (utilising transfer functions and modern analogue techniques), magnesium/calcium (Mg/Ca) palaeothermometry and alkenone palaeothermometry. The techniques used at each site, the locations of each site and the final derived temperatures (as anomalies to modern) can be found in Dowsett et al. (2010b). Modern anomalies were produced by deriving a core top estimate from the work on modern SSTs by Reynolds & Smith (1995) which was produced using ship and buoy records and later satellite observations for SSTs. The MASST data is available as both Pliocene absolutes and 'Pliocene minus modern' anomalies.

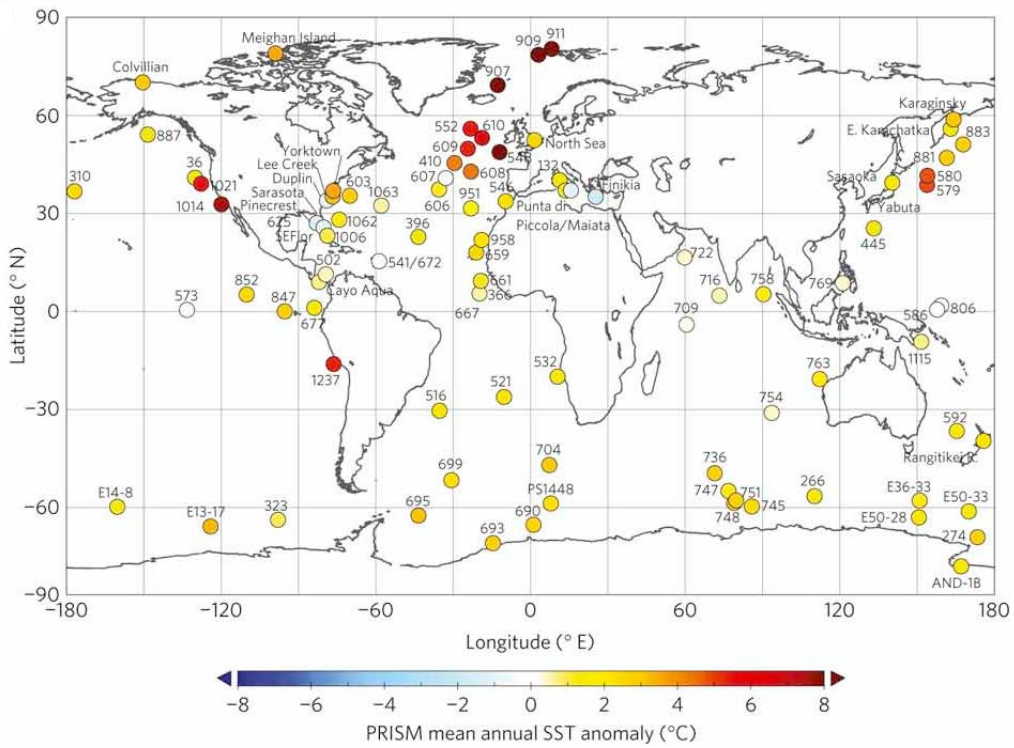


Figure 2.6. The geographical location of each data point in the PRISM3D MASST reconstruction and the reconstructed *"Pliocene minus Modern"* proxy temperature anomaly for each data site. Figure reproduced from Dowsett et al. (2010b).

2.5.4. Data-Model Comparisons: Surface Air Temperature

Surface air temperatures (SATs) were produced using an updated version of the Salzmann et al. (2008) TEVIS database (see Section 2.5.5). Of the 202 sites from the TEVIS database, ~50 included information on temperature. Sites in Iceland and Antarctica were removed along with any sites with an altitude greater than 1000m above sea level or from a marine core further than 250 km offshore (Salzmann et al., 2013). Sites in Canada, Russia, Turkey and Europe added (Salzmann et al., 2013). These qualifications reduced the size of the dataset to 45 sites for use in data-model comparison. As with the original database (Salzmann et al., 2008), the initial authors interpretation and SAT estimate was used for all sites. If temperature data was not published originally, the Coexistence Approach (Mosbrugger & Utescher, 1997) was used where possible (Salzmann et al., 2013). As with the PRISM3D MASST dataset, the ranking will be determined using the RMSE scores for each ensemble member.

2.5.5. Data-Model Comparisons: Vegetation Biome Reconstruction

The outputs of the BIOME4 model (Section 2.3.2) are compared to the palaeo-vegetation biome reconstruction of Salzmann et al. (2008) using Cohen's Kappa statistic (Cohen, 1960). Cohen's Kappa statistic 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 (P_e) from the proportion of observed agreement (P_o) and this result is normalised through dividing it by the maximum possible difference ($1-P_e$) to generate the Kappa statistic:

$$k = (P_e-P_o)/(1-P_e) \quad \text{(Eq.2.5)}$$

The P_e 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 (no agreement/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 an agreement that is beyond chance within the DMCs for each simulation. Weaknesses of the Kappa statistic include that the cause of the difference is not specified, it could be one site or several 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 smaller the data-model mismatch. The Kappa statistic does not make any allowance for how large an individual site mismatch is. 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. The results of the application of Kappa statistics are rarely comparable across studies (Salzmann et al., 2013). The vegetation dataset was released in Salzmann et al. (2008) and consists of 202 marine and terrestrial sites (Figure 2.5b). These were used to create a Tertiary Environment and Vegetation Information System (TEVIS) database of vegetation types for the Piacenzian age (3.6 to 2.6 Ma BP) for input into ArcGIS software (Salzmann et al., 2008). The database compiled data derived from the sites from fossil pollen, leaves, wood and palaeosol carbonate (Salzmann et al., 2008) into a classification scheme that allowed the production of biomes that matched those programmed into the BIOME4 model (Salzmann et al., 2008).

2.6. Methodological Decisions

2.6.1. Impact of Changing the HPC Resources Used.

As discussed in Section 2.2 *i*, the initial results ensemble was run on the supercomputers at the UK Met Office, however the modern, PRISM2 and PRISM3D full ensembles along with the CO_2 sub ensembles were run on the ARC1 high performance computing (HPC) facilities at the University of Leeds. The change in computing resources had two direct effects upon the project. There were issues that arose from the transfer of the project from the Met Office to Leeds with respect to the technical challenges that arose and the impact, scientifically of changing HPC machines.

The process of moving from the Met Office to ARC1 produced a number of technical issues, primarily converting Met Office HadCM3 model set ups to run on non-Met Office machines. A number of changes are made to model setup, specifically around the reconfiguration procedures and the submission of jobs. However, these changes were unlikely to have caused science changes to the output from the simulations as the changes mainly influenced the processes of submitting the jobs to the HPC facility. The ARC1 facility required additional steps for submitting the initial model simulation and a manual resubmitting of jobs when the ARC1 time limit elapsed. The ARC1 HPC facility fixes a maximum 48 hour run length, so model runs were run in 50 year segments (about 45 hours wall clock time), with resubmission required to restart the job using the finishing dump file. Analysis of the model time series of surface air temperatures (i.e. Figure 2.2) indicated that there was no noticeable effect of the regular resubmission on ensemble member output.

The process of porting the model files from the Met Office led to the corruption of three of the flux correction files, for ensemble members A, C & R. The flux adjustments are required to prevent drift in the ensemble members due to the top of atmosphere imbalance created by the perturbed parameters (Section 2.4.3). For members A, C & R the flux adjustments have not corrected the imbalance as intended and the simulations have decadal changes in global annual mean surface air temperature of +/- 1°C. As a result, these ensemble members failed to achieve an acceptable model state and were removed from the final ensemble. The loss of ensemble member R, was the most significant to the project, as R was the ensemble member that was used as the Low Sensitivity member in the initial results ensemble (Chapter 3). Comparison to this would have been useful for assessing intra-ensemble patterns in the initial and full ensembles between the ensemble end members. The final ensemble size is 14 members, including the HadCM3 Standard

simulation instead of the intended 17 member ensemble (Figure 2.1). However, the perturbed physics ensemble is larger than the PlioMIP MME (Haywood et al., 2013a) and comparable in size to the only other PPE conducted in a warmer than modern palaeoclimate, the Sagoo et al. (2013) Eocene PPE run using FAMOUS, the low resolution version of HadCM3.

The change in HPC resources also created a disconnect between the initial results of Chapter 3 (run at the Met Office) and the full ensembles and CO_2 sub-ensembles completed on the ARC1 resources (Chapters 4 & 5). Owing to the change in HPC resource, it is not possible to directly compare the simulations run at the Met Office with those run on ARC1. The change between different HPC architectures, hardware and the compilers can lead to a change in the climate means of the same model simulation (Lunt et al., 2012a). This has been attributed to non-standard programming practises, such as the use of multiple data statements in model subroutines (Steenman-Clark, 2009; Lunt et al., 2012a).

The inability to directly compare between the initial ensemble and the full ensembles was a disappointment scientifically, however the technical issues around developing the ensemble set up for the ARC1 HPC resources and the loss of three ensemble members from the full ensemble had a greater impact, both in terms of lost time from the project and scientifically by losing ensemble members. Whilst members A, C & R represent members with lower Charney sensitivity that were unlikely to influence the rankings of ensemble members with respect to being the strongest members, they would have been useful for assessing the performance of members with lower Charney sensitivities.

Overall, the change of HPC platform had very little effect on the experimental design for the thesis, only the loss of three members making a noticeable change. However, the movement away from the Met Office HPC resources removed the opportunity for comparisons between mPWP, transient 21st century QUMP and Palaeo-QUMP (Last Glacial Maximum and mid-Holocene) PPEs with comparable ensemble members.

2.6.2. Why Follow the QUMP Methodology?

The methodology used to select the QUMP ensemble members is not the only method for sampling model parameter space to select PPE members. Other methodologies for selecting ensemble members include: emulation (Sanderson et al., 2008b; Rougier & Sexton, 2007; Rougier et al., 2009), Ensemble Kalman Filters (Annan et al., 2005a,b; Hargreaves & Annan, 2009), Latin Hypercube Sampling (Gregoire et al., 2011; Sagoo et al., 2013) and Monte Carlo sampling (Sanderson et al., 2008b). Ensembles created using these methods are discussed in Chapter 1 (Section 1.3.3.2). Unlike these techniques the QUMP

methodology does not sample the full parameter space of the model, sampling only what was considered *"uncertain parameters"* through expert solicitation (Murphy et al., 2004; Barnett et al., 2006). 32 parameters (out of over 100) were considered to be uncertain and were perturbed in the atmospheric component of HadCM3. All QUMP work, including the work here, has focussed on this particular region of the parameter space known to be uncertain. The space was sampled thoroughly (whittling ~4 million potential ensemble members down to the best 16 for HadCM3 – Section 2.4.3). However, it has not investigated any other regions of the parameter space within the atmospheric component of HadCM3. The result is confidence in the quality of each ensemble member as a viable version of the HadCM3 model. However, the ensemble could be missing equally valid or better versions of the model that are in the un-sampled parameter space, which another selection method could have chosen.

An assumption being made by following the QUMP methodology is that the parameters that are perturbed in the modern are the *"correct parameters"* to be perturbing in the Pliocene. The assumption is that there have not been significant changes in the Earth system that would mean the perturbed physics ensembles would have been better suited to investigating a different parameter set rejected for the modern. It is not possible to observationally study the key features and processes driving the Pliocene climate system, however there is little evidence from palaeo-data that would suggest the mPWP climate was significantly different to the present day. It is unlikely that processes such as cloud microphysics have changed between the Pliocene and the modern day. The main debate for a different climate structure is centred on the existence of a permanent El Niño state in the Pacific. Some research has suggested that a permanent El Niño state existed in the Pliocene Pacific Ocean and the resulting teleconnections would affect global climate (i.e. Federov et al., 2010; 2013). However, proxy reconstructions and other modelling studies (i.e. Bonham et al., 2009; Scroxton et al., 2011) suggest that a more enhanced ENSO but not permanent conditions existed. Bonham et al. (2009) indicated that this enhanced ENSO was a feature that was modelled in HadCM3 and therefore would be included within the PPE. Similarities such as the continental configuration, suggest that the mPWP also shared crucial climatic features with the current climate such as a modern like thermohaline circulation. These features indicate that there are apparent similarities between the mid-Pliocene and late 21st Century climates. As a result, the mPWP has become an important focus of palaeoclimate research with possible indications of the impact of anthropogenic climate change (IPCC, 2007; Haywood et al., 2009a).

The perturbed physics ensembles, in line with most early QUMP work, has focussed on perturbing parameters in the atmospheric component of HadCM3. Primarily, QUMP

projects focussed on climate change projections for the 21st century and initially used atmosphere and slab ocean versions of HadCM3. With the development of the HadCM3 QUMP PPE, other QUMP variants have perturbed parameters in the oceanic component (Collins et al., 2007, Brierley et al., 2010), sulphur cycle (Ackerley et al., 2009) and the carbon cycle (Booth et al., 2012). Choosing to focus on the atmospheric parameters in HadCM3, the ensembles could be ignoring uncertainties in the other model components, especially in the ocean. Oceanic components do not exert a large forcing on climate in 21st century climate projections (Collins et al., 2007). However, in palaeoclimate simulations with longer simulation length, these perturbations could exert a stronger influence on climate model simulations. For sulphur and carbon cycle parameterisations, a concern is the understanding of key boundary conditions, especially vegetation. While atmospheric parameters are influenced by uncertainty in the vegetation reconstruction, the impact of this uncertainty will be greater for carbon cycle parameter sets. For the perturbed physics ensembles, maintaining a focus on atmospheric perturbations allows a thorough investigation of the uncertainty in one region of model parameter uncertainty.

The motivation for choosing the QUMP ensemble methodology, compared to other methods for selecting parameters for a PPE was the strong development of the method through the QUMP and climate*prediction*.net projects (i.e. Murphy et al., 2004; Stainforth et al., 2005; Collins et al., 2006; 2011). The mid-Pliocene represents the most parsimonious palaeoclimate period in comparison with the climate predicted for the late 21st century. The QUMP project has produced ensembles focussing on the climate of the 21st century, therefore testing the QUMP parameter choices in the closest palaeoclimate is a logical modelling strategy. Additionally, the use of the QUMP parameters allowed validation of the methods and checks to ensure the parameter perturbations and flux adjustments had been applied properly. These decisions to remain within the original QUMP framework are why no other parameters were perturbed, specifically ocean parameters. Perturbing parameterisations in other components is a potential development of the perturbed physics ensembles.

Work undertaken as part of the Japan Uncertainty Modelling Project (JUMP) has tested the reliability of single model ensembles (a.k.a perturbed physics ensembles) and multi-physics ensembles (MPEs). Yokohata et al. (2012) analysed two MMEs and four PPEs, with the analysis later extended to four MMEs, four PPEs and one MPE (Yokohata et al., 2013). Rank histograms were produced for the ensembles, which indicate how reliable an ensemble performs relative to the observational data being used. The processes involved and the ensembles used are discussed in more detail in Chapter 1 (Section 1.4.3.2). The results indicated that a PPE is less reliable than an MME across a range of variables,

however for SAT and precipitation, PPEs can be considered reliable. As a result, a PPE is still a useful tool to investigate model uncertainty. However, it is important to test the reliability of the PPE being used (Yokohata et al., 2012). The HadCM3 QUMP ensemble was analysed in the Yokohata et al. (2012) study and determined it was close to the threshold of reliability. The result indicates it is a suitable PPE to use based on the conclusions of their work.

The perturbed physics ensembles do not fully investigate the parameter space in the atmospheric component of HadCM3 nor seek to investigate the ocean, carbon cycle or sulphur cycle parameter space. However, it represents the most comprehensive sampling of both parameter and boundary condition uncertainty yet undertaken in a warmer than modern palaeoclimate. The aim of the perturbed physics ensembles is to facilitate investigating the uncertainty in mPWP modelling simulations caused by parameter uncertainty. By utilising a thoroughly investigated parameter perturbation methodology in the form of the QUMP ensemble, it is possible to understand the parametric uncertainty and the response of these parameterisations to reproducing warm climates. The ensembles will also encompass boundary condition uncertainty between two potential boundary conditions for the mPWP. The outputs from the ensemble can be tested against palaeo-data to determine the skill of each ensemble member for modelling the mPWP and tackling the existing data-model mismatch. These PPEs will produce the most detailed investigation of model parameter and boundary condition uncertainty in a warmer than modern palaeoclimate.

2.6.3. Use of Pliocene Community Practises

Some elements of the experimental design within this thesis represent existing Pliocene community practises. These practises include:

- Using 400 ppmv as the default Pliocene CO_2 value
- Using pre-industrial concentrations for all other greenhouse gases
- Using a modern orbital configuration to represent the mPWP time slab
- Using site-by-site data-model comparisons as the primary analysis of ensemble member performance.

The chosen greenhouse gas concentrations and the selected orbits influence the radiative forcing exerted upon the model for the mPWP simulations. However, are these appropriate choices for modelling the climate of the mPWP? To ensure the simulations generated by the perturbed physics ensembles were comparable with existing mPWP modelling studies, these existing practises were utilised. They have been formally

outlined within the PlioMIP Experimental Design (Haywood et al., 2010; 2011a) and subsequent publications analysing the results (Dowsett et al., 2012; 2013; Salzmann et al., 2013). However, it is important to assess whether any of these Pliocene community practises are potential weaknesses with the experimental design for the perturbed physics ensembles and therefore this thesis.

i) Greenhouse Gas Concentrations

Section 1.3.4 (Chapter 1) discussed the published range of mPWP atmospheric CO_2 concentrations (Table 1.5, Chapter 1). The range, from 280 ppmv to 435 ppmv is broad and provides a significant challenge to modelling the mPWP. Using the equations for calculating radiative forcing from the IPCC TAR (IPCC, 2001 – RF = 5.35*ln(435/280)), the change from 280 to 435 ppmv represents a change in radiative forcing of 2.36 Wm^{-2} between these two potential end members of mPWP CO_2.

However, the majority of the mPWP atmospheric CO_2 concentrations focus around 360-400 ppmv (Table 1.5, Chapter 1). 400 ppmv is used as the standard value despite being the top end estimate, because of the unavailability of a reliable proxy for other greenhouse gases and aerosols. With no estimate available other greenhouse gases, they are set to pre-industrial values in model simulations. However, in line with the links between CO_2 and CH_4 in Quaternary ice cores, the top end estimate is used to allow for an increase in the other greenhouse gases (Haywood et al., 2010). The choice represents a significant decision in the experimental design. However, without the ability to determine other greenhouse gases, the use of a CO_2 value from towards the top end of the estimate enables the PPE to account for this boundary condition uncertainty. As compensation for this choice in the experimental design for the full ensembles, the decision was taken to include other values for atmospheric CO_2. As indicated in Figure 2.1, the experimental design of these perturbed physics ensembles allows for the possible impact of this community practise. Through running sub-ensembles at 300 ppmv and 350 ppmv, encompassing the possible range of mPWP atmospheric CO_2 concentrations the impact of the choice of CO_2 value is included within the experimental design, tackling a forcing boundary condition uncertainty and assessing its impact on the PPE.

ii) Orbital Configuration and the time slab

To date, mPWP modelling either using AGCMs or AOGCMs has used the SST reconstructions of the PRISM group either for data-model comparisons with AOGCM output, or to provide prescribed SSTs for the AGCM experiments. These SSTs are created as part of the PRISM palaeo-environmental reconstruction, which along with the

vegetation and orographic reconstructions provide both the mPWP physical boundary conditions for models and datasets for testing their performance. The PRISM data has been reconstructed as a 300 Kyr time slab, primarily through the SST data which was constructed as an average of representative warm peaks (see Chapter 1, Section 1.2.3 *ii*). However each of these peaks represents a different orbital configuration throughout the mPWP (Haywood et al., 2013b). When modelling the time slab, a modern orbital configuration has been used traditionally as at 65°N it represents the average summer mPWP orbital forcing (Haywood et al., 2010). The 65°N orbital forcing is used, as it is considered an important latitude for determining the global response to orbital forcing (Haywood et al., 2010).

Recent work investigating the first Pliocene Time Slice (Haywood et al., 2013b) which developed from work by Dolan et al. (2011), focuses on the effect of different orbits during the mPWP 300 Kyr time slab. During the time slab, it is well established that there were changes in the orbit (i.e. Lasker et al., 2004) and the climate (i.e. Haywood et al., 2013b) and ice sheets (i.e. Dolan et al., 2011). Work underway (Prescott et al., 2014), has shown the impact of different orbital configurations on the simulated mPWP climate, and the importance of using a single warm peak within the mPWP to understand the climate of the period instead of an average of these peaks. The time slice represents a significant change in the methodology of modelling the mPWP, and the methodology for creating the palaeo-datasets used to assess the model performance. It will result in changes to the experimental design of phase 2 of the PlioMIP project and future mPWP modelling studies.

However, while this change represents an improvement to the methodology, it is developing too late for the perturbed physics ensembles (and phase 1 of the PlioMIP project). The presently available palaeo-data uses the time slab approach and therefore the ensemble members must utilise this data. The impact of this choice on the perturbed physics ensembles comes from two perspectives, the effect of the chosen orbit on the simulated climate and the use of warm peak averaged palaeo-data. The first Pliocene time slice (Haywood et al., 2013b) has selected a broad warm peak within the mPWP record with a near modern orbital configuration, which is similar to the orbit used by the perturbed physics ensembles. Comparing a PRISM3D time slab HadCM3 simulation and a time slice simulation, Haywood et al. (2013b), showed minimal differences between mean annual SATs. Therefore, it is possible to be confident that the use of the modern orbital configuration will ensure the results from the perturbed physics ensembles remain relevant as the time slice work develops. The second possible effect on the experimental design, is the use of the time slab derived palaeo-data, particularly the SST data. The warm peak averaging technique used by the PRISM team, means that different orbital

configurations could be driving the warmth at different sites. Therefore, by using the time slab approach, it may be impossible for ensemble members to resolve neighbouring data sites. When analysis of the SST results is undertaken as data-model comparisons, this potential impact should be accounted for in the analysis. However, as the weakness applies to all ensemble members it will not cause a bias in the results for the perturbed physics ensembles. The orbital forcing represents another forcing boundary condition uncertainty within the PPEs produced. The comparison of the results from the PPE to orbital time slices is another potential avenue for future development of the work using Pliocene PPEs.

iii) Site-By-Site Data-Model Comparisons

Site-by-site data-model comparisons have been used extensively, either plotted as site anomalies on global maps (i.e. Pope et al., 2011; Dowsett et al., 2012) or as x/y plots (i.e. Haywood & Valdes 2004), with root mean square error used to rank the performance of models at these comparisons. However, it could be suggested that testing a climate model simulation on its ability to reproduce relatively sparse site-by-site palaeo-data is not a realistic test. Site-by-site DMCs represent the comparison with a signal localised to a small region of the land surface or water column within a model (e.g. HadCM3) grid box with an area of ~20,000 km^2 (ocean) and ~100,000 km^2 (land) at the equator. The methodology to create the DMC takes the value from the nearest grid box to the co-ordinates of the data site. When a site falls on a grid box boundary, the script interpolates using these nearest boxes. The result is that for a marine site, potentially 80,000 km^2 of simulated climate is influencing the DMC at a single locality. Another approach to undertaking data-model comparisons is through the creation of zonal annual mean temperature reconstructions. By producing and assessing both DMC methodologies using the PPE we are able to gain an understanding of the relative strengths and weaknesses of each approach.

The zonal mean is created by averaging the global SST field over all longitudes to create the zonal mean SST for each member of the ensemble. A regional analysis, using only SSTs for the North Atlantic is also created. To create the palaeo-data input for the zonal annual mean data-model comparison, SSTs from PlioMIP Experiment 1 (Atmosphere only GCMs – Haywood et al., 2010) were used to produced a zonal mean representation of the PRISM SST data. The Experiment 1 SST field was created using the February and August PRISM3D SST reconstructions which are then fitted with a sinusoidal curve to create 12 months for each site (Dowsett, 2007; Haywood et al., 2010). The reconstruction was made global by fitting the PRISM SSTs to modern SST contours based on the assumption that surface current systems have not changed greatly between the Pliocene and present day

(Dowsett et al., 1999). The method used represents an oversimplification of the seasonal pattern in SSTs (Dowsett et al., 2011) based on analysis of the seasonal trends in the Reynolds & Smith (1995) dataset. The palaeo-data SST field is treated the same as ensemble SST output to create a zonal annual mean palaeo-data for the DMCs for both North Atlantic and global comparisons.

The use of the zonal mean annual plots for data-model comparisons attempts to address this issue, investigating whether the site-by-site approach is biasing the way models. To date, zonal mean annual plots for both model simulations and palaeo-data have not been used in mPWP data-model comparison studies. Within Chapter 4 a comparison of these two methods for conducting DMCs is undertaken with discussion of any differences they indicate in the performance of ensemble members, and assess the impact of this methodological decision.

2.6.4. Use of the Modern Standard in Data-Model Comparisons

Pliocene palaeo-data and models are compared using their anomalies to modern. Palaeo-data is routinely produced as an anomaly to the modern to reduce differences inherent in calibrating proxy datasets using multiple proxies (Dowsett et al., 2010b). For the ensembles created for this project, there are two possible ways of creating this anomaly:

- Pliocene minus modern ensemble member
- Pliocene ensemble member minus modern Standard

These two methods are defined in Section 2.5.1. The project has made use of the second option, "Pliocene ensemble member minus modern Standard", however this could be interpreted as privileging the Standard configuration of HadCM3. The rationale for choosing this method is that the data-model comparisons will include both the effects of the perturbed parameters as well as boundary condition changes within the ensemble members. If the "Pliocene minus modern ensemble member" comparison is used, the anomaly between Pliocene and modern for each ensemble member is dominated by the differences between the boundary conditions from Pliocene to modern. This is highlighted by Table 2.2.

Ensemble Member	Charney Sensitivity	Pliocene Minus Modern Ensemble Member	Pliocene Ensemble Member Minus Modern Standard
B	2.42	2.348	1.908
D	2.88	2.687	2.069
F	3.75	2.724	2.205
H	3.44	2.465	1.938
I	4.4	2.641	3.031
J	3.9	3.054	2.253
K	4.44	2.427	2.855
L	4.88	2.848	3.223
M	4.54	2.748	3.256
N	4.62	2.605	3.055
O	4.79	2.637	2.872
P	5.4	2.893	3.528
Q	7.11	2.976	4.318
STD	3.3	2.532	2.532

Table 2.2. Comparison of global mean annual surface air temperatures (SAT – C) for "Pliocene minus modern ensemble member" and "Pliocene ensemble member minus Modern Standard" for results from the PRISM2 full ensemble.

Table 2.2 shows the changes in global mean annual SATs for both types of Pliocene minus modern anomaly. While Table 2.2 indicates there is some perturbed parameter induced variation around the Standard comparison for the "Pliocene minus modern ensemble member" comparison, it is relatively small. The indication is that the dominant forcing on the SAT anomaly is the boundary condition changes and not the perturbed parameters. Compared to the "Pliocene ensemble member minus modern Standard" comparison, which includes both the influence of the boundary condition changes from Pliocene to modern, there is much greater variation around the Standard value, indicating the influence of the perturbed parameters on the model anomaly. The variation caused by the perturbed parameters is the focus of the investigation within this thesis, therefore for that reason, the "Pliocene ensemble member minus modern Standard" was used for the data-model comparisons.

Chapter 3: Initial Results

3.1. Background

As discussed in Chapter 1, a number of studies have been conducted using the various evolutions of the Quantifying Uncertainty in Model Predictions (QUMP) methodology for creating Perturbed Physics Ensembles (PPE) for the HadCM3 family of models. However, the Plio-QUMP ensembles were the first use of a QUMP ensemble in a climate warmer than the present day. For this reason, it was decided to undertake a test sub-ensemble to investigate how the PPE was likely to perform in the Pliocene. The sub-ensemble also enabled the data-model comparison methods to be tested to see if any developments would be required for the full ensemble. The temperature, precipitation and 'evaporation minus precipitation' intra-model comparisons and both data-model comparisons represented in this chapter were published in:

Pope, J.O., Collins, M., Haywood, A.M., Dowsett, H.J., Hunter, S.J., Lunt, D.J., Pickering, S.J. & Pound, M.J. 2011. Quantifying Uncertainty in Model Predictions for the Pliocene (Plio-QUMP): Initial Results. *Palaeogeography, Palaeoclimatology, Palaeoecology*. **309.** 128-140.

Additionally, metrics for soil moisture content (SMC), mean sea level pressure (MSLP), sea ice depth and sea ice fraction were added for this chapter to develop the investigation into the initial results. The format of the data-model comparisons has changed for Chapter 3 compared to the Pope et al. (2011) publication, with data-model comparisons now calculated as *"model output minus palaeo-data"* to minimise the visual confusion of discussing a model being too cold, with the figure showing the site as a red colour.

3.2. Ensemble Member Specifics

The initial ensemble members used here are the Standard version of HadCM3 and two ensemble members. All three ensemble members were run at the UK Met Office on the IBM supercomputer between November 2009 and October 2010. The Low Sensitivity member was chosen as it has the lowest Charney sensitivity (temperature response for a doubling of CO_2) of the QUMP ensemble members (2.19°C), the High Sensitivity member was chosen as it had the highest Charney sensitivity (7.11°C – HadCM3 Standard has a Charney sensitivity of 3.3°C). The parameters adjusted are shown in Table 3.1 (a sub-table of Table 2.2 from Chapter 2). The High Sensitivity member refers to member Q from the full ensemble and the Low Sensitivity member refers to member R.

	Label	Identifier/Properties	Low Sensitivity	Standard	High Sensitivity
		Charney Sensitivity	2.19	3.3	7.11
1	VF1	Ice Fall Speed (ms^{-1})	0.93976	1	0.54306
2	CT	Cloud Droplet to Rain Conversion Rate (S^{-1})	1.63E-04	1.00E-04	3.50E-04
3	RHCrit	Threshold of Relative Humidity for Cloud Formation	0.84077	0.7	0.7
4a	CW_{land}	Cloud Droplet to Rain Conversion Rate (kg m^{-3})	1.75E-04	2.00E-04	1.41E-04
4b	CW_{sea}	Cloud Droplet to Rain Conversion Rate (kg m^{-3})	4.25E-05	5.00E-05	3.23E-05
5	MinSIA	Minimum Sea Ice Albedo - Dependence of SIA on Temperature	0.50666	0.5	0.60104
6	ice_tr	Sea Ice Minimum Temperature	9.5243	10	3.836
7	Ent	Entrainment Rate Coefficient	2.89934	3	2.37726
8	Icesize	Ice Particle Size (μm)	30.761	30	28.099
9	Cape	CAPE Closure - Intensity of Convective Mass Flux	Off	Off	Off
10	flux_g0	Boundary Layer Flux Profile	16.9231	10	6.9039
11	Charnoc	Charnock Constant - Roughness Lengths & Surface Fluxes Over Sea	0.0128	1.20E-02	0.0134
12	soillev	Number of Soil Levels Accessed for Evapotranspiration	2	4	4
13	lambda	Asymptotic Neutral Mixing Length Parameter	0.3356	0.15	0.135
14	cnv_rl	Free Convective Roughness Over Sea	4.14E-03	1.30E-03	1.67E-03
15	oi_diff	Ocean Ice Diffusion	3.67E-04	3.75E-04	3.73E-04
16	dyndel	Order of Diffusion Operator	4	6	6
17	dyndiff	Diffusion *e*-folding Time	8.513	12	8.5
18	eacfbl	Cloud Fraction at Saturation Boundary Layer	0.79598	0.5	0.51262
19	eacftrp	Cloud Fraction at Saturation Free Troposphere Value	0.64799	0.5	0.50631
20	k_gwd	Surface Gravity Wavelength	1.98E+04	2.00E+04	1.67E+04
21	k_lee	Trapped Lee Wave Constant	2.97E+05	3.00E+05	2.50E+05
22	gw_lev	Starting Level for Gravity Wave Drag	3	3	3
23a	s_sph-sw	Non-spherical Ice Particles Shortwave Radiation Properties	2	2	7
23b	s_sph-lw	Non-spherical Ice Particles Longwave Radiation Properties	1	1	7
24a	c_sph-sw	Non-spherical Ice Particles Shortwave Radiation Properties	3	3	7
24b	c_sph-lw	Non-spherical Ice Particles Longwave Radiation Properties	1	1	7
25	rhparam	Flow Dependant RHcrit	0	0	1
26	vertcld	Vertical Gradient of Cloud Water	1	0	0
27	canopy	Surface Canopy Energy Exchange	0	0	1
28	cnv_upd	Convective Anvils Updraft Factor	0.51016	Off	Off
29	anvil	Convective Anvils Shape Factor: Radiative Properties of Convective Cloud	2.59138	Off	
30	sto_res	Dependence of Stomatal Conductance on CO_2	0	1	1
31	f_rough	Forest Roughness Lengths	0	0	3
32	sw_absn	Shortwave Water Vapour Continuum Absorption	1	0	0

Table 3.1. The perturbed parameter labels, identifiers and values in the three simulations (High Sensitivity, Standard simulation and Low Sensitivity). Parameter names and identifiers are as in Table 2.1a.

The methodology for modelling and testing the ensemble members is the same as that set out in Chapter 2 with the exceptions of:

- Data-model comparisons using the PRISM3D MASST dataset were calculated using absolute climate data (Pliocene Model minus Pliocene Data).
- The vegetation derived surface air temperature data (Salzmann et al., 2013) was not used.
- BIOME4 was run in the absolute climate mode with geographically constant soils.

The altered methods described in Chapter 2 are the result of discussions driven by the initial results paper and recent developments in the palaeo-data.

3.3. Intra-Model Comparisons

3.3.1. Pliocene Minus Modern

The Standard Pliocene simulation was compared to a modern transient simulation (run at the UK Met Office as part of the QUMP project) with the period 1950 to 1985 averaged to create the modern results. The period corresponds with the time period of Reynolds & Smith (1995), whose data was used to produce the core top estimates of the Pliocene Research Interpretations and Synoptic Mapping (PRISM) 3D Mean Annual Sea Surface Temperature (MASST – Dowsett et al., 2010b) dataset.

Simulation	Global Mean Annual Temperature Anomaly	Global Mean Annual Precipitation Anomaly
'Pliocene Minus Modern' (High Sensitivity)	3.3°C	0.15 mm/day
'Pliocene Minus Modern' (Standard simulation)	2.5°C	0.18 mm/day
'Pliocene Minus Modern' (Low Sensitivity)	2.4°C	0.12 mm/day
'Pliocene High Sensitivity Minus Standard'	1.5°C	-0.18 mm/day
'Pliocene Low Sensitivity Minus Standard'	-0.1°C	-0.06 mm/day

Table 3.2. – The mean annual temperature and precipitation values for 'Pliocene minus modern' for the High Sensitivity, Low Sensitivity and Standard simulations and for the Pliocene 'High Sensitivity minus Standard' and Pliocene 'Low Sensitivity minus Standard' simulations.

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 3.2.

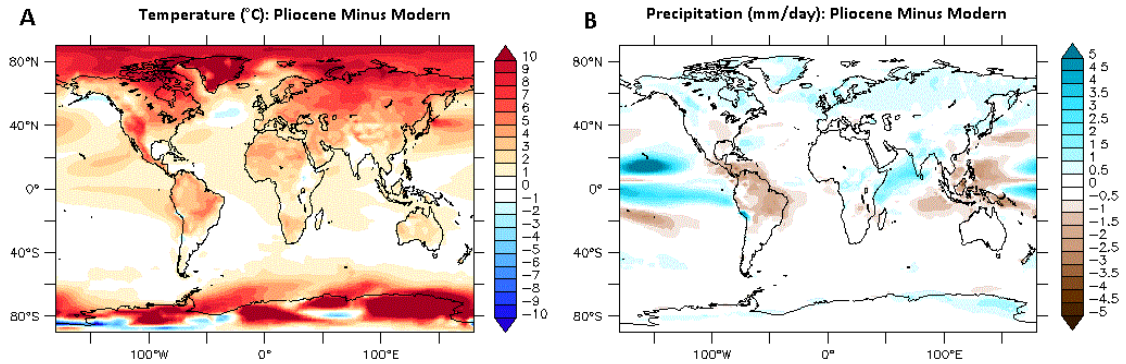


Figure 3.1. Global annual mean Pliocene minus modern plots for the Standard simulation for A) temperature and B) precipitation.

The Pliocene is modelled as being warmer than modern. Figure 3.1a 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 reduced equator to pole temperature gradient in the Pliocene compared to the modern (Dowsett et al., 2010a). PRISM3D MASST data compared to Reynolds & Smith (1995) core top data for sites ODP 911 (80°N) and ODP 847 (0.18°N) shows a temperature gradient of 24.1°C in the modern compared to 14.9°C in the Pliocene. Figure 3.1b 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 precipitation (e.g. Amazonia and Indonesia).

Table 3.2 reinforces the global plots shown in Figure 3.1 with the global mean averages for the 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 across the ensemble members. These differences are driven by the difference in boundary conditions - i.e. orography (Rocky Mountains 50% lower than modern) and CO_2 values as the modern simulations are the corresponding PPE member. The interaction between the difference in the Pliocene and modern boundary conditions will have caused the variation in global means (Table 3.2) and regional differences (Figure 3.1) displayed between the three simulations.

3.3.2. Perturbed Physics Simulations ("Ensemble Member Minus Standard")

i) Surface Temperature

As shown in Table 3.1, the High Sensitivity simulation has a Charney Sensitivity of 7.11°C compared to the 3.3°C of the Standard simulation and the 2.19°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 'High minus Standard' comparison (Figure 3.2a) 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' plot (Figure 3.2b) shows a cooling anomaly over oceanic regions and high latitudes with some areas of warm anomaly on continental areas.

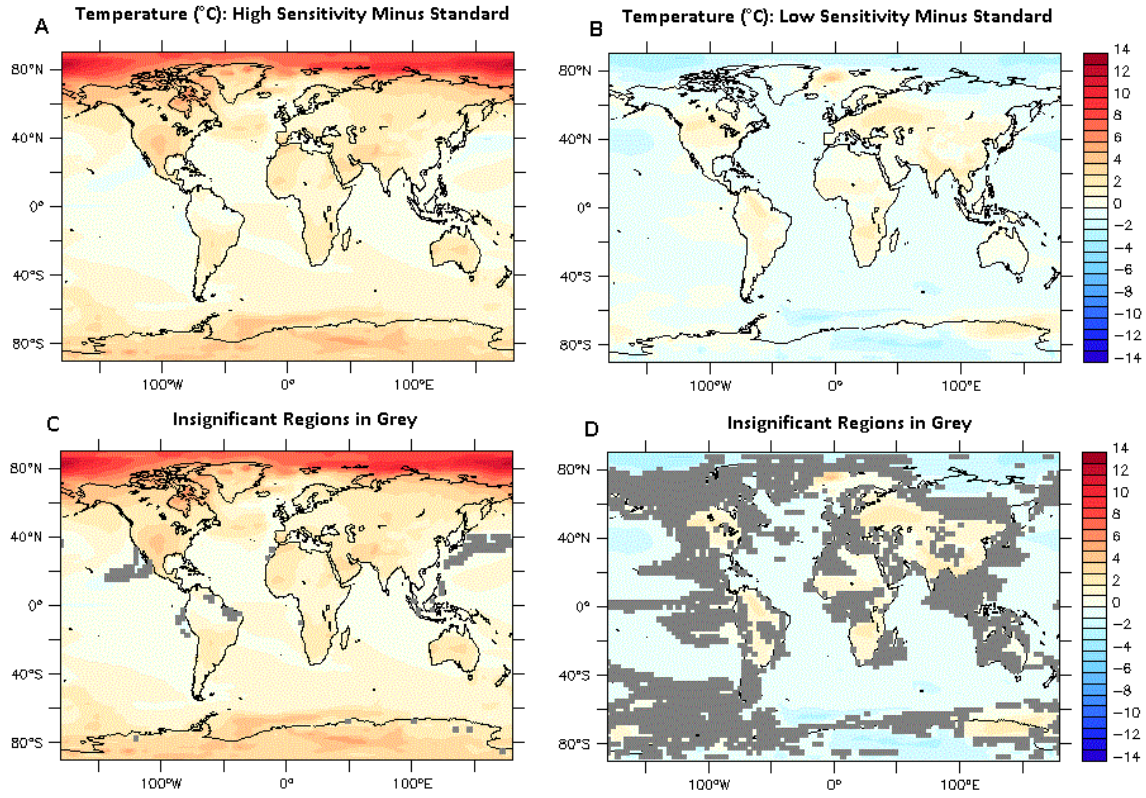


Figure 3.2. Global annual mean temperature plots for A) 'High Sensitivity minus Standard' and B) 'Low Sensitivity minus Standard' in °C. Student's T-test was applied to the comparisons and the insignificant regions, are plotted in grey and overlain over the (C) 'High minus Standard' and (D) 'Low minus Standard' plots.

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 (Figure 3.2c), which are the two large cool anomalies on the plot (Figure 3.2a). For the 'Low minus Standard' anomaly there are slightly more areas where the anomaly was not significant (Figure 3.2d), 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 while only showing minimal warming in the tropics (Figure 3.2a). 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 (Figure 3.8a). In continental areas, the model predicts a warm anomaly in Australia of 2 to 4°C and in 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.

ii) Precipitation

The precipitation patterns in the two plots display a greater range in the 'High minus Standard' (Figure 3.3a) than in the 'Low minus Standard' anomaly (Figure 3.3b). In both experiments there is little change in global average precipitation (Table 3.2), 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 while tropical and extra-tropical areas see a decrease in precipitation. In the 'Low minus Standard' plot the general trend is for a slight reduction in precipitation, but there are fewer and weaker areas of major reduction.

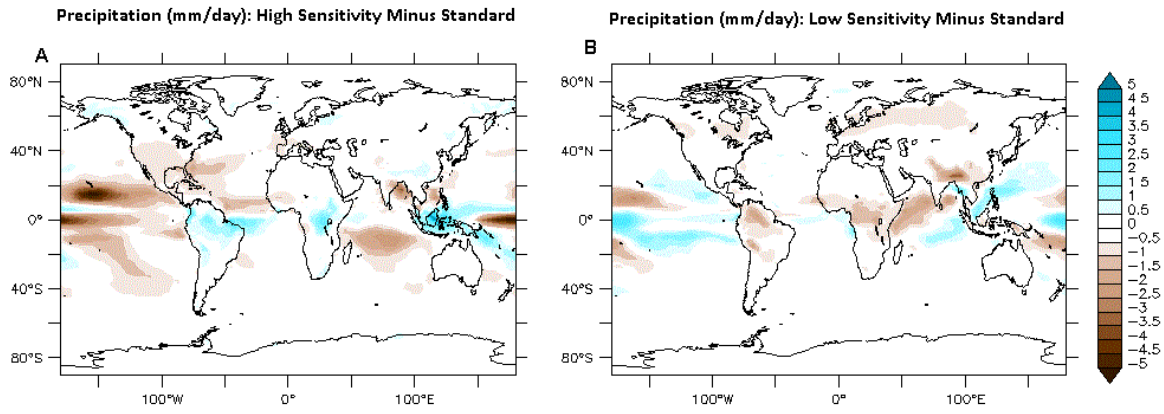


Figure 3.3. – Global annual mean precipitation plots for A) 'High Sensitivity minus Standard' and B) 'Low Sensitivity minus Standard' in mm/day.

The 'High minus Standard' precipitation shows an important reduction in the daily average rainfall for the continental USA and Australia in regions that have some of the largest continental warming anomalies (Figure 3.2). 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 plots with

increases of approximately 2 mm/day through northern South America, central Africa and Indonesia.

iii) Evaporation Minus Precipitation

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 precipitation over most continental areas and high latitudes oceans with a net increase in evaporation from tropical oceans. The patterns for both the 'High minus Standard' (Figure 3.4a) and 'Low minus Standard' anomaly (Figure 3.4b) are very similar. The 'High minus Standard' has greater areas of net evaporation over land than the 'Low minus Standard', but at the same time has more intense areas of net precipitation in tropical forest regions.

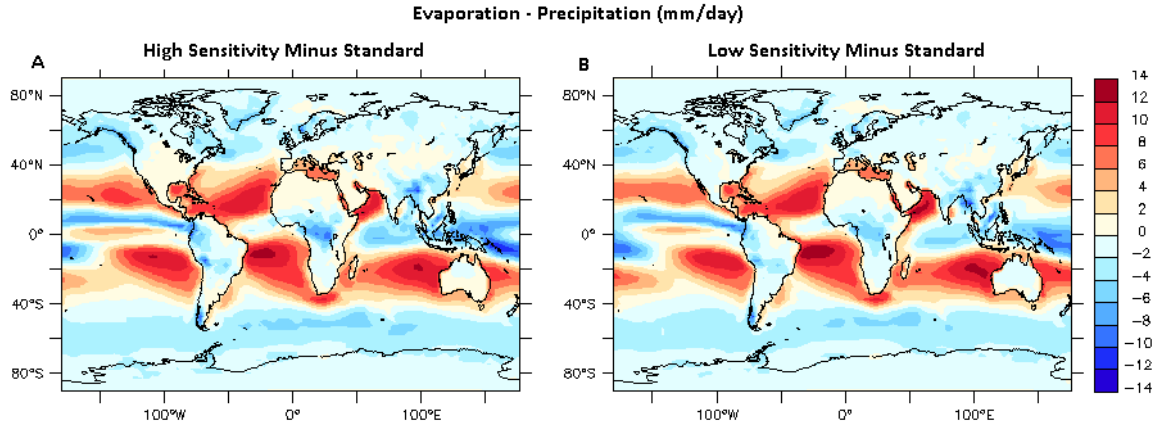


Figure 3.4. Global annual mean evaporation-precipitation plots for A) 'High Sensitivity minus Standard' and B) 'Low Sensitivity minus Standard' in mm/day with negative numbers indicating that there has been net precipitation and positive numbers indicating that there has been net evaporation. These anomalies were calculated using 'high minus (negative Standard)' and 'low minus (negative Standard)'.

iv) Soil Moisture Content (SMC)

In the 'High minus Standard' (Figure 3.5a), the overall trend is for little change in the soil moisture anomaly, with anomalies in most areas less than 30 kg/m^3. There is a pattern of a reduction in soil moisture over lower latitudes with a slight increase at high latitudes. The only difference to this is an area of increasing SMC in north-east South America. There is also a pocket of drying on the east coast of the USA and also in southern Europe.

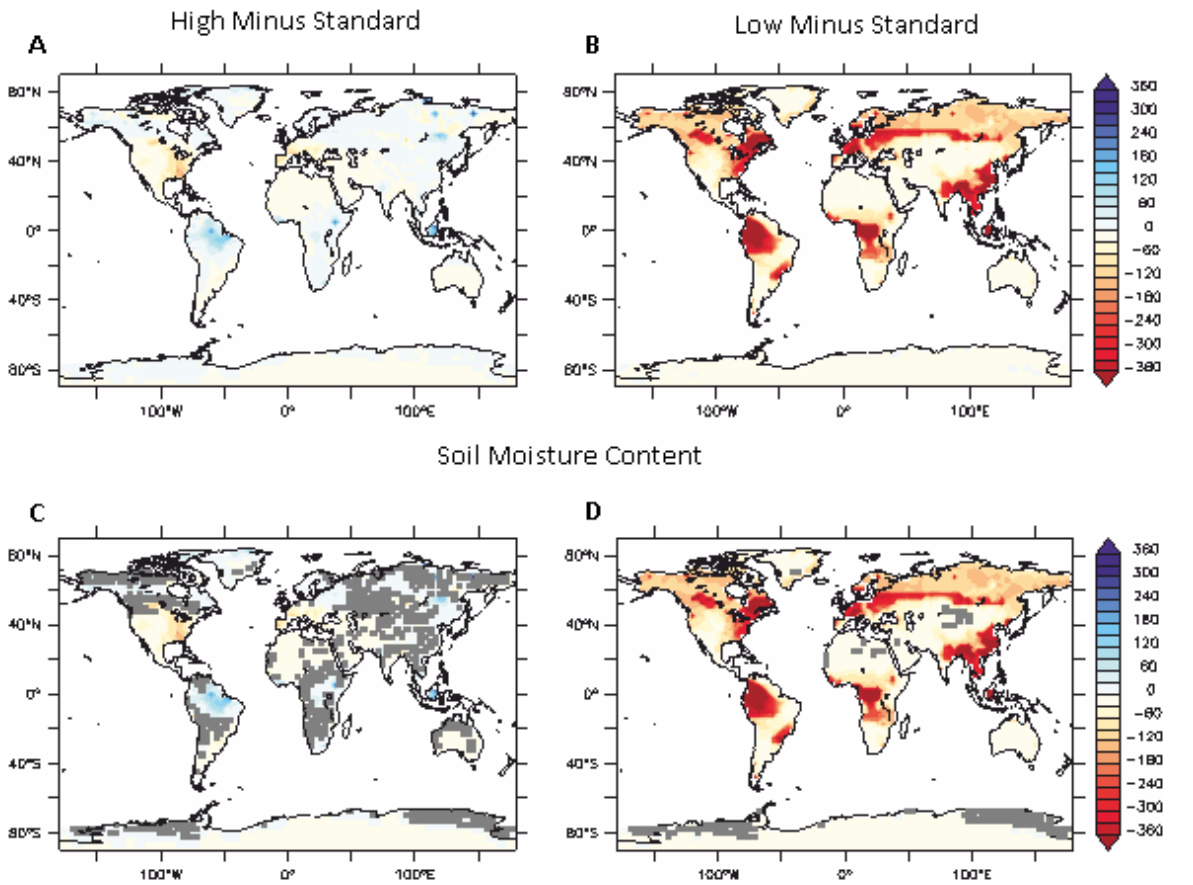


Figure 3.5. Global annual mean soil moisture content (SMC) plots for A) 'High Sensitivity minus Standard' and B) 'Low Sensitivity minus Standard' in kg/m². Student's T-test was applied to the comparisons and the insignificant regions, are plotted in grey and overlain over the (C) 'High minus Standard' and the (D) 'Low minus Standard' plots.

In the 'Low minus Standard' (Figure 3.5b), there are no areas of increased SMC, with intense areas of reduced soil moisture in tropical areas and at high latitudes. There is a particularly intensive band of reduced soil moisture laterally across northern and Eastern Europe and Russia, at about 60°N. The 'High minus Standard' has more areas of insignificant data, but the main areas of intensive change in both plots are significant.

v) Mean Sea Level Pressure (MSLP)

At polar latitudes the mean sea level pressure (MSLP) patterns are at opposites between the 'High minus Standard' (Figure 3.6a) which shows a reduction in pressure at both poles and the high latitudes and 'Low minus Standard' (Figure 3.6b) which shows an increase at the high latitudes compared to the standard. Both anomalies show more intensive change over Antarctica than at the Arctic. The only other noticeable feature in the MSLP plots, is a region of intense increased pressure anomaly in the South Pacific in the 'High minus Standard' plot, that is mirrored in the North Pacific albeit with a less intense anomaly.

Continentally, the trend is for a slight decrease in pressure between both ensemble members compared to the standard.

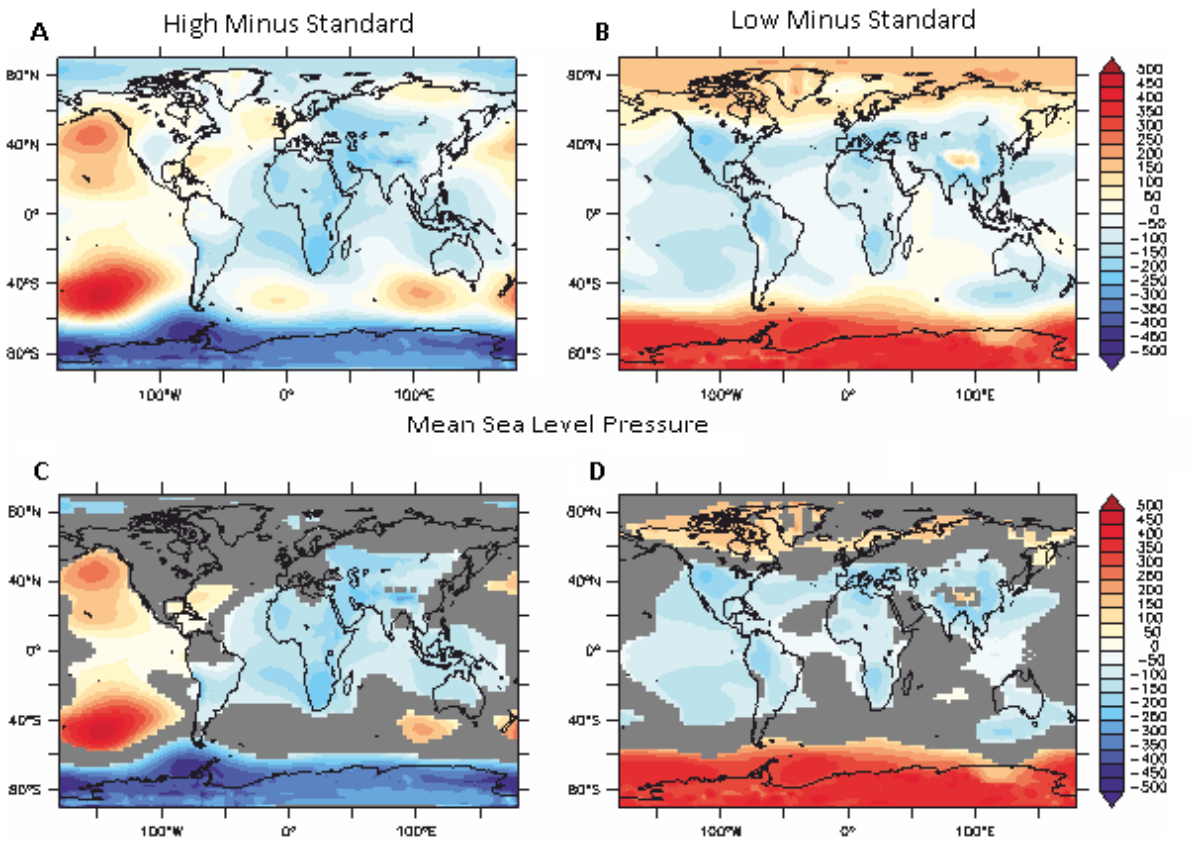


Figure 3.6. Global annual mean plots of mean sea level pressure (MSLP) for A) 'High Sensitivity minus Standard' and B) 'Low Sensitivity minus Standard' in Pascal's (Pa). Student's T-test was applied to the anomaly plots and the insignificant regions, which are plotted in grey and overlain over the (C) 'High minus Standard' anomaly and the (D) 'Low minus Standard' anomaly.

The strong changes compared to the standard in both simulations over the Antarctic and the strong regions of increased pressure in the 'High minus Standard' plot are significant, as is the reduced MSLP in the North Atlantic 'Low minus Standard' plot. However the reduced pressure in the North Atlantic in the 'High minus Standard' plot was shown as being insignificant.

vi) Sea Ice

Sea ice responds as expected to the different forcings between the ensemble members, the effects of which are highlighted by the SATs (Figure 3.2). In the Arctic for the 'High minus Standard', the fraction (Figure 3.7a) and depth anomalies (Figure 3.7c) for sea ice are all greatly reduced. For the 'Low minus Standard' in the Arctic, the fraction (Figure 3.7b) and depth anomalies (Figure 3.7d) show generally increased sea ice. Compared to the Standard, this is consistent with the temperature plots which shows a significant warming

in the 'High minus Standard' comparison (Figure 3.2a) and cooling for the 'Low minus Standard' (Figure 3.2b). This indicates that in the High Sensitivity simulation, less sea ice covers each grid box than in the Standard simulation and the sea ice that exists is thinner. The implication is that the model simulates widespread summer melting, ensuring that mainly young thin sea ice is all that is able to form (Stroeve et al., 2012), with the opposite applying for the Low Sensitivity simulation.

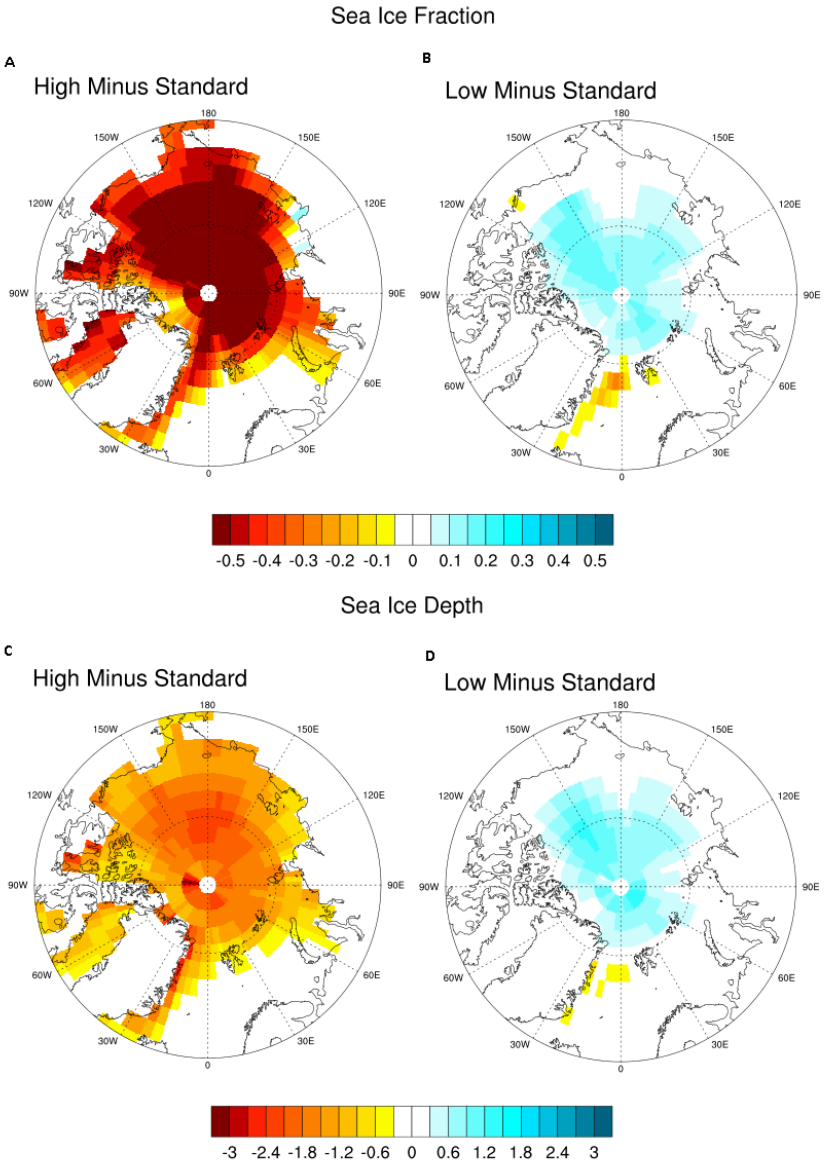


Figure 3.7. Global annual mean Arctic sea ice cover fraction plots for A) 'High Sensitivity minus Standard' and B) 'Low Sensitivity minus Standard' and sea ice thickness (in metres) for (C) 'High Sensitivity minus Standard' and (D) 'Low Sensitivity minus Standard'.

3.4 Data-Model Comparison

3.4.1 Data-Model Comparison – Sea Surface Temperature (SST)

Figure 3.8a 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 et al., 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 mismatch 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 (Figure 3.8b) 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 Charney sensitivity of 7.1°C. The general pattern observed in Figure 3.2a 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 Figure 3.2b (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 data-model comparison at one site is improved by about 1 to 2°C, to within the uncertainty in the data values.

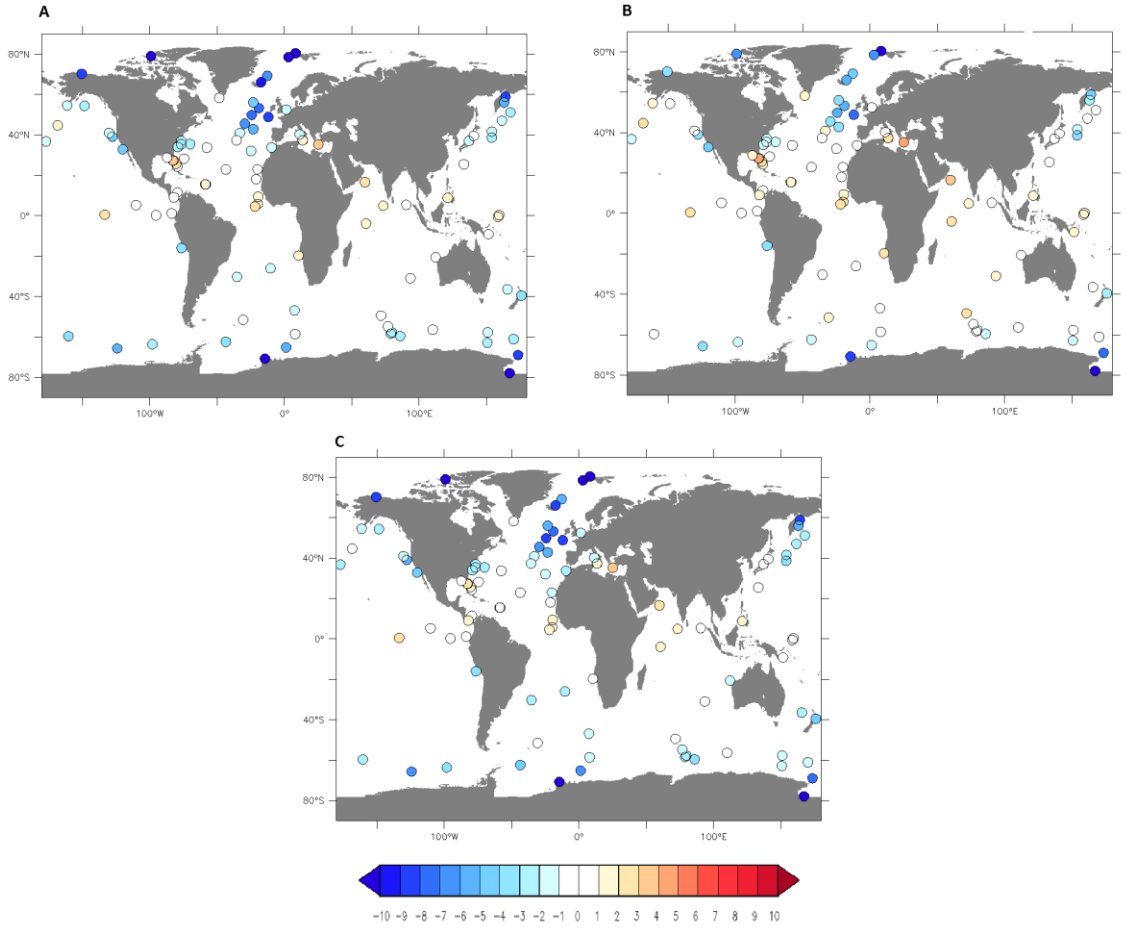


Figure 3.8. Data-model comparison using the PRISM3D Mean Annual Sea Surface Temperature (MASST) dataset. A) Standard simulation minus PRISM3D MASST, B) High Sensitivity simulation minus PRISM3D MASST and (C) Low Sensitivity simulation minus PRISM3D MASST in °C. Root mean square errors (RMSE) were calculated for each comparison as A) 4.37°C, B) 3.25°C and (C) 4.38°C.

In areas away from the North Atlantic, all three simulations perform very similarly in the data-model comparison. The Standard simulation data-model comparison shows that the model over-estimates warming in the tropical regions, causing a negative data-model mismatch, 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 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 while 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 ensemble member at recreating the conditions of the mPWP based on the Root Mean Square Errors (RMSE). These were 3.25°C for the High Sensitivity simulation compared to 4.37°C for the Standard simulation and 4.38°C for the Low Sensitivity simulation.

3.4.2. Data-Model Comparison – Vegetation Biome Reconstruction

The outputs from the BIOME4 model were compared with the Piacenzian age palaeobotanical database of Salzmann et al. (2008) using Cohen's Kappa statistic (Cohen, 1960 - Chapter 2, Section 2.5.5). 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 0.220 for the full Biome classification and 0.338 for the Megabiome (see Section 2.5.5, Chapter 2) classification (Figure 3.9a), with scores of 0.216 (full) and 0.303 (mega) for the High Sensitivity simulation (Figure 3.9b) and scores of 0.183 (full) and 0.297 (mega) for the Low Sensitivity simulation (Figure 3.9c). 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.

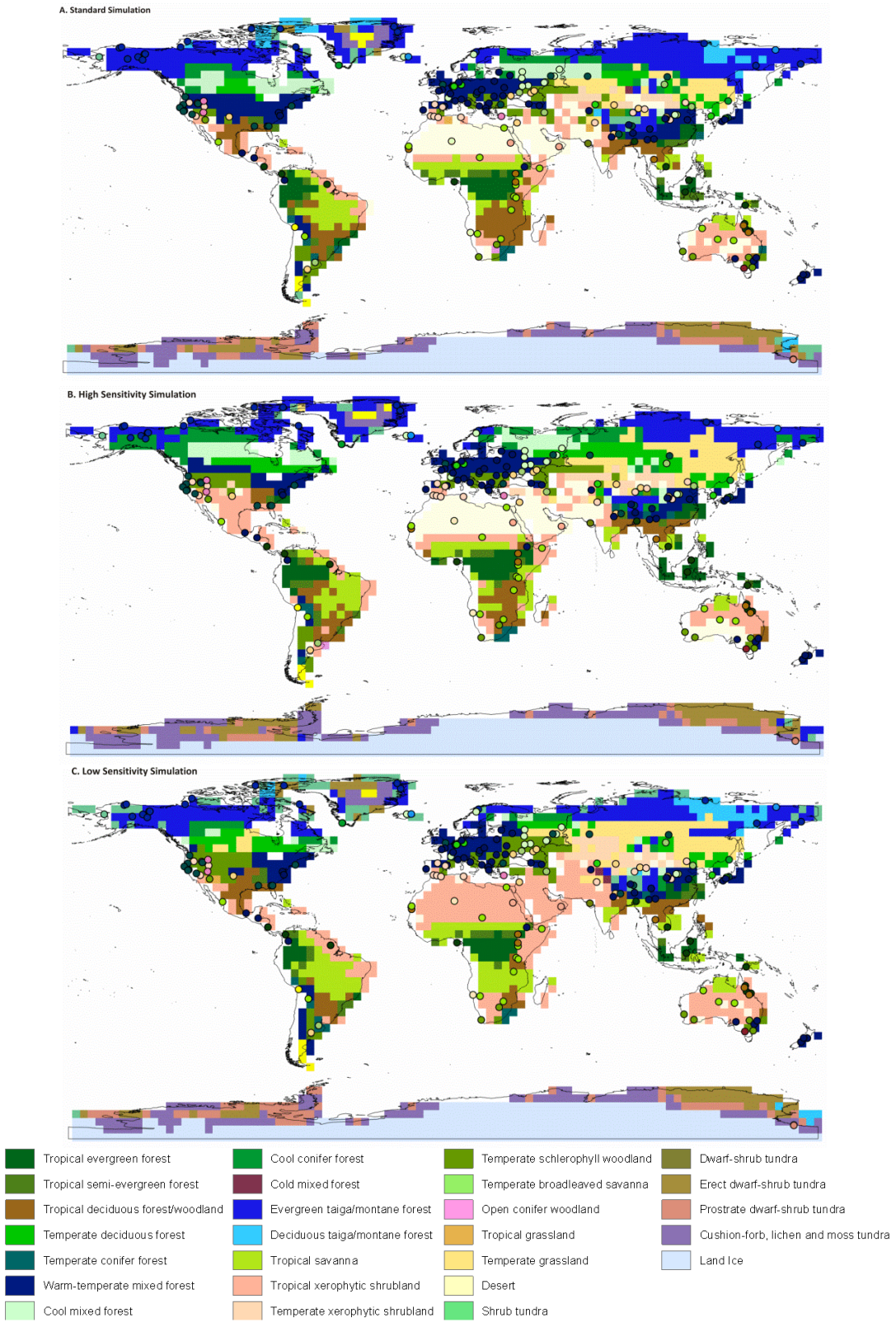


Figure 3.9. BIOME4 outputs for A) the Standard simulation, (B) High Sensitivity simulation and (C) the Low Sensitivity simulation. The Kappa statistics for each simulation where Standard full Biome classification 0.201 and for the megabiome 0.229, High Sensitivity 0.186 (full) and 0.172 (Mega) and for the Low Sensitivity simulation 0.120 (Full) and 0.162 (Mega).

Regional comparisons 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 palaeobotanical data shows 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 indicates the presence of tropical savanna, whereas the model predicts warm-temperate to temperate forest. In southern South America the data again disagrees 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 Sensitivity simulations predict temperate needle-leaf forests and temperate sclerophyll woodland for this region, whereas the High Sensitivity simulation predicts tropical xerophytic shrubland and desert. South America is one of the areas that are 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 skillful 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 indicates 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 for 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 displays 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 skillful 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 further 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, generating a drier biome than that which the data indicate existed at the time.

3.5. 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., 2000a) and later for data-model comparisons with fully coupled AOGCMs (Dowsett et al., 2011). While the dataset has developed into an ever more detailed palaeoenvironmental reconstruction for the mPWP, with the addition of a detailed vegetation biome reconstruction (Salzmann et al., 2008), the SST dataset is the only globally extensive quantitative temperature reconstruction for the period. This has led to a situation where the aim of improving mPWP 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 while not weakening the areas of good data-model agreement. The results from the 'PRISM3D MASST minus High Sensitivity simulation' data-model comparison (Figure 3.8b), show that a simulation with a higher Charney sensitivity (than the Standard version) 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 (Figures 3.2a, 3.3a & 3.4a) 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 biome 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. Additionally a model output independent of the BIOME4 inputs, the soil moisture content 'High minus Standard' anomaly (Figure 3.5a) supports

this drying of North America, with it's main area of decreasing SMC anomaly being in the continental USA.

A feature of the HadCM3 model has been the drying out of tropical forest regions (i.e. the Amazon - Cox et al., 2004). Yet despite increasing the Charney sensitivity in the High Sensitivity version of the model, the HadCM3 model simulated a small (<1°C) warming anomaly (to the Standard version (Figure 3.2a)) and increased precipitation (Figure 3.3a) and E-P (Figure 3.4a) anomalies to the Standard. One change for this ensemble member was the increasing of surface roughness length which has been shown (Betts et al., 1997, 2004) to increase precipitation and reduce the temperature effects in the Amazon region. It will be useful to look at whether this effect is replicated throughout the rest of the QUMP ensemble members which make this change, as such an effect would be useful for decreasing the data-model mismatch in the high latitude SST comparisons, while not increasing the mismatch to terrestrial vegetation.

The response of sea ice to the parameterisations can be measured both as an response to the changes in the climatology, but also to changes in the parameterisations. Sea ice is one of the most dynamic features of the model, responding quickly to forcing changes and exerting its own forcing change on the simulated climate. The changes to the Low Sensitivity version compared to the Standard simulation are relatively minor, however the changes in the High Sensitivity simulation could have aided the response of the sea ice to the warmer climate. The changes to the sea ice parameters are focussed on: minimum sea ice albedo (*MinSIA*), temperature of maximum albedo for sea ice (*ice_tr*), and the ocean-ice heat diffusion coefficient (*oi_diff*), parameters 5, 6 & 15 respectively in Table 3.1. The relationship in the sea ice model component between temperature and albedo was described in Chapter 2 (Section 2.3.1). The parameter changes in the High Sensitivity simulation to *'MinSIA'* and *'ice_tr'* do not seem to lead to a significant change in the sea ice response due to these specific parameters. It causes only a sharper response from the maximum to minimum albedo due to temperature due to a warmer maximum sea ice albedo temperature. This would largely indicate the changes in sea ice and the high latitude warming are a response (in the High Sensitivity simulation) to other warming factors and not due to parameter sets in the sea ice component that promote melting of sea ice. Similarly the High Sensitivity settings for *'oi_diff'* are a marginal change from the Standard settings towards reducing ice melt as the coefficient of heat diffusion from ocean to ice has been slightly reduced. The cause(s) of high latitude warming in the ensemble is important for understanding the improvements that could be made to the existing data-model mismatch.

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 simulation showed an interesting contrast between the land and the ocean (Figure 3.2b). 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 on continental areas. 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. This would also hopefully reduce the SMC drying anomaly which is very apparent in the 'Low minus Standard' anomaly in a number of key locations (i.e. tropical forest belts, northern Europe). This fits with the BIOME4 output for the Low Sensitivity simulation, which performed weakest on the Kappa statistical test. The moisture budgets are the key weak link for many of the regions of poor data-model comparison.

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., 2011), is the next stage for this investigation. 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. It does not allow a focus on one aspect of data-model comparison defining whether an ensemble member has improved the comparison. 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 skillfully in comparison with the available proxy data and where it is failing.

The data-model comparison of Figure 3.8b illustrates that the High Sensitivity simulation was unable to achieve all the necessary warmth in the higher latitudes to align data and model results. It modelled a large reduction in the fraction and thickness of sea ice (Figure 3.7a,c) and continental areas that were too warm and dry 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 three 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 warming influence (Hill et al., 2011). Like topography, ocean bathymetry could be controlling the flows of heat into the Arctic Ocean. 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). The 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. The final boundary condition that could affect the results is the greenhouse gas concentrations in the model. For Pliocene modelling this focuses on CO_2, as there is no proxy for any other greenhouse gas, so they are assumed to be at pre-industrial levels. CO_2 has been shown to range between 300 and 425 ppmv during the mPWP (see Chapter 1, Section 1.3.4), and a combination of a lower CO_2 value with higher sensitivity simulation may produce a better data-model fit than using 400 ppmv in the model. Combining this work with the work of these ensembles could further reduce mismatches 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 these initial results represent 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., 2011). They both perform within the range of validation tests that were undertaken by Collins et al. (2011) and for that reason they were considered acceptable simulations for use in this work. 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 results be known. While 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.

The use of Student's T-test were inconclusive in the initial results, with it producing regions of insignificance for temperature, soil moisture content and mean sea level pressure anomaly plots. 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. This pattern of plots with more areas of larger anomalies being found to be more statistically significant was repeated across all the successfully T-tested intra-model comparison metrics. The variation in the Charney 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.

3.6. Conclusions & Development

3.6.1. Conclusions

The initial results show that PPEs offer a tremendous opportunity to address the mismatches 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 thesis 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 Pliocene Research Interpretation and Synoptic Mapping (PRISM) 3D Mean Annual Sea Surface Temperature (MASST) dataset (Dowsett et al., 2010b), especially in the North Atlantic and Arctic Oceans. However, it was less skillful in comparison with the palaeobotanical reconstruction than the mPWP Standard simulation. The Low Sensitivity simulation was less skillful in both data-model comparisons than the High Sensitivity simulation and the Standard simulations.

The High Sensitivity simulation performed less skillfully in the vegetation biome data-model comparison because it warmed both areas over ocean and land by large amounts. While this improved the PRISM3D 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, a conclusion supported by the soil moisture content data.

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. The 'Low minus Standard' simulation (Figure 3.2b) displayed a contrasting temperature pattern between the ocean and land. Unlike the final full PPE, the initial three-experiment ensemble is not sufficiently large to investigate the parameters that cause this contrast. The thesis 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.

It is evident that while the atmospheric parameter PPEs are a good starting point for this investigation, the work will not be complete without analysis (where possible) of other causes of uncertainty, such as perturbed ocean parameters, or from analysing the uncertainty created by key boundary conditions such as topography or greenhouse gas concentrations. Part of the development of this work will be to determine whether it is oceanic parameters or boundary conditions that are investigated in experiments beyond those that make up the full ensemble. While the oceanic parameters would investigate fully uncertainty in key components of HadCM3, these experiments would require long integrations which may not be feasible within the time-frame of the investigation, whereas boundary condition ensemble experiments could be performed in a similar time scale to the ensembles of atmospheric parameterisations.

3.6.2. Developments Arising from the Initial Results

The initial results indicated that there was a use for a PPE in the Pliocene for working on reducing the data-model mismatch and potential for quantifying the uncertainty in the Pliocene. The initial results displayed the need for a larger ensemble for understanding the effects of parameterisations (such as the cause of high latitude warming). From the tests used here, developments to the data-model comparisons will include the use of the anomaly mode for BIOME4 and for the MASST data-model comparison to be undertaken using 'Pliocene minus modern' anomalies. The use of anomalies increases the robustness

of the results as it removes any biases in the model output either due to the model configuration or the parameter perturbations. It is aimed that these developments on the initial results will lead to better analysis of the results with greater statistical strength in the tests used to determine the value of an ensemble member.

Chapter 4: Intra-Model Assessment of mid-Pliocene Climate: Can Geological Data and Climate Models be Reconciled?

4.1. Introduction

Chapter 3 outlined the first application of and the initial results from a perturbed physics ensemble (PPE) for the mid-Pliocene Warm Period (mPWP), a warmer than modern climate. These results were generated from a sub-ensemble consisting of the Standard version of HadCM3 and the full ensemble outliers, the High Sensitivity and Low Sensitivity simulations. The sub-ensemble was used to check the validity of using a PPE in the mPWP and testing the methods for the data-model comparisons which led to further development and improvement to the methods.

Chapter 4 presents the intra-model comparisons from ensembles using both the full perturbed physics ensemble and the CO_2 sub-ensemble forced with the PRISM3D boundary conditions. These ensembles will be used to produce an assessment of whether it is possible to reconcile mid-Pliocene geological data with climate model simulations of the mPWP. Chapter 4 will only assess the intra-model variation of ensembles forced by the PRISM3D boundary conditions. Analysis of the results from both the PRISM2 and PRISM3D ensembles highlighted that there was a similar intra-model comparison for both ensembles. As the PRISM3D ensemble represents the latest boundary conditions, which will be used in future mid-Pliocene modelling studies, these results have been presented here. The performance of the PRISM3D boundary conditions in comparison to the PRISM2 boundary conditions will be assessed in Chapter 5. The goal of Chapter 4 is to answer questions about the perturbed physics ensembles and the reconciliation of data and models and what the impact of this will be on future modelling studies of the mPWP.

4.2. Large-Scale Features of the Perturbed Physics Ensemble

4.2.1. Pliocene Standard Minus Modern Standard

The Standard simulation highlights the main difference between the simulated Pliocene and modern climates. Pliocene global mean annual warming is 2.94°C (Table 4.1), which can be attributed to a combination of physical boundary condition and CO_2 changes.

		CO_2 400 ppmv			CO_2 350 ppmv	CO_2 300 ppmv
Ensemble Member	**Charney Sensitivity**	**Surface Air Temperature**	**Sea Surface Temperature**	**Precipitation**	**Surface Air Temperature**	**Surface Air Temperature**
B	2.42	1.87	0.548	0.064	1.16	0.58
D	2.88	2.10	0.69	0.03	--	--
F	3.75	2.65	0.97	0.11	--	--
H	3.44	2.62	1.08	0.19	--	--
I	4.40	4.06	1.70	0.13	--	--
J	3.90	2.30	0.67	-0.07	--	--
K	4.44	3.68	1.59	-0.10	--	--
L	4.88	4.05	1.65	0.06	--	--
M	4.54	4.02	1.70	0.22	2.95	2.05
N	4.62	3.84	1.51	0.13	--	--
O	4.79	3.84	1.70	0.13	--	--
P	5.40	4.46	1.94	0.18	3.33	2.54
Q	7.11	5.27	2.30	0.05	4.04	3.16
Standard	3.30	2.94	1.25	0.18	2.10	1.43
	B to Q SAT Range	3.40	1.75		2.88	2.58

Table 1. Global mean annual values of *"Pliocene ensemble member minus modern Standard"* for surface air temperature (SAT - °C), sea surface temperature (SST - °C), Precipitation (mm/day) for the full perturbed physics ensemble run with atmospheric CO_2 at 400 ppmv. Additionally, the SATs for the CO_2 sub-ensemble members (B, M, P, Q & the Standard) are displayed. For reference the Charney sensitivity of each ensemble member is shown as calculated in Collins et al. (2006).

126 As discussed in Chapter 2 (Section 2.4) there are differences in the physical model boundary conditions between the Pliocene and modern that include changes in the land surface cover, ice sheets and topography that are required for the model simulations. Reductions in the extent of the Greenland Ice Sheet (GrIS) and East Antarctic Ice Sheet (EAIS) lead to consistent high latitude warming across all ensemble members of 7 to 10°C (Figure 4.1), in line with previous Pliocene modelling studies (Bragg et al., 2012). There is warming of 0.5 to 1.5°C across the majority of the global oceans with exceptions through the North Atlantic, Arctic Ocean and Weddell Sea (Figure 4.1, Appendices A1 & A2). North Atlantic warming reaches 4 to 5°C, whilst across

the Arctic Ocean and the high northern continental latitudes warming is enhanced compared to the equatorial and tropical latitudes with warming of 3 to 5°C. The Weddell Sea displays warming of 6 to 7°C.

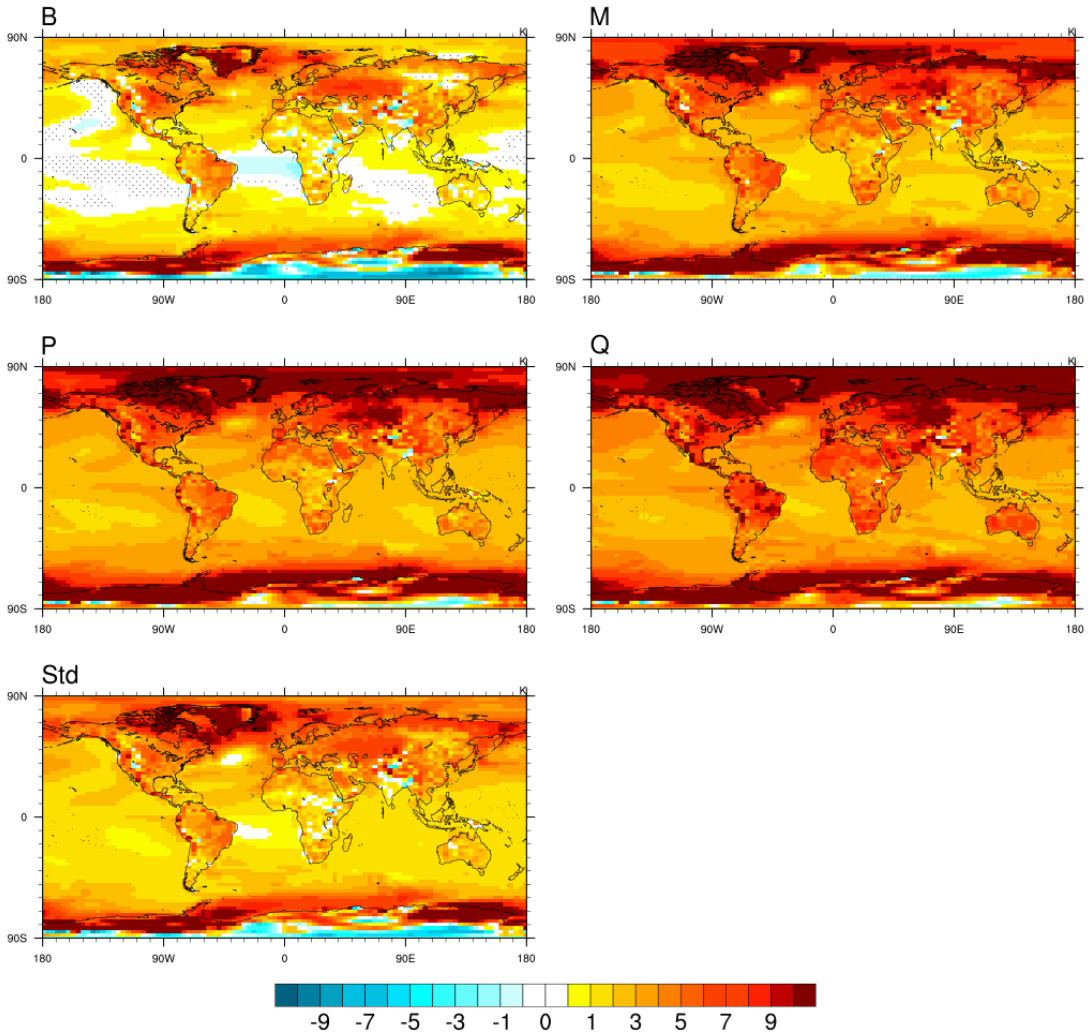


Figure 4.1. Global annual mean plots for *"Pliocene ensemble member minus modern Standard"* for surface air temperature (SAT - °C) for ensemble members B, M, P, Q & the Standard. The remaining ensemble members are displayed in Appendix A1. The plots for Sea Surface Temperatures (SST- °C) are displayed in Appendix A2.

Global mean precipitation is increased (0.175 mm/day – Table 4.1) in the Pliocene compared to the modern. Precipitation increases over the GrIS and reduces over the EAIS by 0.25 to 1 mm/day (Figure 4.2). Globally, precipitation rates reduce (1 to 2 mm/day) over the northeast coast of South America, and in the tropical South Pacific and South Atlantic (0.5 to 1.5 mm/day). There is also a small reduction in precipitation over continental USA (0.25 mm/day). Precipitation increases over Indonesia (1 to 2 mm/day), Western Europe (0.5 to 1.5 mm/day), and along the equator (0.5 to 1 mm/day). In other areas there is minimal change between the Pliocene and modern precipitation.

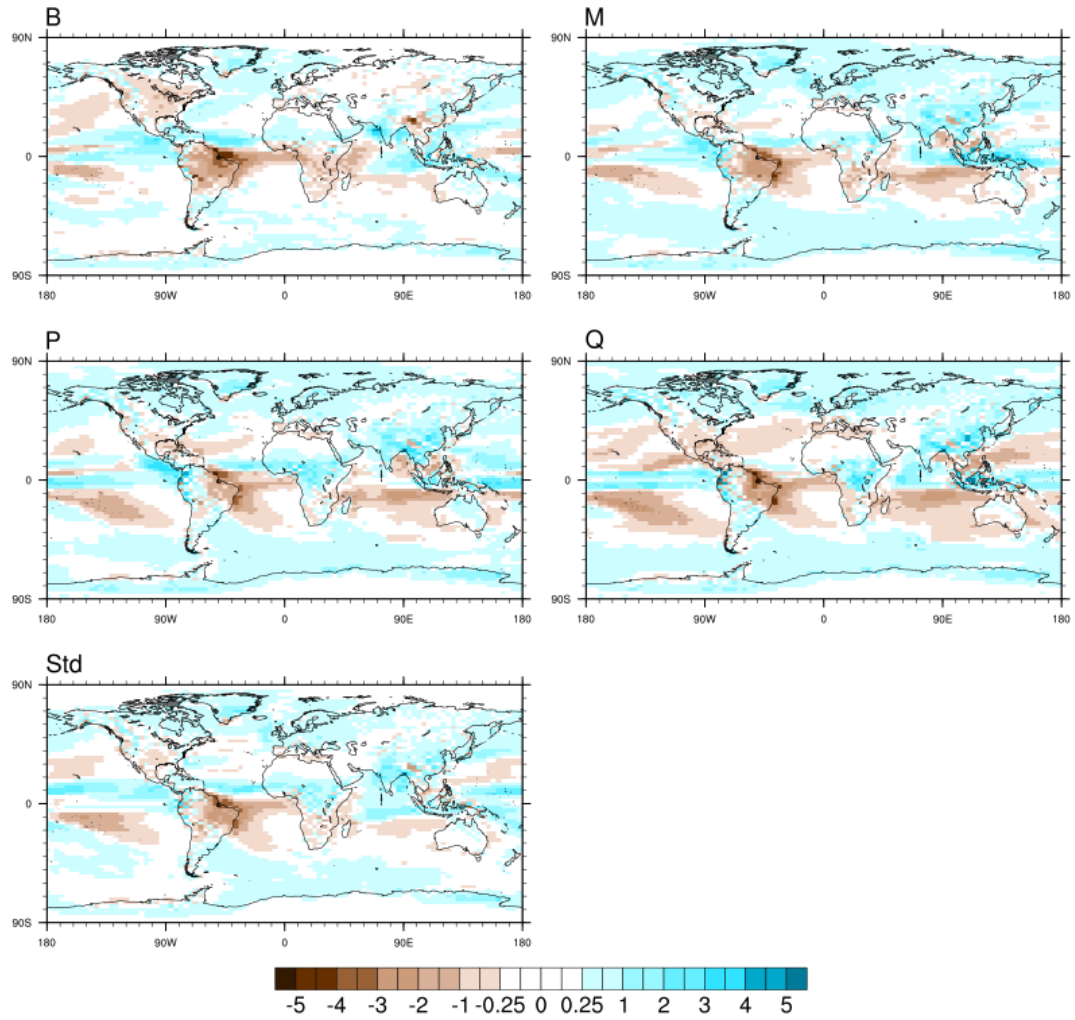


Figure 4.2. Global annual mean plots for *"Pliocene ensemble member minus modern Standard"* for precipitation (mm/day) for ensemble members B, M, P, Q & the Standard. The remaining ensemble members are displayed in Appendix A3.

Overall, the Pliocene is simulated as a warmer and wetter climate than the present day, which is in keeping with the palaeo-environmental reconstructions produced from proxy data (Dowsett et al., 2010a; Salzmann et al., 2008).

4.2.2. Pliocene Ensemble Member Minus Modern Standard

The climatic response within the PPE displays the influence of perturbing parameters on the simulated Pliocene climate. By utilising the modern Standard, it is possible to observe the impact of changing the parameters on the Pliocene simulations. The response will include interactions between boundary condition changes from Pliocene and modern. These responses are limited across the Pliocene ice sheets which displays a consistent response across the ensemble members. Perturbed parameters do interact with the changes in vegetation and CO_2 to generate a range of ensemble results for the Pliocene compared to the modern Standard.

The Pliocene ensemble warming ranges from 1.87 to 5.27°C (Table 4.1). The 'Pliocene minus modern' SAT warming range for the PPE is greater than the range displayed by the PlioMIP MME (1.84 to 3.60°C – Haywood et al., 2013a). Table 4.1 highlights two distinct temperature groupings amongst the ensemble, 'colder than Standard' members (B, D, F, H & J - Appendix A1) and 'warmer than Standard' members (I, K, L, M, N, O, P & Q –Appendix A1). Across the ensemble, all members display enhanced high latitude warming compared to the tropics (Figure 4.1 & Appendix A1). However, the 'warmer than Standard' members display a greater degree of high latitude warming than both the Standard and 'colder than Standard' members. The warming is primarily across the Arctic and is most pronounced (7 to 10°C) in members L, M, N, P & Q.

	PRISM3D			300 ppm			350 ppm		
Ensemble Member	**Polar SAT**	**Global SAT**	**Polar Amplification**	**Polar SAT**	**Global SAT**	**Polar Amplification**	**Polar SAT**	**Global SAT**	**Polar Amplification**
B	4.37	1.87	2.33	2.55	0.58	4.40	3.75	1.16	3.23
D	5.58	2.10	2.66	--	--	--	--	--	--
F	6.48	2.65	2.45	--	--	--	--	--	--
H	6.95	2.62	2.65	--	--	--	--	--	--
I	9.76	4.06	2.40	--	--	--	--	--	--
J	5.85	2.30	2.54	--	--	--	--	--	--
K	7.90	3.68	2.15	--	--	--	--	--	--
L	10.86	4.05	2.68	--	--	--	--	--	--
M	10.31	4.02	2.56	6.03	2.05	2.94	8.19	2.95	2.78
N	10.18	3.84	2.65	--	--	--	--	--	--
O	9.08	3.84	2.36	--	--	--	--	--	--
P	11.82	4.46	2.65	7.45	2.54	2.93	9.44	3.33	2.83
Q	12.98	5.27	2.46	8.72	3.16	2.76	10.60	4.04	2.62
Standard	7.31	2.94	2.49	4.38	1.43	3.06	5.91	2.10	2.81

Table 4.2. Global mean annual surface air temperatures (SAT - °C), polar SATs (°C) and polar amplification ratios (Polar SAT/Global SAT) for the full perturbed physics ensemble run with atmospheric CO_2 at 400 ppmv and the members of both the 300 ppmv and 350 ppmv CO_2 sub-ensemble members (B, M, P, Q & the Standard)

The pattern of polar amplification displayed across the ensemble is important in the context of previous mPWP DMCs, which have highlighted a lack of high latitude warming in models (Chapter 1, Section 1.3.4). Across the PPE, polar amplification ratios range from 2.15 to 2.68 (Table 4.2), with ensemble members D, H, L, N & P has the highest ratios (2.65 to 2.68). Ensemble member K has the lowest ratio (2.15) and the Standard a ratio of 2.49. Within the CMIP5 MME, polar amplification ranges from 1.45 to 2.67 (IPCC, 2013). The polar amplification in the PPE falls within the upper estimates of the CMIP5 ensemble, and all but members B, K & O display a ratio greater than the CMIP5 multi-model mean (2.24 to 2.36).

Parameterisations within the ensemble members will be driving this increased ratio of polar to global warming. Analysis of the sea ice parameterisations (Tables 2.1a,b, Chapter 2 & Howell et al., 2014) indicates that they are designed to make the melting of sea ice harder than in the HadCM3 Standard. If sea ice is harder to melt in the parameter sets created, it would indicate that these parameter sets are designed to make polar amplification less likely. The sea ice parameterisations act as a brake on the simulations, constraining the polar amplification to the top end of the CMIP5 range. Without the analysis of individual parameter perturbations it is not possible to ascertain the driving parameters, but it is likely to be parameters affecting high latitude clouds that will be driving the increased polar amplification ratios.

There is a strong terrestrial warming signal (5 to 7°C) across North America and Western Europe in the 'colder than Standard' members, with other terrestrial warming ranging from 0 to 2.5°C. Terrestrial warming in the 'warmer than Standard' members is more consistent ranging from 4 to 6°C across most continental regions with high northern latitudes in excess of 8°C. Terrestrial warming is weakest across Southern Siberia, China, India and tropical Africa in all ensemble members (between 0.5 to 2.5°C less than continental USA or Western Europe).

Global mean annual precipitation ranges from -0.099 to 0.216 mm/day (Table 4.1) with all but two members (J & K) displaying an increase in precipitation compared to the modern Standard. Across the ensemble, higher latitudes tend to display increased precipitation compared to the modern Standard (1.5 to 3 mm/day), with the tropics displaying decreased precipitation (0.25 to 2 mm/day - Figure 4.2 & Appendix A3). Precipitation is reduced across Amazonia (0.5 to 2 mm/day). The smallest decrease in precipitation is observed in members I, L & P and the greatest in B, D, J & K. An increase in precipitation is observed across southern South America (1 to 2 mm/day) in most ensemble members. Members J & K display no increase in precipitation over this region. However, members M, N & O show the largest area of increased southern South American precipitation. However, the magnitude is consistent with the rest of the ensemble members displaying a change.

Decreased precipitation (0.5 to 1.5 mm/day) is observed in members B, D, F, J, K, M & O across southern Africa, with the rest of the ensemble displaying minimal change in precipitation over the region. Members H, I, L, N, P & Q display an increase in precipitation (1 to 2 mm/day) across equatorial and northern Africa. The remaining ensemble members display a thin band of equatorial increased precipitation (1 mm/day), with no expansion into northern Africa. The thin band of precipitation

represents a response to a physical boundary condition change of reduced extent of the Sahara desert in the Pliocene compared to the modern (Salzmann et al., 2008). The desert is replaced by tropical savannah and enhances precipitation over this region. East of the Rockies, North America shows decreased precipitation across the ensemble of 0.25 to 2 mm/day. The decrease is strongest (2 mm/day) in the 'cooler than Standard' ensemble members and decreases to minimal decrease in 'warmer than Standard' members.

Overall, the ensemble members maintain the trend observed in the ensemble Standard of simulating a Pliocene climate that is warmer and wetter than the present day. Two members (J & K), simulate a warmer and marginally drier climate. The ensemble members display a higher polar amplification ratio than the CMIP5 MME and a broader Pliocene minus modern temperature range than the PlioMIP MME, particularly at the upper boundary of the SAT range. Therefore, it is possible from these ensemble members that improvements will be made to the existing data-model mismatch.

4.2.3 Large-Scale Features of the CO_2 Sub-ensembles

Two sub-ensembles were produced comprising three 'warmer than Standard' ensemble members (M, P & Q), the coolest ensemble member (B) and the ensemble Standard. The purpose of these sub-ensembles was to investigate the influence of CO_2 uncertainty on the DMCs (e.g. Pagani et al., 2010; Raymo et al., 1996; Seki et al., 2010). The sub-ensembles were run at two additional CO_2 concentrations, 300 ppmv and 350 ppmv (the full ensemble was run at 400 ppmv) and compared to the modern Standard.

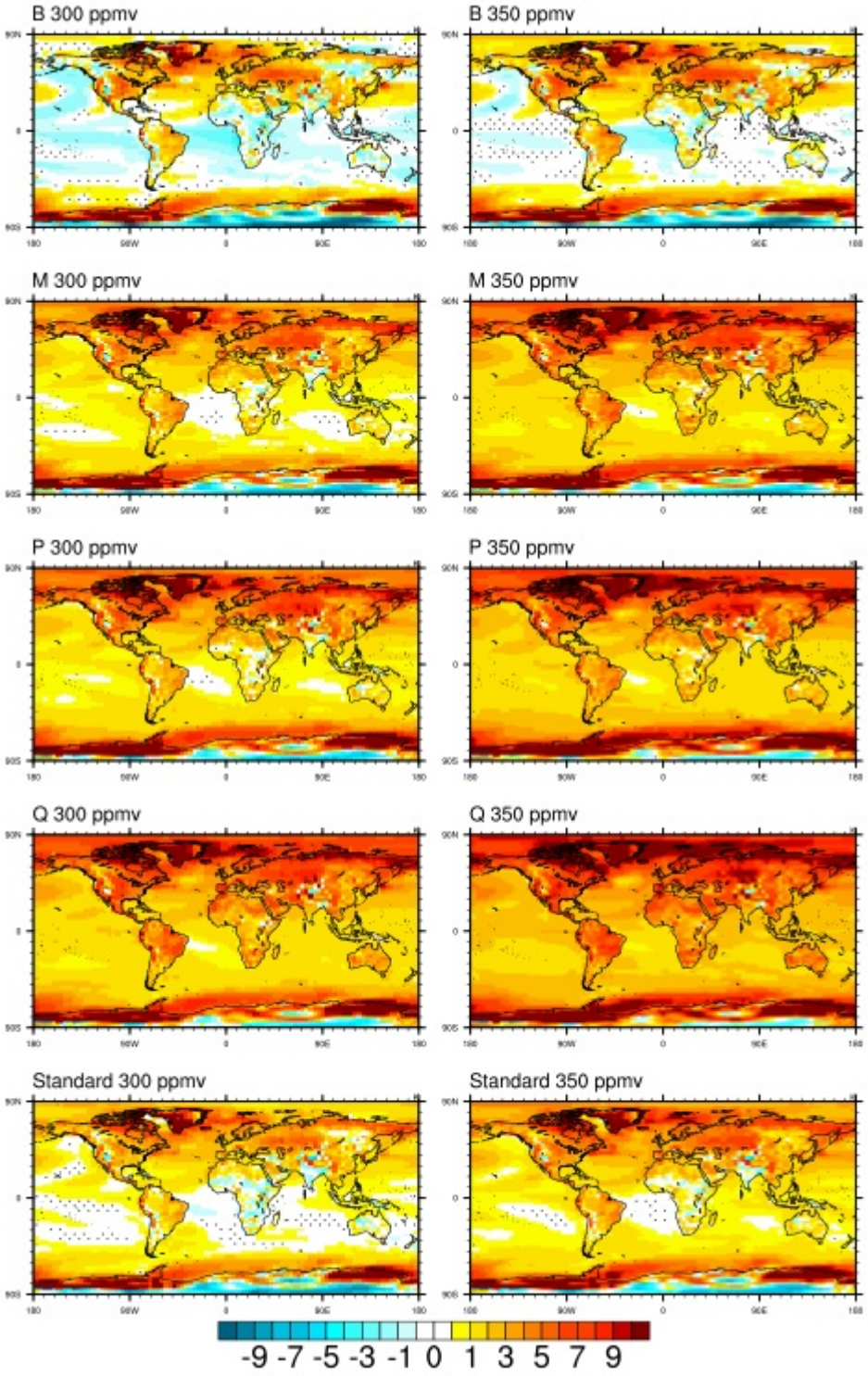


Figure 4.3. Global annual mean plots for *"Pliocene CO_2 sub-ensemble member minus modern Standard"* for surface air temperature (SAT - °C) for the 300 ppmv and 350 ppmv CO2 sub-ensembles which comprise ensemble members B, M, P, Q & the Standard.

As expected, lower concentrations of atmospheric CO_2 result in reduced global mean annual temperature anomalies between Pliocene and modern compared to the equivalent full ensemble members (Table 4.1). Despite the reduced CO_2, the temperature responses between mPWP and modern physical boundary conditions (Section 4.2.1) still exert warming on the mPWP climate in all sub-ensemble members,

consistent with previous mPWP sensitivity studies (i.e. Hill et al., 2011). The temperature responses due to changes in physical boundary conditions are consistent with respect to the magnitude and spatial distribution between members of both sub-ensembles and the full ensemble, for example over the GrIS and EAIS (Figure 4.3).

Areas of warming linked to the perturbed parameters in the full ensemble are apparent in the sub-ensembles. However, the magnitude of warming from these features is reduced. Members M, P & Q in both the 300 and 350 ppmv sub-ensembles display a polar amplification trend. Polar amplification is observed in the full ensemble with a range of 2.14 to 2.68. For the 300 ppmv sub-ensemble the ratio ranges from 2.93 to 4.40 and 2.62 to 3.23 for the 350 ppmv sub-ensemble (Table 4.2). Despite decreases in the polar temperatures in both sub-ensembles in comparison to their full ensemble equivalents, these decreases are less than the reductions in global mean temperatures. As a result, both sub-ensembles display enhanced polar amplification ratios compared to the full ensemble. The polar warming to global warming in the sub-ensemble members can be observed in Figure 4.3.

The temperature range between the coldest ensemble member (B) and the warmest member (Q) is slightly reduced in each sub-ensemble (300 ppmv range: 2.58°C, 350 ppmv range: 2.88°C) compared to the full ensemble (3.40°C - Table 4.1). Unsurprisingly increasing Charney sensitivity results in a greater sensitivity to changes in CO_2. Most of the reduced temperature range originates from the greater decrease in mean annual temperature for member Q compared to member B in both the 350 and 300 ppmv sub-ensembles. Member Q has the highest Charney sensitivity of the PPE and for member Q the SAT difference reduced from 5.27°C (full ensemble) to 3.16°C (300 ppmv sub-ensemble). In comparison member B, with the lowest Charney sensitivity has an SAT decrease of 1.87°C to 0.58°C.

In summary, the PPE and physical boundary condition features observed within the full ensemble are observed within the sub-ensembles. Boundary condition related changes, such as the distribution of ice sheets results in a similar magnitude of SAT difference between Pliocene and modern as in the full ensemble. Polar amplification ratios are increased in both sub-ensembles compared to the full ensemble. Despite decreases in polar temperature in both sub-ensembles, global mean annual temperatures decrease by a greater amount. It has been hypothesised that increased polar amplification ratios will lead to improved data-model comparisons. However, despite higher polar amplification ratios the sub ensembles have reduced polar SATs. The CO_2 sub-ensembles will provide a good assessment of this hypothesis, offering

ensemble members able to determine whether polar SATs or increased polar amplification ratios produce the strongest DMCs. Owing to the uncertainty in the greenhouse gas concentrations during the mPWP, the CO_2 sub-ensembles are also important for understanding the impact of lower greenhouse gas concentrations on future mPWP modelling studies, should the estimates be revised downwards.

4.3. Zonal Annual Mean Data-Model Comparison

As discussed in Chapter 2 (Section 2.6.3 *iii*) traditional DMCs for palaeoclimate studies of the mPWP have used a *"site-by-site"* comparison for SST and SAT reconstructions (e.g. Haywood et al., 2013a; Salzmann et al., 2013) and are presented here (Section 4.4). The model skill in comparison to the data is assessed using metrics such as the root mean square error (RMSE) based on the scale of the mismatch between the palaeo-data locality and the corresponding model grid box. However, it could be suggested that testing a climate model simulation on its ability to reproduce relatively sparse site-by-site palaeo-data is not a realistic test. An alternative DMC approach, using a zonal mean DMC is produced (Section 2.6.3, Chapter 2) to avoid comparing localised sites with large grid box areas. Both *"site-by-site"* and zonal DMCs are produced and assessing both DMC methodologies using the PPE to gain an understanding of the relative strengths and weaknesses of each approach and the impact this has on the standard methodology for mPWP data-model studies.

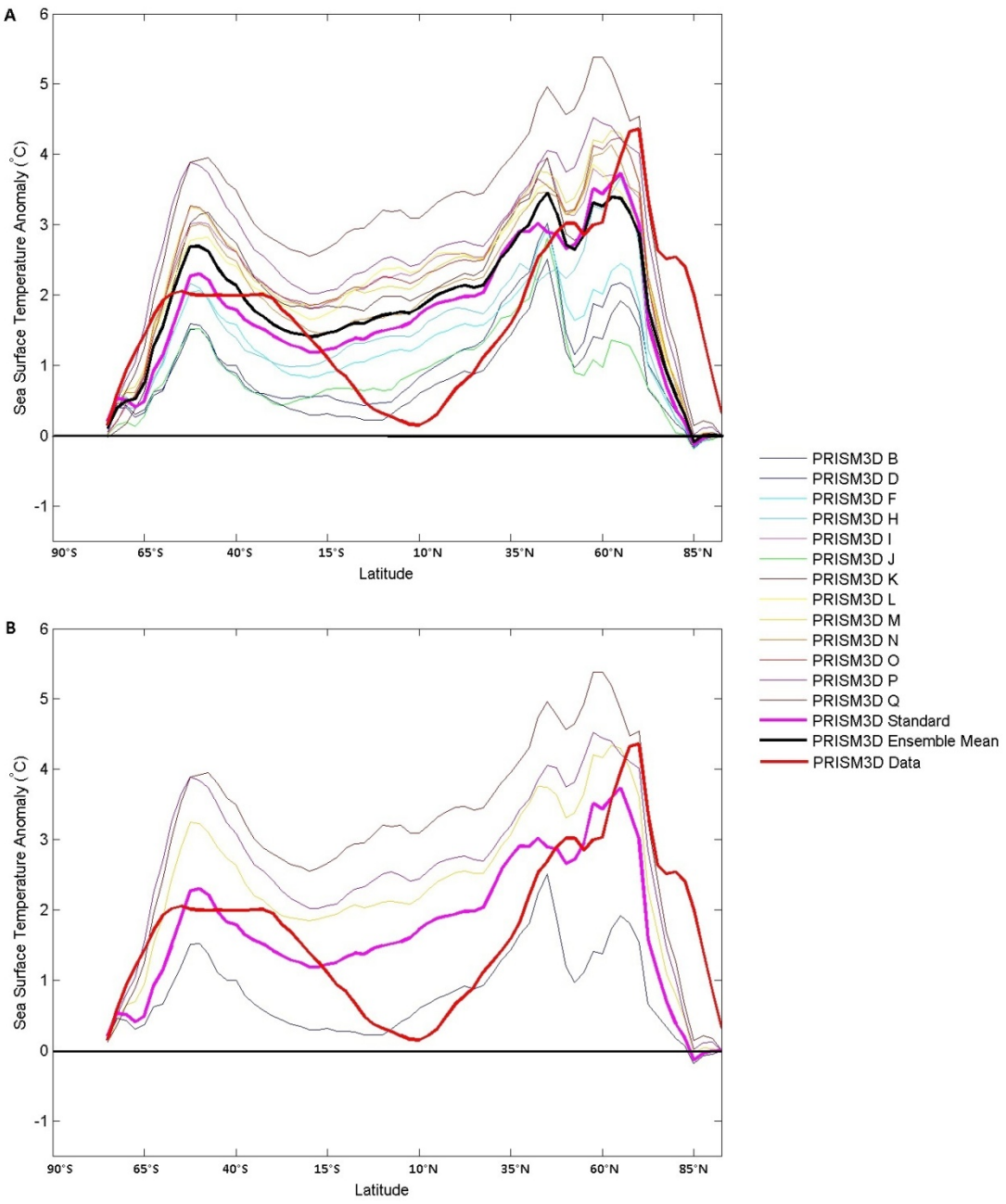


Figure 4.4. Global zonal annual mean *"Pliocene ensemble member minus modern Standard"* sea surface temperatures for A) all ensemble members and palaeo-data and B) a subset of ensemble members B, M, P, Q, the Standard and the palaeo-data.

The global zonal annual mean DMC (Figure 4.4) highlights the existing SST data-model mismatch (Chapter 1, Section 1.3.4) noted through the tropical and high northern latitudes. In the southern high latitudes, most ensemble members are warmer than the palaeo-data, a feature of warming induced by Antarctic boundary condition changes (Figure 4.1 & Appendices A1 & A2). Through the tropics no ensemble member can replicate the near zero anomaly shown by the SST data with member B (the coolest ensemble member) coming closest. Through the Northern Hemisphere, ensemble members are all warmer than the data through the tropics, whilst into the mid-

latitudes, the 'warmer than Standard' members remain warmer than the data despite a slight dip in SSTs around 45 to 50°N. The palaeo-data peaks at 70°N, whilst the ensemble members peak at 65°N. The difference in the latitude of the temperature peak leads to a significant high latitude data-model mismatch. In the 'cooler than Standard' ensemble members, the slight dip in SSTs observed at around 45 to 50°N is more pronounced resulting in a colder than data mismatch of 2 to 2.5°C.

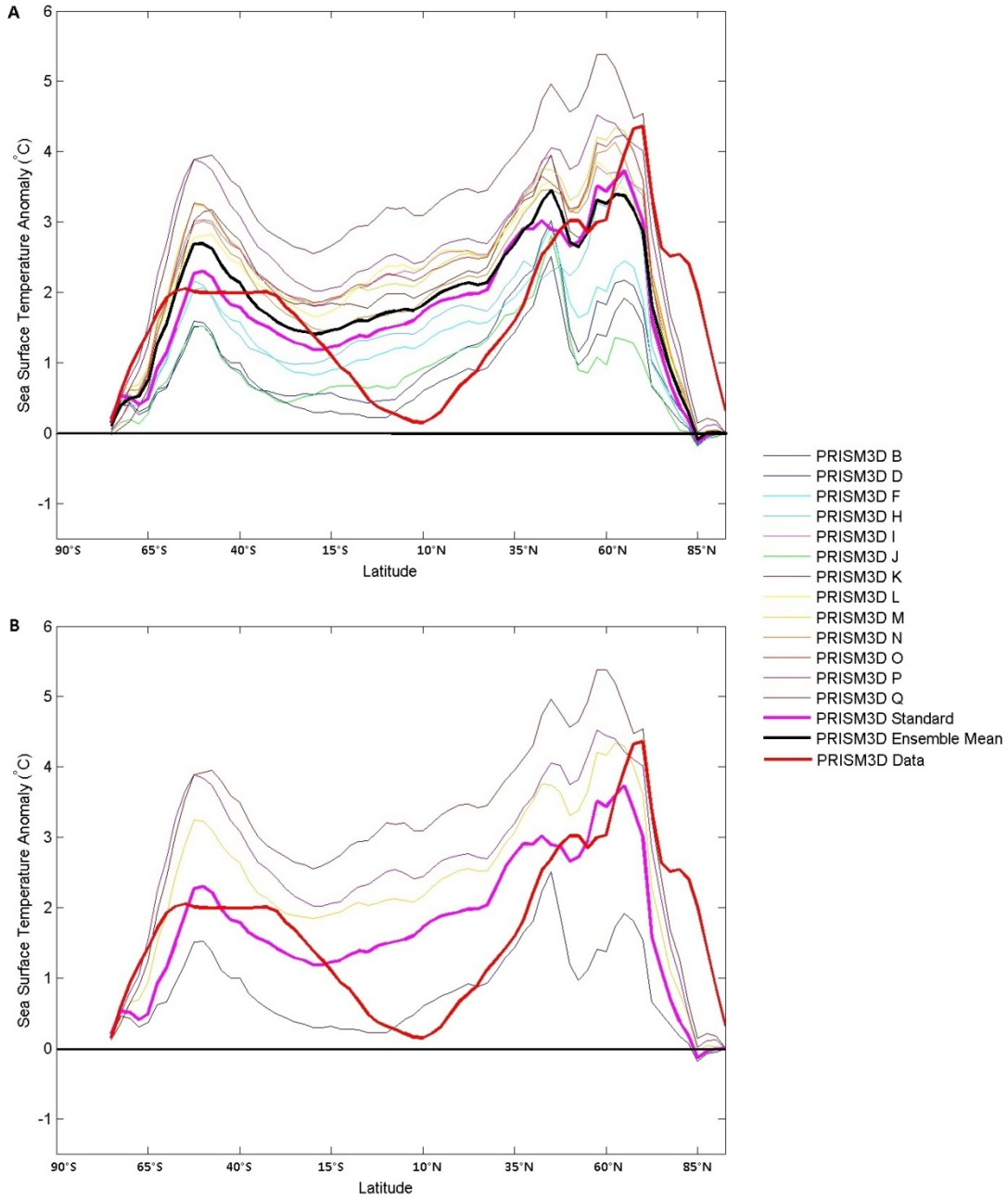


Figure 4.5. North Atlantic zonal annual mean *"Pliocene ensemble member minus modern Standard"* sea surface temperatures for A) All ensemble members and palaeo-data and B) A subset of ensemble members B, M, P, Q, the Standard and the palaeo-data.

Previous work (i.e. Dowsett et al., 2012) has highlighted the North Atlantic as a region of particular difficulty for climate models to reproduce palaeo-SST. Signals in other high latitude regions such as the North Pacific show much smaller data-model mismatches than the North Atlantic (Dowsett et al., 2012; 2013; Haywood et al., 2013a). Through the North Atlantic (Figure 4.5), the ensemble members display the same latitudinal peak as in the global zonal mean (Figure 4.4) at 45°N and 60°N. The cause of these two peaks is likely related to the representation of ocean currents and the Greenland ice sheet boundary condition change. Zhang et al. (2013), investigating Atlantic meridional overturning circulation (AMOC) in the PlioMIP MME, concluded that AMOC causes a peak temperature at 45°N in HadCM3, which is observed in the zonal mean SSTs for the ensemble members as the lower latitude (smaller magnitude) peak. The influence of warming from changes in the GrIS (Figure 4.1, Appendices A1 & A2) causes the peak in SSTs at 60°N. However, in the 'cooler than Standard' members, this peak is reduced. Through the North Atlantic the palaeo data peaks later at 80°N (Figure 4.5) compared to 70°N in the global comparison (Figure 4.4). The ensemble members SST profile through the North Atlantic does not change as significantly as the palaeo-data profile and results in an increased data-model mismatch through the North Atlantic compared to the global zonal mean comparison.

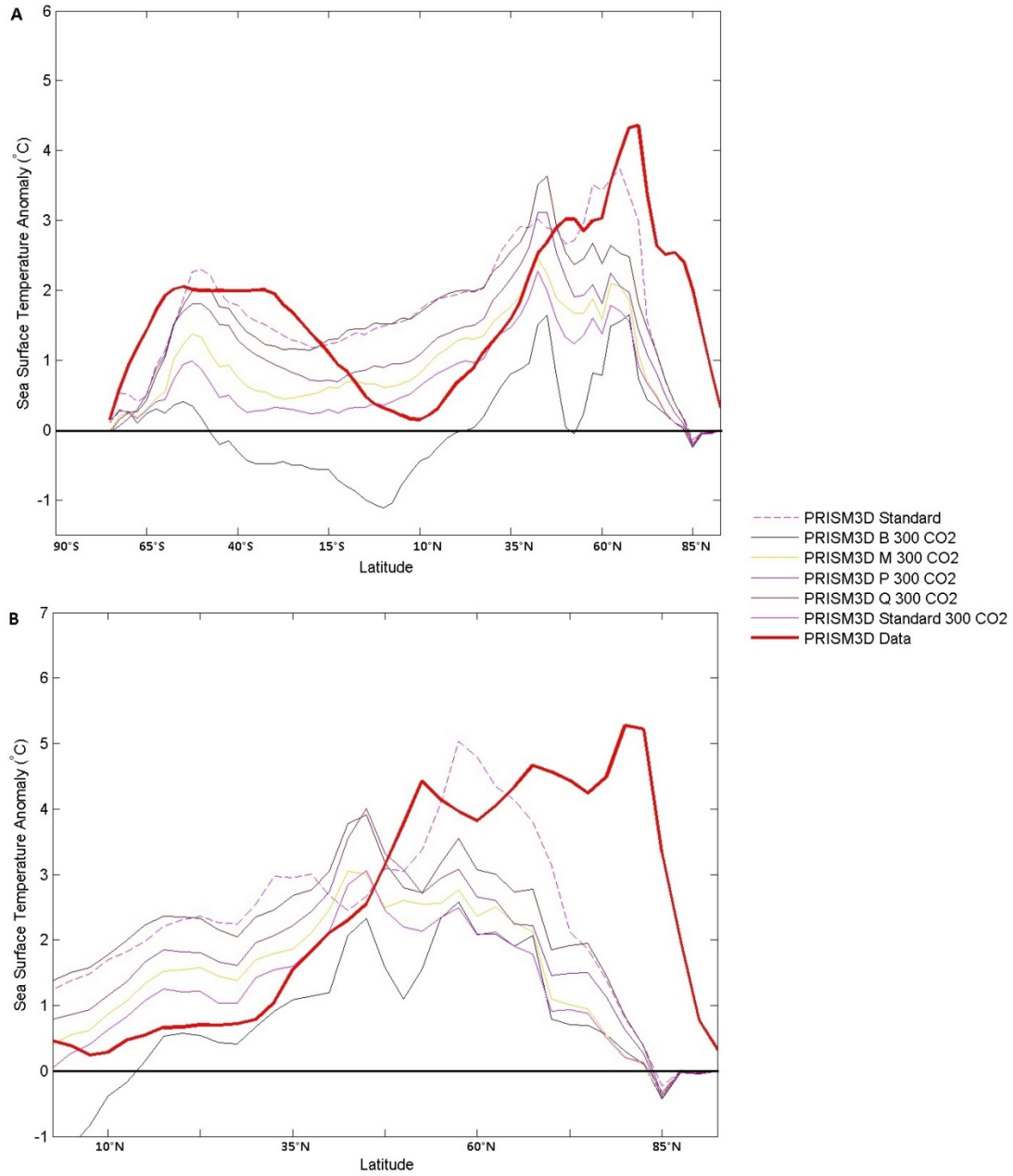


Figure 4.6. Zonal annual mean *"Pliocene ensemble member minus modern Standard"* sea surface temperatures for the 300 ppm CO_2 ensemble for A) global ocean and B) North Atlantic ocean only. The dashed line represents the full ensemble (400 ppmv) Pliocene Standard and the thick red line the palaeo-data.

The CO_2 sub-ensembles of 300 ppmv (Figure 4.6) and 350 ppmv (Figure 4.7) improve the tropical match between ensemble members and proxy data. The 300 ppmv sub-ensemble members display zonal means that are cooler than the full ensemble (400 ppmv) Standard. In both the global and North Atlantic zonal means, the sub-ensemble range brackets all the tropical palaeo-data. At high latitudes there is an increase in the mismatch between the sub-ensemble members and palaeo-data. Through the North Atlantic, the sub-ensemble members are weaker than the full ensemble Standard.

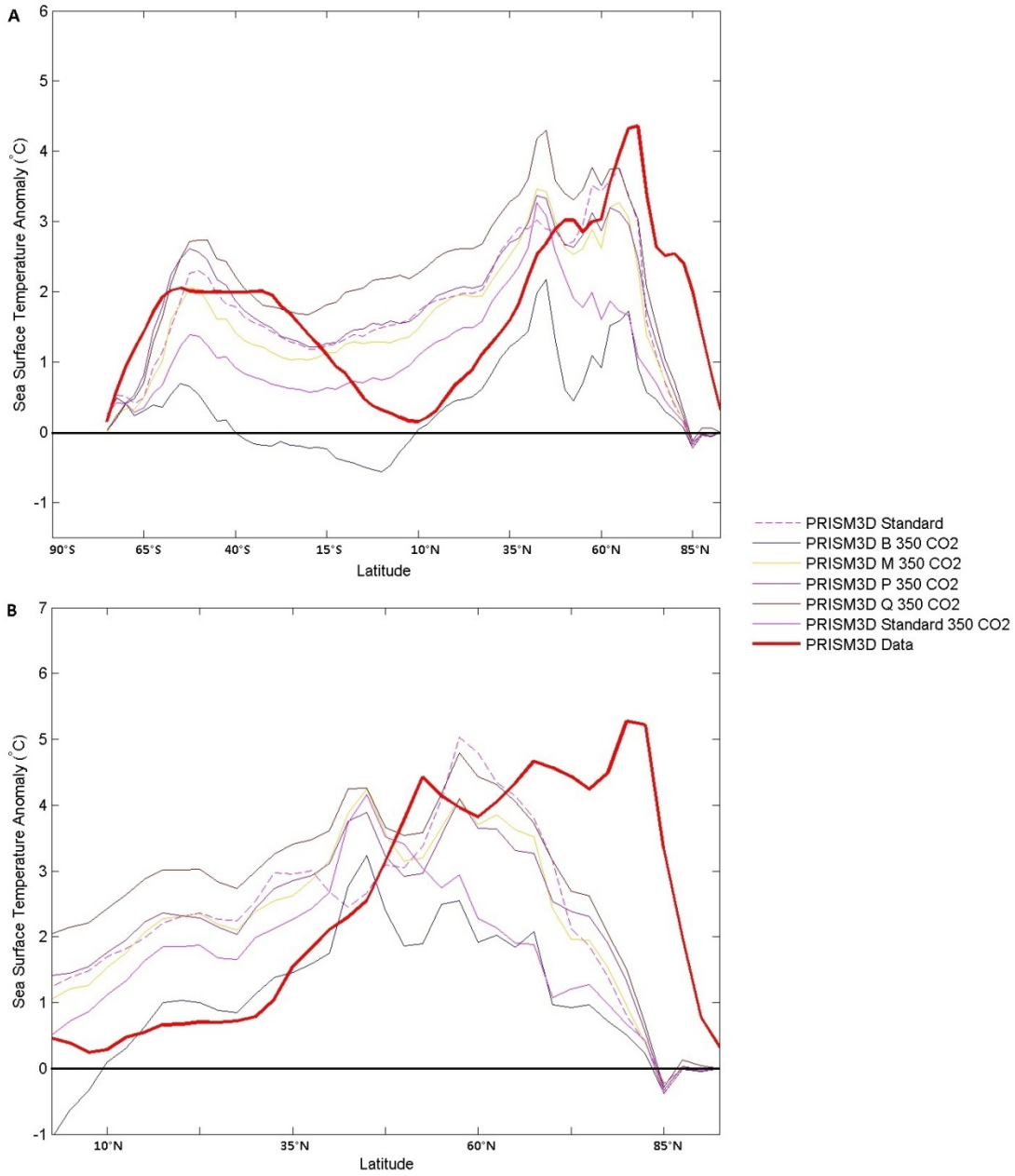


Figure 4.7. Zonal annual mean *"Pliocene ensemble member minus modern Standard"* sea surface temperatures for the 350 ppm CO_2 ensemble for A) global ocean and B) North Atlantic ocean only. The dashed line represents the full ensemble (400 ppmv) Pliocene Standard and the thick red line the palaeo-data.

For the 350 ppmv sub-ensemble, the members display zonal means that are warmer than the full ensemble Standard. However, unlike the full ensemble, only member Q of the 350 ppmv sub-ensemble has sufficient warming to resemble the palaeo-data. Up to ~60°N, the 350 ppmv sub-ensemble brackets the palaeo-data for the global DMC (Figure 4.7a) and (with the exception of a small tropical mismatch), also achieves this in the North Atlantic. At high latitudes, the performance of the 350 ppmv sub-ensemble is enhanced compared to the 300 ppmv sub-ensemble. However, performance is little better than the full ensemble Standard. The AMOC related zonal mean SST peak at

~45°N observed in the full ensemble is apparent in both sub-ensembles. However, the 60°N peak observed in the full ensemble due to the GrIS changes is less pronounced in the sub-ensemble members. Based on the zonal DMCs, the reduction in polar temperatures in the sub-ensemble members compared to their full ensemble equivalents has resulted in a weaker data-model comparison.

In summary, maximum warming across the full ensemble indicates a peak in the model SSTs to the south of the palaeo-data localities. This feature is clearest in the North Atlantic (Figure 4.5). The peak in member Q is warmer than the palaeo-data peak by ~1°C, and north of this the ensemble members underestimate SSTs compared to the palaeo-data. The PPE is able to simulate the magnitude of Pliocene warming, but is unable to simulate the spatial distribution of this warming. Both sub-ensembles achieve little additional reconciliation with the palaeo-data. If Pliocene CO_2 was determined to be below 350 ppmv (with other greenhouse gases kept at pre-industrial levels), it would be even more difficult to reconcile models and data.

4.4. Site-by-Site Data-Model Comparisons

Site-by-site DMCs represent the more traditional comparison of palaeo-data and climate models. Two site-by-site DMCs are presented here, utilising the SST and SAT palaeo-data, the RMSE scores for each DMC for all ensemble members are shown in Table 4.3. For the SST DMC, ensemble member N has the lowest RMSE score along with members L & M. Nine ensemble members (F, H, I, K, L, M, N, O & P) are an improvement on the HadCM3 Standard DMC (Table 4.3).

	Sea Surface Temperatures			Surface Air Temperatures		
Ensemble Member	**PRISM3D**	**300 ppmv**	**350 ppmv**	**PRISM3D**	**300 ppmv**	**350 ppmv**
B	2.70	3.29	3.04	5.73	6.26	5.72
D	2.67	--	--	5.33	--	--
F	2.39	--	--	5.05	--	--
H	2.41	--	--	5.04	--	--
I	2.39	--	--	4.03	--	--
J	2.54	--	--	5.41	--	--
K	2.38	--	--	4.66	--	--
L	2.27	--	--	3.95	--	--
M	2.36	2.60	2.40	3.92	4.91	4.19
N	2.25	--	--	3.75	--	--
O	2.38	--	--	4.23	--	--
P	2.39	2.41	2.42	3.70	4.38	3.91
Q	2.71	2.31	2.42	3.79	4.28	3.85
Standard	2.42	2.80	2.59	4.64	5.53	5.03

Table 4.3. Root mean square error (RMSE) scores for the site-by-site data-model comparisons to the sea surface temperature (SST) and surface air temperature (SAT) DMCs for full perturbed physics ensemble run with atmospheric CO_2 at 400 ppmv and the members of both the 300 ppmv and 350 ppmv CO_2 sub-ensemble members (B, M, P, Q & the Standard).

Figure 4.8, shows ensemble members B, N & Q and the Standard, with the rest of the ensemble members displayed in Appendix A4. Member N was the best member, with B and Q for different reasons the weakest members. Member N has the lowest RMSE score, reducing the high latitude data-model mismatch in comparison with the Standard by up to 3°C. Member Q achieves greater reductions than member N in high latitude sites compared to the Standard. However, compared to both N and the Standard, member Q increases warming mismatches through the tropics.

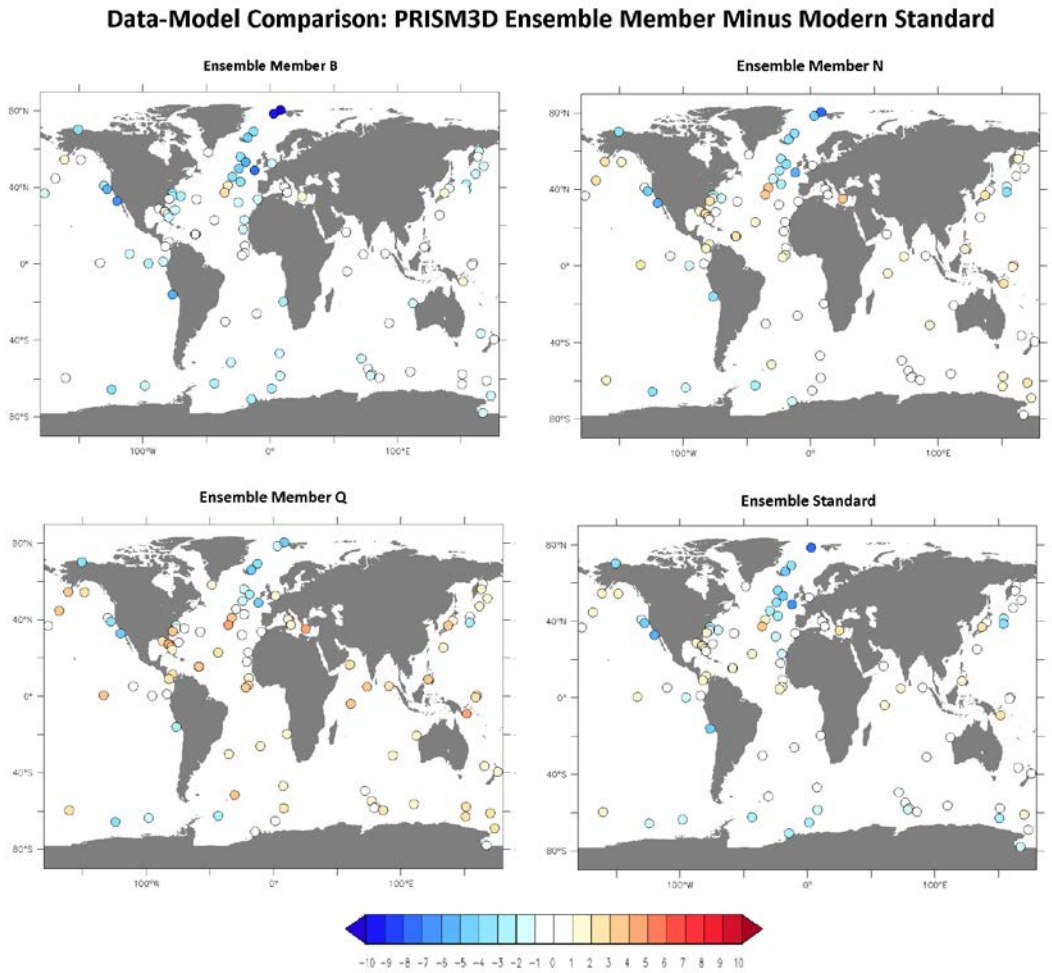


Figure 4.8. Site-by-site *"Pliocene ensemble member minus modern Standard"* sea surface temperature data-model comparisons for ensemble members B, N, Q & the Standard. The remaining ensemble members are displayed in Appendix A4.

These warming mismatches eliminate the improvements observed in the high latitudes and lead to a weakened RMSE score for Q. In comparison, member B which has a similar RMSE to member Q, produces a strong tropical DMC with many sites indicating minimal difference between the model and data. However, high latitude mismatches are in excess of 8 to 10°C. As observed in the zonal mean DMCs, ensemble members that produced strong simulations of high latitude SSTs, struggled to not simulate too much tropical warming. Members B & Q highlight that the ensemble members producing the best DMC in one region (i.e. tropics or high latitudes) weaken the overall DMC by producing the worst comparison in another region. Member N (and other high ranking members such as L & M), produce a DMC that is not best in a region, but simply achieves a balance between warming high latitudes and maintaining tropical temperatures as close to the modern as possible.

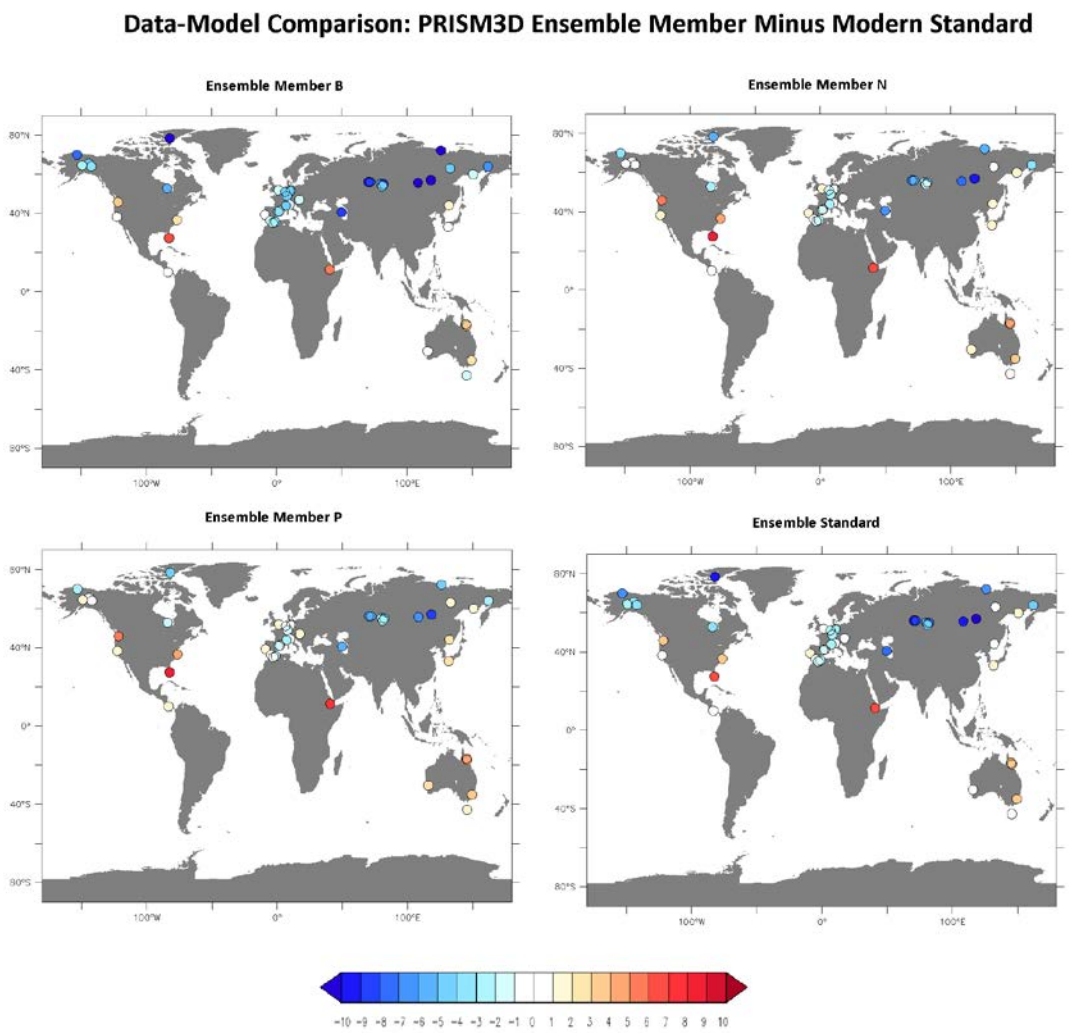


Figure 4.9. Site-by-site *"Pliocene ensemble member minus modern Standard"* surface air temperature data-model comparisons for ensemble members B, N, P & the Standard. The remaining ensemble members are displayed in Appendix A5.

For the SATs, member P produces the strongest RMSE with members N & Q. Member B ranks lowest whilst seven ensemble members (I, L, M, N, O, P & Q) outperform the Standard (Table 4.3). Figure 4.9 (and Appendix A5) highlights the differences from these rankings across the sites. Member B (the weakest performing member), P (the strongest), N and the Standard are shown to highlight the performance across the ensemble. Member P reduces the data-model mismatches throughout Western Europe and Alaska compared to the Standard, removing the mismatch completely in some regions whereas the Standard indicates a 2 to 5°C mismatch. Small improvements are also made throughout Siberia and eastern Eurasia. Member P displays a similar performance to the Standard across other sites in the USA, Africa and Australia. Member N shows a similar performance to member P, however it displays reduced improvement (by 0.5 to 1°C) across the Western Europe sites compared to member P. As a result it performs well for this DMC, but not as strongly as member P. Member B produces the best representation of sites in the USA displaying minimal mismatch in

these sites. However, in comparison to the Standard mismatches across Alaska, Siberia and Eurasia are increased by 3 to 4°C, with a smaller increase (~1°C) in Western Europe mismatches. These mismatches combine to produce a much weaker RMSE score than the Standard.

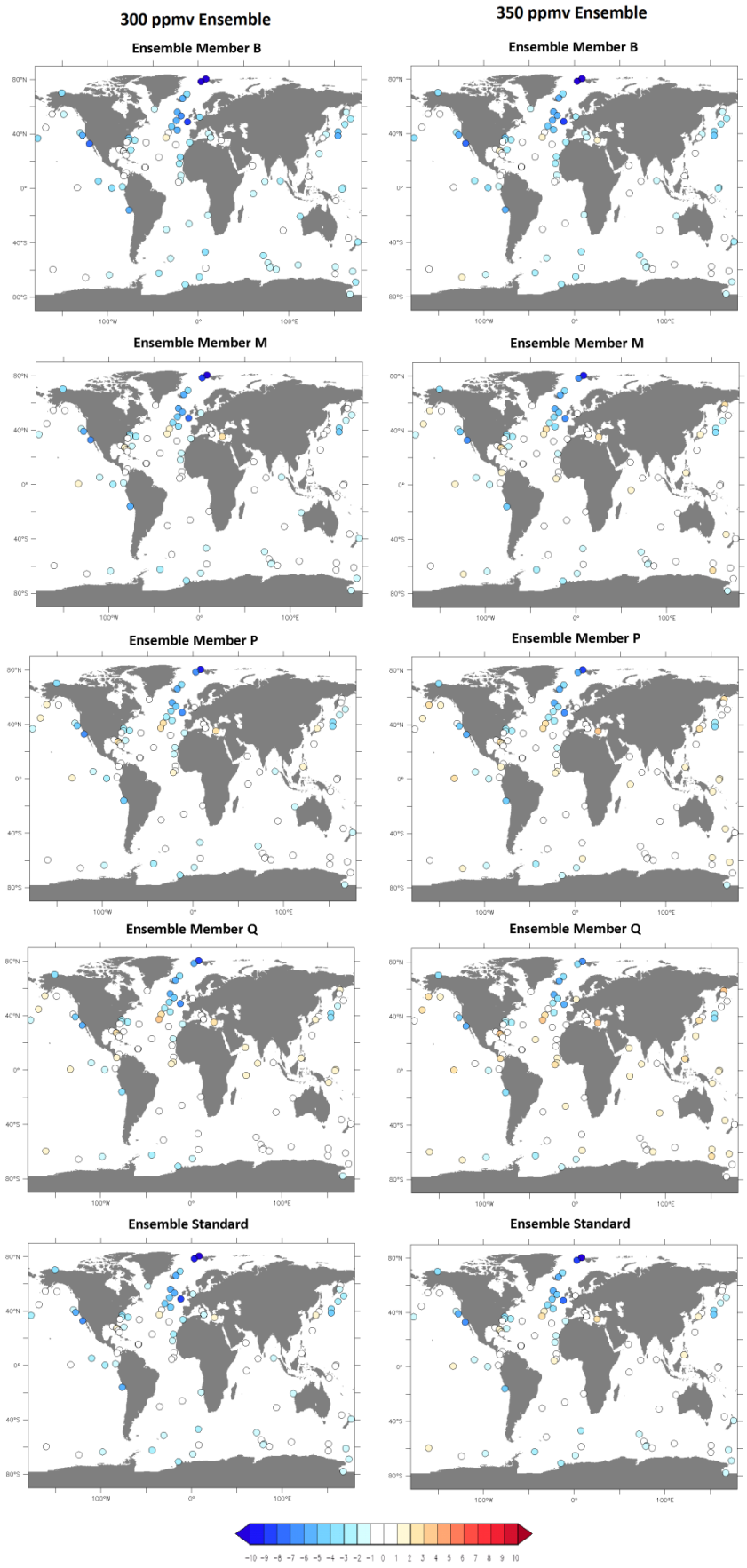


Figure 4.10. Site-by-site *"Pliocene ensemble member minus modern Standard"* sea surface temperature data-model comparisons for both CO_2 sub-ensembles.

Both CO_2 sub-ensembles struggle to improve the zonal DMC to the SST data at high latitudes, despite the elevated polar amplification ratios across all sub-ensemble members compared to the full ensemble. A similar result is returned for the site-by-site DMCs, with the RMSE scores for both the 300 ppmv and 350 ppmv sub-ensembles generally performing worse than their full ensemble equivalents (Table 4.3). The only exceptions are the SST DMC for member Q in both sub-ensembles which show an improved RMSE score and the SAT for member B in the 350 ppmv sub-ensemble which performs the same as its 400 ppmv equivalent (Table 4.3). For the SSTs, the improvement in member Q in both sub-ensembles is based around the representation of tropical sites. Member Q from the full ensemble achieves the greatest reduction in the high North Atlantic sites, but at the cost of the weakening the performance of the model in comparison to the tropical data sites (Figure 4.8). Member Q from both sub-ensembles does not achieve the high latitude improvements of its full ensemble equivalent, but simply improves the data-model mismatch through the tropics (Figure 4.10). In the SATs, slight improvements in the 350 ppmv member B in tropical sites compared to its 400 ppmv equivalent are matched by a reduced performance in Western Europe resulting in the same RMSE score for both ensemble members (Figure 4.11). In summary both the site-by-site and zonal DMCs indicate that despite the increased polar amplification ratios, both sub-ensembles do not improve data-model comparisons, because they are unable to simulate the required polar temperatures. Any improvements observed in these DMCs are more a reflection of improved tropical comparisons rather than the effect of increased polar amplification ratios.

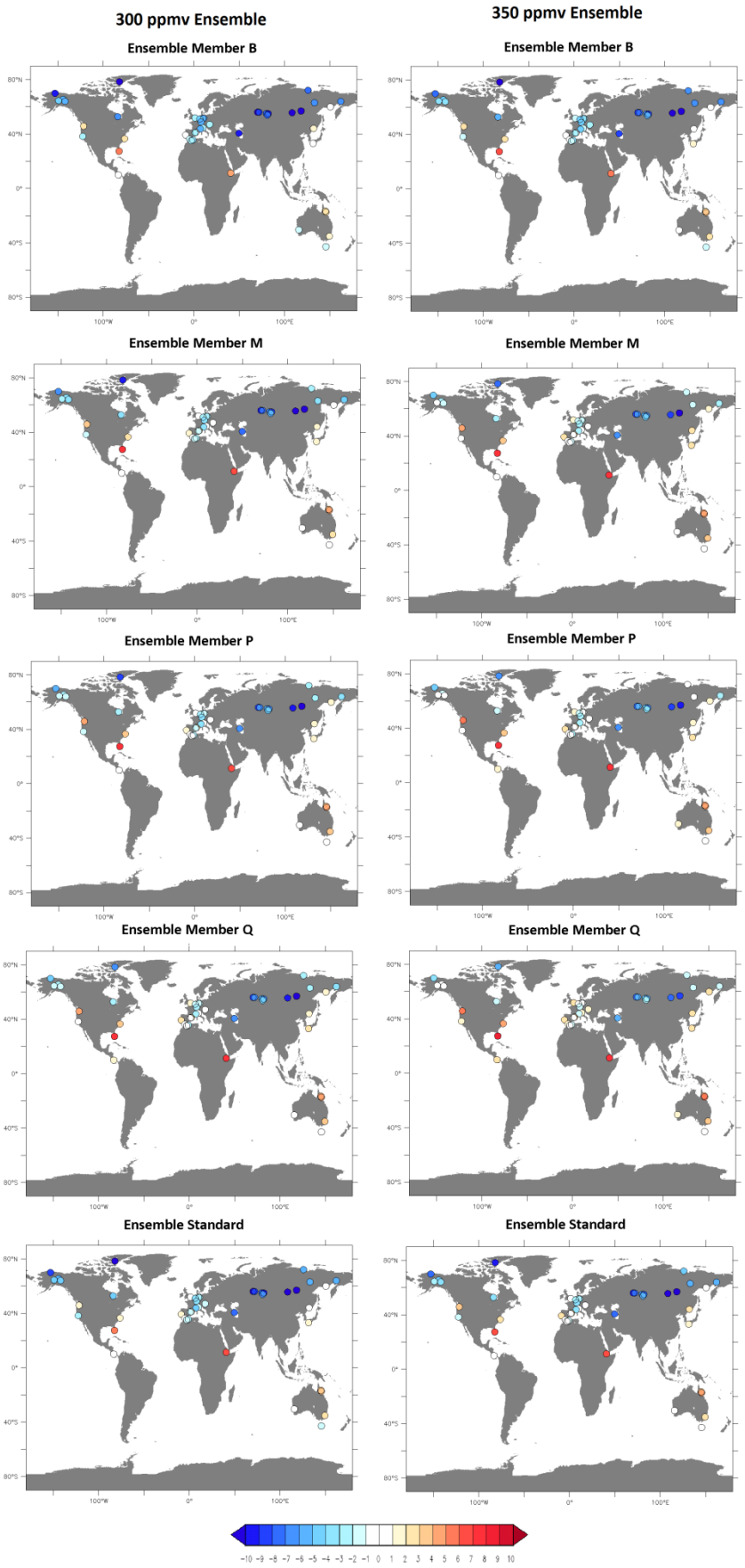


Figure 4.11. Site-by-site *"Pliocene ensemble member minus modern Standard"* surface air temperature data-model comparisons for both CO_2 sub-ensembles.

4.5. Discussion

4.5.1 Discrimination of the Model Ensemble Using Geological Proxy Data

The zonal annual mean palaeo-data both globally and for the North Atlantic (Figures 4.4 & 4.5), and site-by-site DMCs to both SSTs and SATs (Figures 4.8 & 4.9), indicates a high latitude temperature pattern that is irreconcilable with the Standard version of HadCM3, a well established data-model mismatch in mPWP studies. However, nine PPE members for the SSTs and seven members for the SATs rank higher than the Standard in these DMCs, improving the data-model comparison. Generally, these members are from the 'warmer than Standard' members which have higher than Standard Charney sensitivities. These members are able simulate warmer high latitude temperatures and therefore are able to better replicate the magnitude of high latitude mPWP warming in the SSTs (Figures 4.4, 4.5 & 4.8) and SATs (Figure 4.9).

However, there is a limit to viability of generating simulations with greater high latitude warming. The largest reduction in the high latitude SST data-model mismatch occurs in member Q (Figure 4.8), yet the SST RMSE score for member Q is the lowest within the ensemble (Table 4.3). Member B, which has a similar SST RMSE score as member Q produces a strong DMC at the tropical latitudes, but is unable to simulate the requisite high latitude temperatures. These trends can be observed in the zonal mean DMCs (Figure 4.4 & 4.5) as well as the site-by-site comparison (Figure 4.8). Both highlight that ensemble members reducing high latitude data-model mismatches generate a tropical data-model mismatch where the model is warmer than the data. For the SST DMC, it is the ensemble members that produce the best balance between high latitude warming with minimal tropical warming that produce the strongest performance compared to the data. For the SSTs, member N produces the best fit. Similarly through the site-by-site SAT DMC, ensemble member P outperforms member Q through achieving the balance of reducing high latitude mismatches without incurring too much tropical warming.

The CO_2 sub-ensembles do not contribute stronger performing members to this analysis. Whilst they display enhanced polar amplification ratios compared to the full ensemble, they lack the high latitude warming to improve the data-model comparison. Similar to member B in the full ensemble, they produce strong representations of the tropical mPWP palaeo-data, but very poor comparisons to the high latitude data (Figures 4.6, 4.7 & 4.10). Table 4.3 highlights that it is the combination of polar amplification ratio and polar warming that ensures that an ensemble member performs well in comparison with the SST data. Members M (2.56), N (2.65) and P (2.65) all have

high polar amplification ratios and also have high polar temperatures, indicating it requires both elements to ensure a strong DMC. Based on the RMSE scores, it can be determined that members M (Charney sensitivity = 4.54°C), N (4.62°C) and P (5.40°C) run with 400 ppmv atmospheric CO_2 produces the strongest combined DMC for both SSTs and SATs (Table 4.3).

Overall, the full ensemble zonal means indicate that the temperature range across the model versions produced by the PPE replicates the palaeo-data with the exception of a region in the Northern Hemisphere tropics and the high northern latitudes. The results suggest that the PPE is able to replicate the magnitude of the high latitude warming in the palaeo-data, but can not reproduce the spatial distribution of the warming. The weakness in the distribution is specifically linked to the exact location of peak Northern Hemisphere zonal SSTs, with the ensemble members unable to generate it far enough north to match the palaeo-data. The site-by-site DMCs highlight the data sites where the PPE is still unable to achieve the required warming primarily through the North Atlantic. Some of these sites are at latitudes where there zonal DMCs indicate that the model can simulate the magnitude of the palaeo-data. The disparity that arises highlights the importance of the specific regional location of the model warming with respect to the palaeo-data, and the methodology utilised for comparing model and data.

Salzmann et al. (2013) highlighted the impact of taking into account different types of uncertainty (associated with both the model ensemble used and the proxy-derived SAT data) when performing DMCs. Using one approach, they considered which model within the PlioMIP ensemble produced the minimal data-model mismatch for each data point, in order to determine whether the PlioMIP ensemble enveloped the range of proxy-derived SAT estimates or whether a data-model discrepancy remained for each site. Such an analysis has been repeated here, with data uncertainty included based on estimates from Dowsett et al. (2013 - SSTs) and Salzmann et al. (2013 - SATs). The analysis creates an *"idealised DMC"* as it is not a DMC composed of a single model or multi-model mean, but selects the ensemble member giving the best (most ideal) DMC result at each individual site, which are then compiled into a single data-model comparison.

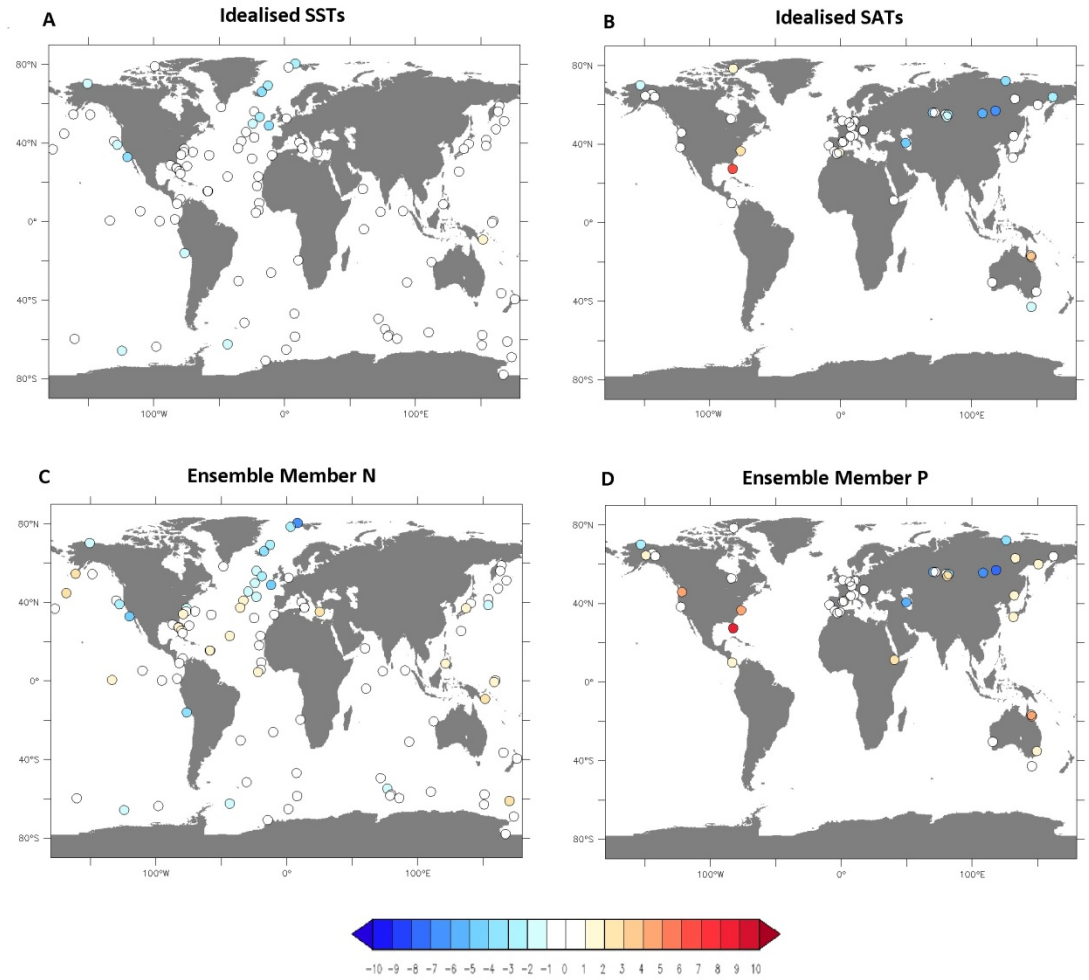


Figure 4.12. Distribution of the site-by-site data-model mismatches for the *"PRISM3D ensemble member minus modern Standard"* selecting across the ensemble members to choose the best (smallest) data-model mismatch for each site A) idealised sea surface temperatures including data uncertainty and B) idealised surface air temperatures including data uncertainty. The strongest performing individual ensemble members (based on an RMSE ranking) for C) SSTs (member N) and D) SATs (member P) are also displayed as a comparison to the idealised DMCs.

In the idealised SST DMC, data-model mismatches less than 1°C (as per Salzmann et al., 2013) are removed for all sites in the SST dataset leaving only twelve sites (Figure 4.12a). Of the sites that have mismatches greater than 1°C, three are in upwelling zones, two in coastal regions and two are in regions of strong oceanic currents and gradients in temperature. Five remaining sites are in a region where models are known to be poor at reproducing the data, the North Atlantic. The site-by-site RMSE scores indicated that ensemble member displaying strong simulation of one regional climate were weaker in another region, a conclusion supported by the zonal mean DMCs. The zonal mean DMCs highlighted that the ensemble members were unable to replicate the spatial distribution of the Pliocene warming, either failing to match the warming at the tropics or the high northern latitudes. The idealised DMC is able to avoid this problem

through sampling the smallest mismatches from across the ensemble members at both the high latitudes and the tropics. The advantages of the idealised DMC (Figure 4.12a) are highlighted when they are compared to the strongest performing ensemble member, N (Figure 4.12c). Member N performs comparably with the idealised DMC through the North Atlantic, but is unable to resolve the tropical temperatures, with a number of sites displaying a warming mismatch of greater than 1°C. Member N has an RMSE score of 2.25 (Table 4.3), but the idealised DMC reduces this to an RMSE of 0.88. The ability of the idealised DMC to sample the strongest performing ensemble member at each data locality and include the range of palaeo-data uncertainty makes a massive improvement to the DMC score.

Figure 4.12b displays the idealised SAT DMC which indicates that the main data-model mismatch remains through the sites in southern Siberia, with up to 10°C mismatch remaining at some sites. However, in comparison to the strongest ensemble member, P (Figure 4.12d), the idealised DMC sees the mismatch at sites through Western Europe and across North America reduced to less than 1°C, whereas for member P there is a slight (1 to 2°C) mismatch. However, over Siberian sites the performance of member P is similar to the idealised DMC. For both idealised DMCs, large discrepancies (>5°C) remain between the model and data showing that the perturbed parameters within this ensemble are not able to fully reproduce the palaeo-data. The idealised SAT DMC has an RMSE score of 2.14 compared to 3.70 for member P (Table 4.3).

The combination of zonal mean, site-by-site and idealised DMCs highlights three key elements of the PPE through data-model comparison:

1. The PPE is able to simulate the magnitude of required mPWP high latitude warming but without the appropriate spatial distribution.
2. The strongest performing ensemble members achieve a balance between increasing high latitude warming and minimising changes at the tropics.
3. The existing data-model mismatch is irreconcilable with the HadCM3 Standard and this PPE even under the most favourable conditions.

These findings are in common with analysis of the PlioMIP MME (Dowsett et al., 2013; Haywood et al., 2013a; Salzmann et al., 2013)

4.5.2 Are There Implications for Model Estimates of Charney Sensitivity?

The ensemble represents a range of potential versions of HadCM3, with a range of Charney sensitivities (2.4 to 7.1°C, Standard = 3.3°C). A key consideration of these ensembles is whether changes in the model Charney sensitivity leads to a better

representation of the mPWP climate?

Sections 4.3 & 4.4 discussed the performance of the ensemble members in comparison with the palaeo-data. The PPE was able to simulate mPWP climates that reduced the data-model mismatch at high latitudes in the site-by-site DMCs. Meanwhile zonal DMCs suggested that members of the PPE are able to simulate the required magnitude of high latitude warming observed in the proxy derived SSTs. However, the ensemble members were unable to replicate the spatial distribution of the Pliocene warming. Specifically ensemble members that could simulate appropriate high latitude warming were unable to simulate the minimal warming through the tropics. As a result, the strongest performing ensemble members in the RMSE scores from the site-by-site DMCs were ensemble members that achieved a balance between high latitude warming and minimal tropical warming. The balance of warming in the high latitudes was observed in ensemble members from the 'warmer than Standard' grouping, a trend observed in both the zonal mean and site-by-site DMCs. Section 4.4 showed that overall the highest ranking ensemble members for the combination of DMCs were members M, N & P, with Charney sensitivities of 4.54°C, 4.62°C & 5.40°C respectively.

The IPCC defines a likely range with high confidence for Charney sensitivity of 1.5 to 4.5°C (IPCC, 2013), therefore the strongest performing Pliocene PPE members are just above the likely top end estimate for Charney sensitivity. These results suggest that to best simulate the warming of the mPWP with an atmospheric CO_2 concentration of 400 ppmv requires models with a higher value of Charney sensitivity. However, it is evident from the zonal means, site-by-site and idealised DMCs, that it is not just generating the magnitude of warming, but the spatial distribution of this warming that is crucial to being able to reconcile models and data for the mPWP. The DMCs indicate that the PPE can simulate the requisite warming in the Pliocene climate at high latitudes. However, the weakness of these data-model comparisons for the members simulating the required high latitude warming are that too much warmth is simulated in the tropical latitudes. Overall, it is not that the PPE is unable to simulate the warmth of the mPWP, but that it is unable to adequately simulate the Earth system and therefore the spatial distribution of mPWP warming. From the DMCs it can be confidently determined that a value for Charney sensitivity cooler than the present HadCM3 Standard is highly unlikely to produce a stronger simulation of the mPWP climate. However, it can also be determined that increasing the value for Charney sensitivity much above the upper end of the IPCC estimate will not generate an improved representation of mPWP climate. The idealised DMC along with the site-by-site and zonal DMCs displays that it is not the magnitude of the warming, but the spatial

profile of this warming that the ensemble is unable to simulate. The profile is not a feature of the Charney sensitivity but more a limitation elsewhere within the HadCM3 model, the ensemble design, experimental design or uncertainty within the palaeo-data reconstruction. These results indicate that the Pliocene PPE offers no conclusive evidence to suggest that models with Charney sensitivities in excess of the upper bound of the IPCC likely range will result in improved modelling of warmer world climates.

4.5.3 Can The Model Be Reconciled With mPWP Proxy-Data?

Based on the analysis discussed in this chapter, the main issue of comparing models to palaeo-data is that HadCM3 is unable to simulate the global climate profile as reconstructed from the palaeo-data. However, this is not solely a feature of the Charney sensitivity of the model and resulting warming response. The zonal means indicate that the magnitude of the warming can be reproduced in ensemble members with higher Charney sensitivities. However, the spatial distribution of this warming is not compatible with the palaeo-data (Figures 4.4 & 4.5). These results would indicate that the solution to resolving data-model mismatches for the mPWP lie not with the sensitivity of the model, but within the realm of the model design, the design of this ensemble and uncertainty from orbital forcing boundary conditions and the palaeo-data. Potential factors include:

- The parameter space is not fully sampled
- Structural uncertainty is not included
- Poorly constrained or misrepresented boundary conditions
- Unaccounted for data uncertainty

The QUMP methodology utilised in this work produces parameter sets that sample a small number of points in a high dimensional parameter space (Williamson et al., 2013), focussing on parameterisations known to be uncertain (Collins et al., 2006; Murphy et al., 2004). However, it is possible that a more complete sampling of the parameter space using alternative sampling methods (e.g. Latin Hypercube) would generate ensemble members which could produce a better representation of the mPWP climate. The 100 member Latin Hypercube ensemble design of Gregoire et al. (2011) for FAMOUS, the low resolution version of HadCM3 used was later applied to the Early Eocene (56 to 47.8 Ma) by Sagoo et al. (2013). Across the Sagoo et al. ensemble, the spatial distribution of Eocene zonal mean SATs and SSTs displays a similar pattern despite a large offsets across the ensemble. The offset is caused by the ensemble members with higher tropical temperatures also having elevated polar temperatures

and results in equator to pole gradients being broadly similar across the ensemble members (Sagoo et al., 2013). As a result, the Eocene ensemble displays similar difficulties to the mPWP ensemble, being unable to reproduce high latitude warming without increasing tropical data-model mismatches.

Other QUMP ensembles, focussing on perturbing parameters in the ocean (Brierley et al., 2010; Collins et al., 2007), sulphur (Ackerley et al., 2009) and carbon cycles (Booth et al., 2012) have also been created. If these parameter perturbations were included in future ensembles generated to simulate the mPWP they might have the potential to improve the data-model mismatch. For example, non-flux adjusted PPEs, are able to investigate ocean currents such as AMOC such as those by Hodson et al. (2013), Vellinga & Wu (2008) and Williamson et al. (2013) might include parameter sets which improve the representation of the North Atlantic sites for the mPWP.

The data-model mismatch through the North Atlantic sites (Figure 4.8 & 4.12a) may represent an inability of HadCM3 to suitably enhance ocean heat transport. Hodson et al. (2013) identified ocean heat transport as a structural variable, which does not change between HadCM3 PPE members. Williamson et al., (2013) found systematically different behaviour in AMOC when comparing PPEs sampling an observationally constrained parameter space compared to the unconstrained parameter space. AMOC was shown to have a greater average reduction within the constrained parameter space as a result of increased CO_2 forcing. Both the work of Hodson et al. (2013) and Williamson et al. (2013) could not be represented in this PPE. For sites in the North Atlantic, it is possible that enhanced ocean heat transport and other structural variables could result in a closer approximation for some of these sites. Within the PlioMIP ensemble, the COSMOS, MIROC or NorESM simulations display enhanced warming through this region (Haywood et al., 2013a). However Zhang et al. (2013) investigated AMOC and ocean heat transport in the PlioMIP ensemble finding that there was insignificant change compared to the present day. Therefore, the cause of high latitude warming in these members is from another element of their model construction (Zhang et al., 2013). Finally with respect to the modelling, limitations in the physical representation of the boundary conditions in the model could be resulting in the inability of the model to reproduce the palaeo-data. Ivanovic et al. (2013; 2014) showed how modifying the representation of Mediterranean outflow in HadCM3 affects North Atlantic SSTs, Arctic Ocean SSTs and AMOC. Other uncertainties in boundary conditions such as vegetation, topography or ice sheets have been shown in previous work (Hill et al., 2011; Robinson et al., 2011) to exert a strong effect on the climates

simulated by the ensemble in such a way that could improve data-model mismatches through the North Atlantic sites.

Another boundary condition uncertainty arises from the time slab approach (Haywood et al., 2013b). Presently the PRISM palaeo-data is produced as a ~300 Kyr average mPWP climate. The climate is modelled using snapshot simulations with fixed physical boundary conditions and radiative forcing from greenhouse gases and orbital configuration. However, the palaeo-data records climate signals affected by changes to these forcings, therefore adjacent sites may be dominated by samples influenced by climate signals thousands of years apart. These sites represent a warm climate, which is close to the average for the mPWP using the PRISM warm peak averaging methodology (Dowsett, 2007; Dowsett & Poore, 1990). However, the timing of each sites warm peak is not necessarily coeval, and therefore the palaeo-data time slab is not necessarily physically possible within the climate system, specifically due to the influence of orbital forcing. Modelling work presently investigating the mPWP time slice (Haywood et al., 2013b) indicates that at the simulated regional climates are highly sensitive to changes to orbital forcing across a 40 Kyr period. Prescott et al. (2014), highlighted this variability using snap shot simulations 20,000 years either side of the selected mPWP interglacial time slice (3.205 Ma). Regions such as the North Atlantic displayed significant variability in the timing of their peak warming across this 40 Kyr period. In essence, the palaeo-data utilised within the DMCs are susceptible to variations in orbital forcing (e.g. Lawrence et al., 2009), but the ensemble members run with a fixed orbital configuration are not. Limitations of the boundary conditions and the temporal scale of the target for the DMCs could be hampering the ability of the ensemble to reproduce the mPWP palaeo-data reconstructions. The orbital forcings applied at the start of these ensembles may be as much a source of error as the parameters perturbed or the physical boundary conditions applied.

Finally, there is unaccounted for data uncertainty. The cause of the mismatch could be due to data uncertainty from analytical issues (Salzmann et al., 2013). Uncertainties in pre-Quaternary proxy data based temperature reconstructions are primarily caused by non-modern analogue environments, evolutionary changes of ecological tolerance and methodological problems (Prescott et al., 2014). Without a full appreciation of the uncertainty on palaeo-data reconstructions, it is not possible to reliably evaluate the model error as the reconstructed signal may not be robust or the stated uncertainty in that signal may be too small, or both. With respect to CO_2, there is confluence of two of these sources of uncertainty. CO_2 represents a forcing boundary condition for the model, but the CO_2 value is analytically calculated from proxy data. There are a range

of estimates for mPWP CO_2 from pre-industrial CO_2 at 280 ppmv (lower bound estimate of Tripati et al., 2009) to 435 ppmv (upper bound estimate of Raymo et al., 1996) with a congregation between 350 and 400 ppmv based on a range of techniques. Additionally at some sites different analytical techniques have produced different estimates (Seki et al., 2010). Therefore, the CO_2 value used in modelling the mPWP represents a boundary condition uncertainty, which itself is a product of data uncertainty. Figures 4.6 & 4.7 indicate that high latitude DMCs for the 300 and 350 ppmv sub-ensemble members are much weaker in comparison to the 400 ppmv Standard. If mPWP simulations were undertaken with atmospheric CO_2 below 350 ppmv, it would be highly unlikely that even other boundary condition factors such as orbital forcings would enable models to minimise the high latitude data-model mismatches.

Overall, uncertainty in the proxy-data are as important as uncertainty in the model and the PPE. Work underway to address some of these weaknesses may lead to improved representations of the mPWP based on data-model comparisons.

4.5.4. Which Methodology Produces the Best Comparison of Model and Data?

Through analysing the PPE presented here, zonal annual DMCs and site-by-site DMCs have been utilised. It is important to assess the role of each of these methodologies in producing these DMCs and judgements on ensemble member performance. Both DMCs have advantages and weaknesses. The site-by-site DMC reflects the influence of a large area of modelled climate upon a single data locality, which is unlikely to a realistic representation of the proxy temperature recorded. However, site-by-site DMCs allow for an easy to asses metric of model performance through the RMSE score, which has been used in this work to rank the ensemble members. The zonal DMC avoids the issues of scale around a site-by-site comparison, it is reliant on the palaeo-data being extrapolated from point localities to cover full ocean basins. This extrapolation, based on the assumption that SST patterns have not changed between the mPWP and the present day (Dowsett et al., 1999; Haywood et al., 2010) does add an element of uncertainty to the zonal mean palaeo-data. However, this assumption has been used in a number of mPWP modelling studies using the atmosphere only version of climate models (Haywood et al., 2000; 2013a). Additionally, the development of uncertainty estimates for the site-by-site data (Dowsett et al., 2012; Salzmann et al., 2013) offer a developing dimension to assessing model performance whilst allowing for data uncertainty. At present, it is not possible to provide an RMSE ranking or data uncertainty estimate on the zonal mean DMCs.

Overall, both DMCs are complimentary, the zonal mean DMCs highlighted that the

ensemble members are able to fully bracket the magnitude in the palaeo-data (Figure 4.4). Similarly, the site-by-site DMCs and RMSE scores highlighted that the strongest performing member through the high latitudes was producing a poorer simulation of mPWP climate due to its large tropical data-model mismatch evident in both types of DMC. Neither DMC methodology is perfect, however in combination they produce the best way to assess the performance of a climate model ensemble to palaeo-data. The results within this address the importance of the *"PMIP Triangle"* (Haywood et al., 2013a) that data-model mismatches are a feature of modelling, boundary condition and palaeo-data uncertainty.

4.6 Conclusions

Chapter 4 has presented data-model comparisons (DMCs) for the mid-Pliocene Warm Period (mPWP) comparing the results from a 14 member perturbed physics ensemble (PPE) with two types of palaeo-data: sea surface temperature (Dowsett et al., 2010b) and vegetation derived surface air temperature (Salzmann et al., 2013). Pliocene modelling studies either using single model simulations or multi-model ensembles (MMEs) have consistently shown a data-model mismatch with models able to replicate the tropical data sites but unable to reproduce high latitude sites, especially through the North Atlantic (Dowsett et al., 2012; Haywood & Valdes, 2004; Haywood et al., 2013a). The aim was to investigate the role of parameter uncertainty in these data-model mismatches, to see if model and data can be reconciled.

The PPE has demonstrated that it is possible to reproduce the magnitude of mPWP warming with ensemble members with Charney sensitivities close to and above the upper bound of the IPCC likely range. However, the strongest performing ensemble members were not the simulations generating the greatest high latitude warming, but those achieving the balance between high latitude warming with minimal change at the tropics. Ensemble members M (4.54°C), N (4.62°C) and P (5.40°C) simulate the strongest DMCs based on the combination of SST and SAT data, fitting this requirement. Ensemble member Q, the member with the greatest warming, produced the lowest ranking SST ensemble member due to its inability to blend high latitude warming without tropical warming. Ensemble members with Charney sensitivities near the lower end of the IPCC likely range are unable to simulate the required high latitude warming but did simulate a stronger representation of mPWP tropical climate. The sub-ensemble members with lower values of atmospheric CO_2 improve the tropical DMC (with respect to their full ensemble equivalents), however weakened the high latitude DMC, especially in the 300 ppmv sub-ensemble.

Improved polar amplification ratios were important in ensemble members improving the DMC compared to the HadCM3 Standard. However, the sub-ensemble members, despite the largest polar amplification ratios of the ensemble were unable to improve the high latitude mismatch owing to a lack of high latitude warming. The polar amplification ratio can not alone improve the data-model mismatch.

Overall, while the PPE has shown it is possible to achieve the magnitude of mPWP high latitude warming, it is at present not possible to achieve the spatial distribution of this warming. The results indicate that it is unlikely that mPWP data-model mismatches can be reduced purely through investigating Charney sensitivity, but require exploration of all three vertices of the "PMIP Triangle" boundary condition, model and data uncertainties.

Chapter 5: On The Importance of Accurate Boundary Conditions for Modelling the mid-Pliocene Warm Period

5.1 Introduction

Chapter 4 discussed the whether it was possible to reconcile the palaeo-data and the perturbed physics ensemble (PPE), concluding that the PPE is able to reproduce the magnitude of warming that has been observed in the palaeo-data at high latitudes. However, the ensemble members were unable to represent the spatial distribution of this warming. Ensemble members that generated the appropriate amount of high latitude warming, were unable to avoid overheating tropical latitudes, and vice versa in comparison with the palaeo-data. The sub-ensemble investigating the CO_2 boundary condition within the ensemble, indicated that while this sub-ensemble improved tropical data-model mismatches, the sub-ensemble performed weaker in the high latitudes than the full ensemble members. Discussion of other potential causes of the mismatch and the poor spatial distribution of the warming suggested that the mismatch could be explained using the "*PMIP Triangle*" (Haywood et al., 2013a), a confluence of model, boundary condition and palaeo-data uncertainties that results in weaknesses in the data-model comparisons (DMCs). A forcing boundary condition uncertainty was investigated with the CO_2 sub-ensemble, however the main form of boundary condition uncertainty in the mid-Pliocene is the physical boundary conditions (ice sheets, vegetation and topography), with two sets of reconstructions, the PRISM2 and PRISM3D boundary condition sets.

Chapter 5 presents the comparison of the PRISM3D and PRISM2 ensembles, primarily investigating the impacts of the changes in the boundary conditions in combination with the parameter perturbations. The analysis will utilise both the intra-model comparisons between the two boundary condition ensembles and then use the data-model comparisons to ascertain the performance of each boundary condition ensemble. The changes between the PRISM3D and PRISM2 physical boundary conditions have been discussed in Chapters 1 and 2 (Sections 1.2.3 & 2.4.4), however, it is important to consider how these changes are observed by the model. Figures 5.1 and 5.2 highlight the changes to the model caused by the alterations to the physical boundary conditions.

The main difference between the PRISM3D and PRISM2 orography is seen in the raising of the Rockies in PRISM3D to a near modern elevation. Additionally changes in the elevation of the ice sheets are observed. Figure 5.1 illustrates both the changes in comparison to the modern (which will influence the data-model comparisons, performed as 'Pliocene minus modern'), but also the differences between PRISM3D and PRISM2. This comparison is especially useful to illustrate the changes in the West Antarctic (WAIS), East Antarctic (EAIS) and Greenland Ice Sheets (GrIS) between each set of boundary conditions.

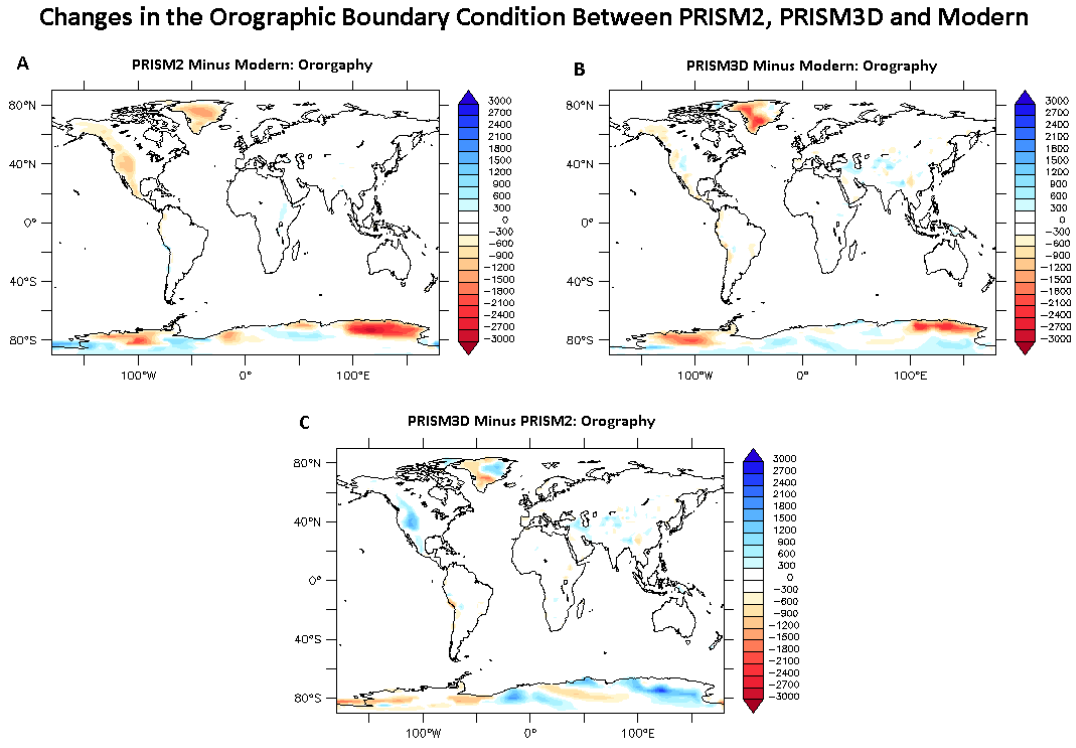


Figure 5.1. Global distribution of A) 'PRISM2 minus modern' B) 'PRISM3D minus modern' and C) 'PRISM3D minus PRISM2' orographic changes. For the model ancillary files, the orography includes the height of ice sheets, changes in the elevation of ice sheets between boundary conditions are displayed as well as changes to mountains.

Both ensembles contain vegetation inputs based on palaeo-data that are associated with a warmer and wetter than modern climate (Dowsett et al., 2010a). However, the PRISM3D reconstruction utilises a 28 biome model-data hybrid approach, compared to the 7 mega-biome PRISM2 approach. The result is between the PRISm3D and PRISM2 ensembles, the main differences in vegetation include significant reductions to high latitude tundra coverage, replaced with forest which extends throughout mid and eastern Europe. Additionally, desert and grassland coverage is reduced across Africa and Australia replaced with tropical savannah and tropical woodland while tropical forest extent is maintained (Dowsett et al., 2010a).

Vegetation is not represented directly in HadCM3 in the form of biomes, instead biomes are translated into inputs for MOSES (Met Office Surface Exchange Scheme – Cox et al., 1999). The biomes are represented as the physical characteristics of plants (such as root depth and roughness length) using look up tables to represent the mPWP vegetation data (Haywood et al., 2010a; Bragg et al., 2012). These characteristics are the physical boundary condition changes within the model that can interact with the parameterisations. The differences in the features representing vegetation between PRISM3D and PRISM2 are shown in Figure 5.2.

Vegetation feedbacks on climate are driven by fire/biomass burning or through changes to moisture, energy, momentum and carbon fluxes (Notaro et al., 2006). For these ensembles carbon fluxes and fire/biomass burning are not applicable as they are not modelled in the used version of MOSES. Moisture availability from vegetation will vary between boundary conditions as the vegetation type changes (Bonan, 2002). Forests have a greater surface roughness than grasslands and this forcing can be modulated by parameterisations of surface roughness and cloud formation. Changes to the surface roughness (either due to the boundary conditions or parameterisations) affects the boundary layer influencing low level cloud formation (through water vapour transport and momentum changes) and wind flows (Sud et al., 1988; Buermann, 2002). The stomatal resistance which influences regional moisture availability (Pollard & Thompson, 1995; Royer, 2001) is controlled by temperature, precipitation and CO_2 and is parameterised within the model, a parameter that is switched off in some ensemble members. Increasing high latitude forest cover exerts a change in albedo between boundary conditions (Betts & Ball, 1997), particularly influencing the snow free albedo within the model.

The physical changes to the land surface scheme (the way the model interprets the changes to the vegetation boundary conditions) will influence ensemble responses comparing PRISM2 and PRISM3D without the influence of the perturbed parameters. Some of the perturbed parameters will directly affect the physical changes (such as changes to surface roughness) and finally the physical changes to the land surface will affect how parameters for cloud formation, surface evaporation and convection respond. The differences caused by these three interactions will be represented across the different ensembles in regions where there is a change to the vegetation boundary condition.

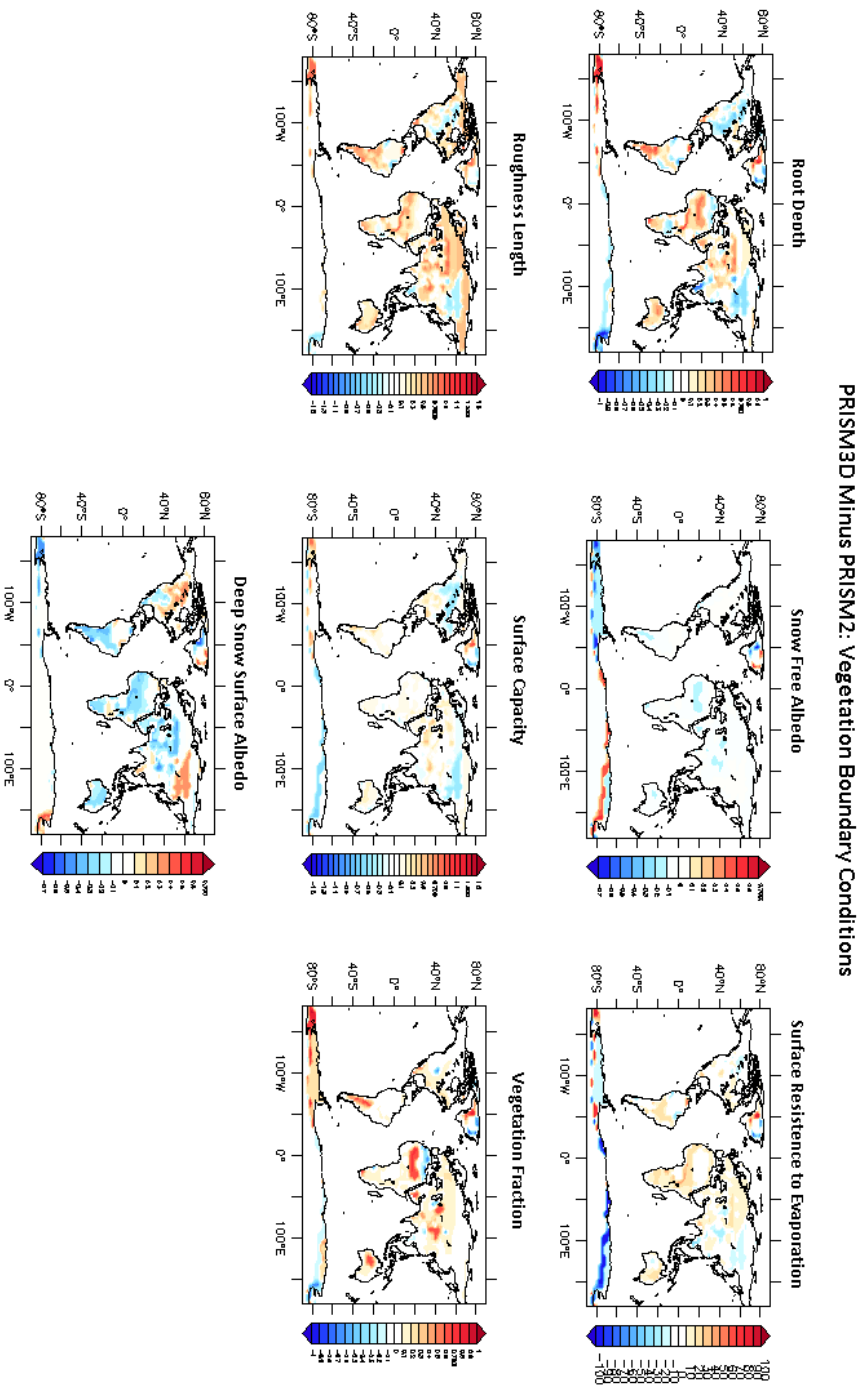


Figure 5.2. Global distribution of 'PRISM3D minus PRISM2' vegetation boundary condition changes for A) root depth; B) snow free albedo; C) surface resistance to evaporation; D) roughness length (prior to any perturbation); E) surface capacity; F) vegetation fraction and G) deep snow surface albedo.

The goal of the chapter is assess which boundary condition type produces the best representation of the mPWP climate, judged on performance of ensemble members in data-model comparisons. It will then assess, what impact, if any this has on future mPWP modelling, before finally assessing the role of boundary condition uncertainty in mPWP data-model mismatches. The goal will be achieved using both intra-model and

data-model comparisons. The IMCs will first assess the changes between the two Standard simulations, to illustrate the impact of the differences between the PRISM3D and PRISM2 boundary conditions on simulating the mPWP. The IMCs will then expand to include the PPE members, to assess how the perturbed physics and the boundary conditions affect each other. The IMCs will also investigate the role of the differences between both Pliocene boundary condition sets and the modern Standard simulation, which is used in the DMCs. The DMCs will then be used to rank the performance of each boundary condition ensemble and assess the impact of the boundary condition changes on modelling of the mid-Pliocene.

5.2. Intra-Model Comparisons

5.2.1. PRISM3D Standard minus PRISM2 Standard

The differences between the PRISM3D and PRISM2 physical boundary conditions on GCM modelling has never been detailed as a direct comparison between the boundary condition sets. Table 5.1 indicates that the PRISM3D Standard is warmer (0.41°C) & wetter (0.024 mm/day) than the PRISM2 Standard, an increase of ~16% for both metrics. However, despite these global mean increases, there are a number of noticeable regional variations (Figure 5.3), which may lead to important influences on the data-model comparisons which will be used to assess the model performance.

		Surface Air Temperatures				Sea Surface Temperatures				Precipitation			
Ensemble Member	**Charney Sensitivity**	**PRISM2**	**PRISM3D**	**Difference**	**Percent**	**PRISM2**	**PRISM3D**	**Difference**	**Percent**	**PRISM2**	**PRISM3D**	**Difference**	**Percent**
B	2.42	1.98	1.87	-0.11	-5.56	0.37	0.55	0.17	46.72	0.063	0.064	0.001	2.19
D	2.88	2.07	2.10	0.03	1.50	0.50	0.69	0.20	39.32	0.024	0.032	0.008	35.54
F	3.75	2.21	2.65	0.45	20.18	0.54	0.97	0.43	80.40	0.062	0.113	0.051	80.89
H	3.44	1.94	2.62	0.68	35.19	0.55	1.08	0.53	96.41	0.136	0.192	0.056	41.18
I	4.4	3.03	4.06	1.03	33.95	1.02	1.70	0.68	66.80	0.082	0.127	0.045	55.18
J	3.9	2.25	2.30	0.05	2.09	0.44	0.67	0.23	51.67	-0.079	-0.071	0.008	-10.13
K	4.44	2.86	3.68	0.83	28.90	0.97	1.59	0.62	63.87	-0.147	-0.099	0.048	-32.65
L	4.88	3.22	4.05	0.83	25.66	1.03	1.65	0.63	60.98	0.006	0.056	0.050	810.57
M	4.54	3.26	4.02	0.76	23.46	1.11	1.70	0.60	53.85	0.154	0.216	0.062	39.93
N	4.62	3.06	3.84	0.79	25.70	0.92	1.51	0.59	63.30	0.008	0.126	0.118	1405.48
O	4.79	2.87	3.84	0.97	33.70	1.00	1.70	0.70	70.82	0.050	0.132	0.082	166.61
P	5.4	3.53	4.46	0.93	26.42	1.24	1.94	0.70	56.20	0.116	0.175	0.059	50.86
Q	7.11	4.32	5.27	0.95	22.05	1.63	2.30	0.67	41.45	0.003	0.051	0.048	1395.49
Standard	3.3	2.53	2.94	0.41	16.11	1.21	1.25	0.04	2.97	0.151	0.175	0.024	15.59

Table 5.1 – Global mean annual values for surface air temperature (SAT - °C**), sea surface temperature (SST -** °C**) and precipitation (mm/day) PRISM2 and PRISM3D ensemble member minus modern Standard and the difference between PRISM2 and PRISM3D ensemble members.**

Figure 5.3a highlights the effect of the changes in boundary conditions between PRISM2 and PRISM3D. Key features of the change from PRISM3D to PRISM2 are:

- Impacts of changes in orography of mountains and ice sheets
- Regions of cooling over the Arctic Ocean, Eurasia & sub-Saharan Africa
- Southern Hemisphere warming, particularly over South America, Australia and the Southern Ocean.

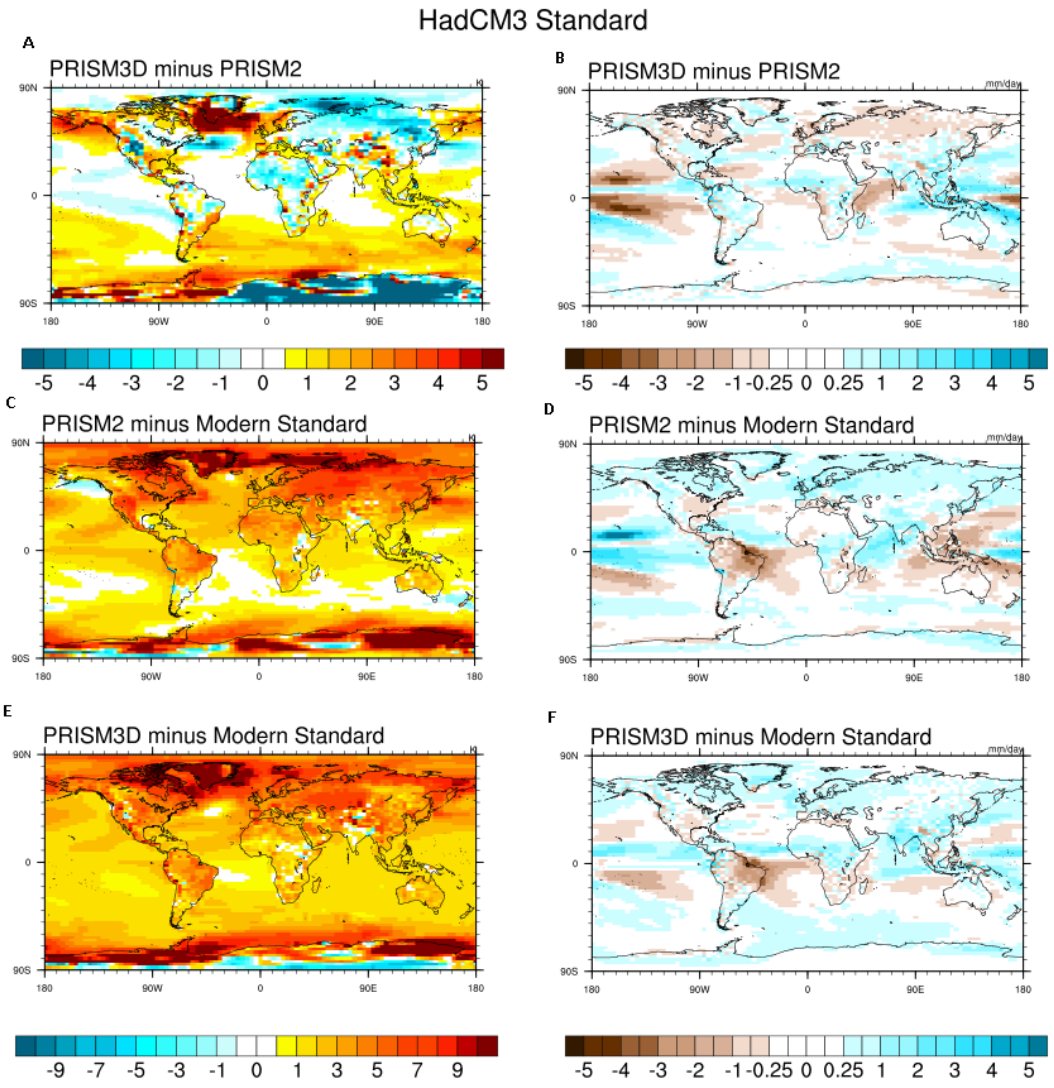


Figure 5.3 – Global mean annual intra-model comparisons for the Ensemble Standard for PRISM3D minus PRISM2 - a) surface air temperature (SAT - °C) & b) Precipitation (mm/day), PRISM2 minus modern Standard c) SAT & d) Precipitation and PRISM3D minus modern Standard e) SAT & f) Precipitation.

The temperature and precipitation changes between PRISM3D and PRISM2 Standards are shown in Figures 5.3a and 5.3b respectively. Orographic changes result in both patterns of warming and cooling. The raising of the Rockies causes a cooling of 2.5 to 3.5°C, and adjustments to the East Antarctic Ice Sheet (EAIS), which increases its eastern extent in the PRISM3D reconstruction compared to PRISM2 (Figure 5.1) results in up to 5°C cooling. Likewise reduction in the western portion of the EAIS results in warming of 2 to 4.5°C around the West Antarctic peninsula. Changes to the Greenland Ice Sheet, result in cooling (3°C) across Northern Greenland were the GrIS is elevated in PRISM3D compared to PRISM2 (Figure 5.1) and warming (5°C) were the GrIS is reduced. The changes in the distribution of the Antarctic ice sheets leads to small regional changes in precipitation around the Antarctic coasts (Figure 5.3b). These changes are primarily a decrease (0.25 mm/day) over East Antarctica on the boundary

of the increased EAIS elevation, whilst coastal East Antarctica and over the West Antarctic peninsula displays an increase in precipitation (0.25 mm/day). Over the GrIS, the change in elevation and resulting influence on SATs also has a corresponding effect on precipitation with changes of 0.25 mm/day increase or decrease for regions of warming or cooling respectively. Similarly a decrease in precipitation over the increased elevation of the Rockies (0.5 mm/day) is observed.

Across sub-Saharan Africa, there is a decrease in SATs (2.5°C) and an increase in precipitation (0.5 to 1 mm/day), although the temperature decrease covers a greater spatial area than the precipitation increase. The driver for this feature is the change in vegetation across the region. The PRISM3D vegetation increases the vegetation fraction in sub-Saharan Africa resulting in increased root depth and surface roughness (Figure 5.2) representing a change from desert to savannah (Salzmann et al., 2008). The PRISM2 Standard is warmer over the Arctic Ocean and eastern Eurasia than the PRISM3D Standard, resulting in extensive cooling of 1.5 to 3°C, with small regions of 4.5°C cooling such as the Barents Sea (Figure 5.3a). There is a corresponding decrease in terrestrial precipitation of 0.5 mm/day over eastern Eurasia. The Eurasian feature is a result of the alteration in the orography of the Rockies. Modelling sensitivity studies adjusting the height of the Rockies have indicated that by raising the Rockies, Northern Hemisphere westerly winds are deflected (Seager et al., 2002; Foster et al., 2010). The deflection removes zonal wind structures (as seen in the Southern Ocean around Antarctica) and replaces them with Rossby Wave structures, which shifts the jet stream flow (Seager et al., 2002; Foster et al., 2010). These waves continue around the Northern Hemisphere contributing to North Atlantic warming and establishing the colder conditions over Eurasia. The pattern of the Northern Hemisphere winds influences storm tracks and the loss of a zonal pattern leads to reduced precipitation over continental interiors, as observed across Eurasia (Figure 5.3b & Broccoli & Manabe, 1992; 1997; Foster et al., 2010). The same processes influencing Eurasian climate, reduces northwards heat transport leading to the cooling observed in the Arctic Ocean. Previous work comparing PRISM2 orography to modern day orography (Hill et al., 2011) has shown that raising the Rockies reduces northward heat flow increasing the equator to pole temperature gradient.

The most spatially extensive region of warming in the PRISM3D Standard compared to PRISM2 is across the Southern Hemisphere with warming between 1 and 2.5°C (Figure 5.3a). The warming, concentrated in the South Atlantic, appears to be a confluence of effects of boundary condition changes around the orography in West Antarctica (Figure 5.1) and vegetation changes in southern South America (Figure 5.2). Lowered

elevation of the WAIS, near its boundary with the EAIS results in warming around the West Antarctic peninsula and the Weddell Sea (Figure 5.3a). The vegetation changes across South America increase root depth, surface roughness and reduce albedo, which lead to an increase in SATs (Figure 5.3a). The warming protrudes off South America into the Atlantic sector of the Southern Ocean where it combines with warming from the changes to the WAIS to enhance the Southern Hemisphere warming.

5.2.2. Pliocene Standards minus Modern Standard

The comparison of *"PRISM3D Standard minus PRISM2 Standard"* is not utilised in the data-model comparisons, but instead the comparison of each ensemble member to the modern Standard is used. This section describes the key differences between PRISM3D and PRISM2 Standards to the modern Standard. Table 5.1 highlights the comparison of the PRISM2 and PRISM3D Standard simulations to the modern Standard. The PRISM3D Standard is warmer and wetter than the PRISM2 Standard compared to the modern, as expected based on the comparison between the two Standards. Spatial variation within these comparisons is displayed in Figures 5.3c & 5.3e showing the comparison for both ensemble Standards to the modern Standard. The PRISM3D Standard with near modern orography over the Rockies displays minimal orographic response, whilst the PRISM2 Standard displays 5°C warming. The orographic warming between PRISM2 and modern is the same as observed between PRISM2 and PRISM3D. For both PRISM2 and PRISM3D ensembles, the GrIS, WAIS and EAIS are adjusted compared to modern (Figure 5.1a & b). The adjustments in the ice sheets, which is primarily represented through a change in orography, show consistent variations between both ensemble Standards. Warming of 9 to 10°C is observed over regions where these ice sheets are reduced in comparison to the modern. The differences between the two ensembles are displayed through the reduced extent of EAIS in PRISM2 compared to PRISM3D, leading to a smaller area of warming in the PRISM3D comparison to modern. However, there is greater warming over the WAIS in PRISM3D than observed in PRISM2. Over the GrIS, the PRISM2 comparison displays uniform warming of 8 to 10°C across Greenland compared to the modern. The PRISM3D Standard displays more intense warming 9 to 10°C over most of Greenland, but a small region where warming is approximately 5°C. The difference is a result of the changes in different distributions of the GrIS between PRISM2 and PRISM3D (Figure 5.1). Whilst there are spatial differences between the PRISM3D and PRISM2 ice sheets, where there are changes compared to modern these differences are consistent in terms of magnitude.

Over the Arctic Ocean and Eurasia, the PRISM3D Standard is cooler than the PRISM2 Standard and this cooling is also observed as reduced Pliocene warming over this region in the comparisons to the modern Standard. Warming over the Arctic and Eurasia is 2 to 6°C in the PRISM3D Standard and 3 to 7°C in the PRISM2 Standard. The main differences in precipitation in the comparison with the modern Standard is that the PRISM2 Standard is drier over Amazonia (2 to 2.5 mm/day decrease) compared to PRISM3D (0.5 to 1.5 mm/day increase) with respect to the modern Standard. The PRISM2 Standard is wetter over Eurasia (0.25 to 1 mm/day), with the PRISM3D Standard showing little change in precipitation between the Pliocene and modern. In general, the trends observed in the comparison between the PRISM3D and PRISM2 Standards can be translated over to the comparisons between PRISM3D and PRISM2 to modern.

5.2.3. Impact of Parameter Uncertainty on Boundary Condition Changes

a) PRISM3D minus PRISM2 Ensemble Member

The differences between the PRISM3D and PRISM2 boundary conditions have been described using the comparison of the ensemble Standard forced with each boundary condition set. However, while the analysis of the Standard reveals important changes to the simulated mPWP climate by adopting the PRISM3D boundary conditions, as discussed previously parameter uncertainty results in a range of potential simulated mPWP climates (Pope et al., 2011 & Chapter 4). Through this section, the result of the confluence of parameter and boundary condition changes between PRISm3D and PRISM2 will be outlined.

PRISM3D ensemble members are warmer and wetter than their PRISM2 ensemble equivalents (Table 5.1) with the exception of ensemble member B. The greatest mean annual warming is in ensemble members with higher Charney sensitivities than the ensemble Standard. Analysis of the Pliocene PPE in Chapter 4 (Section 4.2.1), highlighted that the ensemble forced with PRISM3D boundary conditions included two broad groupings, the 'colder than Standard' members (B, D, F, H & J) and the 'warmer than Standard' members (I, K, L, M, N, O, P & Q). The same groupings exist in the PRISM2 ensemble (Table 5.1), and the increase in temperature is generally split into these two groups. The 'warmer than Standard' members show a 0.8 to 1.0°C increase, whilst the 'colder than Standard' members show a -0.1 to 0.7°C increase from PRISM2 to PRISM3D (Table 5.1). Surface air temperature trends are replicated in the sea surface temperatures across the ensemble members (Table 5.1).

The comparison of the ensemble Standards forced with the two sets of boundary conditions highlighted a number of climate responses to the changes. Through utilising the PPE members it reveals more detail about these boundary condition responses. Across the ensemble members it is evident that the boundary condition changes result in two main types of response, direct and indirect. Direct features represent the responses that occur at the location of a boundary condition change. Across the ensemble they display consistent magnitude and spatial distribution. Indirect features represent responses that occur downstream of a boundary condition change, and have been modified by the interaction with other boundary conditions or perturbed model parameters. Indirect features display variation in magnitude and/or spatial distribution across the ensemble. The differentiation of these two response types is important as the variation across the ensemble could influence the performance of ensemble members in the DMCs.

Direct responses to the boundary condition changes are evident at regions where the orography has been adjusted such as the Rockies and the elevation of the ice sheets (Figure 5.4). Across the ensemble members, the SAT decrease of 2.5 to 3.5°C is consistent with those observed in the ensemble Standard (Figure 5.3c,e). Similarly, a 0.5 mm/day decrease in the precipitation response over the elevated Rockies is observed in all ensemble members (Figure 5.5) and the Standard.

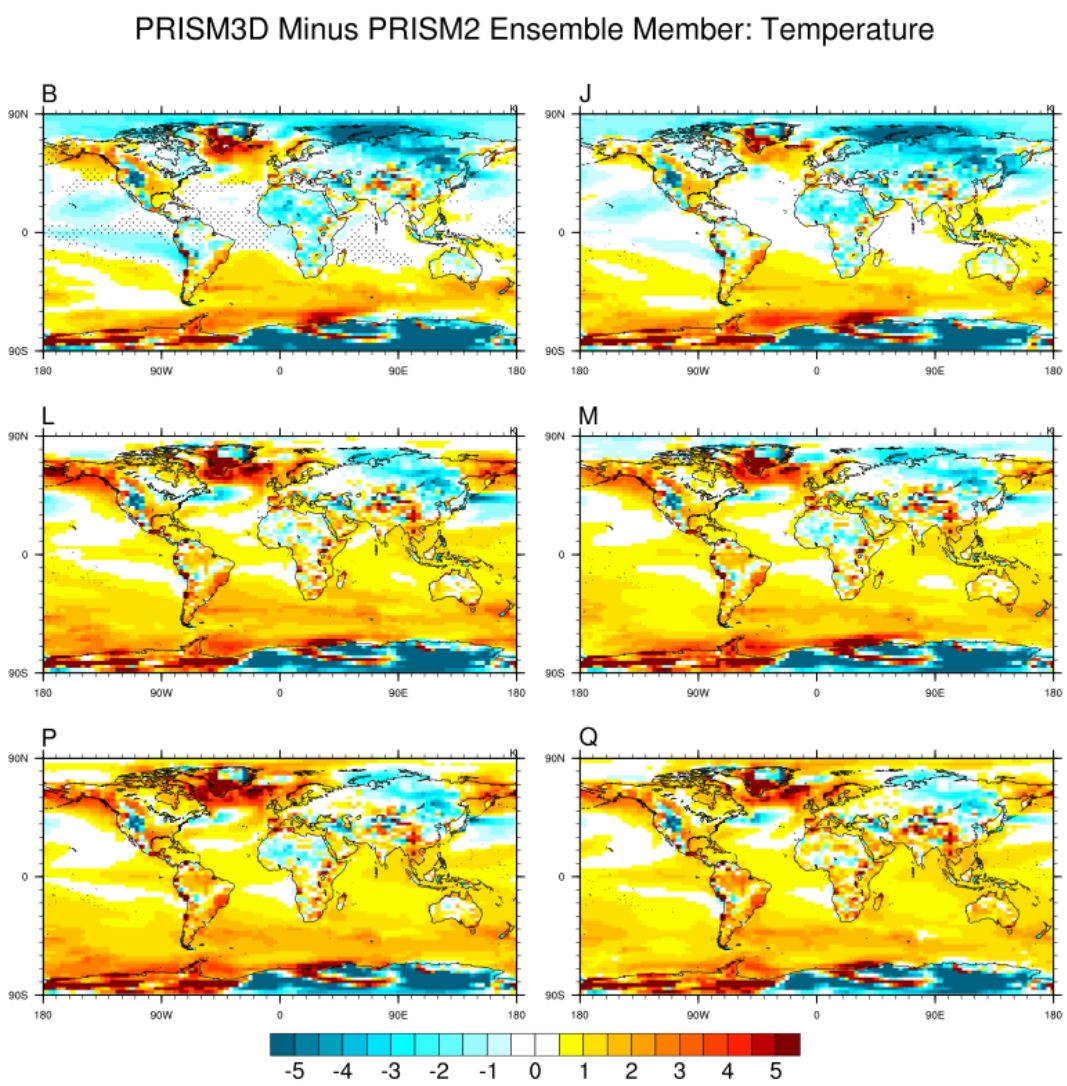


Figure 5.4 – Global mean annual intra-model comparisons for PRISM3D minus PRISM2 ensemble member surface air temperature (SAT - °C) for ensemble members B, J, L, M, P & Q. The remaining ensemble members are displayed in Appendix B1.

However some of these direct features, while consistent across the ensemble members display indirect features evident in this ensemble. The other SAT and precipitation responses observed between the two boundary condition sets in the Standards represent indirect features. The variation in the Arctic and Eurasian temperature response is pronounced across the ensemble and is representative of indirect features with variation in the magnitude and spatial dimensions of the response. In the 'cooler than Standard' members, there are extensive areas of cooling in the Arctic Ocean and over Eurasia, of 1.5 to 3°C. However, in the 'warmer than Standard' members the Arctic cooling dissipates and is replaced by warming or regions of no SAT difference.

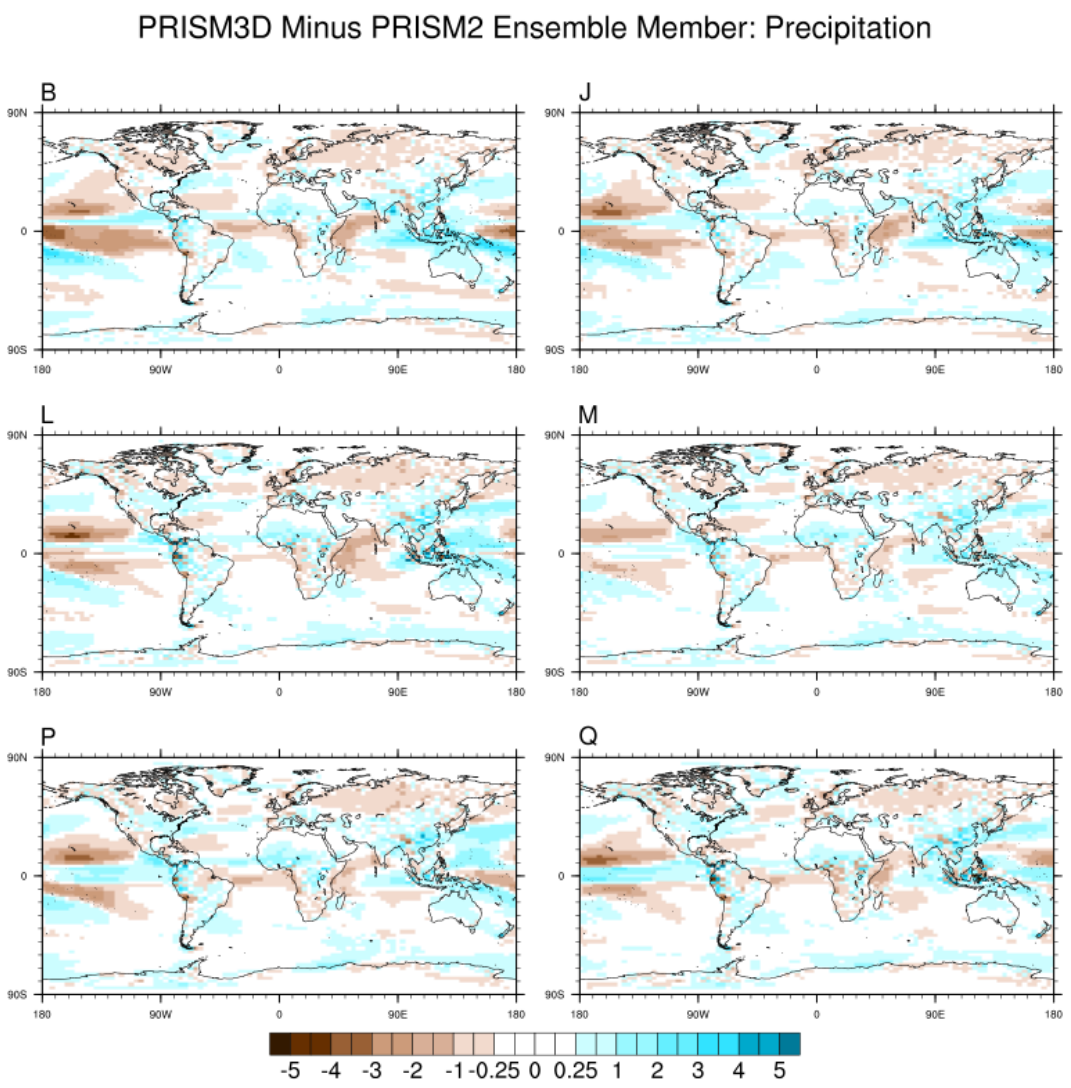


Figure 5.5 – Global mean annual intra-model comparisons for PRISM3D minus PRISM2 ensemble member precipitation (mm/day) for ensemble members B, J, L, M, P & Q. The remaining ensemble members are displayed in Appendix B2.

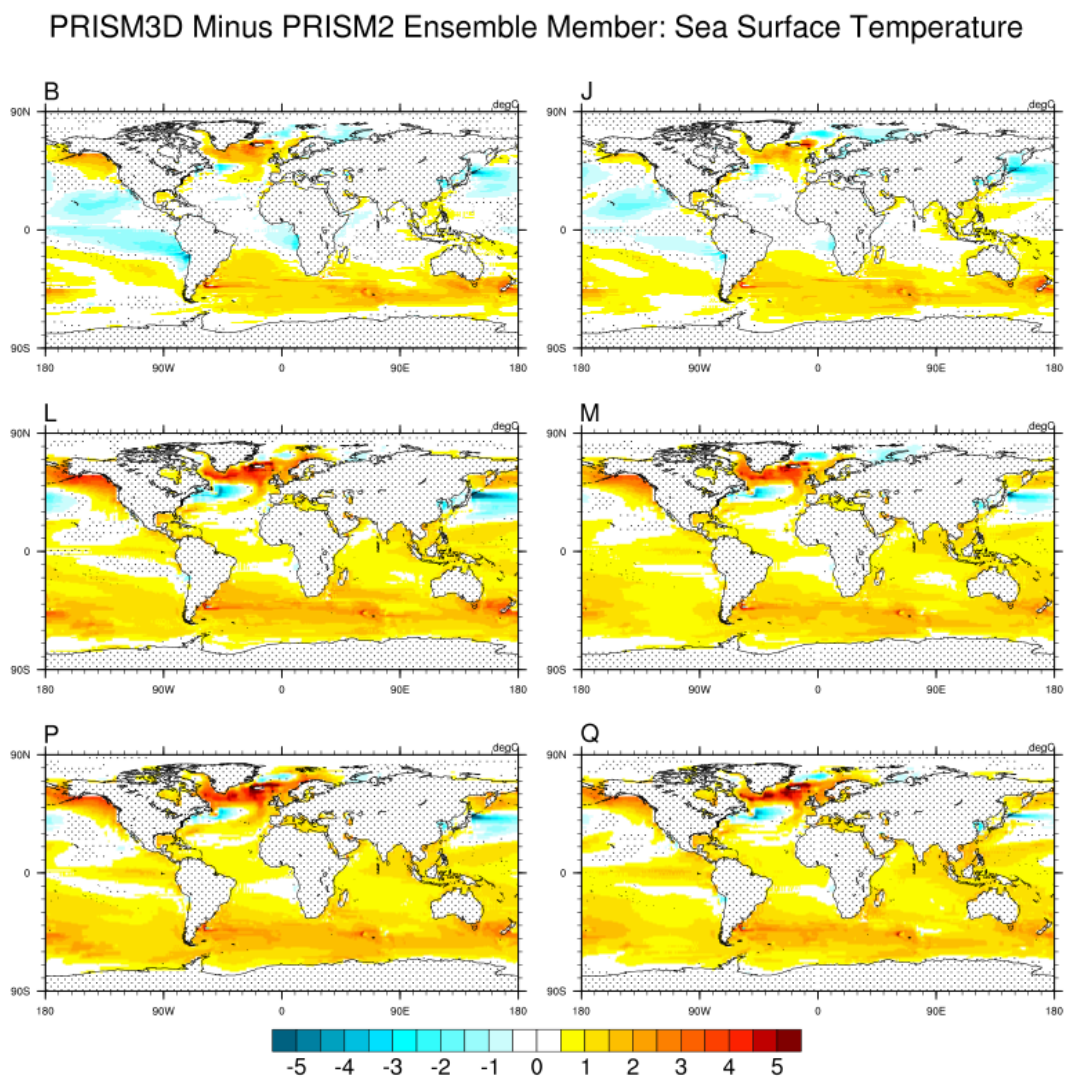


Figure 5.6 – Global mean annual intra-model comparisons for PRISM3D minus PRISM2 ensemble member sea surface temperature (SST - °C) for ensemble members B, J, L, M, P & Q. The remaining ensemble members are displayed in Appendix B3.

There is distinct Southern Hemisphere warming across the ensemble members, with the warming increasing through the 'warmer than Standard' members. The SAT increase is greater in ensemble members with higher Charney sensitivity. The warming is also observed in the SSTs (Figure 5.6). These two warming features combine to generate increased warming through the South Atlantic which propagates across the Southern Oceans with a range of 1.5 to 3°C (Figures 5.4 & 5.6). The components of this warming were observed in the comparison of the ensemble Standards (Figure 5.3).

There is a band of cooling and increased precipitation throughout sub-Saharan Africa in all ensemble members. Whilst the temperature changes vary across the ensemble, the increased precipitation (0.5 to 1 mm/day) is consistent across all the members (Figure 5.7). The reduction in temperature ranges between 2 to 3°C across the ensemble and

the spatial extent varies across the ensemble, reducing in scale through the 'warmer than Standard' ensemble members. The cause of the temperature and precipitation patterns, is the representation of the change in vegetation boundary conditions, represented as an increased vegetation fraction and resulting increases in root depth and surface roughness as a result of the change from desert to savannah (Figure 5.2).

Each ensemble member has been forced by the same changes to the boundary conditions, however there is not a linear response across the ensemble with respect to the mean annual temperature and precipitation values (Table 5.1). The comparison between the PRISM3D and PRISM2 Standards shows a temperature difference of 0.41°C, the warming a result of the change in the physical boundary conditions. Across the ensemble, the general trend is that 'warmer than Standard' members show the greater increases in SATs between the ensembles based on comparisons between equivalent members. However there is variation between the ensemble members. The warmest member with warmest SATs in both ensembles is member Q, which has the highest Charney sensitivity (7.11°C) of any ensemble member (Table 5.1). However, the largest increase in SATs between the PRISM2 and PRISM3D ensembles is member I with 1.03°C. Member I has a Charney sensitivity (4.44°C) which while greater than the Standard, is in the middle of the ensemble range (2.42 to 7.11°C). Elsewhere there are changes in the rankings for some ensemble members between each ensemble. To assess the effect of the boundary condition changes it is important to determine whether the response of the ensemble members is due to the boundary condition changes or the interaction of the perturbed physics and the boundary condition changes.

Temperature							
Ensemble Member	PRISM2	PRISM3D	Difference_PRISM2	Difference_PRISM3D	Difference_IMC	Difference_BC	Boundary Condition Change
B	1.98	1.87	-0.55	-1.07	-0.52	-0.11	0.41
D	2.07	2.10	-0.46	-0.84	-0.38	0.03	0.41
F	2.21	2.65	-0.33	-0.29	0.03	0.45	0.41
H	1.94	2.62	-0.59	-0.32	0.27	0.68	0.41
I	3.03	4.06	0.50	1.12	0.62	1.03	0.41
J	2.25	2.30	-0.28	-0.64	-0.36	0.05	0.41
K	2.86	3.68	0.33	0.74	0.42	0.83	0.41
L	3.22	4.05	0.69	1.11	0.42	0.83	0.41
M	3.26	4.02	0.73	1.08	0.35	0.76	0.41
N	3.06	3.84	0.53	0.90	0.38	0.79	0.41
O	2.87	3.84	0.34	0.90	0.56	0.97	0.41
P	3.53	4.46	1.00	1.52	0.52	0.93	0.41
Q	4.32	5.27	1.79	2.33	0.54	0.95	0.41
STD	2.53	2.94	0.00	0.00	0.00	0.41	0.41

Precipitation							
Ensemble Member	PRISM2	PRISM3D	Difference_PRISM2	Difference_PRISM3D	Difference_IMC	Difference_BC	Boundary Condition Change
B	0.063	0.064	-0.088	-0.111	-0.023	0.001	0.024
D	0.024	0.032	-0.127	-0.143	-0.016	0.008	0.024
F	0.062	0.113	-0.089	-0.062	0.027	0.051	0.024
H	0.136	0.192	-0.015	0.017	0.032	0.056	0.024
I	0.082	0.127	-0.069	-0.048	0.021	0.045	0.024
J	-0.079	-0.071	-0.230	-0.246	-0.016	0.008	0.024
K	-0.147	-0.099	-0.298	-0.274	0.024	0.048	0.024
L	0.006	0.056	-0.145	-0.119	0.026	0.050	0.024
M	0.154	0.216	0.003	0.041	0.038	0.062	0.024
N	0.008	0.126	-0.143	-0.049	0.094	0.118	0.024
O	0.050	0.132	-0.101	-0.043	0.058	0.082	0.024
P	0.116	0.175	-0.035	0.000	0.035	0.059	0.024
Q	0.003	0.051	-0.148	-0.124	0.024	0.048	0.024
STD	0.151	0.175	0.000	0.000	0.000	0.024	0.024

Table 5.2 – The calculation of the differences between the PRISM3D and PRISM2 ensemble members for temperature and precipitation for the changing of the boundary conditions. The calculations used are explained in equations 5.1 to 5.6.

We can break down the impact of the boundary condition and perturbed parameters on the differences between ensembles using analysis of the SATs. Firstly, to allow for the impact of the perturbed parameters on the ensemble member responses displayed between PRISM3D and PRISM2 we assessed the changes through comparison of each ensemble member to its ensemble Standard (Eqs. 5.1 & 5.2).

$$Difference\_PRISM3D_{EnsMem} = PRISM3D_{EnsMem} - PRISM3D_{Std} \quad \text{(Eq 5.1)}$$

$$Difference\_PRISM2_{EnsMem} = PRISM2_{EnsMem} - PRISM2_{Std} \quad \text{(Eq 5.2)}$$

Differences between PRISM3D and PRISM2 boundary conditions can be assessed through the temperature changes between the equivalent ensemble members. To calculate this, we use the difference within the ensemble which reflects the impact of the perturbed parameters (Eq. 5.1) and calculate the changes between these from PRISM3D to PRISM2 (Eq. 5.3). We then utilise the difference due to the boundary conditions, which is calculated using identical ensemble members and was originally shown in Table 5.1 (Eq. 5.4). Finally we subtract these to calculate the response for each ensemble member from the change in boundary conditions (Eq. 5.5)

$$Difference\_IMC_{EnsMem} = Difference\_PRISM3D_{EnsMem} - Difference\_PRISM2_{EnsMem} \quad \text{(Eq. 5.3)}$$

$$Difference\_BC_{EnsMem} = PRISM3D_{EnsMem} - PRISM2_{EnsMem} \quad \text{(Eq. 5.4)}$$

$$Boundary\_Condition\_Change = Difference\_BC_{EnsMem} - Difference\_IMC_{EnsMem} \quad \text{(Eq. 5.5)}$$

From equations 5.3 through 5.5, we can determine that a warming of 0.41°C occurs in each ensemble member as a result of the physical boundary condition changes between PRISM3D and PRISM2 (Table 5.2). However, whilst the influence of the changes in the boundary conditions is constant across the ensemble, the effects from perturbing parameters varies between equivalent ensemble members.

$$Difference\_PPE_{EnsMem} = Difference\_IMC_{EnsMem} - Difference\_BC_{EnsMem} \quad \text{(Eq. 5.6)}$$

As previously outlined, a range of responses exists between ensemble members. Removing the boundary condition warming (Eq. 5.6), the PRISM3D ensemble members range between -0.52°C to 0.62°C compared to PRISM2. The range observed here is the result of the perturbed parameters interacting with the adjusted boundary conditions. For the 'warmer than Standard' ensemble members this interaction results in enhanced warming but for the 'colder than Standard' members it results in a cooling. With the multi-perturbed parameters utilised within this study, it is not possible to identify the particular parameters that are influencing the temperature responses observed within

the ensemble members. One parameterisation, the forest roughness length, directly interacts with the changes to the vegetation, a parameter applied to members H, I, K, L, M, N & Q (Tables 2.1a,b in Chapter 2). However, these members all show a wide variation of parameter based warming, therefore it is unlikely that this parameterisation is influencing the results observed here.

Applying the same equations to the precipitation values from the ensemble we can determine that the precipitation response to the boundary condition change results in the PRISM3D ensemble being wetter by 0.024 mm/day compared to its PRISM2 equivalent. When the boundary condition ensemble effect is removed, precipaition in the ensemble ranges from -0.023 mm/day to 0.094 mm/day, with only members B (-0.023 mm/day), D (-0.016 mm/day) & J (-0.016 mm/day - Table 5.2) drier than their PRISM2 equivalents prior to the boundary condition effect. In both the precipitation and SAT analysis the ensemble members simulating a warmer climate with the exclusion of the boundary condition effect also simulate a wetter climate. However, the magnitude of the warming in an ensemble member is not related to the magnitude of its precipitation increases. Therefore the parameters that control the warming due to boundary condition changes in the ensemble members are not the same as those influencing the change in precipitation in these ensemble members.

b) PRISM3D/PRISM2 Ensemble Member minus Modern Standard

The comparison of the PRISM3D Standard and PRISM2 Standard to the modern Standard highlighted that the IMC differences noticed between the ensemble members were observed in the comparison to the modern Standard. The purpose of this section is to highlight the difference between the boundary condition ensembles and how these differences will potentially affect the DMCs. As in the previous comparisons, two types of response have been observed, direct and indirect responses. Direct responses, such as the changes in orography either over mountains or ice sheets are consistent across all ensemble members. Indirect responses represent the downstream effects of interactions between the perturbed physics and the changes in boundary conditions.

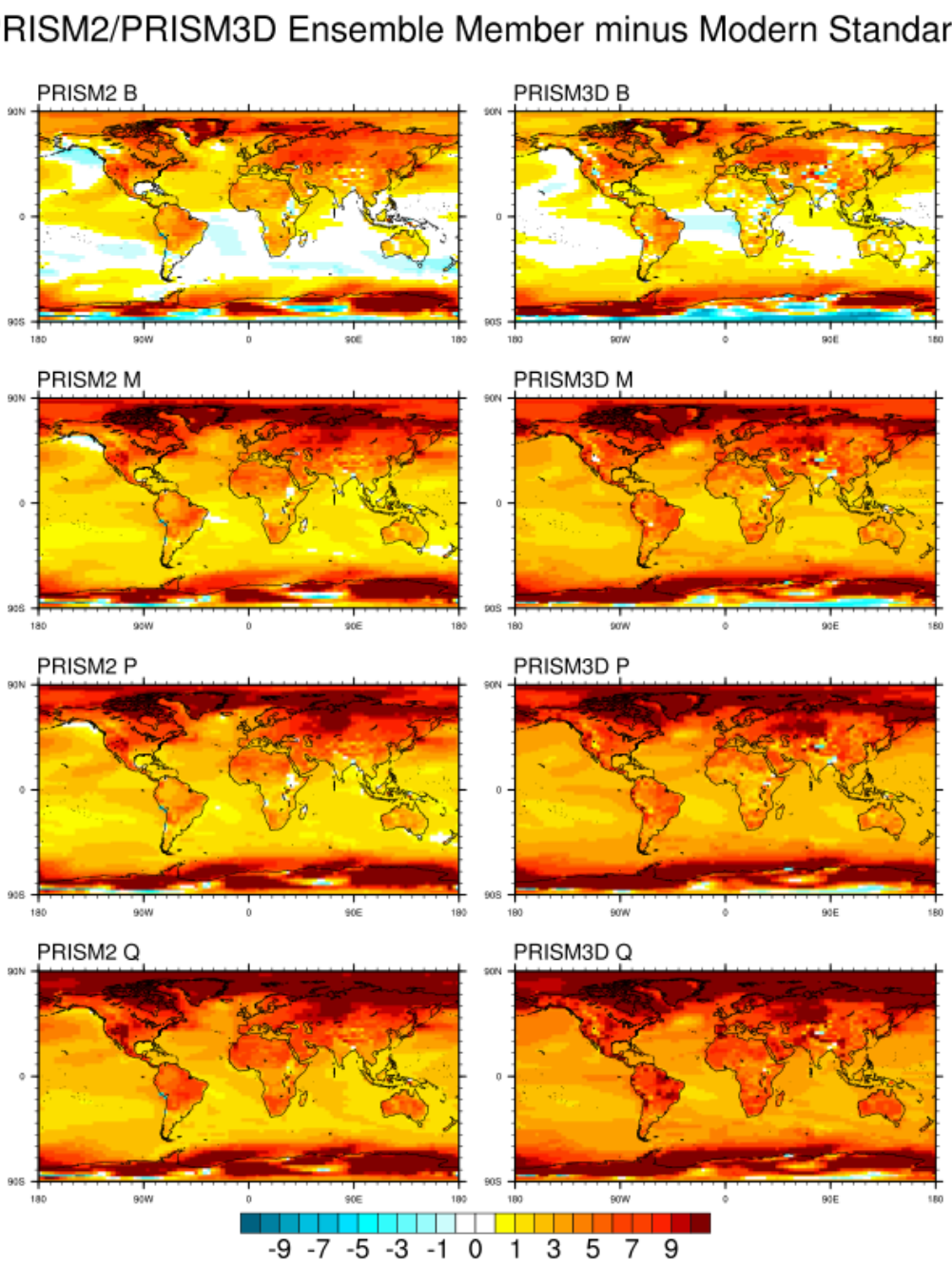


Figure 5.7 – Global mean annual intra-model comparisons for "PRISM2 ensemble member minus modern Standard" & for "PRISM3D ensemble member minus modern Standard" surface air temperature (SAT - °C) for ensemble members B, M, P & Q. The remaining PRISM2 ensemble members are displayed in Appendix B4 and PRISM3D members in Appendix B5.

Differences in the distribution of orography and ice sheets between PRISM3D and PRISM2 (Figure 5.1) are represented in the two comparisons to the modern Standard (Figures 5.7 & 5.8). In the comparisons between PRISM3D and PRISM2 ensemble members, temperature variation was observed across the Arctic Ocean and Eurasia. The cause of this variation is attributed to the indirect response between changes in PRISM3D orography and the perturbed parameters. The 'cooler than Standard'

ensemble members displayed cooling through this region compared to their PRISM2 equivalents (Figure 5.4), and this can be observed in the comparisons to modern. PRISM3D 'cooler than Standard' ensemble members display Arctic and Eurasian temperatures of 0 to 4°C compared to 5 to 7°C in the PRISM2 equivalents (Figure 5.7). The 'warmer than Standard' ensemble members displayed similar or warmer temperatures through Eurasia and the Arctic compared to their PRISM2 equivalents (Figure 5.4).

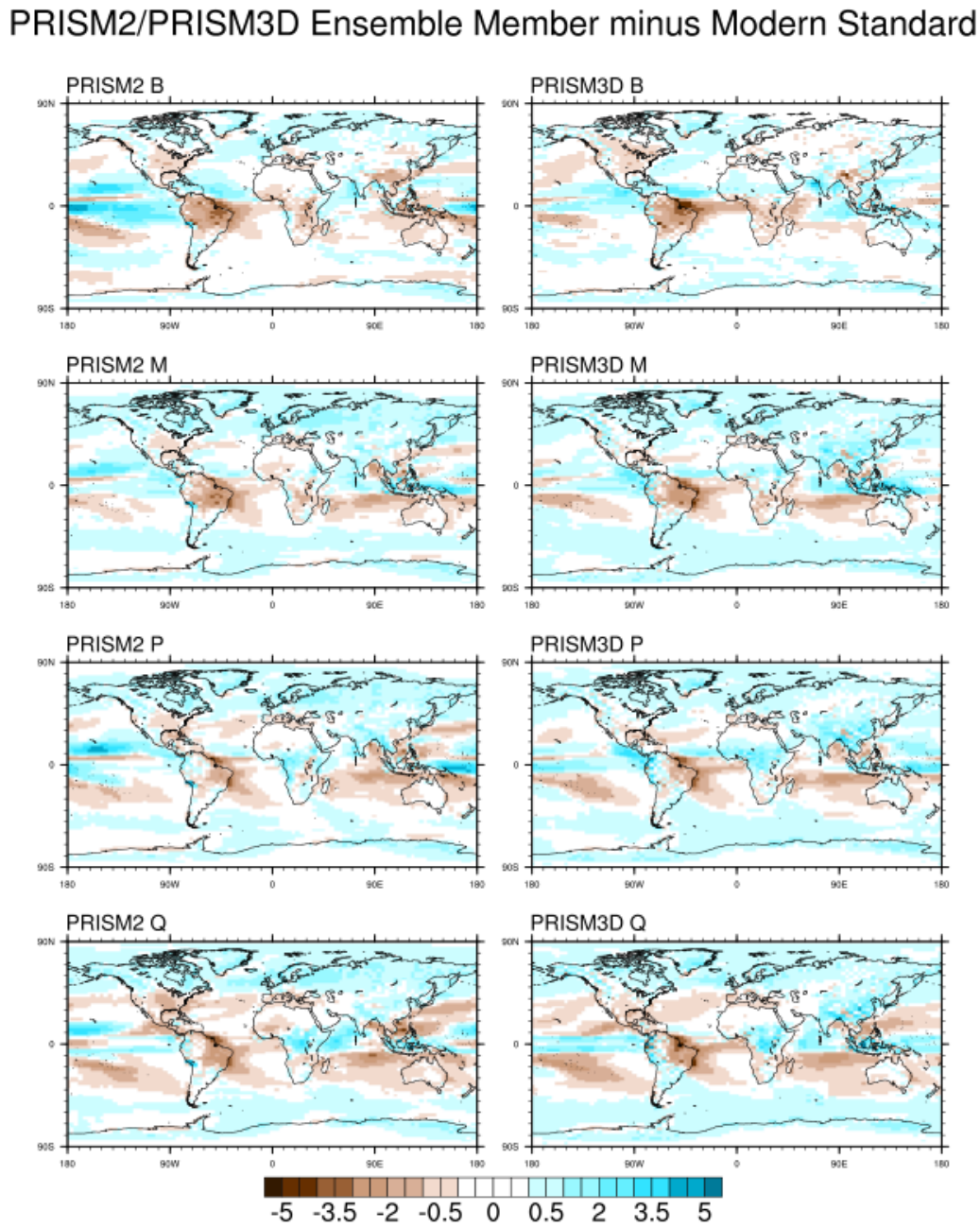


Figure 5.8 – Global mean annual intra-model comparisons for "PRISM2 ensemble member minus modern Standard" & for "PRISM3D ensemble member minus modern Standard" precipitation for ensemble members B, M, P & Q. The remaining PRISM2 ensemble members are displayed in Appendix B6 and PRISM3D members in Appendix B7.

The trend is replicated in the comparison to the modern Standard, with members I, K, L & O displaying comparable warming across this region to their PRISM2 equivalents, whilst members M, N, P & Q are warmer through this region than their PRISM2 equivalents by 1 to 2°C (Figure 5.7). Similarly PRISM2 ensemble members are predominantly slightly wetter by 0.25 mm/day over Eurasia compared to the PRISM3D members (Figure 5.8). As with temperature, this response is reduced through the 'warmer than Standard' ensemble members.

The band of cooling and increased precipitation observed across sub-Saharan Africa can be observed in the comparisons to the modern Standard. As observed in the IMCs the differences in cooling is greatest in the 'cooler than Standard' members than in the 'warmer than Standard' members (Figure 5.4), a feature also observed for the increase in precipitation (Figure 5.5). For both the precipitation and temperatures, the magnitude of the change is equivalent to that observed in the IMCs for each ensemble member (Figures 5.7 & 5.8). The same pattern of response is also observed for the Southern Hemisphere warming, with PRISM3D members warmer than the modern Standard through the Southern Hemisphere by 2 to 3°C with the warming greatest in the warmer ensemble members (M, N, O, P & Q).

c) Polar Amplification across the Ensemble

An important feature observed in analysis of the PRISM3D ensemble (Chapter 4) was the magnitude of polar warming and polar amplification ratios. Elevated polar warming and a high polar amplification ratio was observed as being able to reduce data-model mismatches at high latitudes whilst minimising the warming through tropical latitudes.

		PRISM2			PRISM3D		
Ensemble Member	**Charney Sensitivity**	**Polar Temperature**	**Global Mean Tempeature**	**Polar Amplification Ratio**	**Polar Temperature**	**Global Mean Tempeature**	**Polar Amplification Ratio**
B	2.42	6.11	1.98	3.09	4.37	1.87	2.33
D	2.88	6.38	2.07	3.08	5.58	2.10	2.66
F	3.75	6.95	2.21	3.15	6.48	2.65	2.45
H	3.44	7.08	1.94	3.65	6.95	2.62	2.65
I	4.4	8.41	3.03	2.77	9.76	4.06	2.40
J	3.9	7.59	2.25	3.37	5.85	2.30	2.54
K	4.44	7.02	2.86	2.46	7.90	3.68	2.15
L	4.88	10.18	3.22	3.16	10.86	4.05	2.68
M	4.54	10.33	3.26	3.17	10.31	4.02	2.56
N	4.62	9.48	3.06	3.10	10.18	3.84	2.65
O	4.79	8.51	2.87	2.96	9.08	3.84	2.36
P	5.4	10.70	3.53	3.03	11.82	4.46	2.65
Q	7.11	12.15	4.32	2.81	12.98	5.27	2.46
Standard	3.3	7.47	2.53	2.95	7.31	2.94	2.49

Table 5.3 – Global and polar mean annual surface air temperatures (SATs - °C) and polar amplification ratios for the PRISM2 and PRISM3D ensemble members. The polar amplification ratio is calculated as being the ratio of the polar average temperature (67.5 to 90°N) to the global mean annual temperature.

Table 5.3 displays the polar amplification ratios for the PRISM2 and PRISM3D ensembles. The polar amplification ratios for the PRISM2 ensemble are greater than their equivalent PRISM3D ensemble members (Table 5.3), with a greater range in the PRISM2 ensemble (PRISM2: 2.45 to 3.65, PRISM3D: 2.14 to 2.68). For members B, D, F, H, J, M & the Standard, the PRISM3D global mean SATs are increased compared to PRISM2, however the polar SATs are decreased in comparison to or are the same as in PRISM2. As a result the polar amplification ratio is decreased in these PRISM3D members compared to their PRISM2 equivalents. For members I, K, L, N, O, P & Q, both global mean and polar SATs increase in the PRISM3D members

compared to PRISM2. However, the increase in polar SATs is smaller than the increase in global mean SATs for these members. Again, this results in a smaller polar amplification ratio for the PRISM3D ensemble than the PRISM2 ensemble members.

The reduced polar amplification ratios in the PRISM3D ensemble members could result in a weakened DMC compared to the PRISM2 ensemble. The importance of the balance between polar and tropical warming has been previously discussed with respect to this PPE (Pope et al., 2011 & Chapter 4). In that analysis, ensemble members with high polar warming also displayed increased tropical temperatures. For member Q, this resulted in a weakened data-model comparison compared to the other 'warmer than Standard' members from that ensemble. However, analysis of ensemble members run with lower concentrations (300 & 350 ppmv) of atmospheric CO_2, which displayed a higher polar amplification ratio than the full ensemble members (run at 400 ppmv CO_2) had lower polar SATs. These ensemble members produced weaker data-model comparisons than their full ensemble equivalents, indicating that a balance was needed between increased polar and tropical SATs (Chapter 4). The differences in polar amplification ratios, polar SATs and global SATs between the PRISM3D and PRISM2 ensemble members could influence the DMCs. The balance between these three factors will determine which boundary condition ensemble performs strongest on the data-model comparisons, specifically to the SSTs.

5.3. Data-Model Comparisons

5.3.1. Site-by-Site Data-Model Comparisons

The existing data-model mismatch across a range of mPWP modelling studies (i.e. Haywood & Valdes, 2004; Dowsett et al., 2012; Haywood et al., 2013a) highlights that model simulations display a strong representation of mPWP tropical SSTs but perform poorly at high latitudes, particularly through the North Atlantic. Similar trends exist through the data-model comparisons to the recently developed vegetation derived terrestrial SAT data (Salzmann et al., 2013). The performance of Pliocene PPE members compared to the existing SST data-model mismatch is discussed in Chapter 4. The conclusion from Chapter 4 was that the ensemble was able to achieve the magnitude of the mPWP proxy temperatures, but was unable to replicate the spatial distribution of these temperatures, particularly through the high North Atlantic in all members and the tropics in 'warmer than Standard' members.

a) Data-Model Comparisons for the Ensemble Standards

The RMSE scores for the PRISM3D Standard indicate that this simulation (rather than the PRISM2 Standard) has produced an improved representation of the mPWP climate in comparison with to SST palaeo-data (Table 5.4).

Compared to the PRISM2 Standard, the PRISM3D Standard represents a 13.9% improvement in the RMSE score, from 2.80 to 2.42 (Table 5.4). For the terrestrial vegetation derived SATs, a similar improvement is observed between PRISM3D and PRISM2 Standards as observed in the SST DMCs. The PRISM3D Standard RMSE (4.64) is 10.2% improved on the PRISM2 Standard (5.17). The PRISM3D Standard improves the DMC to the vegetation biomes compared to the PRISM2 Standard, with the Kappa score increasing from 0.22 in PRISM2 to 0.26 in PRISM3D an 18% improvement.

Sea Surface Temperatures				
Ensemble Member	**PRISM2**	**PRISM3D**	**Difference**	**Percentage Change**
B	3.33	2.70	-0.63	-19.0
D	2.95	2.67	-0.28	-9.5
F	2.83	2.39	-0.44	-15.5
H	2.88	2.41	-0.46	-16.1
I	2.60	2.39	-0.21	-8.2
J	2.93	2.54	-0.40	-13.5
K	2.61	2.38	-0.23	-8.7
L	2.60	2.27	-0.33	-12.8
M	2.47	2.36	-0.11	-4.5
N	2.54	2.25	-0.29	-11.3
O	2.60	2.38	-0.22	-8.6
P	2.44	2.39	-0.05	-1.9
Q	2.47	2.71	0.24	9.9
Standard	2.80	2.42	-0.38	-13.7
Surface Air Temperatures				
B	5.83	5.73	-0.10	-1.7
D	5.51	5.33	-0.18	-3.3
F	5.45	5.05	-0.40	-7.3
H	5.74	5.04	-0.70	-12.2
I	4.75	4.03	-0.72	-15.2
J	5.32	5.41	0.09	1.6
K	5.35	4.66	-0.69	-12.9
L	4.40	3.95	-0.45	-10.3
M	4.23	3.92	-0.31	-7.4
N	4.31	3.75	-0.56	-13.0
O	4.82	4.23	-0.59	-12.2
P	4.15	3.70	-0.45	-10.9
Q	3.79	3.79	0.00	0.0
Standard	5.17	4.64	-0.53	-10.2
Kappa Scores				
B	0.17	0.16	-0.01	-5.9
D	0.15	0.15	0.00	0.0
F	0.17	0.17	0.00	0.0
H	0.18	0.23	0.05	27.8
I	0.20	0.20	0.00	0.0
J	0.14	0.11	-0.03	-21.4
K	0.14	0.14	0.00	0.0
L	0.22	0.22	0.00	0.0
M	0.23	0.26	0.03	13.0
N	0.20	0.21	0.01	5.0
O	0.23	0.21	-0.02	-8.7
P	0.22	0.20	-0.02	-9.1
Q	0.18	0.15	-0.03	-16.7
Standard	0.22	0.26	0.04	18.2

Table 5.4 – Root mean square errors (RMSE) for the sea surface temperature (SST - ℃) and surface air temperature (SAT ℃) data-model comparisons and the Kappa statistic for the vegetation biome data-model comparison.

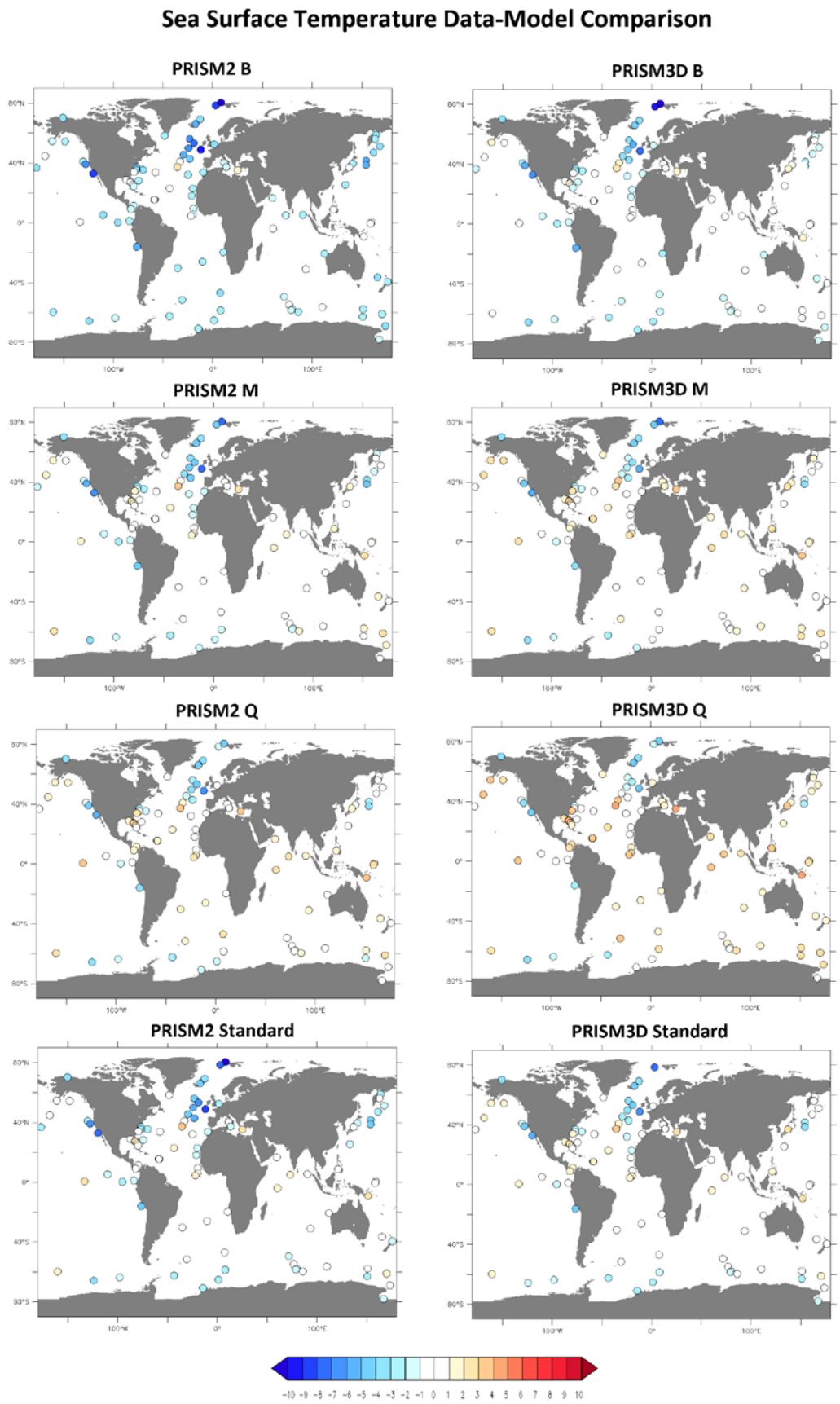


Figure 5.9 – Global distribution of site-by-site sea surface temperature (SST - °C) data-model comparisons for ensemble members B, M, Q & Standard for the PRISM2 and PRISM3D ensembles. The remaining PRISM2 members are shown in Appendix B8 and PRISM3D members in Appendix A4.

The site-by-site SST DMC performance between the PRISM2 and PRISM3D Standards (Figure 5.9) indicates where the improvements occur in the DMC between the two boundary condition sets. The improvements (in PRISM3D compared to PRISM2) are primarily at sites with existing large data-model mismatches, with minimal change through the tropics and Southern Hemisphere sites in both comparisons. The improvements are through the north west Pacific sites south of Siberia (1 to 2°C improvement), the sites off the California coast (2 to 3°C improvement) and the corridor of sites through the North Atlantic (2 to 4°C improvement). However, even with the improvements in these sites, the Standard version of HadCM3 is still displaying up to 7°C mismatch at the very high latitude sites and the 3 to 6°C through the North Atlantic.

As with the SSTs, the SAT DMC indicates improvements (in PRISM3D compared to PRISM2) are observed in high latitude sites, particularly across Eurasia, with some sites displaying a 4 to 7°C improvement. In contrast to the SSTs, where primarily the minimal tropical data-model mismatch was consistent when comparing both boundary condition sets, sites in the tropics show a weakened data-model comparison in the PRISM3D ensemble to the PRISM2 ensemble, a site in Florida displays an 8°C increase in the PRISM3D ensemble, however a number of small (~1°C) improvements in the DMCs across Western Europe across the region with the highest concentration of sites, aid the overall improvement observed in the PRISM3D Standard compared to the PRISM2.

The PRISM3D Standard displays increases in SATs and precipitation compared to PRISM2 (Table 5.1) and these combine to improve the overall representation of vegetation biomes. Improvements are mainly increased high latitude forest extent, increased tropical forest cover and reduced extent of the Sahara desert, in line with the palaeo-data. Overall, the data-model comparisons indicate that the PRISM3D Standard has produced an improved simulation of the mPWP compared to the PRISM2 Standard. The improvement has been seen in the representation of high latitude sites in both the temperature based DMCs (using SSTs and SATs) and improved vegetation reconstructions through high northern latitudes which aid the biome comparison. Primarily the improvements in vegetation between PRISM2 and PRISM3D are through increased high latitude forest extent and decreased extent of the Sahara desert.

b) Ensemble Member Data-Model Comparisons

The intra-model comparisons comparing the boundary condition ensembles to each other and the modern Standard highlighted two types of response. Whilst the direct

responses are consistent in terms of spatial distribution and magnitude, the indirect responses vary in both respects across the ensemble. The variation observed could have an impact on the DMCs and therefore assessing the performance of each ensemble member is required to determine the full range of effects of changing boundary conditions.

For the SST DMC, based on the RMSE scores (Table 5.4) each member of the PRISM3D ensemble performs better than its PRISM2 equivalent, with the exception of member Q. Improvements in the RMSE scores vary across the ensemble from a 19% improvement (member B) to just a 1.9% improvement in member P. Generally, the 'colder than Standard' members show a greater improvement (13.5% to 19% improvement) than the 'warmer than Standard' members (1.9% to 12.8% improvement). The PRISM3D member Q has a 9.9% deterioration in its DMC compared to the PRISM2 ensemble. As with the SST DMC, the SAT DMC generally indicates that PRISM3D ensemble members perform stronger than their PRISM2 equivalents with improvements from 1.7 to 13%. Member J performs worse in the PRISM3D ensemble showing a 1.6% decrease in its RMSE score, whilst member Q indicates no difference in performance between the boundary condition sets. To test the robustness of these RMSE scores, the comparisons were replicated using a change in the averaging period (Chapter 2, Section 2.5.3). The change in averaging period led to small changes in the RMSE comparisons between PRISM3D and PRISM2 across the ensemble, ranging from 0.114 to 0.169. Only members M (0.11) & P (0.05) display differences between PRISM3D and PRISM2 RMSE scores that are smaller than the impact of the using a new averaging period for the same comparison. A similar difference range exists for the SATs (0.18 to 0.30), meaning that all ensemble members bar B (-0.1), D (-0.18), J (0.09) & Q (0.0) are robust. As the majority of the ensemble members are not affected by the change in averaging period, we can be confident that the RMSE rankings represent changes in the performance of the ensemble members SST and SAT data-model comparisons and therefore using them for the ranking of ensemble members and the assessment of the two boundary condition types is acceptable.

For the vegetation DMC, the Kappa scores are more mixed than the temperature based DMCs. Only members H. M & N show an improvement in PRISM3D compared to PRISM2 with members D, F, I, K & L indicating no change between the PRISM3D and PRISM2 ensembles from the Kappa scores. Members B, J, O, P & Q indicate that the PRISM2 ensemble members are stronger than the PRISM3D members. The vegetation biomes DMC is the only DMC were the Standard performs strongly within each

ensemble, with the PRISM3D Standard is joint top of the PRISM3D ensemble biome rankings with member M (Table 5.4).

For the SST DMC, the improvement is due to warming in the SSTs (Figure 5.6) throughout the PRISM3D ensemble, particularly in the 'warmer than Standard' members. The existing data-model mismatch (i.e. Dowsett et al., 2012) is concentrated at the high latitudes, with minimal mismatch at the tropics. The increased temperatures in the PRISM3D ensemble compared to the PRISM2 ensemble (Table 5.1, Figures 5.3, 5.4 & 5.6) causes a greater reduction in the large high latitude data-model mismatches across the PRISM3D ensemble members. The largest reductions being observed in members P & Q. However, the warming in member Q causes an increase in tropical data-model mismatches which weakens the DMC RMSE score for this member. Analysis of the polar amplification between PRISM2 and PRISM3D ensemble members, highlighted that the polar amplification ratio was reduced in the PRISM3D ensemble (Table 5.3). The increased high latitude warmth through 'warmer than Standard' members improves the DMC through the North Atlantic sites across the ensemble members in the PRISM3D ensemble. However the resulting increase in the mean annual temperatures, primarily through tropical latitudes results in a weakening of the DMC in the 'warmer than Standard' members. For member Q, despite the high latitude North Atlantic sites being reduced by 1 to 4°C compared to the PRISM2 member Q, the warming throughout the tropics weakens the DMC overall for PRISM3D member Q. These results highlight the importance of both the polar amplification ratio and the magnitude of polar warming. For the majority of the PRISM3D ensemble members, the greater polar warming, despite the lower polar amplification ratios compared to the PRISM2 ensemble results in a stronger data-model comparison to the SSTs. Only member Q in the PRISM3D ensemble goes against this trend. The conclusion from this result, is that for the mPWP SST data-model mismatch it is more important to generate higher latitude warming as opposed to a higher ratio of polar to global mean warming. However, the results also indicate that there is a balance to be struck with ensuring that tropical temperatures do not increase too much.

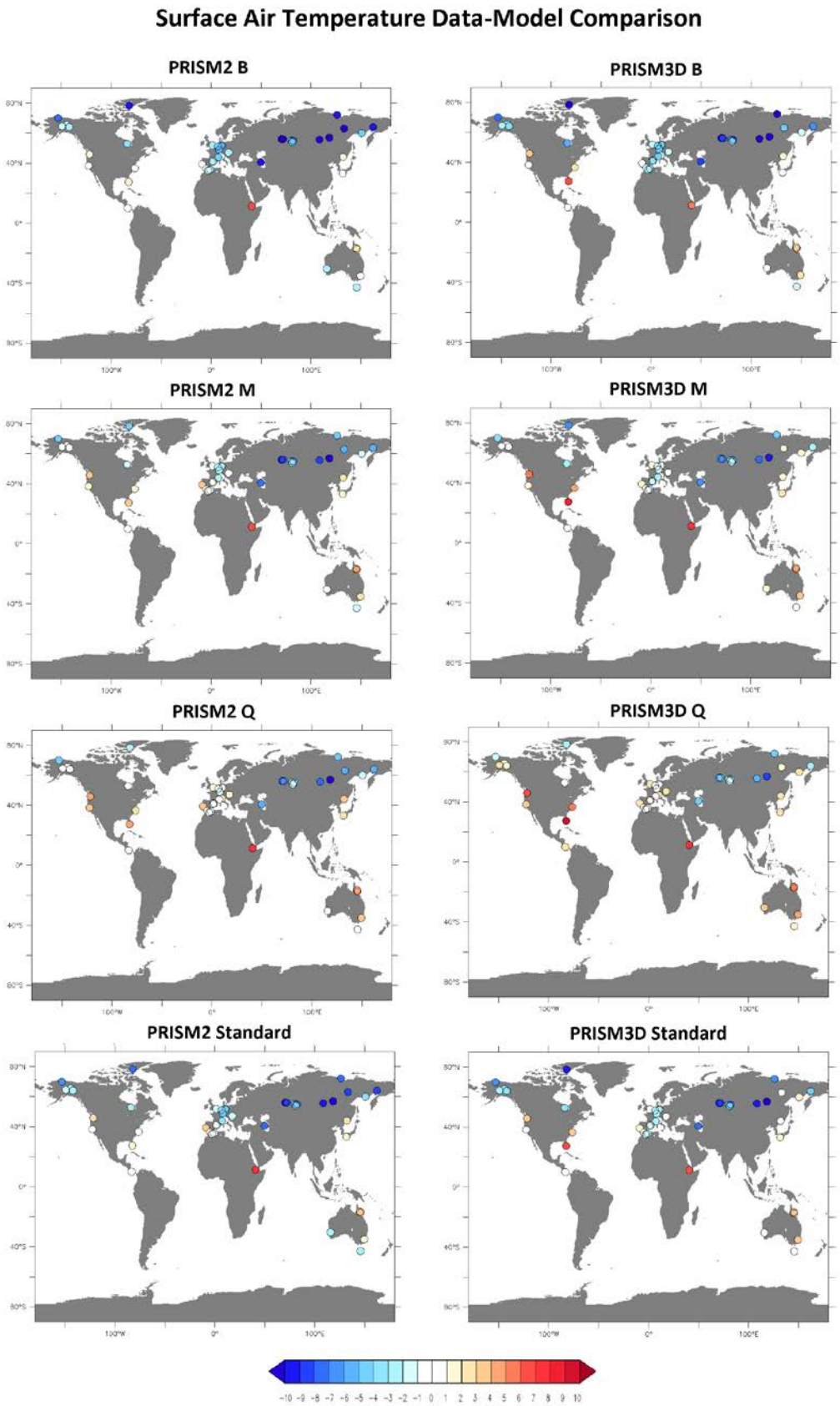


Figure 5.10 – Global distribution of site-by-site surface air temperature (SAT - °C) data-model comparisons for ensemble members B, M, Q & Standard for the PRISM2 and PRISM3D ensembles. The remaining PRISM2 members are shown in Appendix B9 and PRISM3D members in Appendix A5.

As with the SST DMC, high latitude sites in the SAT DMC show the greatest improvement through the 'warmer than Standard' members of the PRISM3D ensemble. For the SATs this is through reductions in the mismatches over sites in western Europe and eastern Siberia (Figure 5.10). Despite these improvements a number of Eurasian sites still display mismatches greater than 10°C (Figure 5.10). In the IMCs between PRISM2 and PRISM3D ensemble member SATs (Figure 5.4), a noticeable affect of the boundary condition changes was cooling over the Arctic Ocean and Eurasia. The location of the data points is along the southern extent of the terrestrial cooling and this location reduces the impact of the cooling in the PRISM3D ensemble relative to the PRISM2 ensemble for the DMC. However evidence of the SAT trends observed over Eurasia is observed in some of the DMCs. PRISM3D ensemble member J displays a weaker RMSE score than its PRISM2 equivalent with intense Arctic and Siberian cooling (Figure 5.4) resulting in a far weaker performance in the PRISM3D ensemble across the Eurasian sites (Figure 5.10). Meanwhile 'warmer than Standard' members, with Eurasian SATs comparable or warmer than their PRISM2 equivalents, reduce the mismatch over Eurasia and this improves their RMSE score for the PRISM3D compared to the PRISM2 ensemble. PRISM3D member Q performs has the same RMSE score as its PRISM2 equivalent. Member Q in the SATs much as in the SSTs, demonstrates a much stronger reconstruction of the high latitude sites compared to its PRISM2 equivalent. However, mid and lower latitude sites show larger mismatches due to the warming in the PRISM3D ensemble members. The high latitude improvements compared to its PRISM2 equivalent are cancelled out in the RMSE score by the weakening of tropical mismatches and results in an RMSE score that is the same for both.

For the vegetation biome DMC, the PRISM3D Standard outperforms both its PRISM2 equivalent and all but member M of the PRISM3D ensemble. As discussed in analysis of the Standards, the PRISM3D Standard increases high latitude forest biomes and reduces the extent of the Sahara desert compared to PRISM2.

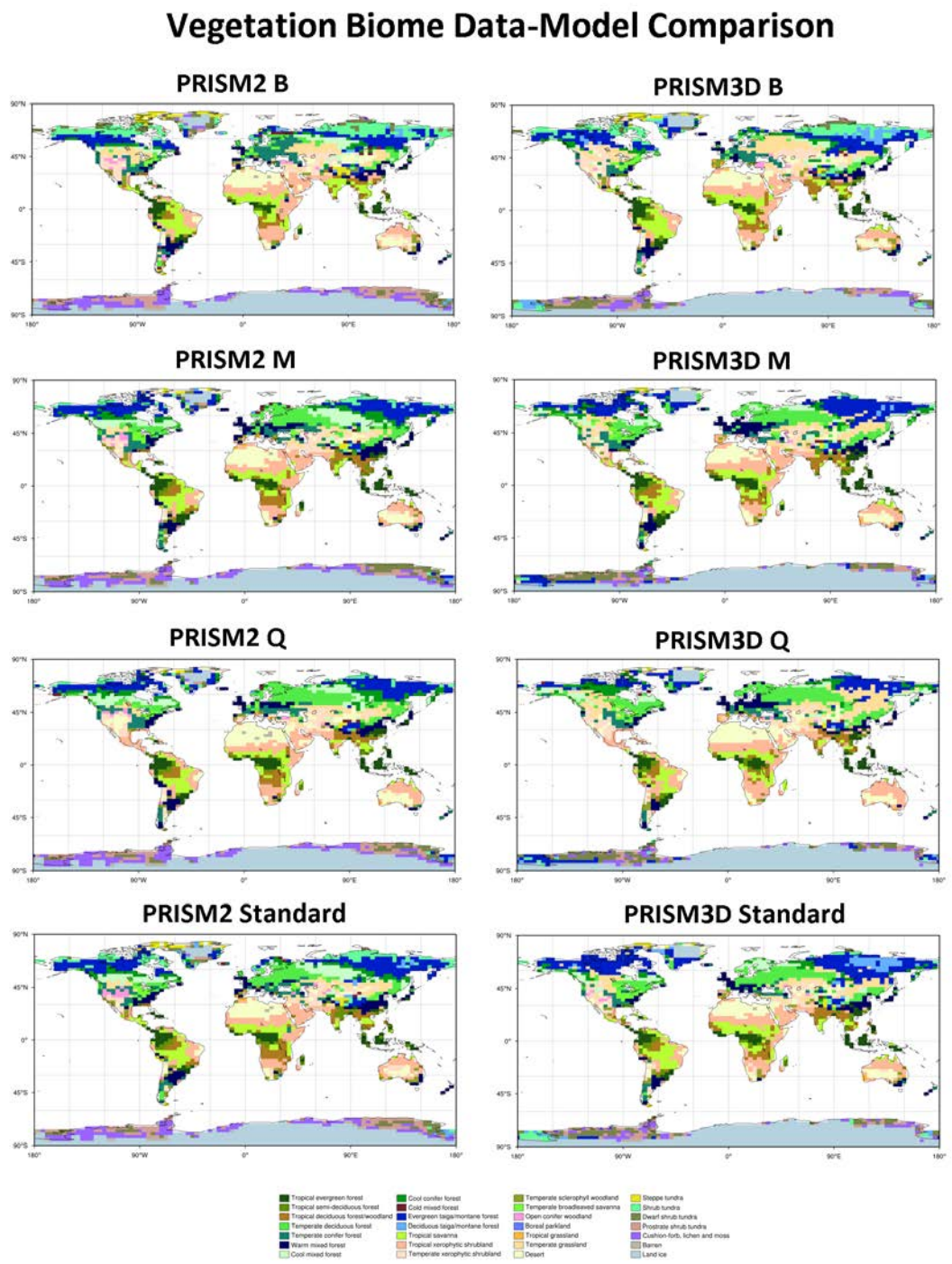


Figure 5.11 – Global distribution of simulated biomes for ensemble members B, M & the Standard for the PRISM2 and PRISM3D ensembles.

For the ensemble members also performing stronger than their PRISM2 equivalent (H, M & N), improvements are again focussed on the extent of high latitude forest in Siberia, the representation of Canadian shrub and grasslands and extended extent of tropical forest biomes in Amazonia and central Africa. The ensemble members showing no change between PRISM2 and PRISM3D (D, F, I, K & L) maintain the high latitude improvements observed in members H, M & N, but produce weaker comparisons through the tropical latitudes than their PRISM2 equivalent. Members H,

M & N in comparison with the modern Standard, are wetter in the PRISM3D ensemble through the tropical latitudes whereas members D, F, I, K & L display similar patterns of precipitation between PRISM2 and PRISM3D (Figures 5.5 & 5.8). Through the tropical latitudes the increased temperatures in the PRISM3D ensemble members (Figures 5.4 & 5.7) results in the drier biomes observed in D, F, I, K & L as the climate is warmer but not wetter. H, M & N display a better warmer and wetter climate balance and as a result simulate a stronger vegetation for the DMC in PRISM3D over PRISM2 (Figure 5.11). Through the high latitudes the similar or slightly cooler temperatures in the PRISM3D ensemble compared to the PRISM2 ensemble with similar precipitation results in enhanced simulated forest distribution for the PRISM3D ensemble members and improved high latitude DMC. These changes combine to result in the ensemble members improving the DMC in PRISM3D (H, M & N) or showing no change (D, F, I, K & L). The members displaying a weakened DMC in the PRISM3D ensemble compared to PRISM2 (B, J, O, P & Q) achieve a similar performance to the no change members through the mid-latitudes. Increased SATs and similar or decreased precipitation through the tropics compared to PRISM2 result in smaller extent of tropical forest biomes and expansion of xerophytic shrubland and desert biomes. At high latitudes there is reduced extent of high latitude forest biomes across Siberia and a reduction in forest through Western Europe. The forest is replaced by grass and shrubland, drier biomes reflecting reduced precipitation. Variation across these members in high latitude SATs influences the type of grassland, however for the Kappa statistic, the simulated biome is either right or wrong.

5.3.2. Zonal Mean Annual Data-Model Comparisons

Zonal mean annual SST DMCs have been produced for members B, M & Q and the Standard for both PRISM2 and PRISM3D boundary condition ensembles. B & Q represent end members within the ensemble with respect to temperature performance, B having the coldest mean annual temperature and Q the warmest (Table 5.1), whilst member M is a consistently high ranking member in the site-by-site DMCs. Two DMCs are plotted, one a global mean annual SST across all ocean basins and a North Atlantic basin only comparison. Versions of these DMCs were discussed in Chapter 4 (Section 4.4), which concluded that the PRISM3D ensemble members were able to achieve the magnitude of warming observed in the palaeo-data, but were unable to replicate the spatial distribution of this warming, especially through northern high latitudes and particularly in the North Atlantic only comparison.

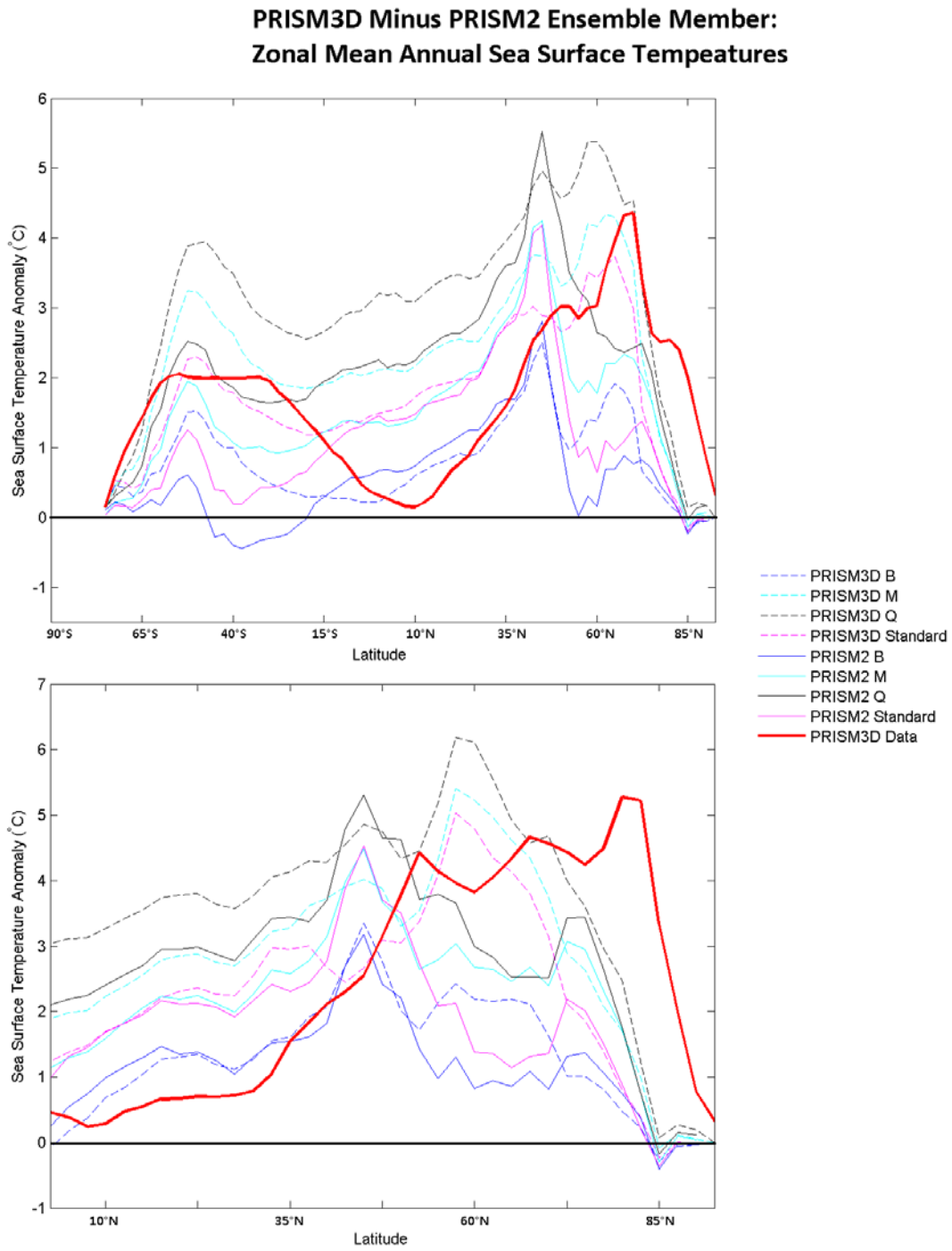


Figure 5.12 – Zonal annual mean sea surface temperatures for a) global oceans and b) North Atlantic ocean for ensemble members B, M, Q and the Standard for the PRISM2 (solid line) and PRISM3D (dashed line). The thick red line represents the palaeo-data.

The PRISM3D ensemble members are warmer than their PRISM2 equivalents in the Southern Ocean (Figure 5.12), the warming results in the PRISM3D Standard SST being comparable to the palaeo-data. The PRISM2 members M & Q are also comparable to the palaeo-data in the Southern Ocean, bracketing the palaeo-data, but the warming in the PRISM3D ensemble results in a 1 to 2°C mismatch for both members M & Q. Moving towards the equator, both ensembles are unable to replicate the trend of the palaeo-data. Differences between the two ensembles are minimal for member B and

the Standard, but quite pronounced for members M & Q. The large mismatch through the tropics, especially in member Q highlights the reason the PRISM2 member Q was stronger than its PRISM3D equivalent in the SST RMSE scores. Despite stronger high latitude performance in the PRISM3D member Q, the increased tropical mismatch weakens the DMC both to other PRISM3D ensemble members and its PRISM2 equivalent. Member M also shows a warming in the PRISM3D member compared to the PRISM2 member, but it is less pronounced than in member Q. Through the Northern Hemisphere, for both the global and North Atlantic comparisons, the ensemble members continue to remain warmer than the data till around 60°N, from where the data becomes much warmer than all the ensemble members. However, both M & Q for both ensembles along with the PRISM3D Standard are able to match the magnitude of the peak warming in the data. In both the global and North Atlantic comparisons, the PRISM3D ensemble members peak further north than the PRISM2 members, which is more in line with the palaeo-data reconstruction. The PRISM2 ensemble weakens the spatial distribution of warming in the zonal means compared to the PRISM3D ensemble.

The zonal mean DMCs reinforce the results of the site-by-site DMC, PRISM3D members perform stronger at high latitudes than their PRISM2 equivalents. PRISM3D members are also warmer through the tropics, but with the exception of member Q, the improvement in North Atlantic data-model mismatches results in stronger performing ensemble members based on the RMSE scores for PRISM3D over PRISM2 ensembles.

5.4. Discussion

5.4.1. Which boundary condition set produces the strongest mPWP simulations?

The PRISM3D palaeo-environmental reconstruction represents the most sophisticated integrated geological realisation of the mid-Pliocene Warm Period, a period of past global warmth (Dowsett et al., 2010a). However, it is trivial to assume that improving the physical boundary conditions will improve model simulations of mPWP climate. Changes to the physical boundary conditions such as the topography, vegetation and ice sheets will interact with the model components and parameters. However, these interactions are not guaranteed to result in improved model simulations of the mPWP solely because the boundary conditions are more geologically accurate. To date, no work has been done to investigate the differences between the two sets of boundary conditions with respect to simulating the mPWP. The climatological differences between the two boundary condition sets, for both the ensemble Standard and the PPE

were assessed using intra-model comparisons (IMCs - Figures 5.3 to 5.8, Tables 5.1 to 5.3). Additionally, through running a suite of data-model comparisons (DMCs - Figures 5.9 to 5.12, Table 5.4) enables an investigation of the impact of the changes on the performance of the model for simulating the climate of the mPWP. The HadCM3 Standard IMCs indicate that the PRISM3D boundary conditions generated a globally warmer and wetter representation of the mPWP climate than the PRISM2 boundary conditions. In each of the three site-by-site DMCs, the PRISM3D Standard ranks above the PRISM2 Standard (Table 5.5). The greatest improvement is in the vegetation reconstruction (18% improvement), followed by the SST (14%) and SAT reconstructions (10%). Based on the ensemble Standards, the PRISM3D boundary conditions represent an improvement in the modelling of the mPWP.

Member	SST Rank	Member	SAT Rank	Member	Kappa Rank	Member	Total Ranking	Average Ranking
N_PRISM3D	1	P_PRISM3D	1	M_PRISM3D	1	M_PRISM3D	9	3.00
L_PRISM3D	2	N_PRISM3D	2	Standard_ PRISM3D	1	N_PRISM3D	13	4.33
M_PRISM3D	3	Q_PRISM2	3	H_PRISM3D	3	L_PRISM3D	14	4.67
K_PRISM3D	4	Q_PRISM3D	3	M_PRISM2	3	P_PRISM3D	19	6.33
O_PRISM3D	4	M_PRISM3D	5	O_PRISM2	3	O_PRISM3D	23	7.67
F_PRISM3D	6	L_PRISM3D	6	L_PRISM2	6	M_PRISM2	24	8.00
I_PRISM3D	6	I_PRISM3D	7	L_PRISM3D	6	Standard_ PRISM3D	24	8.00
P_PRISM3D	6	P_PRISM2	8	P_PRISM2	6	I_PRISM3D	25	8.33
H_PRISM3D	9	M_PRISM2	9	Standard_ PRISM2	6	P_PRISM2	25	8.33
Standard_ PRISM3D	10	O_PRISM3D	9	N_PRISM3D	10	H_PRISM3D	29	9.67
P_PRISM2	11	N_PRISM2	11	O_PRISM3D	10	Q_PRISM2	31	10.33
M_PRISM2	12	L_PRISM2	12	I_PRISM2	12	L_PRISM2	34	11.33
Q_PRISM2	12	Standard_ PRISM3D	13	I_PRISM3D	12	O_PRISM2	35	11.67
J_PRISM3D	14	K_PRISM3D	14	N_PRISM2	12	N_PRISM2	37	12.33
N_PRISM2	14	I_PRISM2	15	P_PRISM3D	12	F_PRISM3D	42	14.00
I_PRISM2	16	O_PRISM2	16	H_PRISM2	16	I_PRISM2	43	14.33
L_PRISM2	16	H_PRISM3D	17	Q_PRISM2	16	K_PRISM3D	43	14.33
O_PRISM2	16	F_PRISM3D	18	B_PRISM2	18	Q_PRISM3D	47	15.67
K_PRISM2	19	Standard_ PRISM2	19	F_PRISM2	18	Standard_ PRISM2	48	16.00
D_PRISM3D	20	J_PRISM2	20	F_PRISM3D	18	D_PRISM3D	63	21.00
B_PRISM3D	21	D_PRISM3D	21	B_PRISM3D	21	J_PRISM3D	65	21.67
Q_PRISM3D	22	K_PRISM2	22	D_PRISM2	22	F_PRISM2	66	22.00
Standard_ PRISM2	23	J_PRISM3D	23	D_PRISM3D	22	K_PRISM2	66	22.00
F_PRISM2	24	F_PRISM2	24	Q_PRISM3D	22	B_PRISM3D	68	22.67
H_PRISM2	25	D_PRISM2	25	J_PRISM2	25	H_PRISM2	68	22.67
J_PRISM2	26	B_PRISM3D	26	K_PRISM2	25	J_PRISM2	71	23.67
D_PRISM2	27	H_PRISM2	27	K_PRISM3D	25	B_PRISM2	74	24.67
B_PRISM2	28	B_PRISM2	28	J_PRISM3D	28	D_PRISM2	74	24.67

Table 5.5 – The combined rankings of all the ensemble members for the three data-model comparisons (SATs, SSTs and Kappa scores). Each ensemble member is given a ranking for individual DMCs and then the combined ranking is produced based on each ensemble members performance in these three comparisons.

Based on the IMCs and DMCs, is it possible to determine whether the PRISM3D boundary conditions aid a climate model to produce improved simulations of the mPWP? Certainly, based on the ensemble Standards, that is correct, across all three DMC metrics produced here the PRISM3D Standard is an improvement on the PRISM2 Standard. However, is the performance of this one version of the model acceptable? Chapter 4 (Section 4.4) determined that ensemble members I, L, M, N, O & P represented better simulations of the mPWP than the ensemble Standard based on their combined RMSE rankings. Additionally, as well as direct responses to the changes in boundary conditions, a number of indirect features were observed and these vary across the ensemble, as a result of interactions with ensemble members and the parameterisations within them. The impact of changing from PRISM2 to PRISM3D boundary conditions was the same across each ensemble member (0.41°C and 0.024 mm/day). However the effect of the perturbed parameters interacting with the boundary condition changes displayed a range of -0.52 to 0.62°C and -0.023 to 0.094 mm/day (Table 5.2). These indirect responses represent the confluence of boundary condition and parameter uncertainty within these studies. As a result, nearly all PRISM3D ensemble members improved their DMC scores with respect to the PRISM2 equivalent (Table 5.4). Across the SST and SAT DMCs the comparison between the two boundary condition types is relatively simple, with only PRISM3D members Q (SSTs) and J (SATs) performing worse than their PRISM2 equivalents. Based on the results of using a different averaging period, it is possible to be confident that differences between PRISM3D and PRISM2 RMSE scores are significant, with only the intra-ensemble differences between M & P for the SSTs and B, J & Q for the SATs showing less variation than that observed for changing the averaging period used. For the vegetation biome DMC, it is a less clear picture, with PRISM3D members B, J, O, P & Q performed worse than their PRISM2 equivalents with the rest of the PRISM3D ensemble matching their PRISM2 equivalent, bar members H, M, N & the Standard which improve the DMC. Two of the DMCs would strongly indicate that the PRISM3D boundary conditions are an improvement, but one comparison suggests they are a small change.

Therefore, based on the investigation of the parameter uncertainty and model performance, it is vital to ascertain whether the PRISM3D boundary conditions are definitely leading to improved representations of the mPWP through assessment of their performance across the ensemble. For each of the site-by-site DMCs a ranking across both ensembles was created. The ranking was generated for each individual DMC, and then the combined unweighted ranking was generated from the cumulative

ranking score. It was not possible to base this final ranking on the RMSE or Kappa scores as this would give too much weighting to one DMC over another. At present there is no quantified weighting on the relative performance of each of these three metrics (SST, SAT and vegetation biomes) which would allow this comparison to be achieved. These rankings of the 28 simulations can be used to assess the performance of both boundary condition sets. The ranking also highlights the strongest performing ensemble member when considering both of parameter and boundary condition uncertainties.

Table 5.5 indicates that the PRISM3D boundary conditions produce stronger model versions for the mPWP in comparison to the available palaeo-data. With the exception of PRISM3D member Q, every ensemble member and the Standard ranks higher than their PRISM2 equivalent when the three DMC rankings are combined. Ensemble member M from the PRISM3D ensemble is the strongest performing member when the ranking scores for each DMC are combined (Table 5.5). Five ensemble members (L, M, N, O & P) outperform the PRISM3D Standard and the rest of the ensemble. Consistent with Chapter 4, these ensemble members had previously been determined to represent an improved simulation of the mPWP. Based on these results, it is possible to confidently conclude that the PRISM3D boundary conditions lead to improved simulations of the mPWP.

5.4.2. What are the Implications of this Result for mPWP Modelling?

The results from the investigation of the two physical boundary condition sets indicate that the PRISM3D boundary conditions produce a stronger representation of mPWP climate. For this reason it is recommended that future mPWP modelling studies only use the PRISM3D boundary conditions. To date, the majority of mPWP modelling has used the PRISM2 physical boundary conditions (i.e. Haywood & Valdes 2004; Lunt et al., 2008; 2010). The PlioMIP experimental designs (Haywood et al., 2010; 2011a) decided that the PlioMIP project would use the PRISM3D physical boundary conditions, leading to a transitional period with mPWP modelling papers using both the PRISM2 boundary conditions (i.e. Dolan et al., 2011; Pope et al., 2011; Lunt et al., 2012a) and the PRISM3D conditions (i.e. Bragg et al., 2012; Haywood et al., 2013a; Howell et al., 2014). As stated, the results indicate that to produce the best simulations of the mPWP climate, all modelling studies should use the PRISM3D boundary conditions. However, Lunt et al. (2012a) outlined some advantages for remaining with the PRISM2 boundary conditions, primarily:

- Consistency with previous mPWP modelling studies

- More developed model simulations with longer spin up times (> 1000 years).

Whilst the move to PRISM3D boundary conditions creates a break between earlier work using PRISM2 boundary conditions, the evidence that using PRISM3D produces stronger, more realistic simulations of mPWP climate, means that this break is an important change to make, the best tools should be used simulating the mPWP climate. Additionally, the developing research into the importance of the orbital forcing for representing mPWP climate, and the especially on the palaeo-data reconstructions, will see a movement towards using set orbital configurations for time slices with in the mPWP instead of the existing time slab (Haywood et al., 2013b; Prescott et al., 2014). Whilst the change from PRISM2 to PRISM3D boundary conditions creates the break between previous mPWP and future work with PRISM3D, the changes to orbital parameters also creates a break from the previous mPWP modelling. Therefore, it is the logical time to fully switch from modelling with PRISM2 to PRISM3D boundary conditions.

Further changes to boundary conditions are in development such as the reconstruction of mPWP soils and lakes (Pound et al., 2013). Refinements on techniques to develop understanding of mPWP ice sheets, sea level and other boundary conditions are future developments to the mPWP boundary conditions which will likely exert an influence on the simulated mPWP climate. Whilst work to date has revealed regional impacts due to changes in mPWP lakes (Contoux et al., 2013; Pound et al., 2013), the work undertaken here demonstrates the importance of assessing the influence of changes in boundary conditions on model performance. It is recommended that future changes to boundary conditions for the mPWP are tested using a range of models. Whilst the change from PRISM2 to PRISM3D boundary conditions had a uniform warming (0.41°C) and precipitation increase (0.024 mm/day) their was a wide range of variation across the ensemble members as a result of the boundary condition change interacting with the different parameter sets (Table 5.2). The variation represents the confluence of boundary condition and modelling uncertainty and it is vital to fully understand the range of impacts from changing boundary conditions on simulations of the mPWP climate. For some ensemble members, the change to PRISM3D boundary conditions resulted in a cooler and drier simulation of the mPWP when the boundary condition constant response was eliminated, whilst other ensemble members showed warmer and wetter climates that were dominated by the response of the perturbed parameters.

Therefore, to ensure that the impacts of boundary condition changes on simulating the mPWP, the assessment of new boundary conditions should utilise modelling

ensembles. As has been shown in this PPE, there are two effects from changing boundary conditions. Whilst a single model assessment will include direct effects of the changes, the indirect effects (which for this ensemble were sometimes in excess of twice the direct effects) will vary depending on model architecture. It is only through assessing the changes to boundary conditions across an ensemble of models that the full range of effects from these changes can be observed and assessed. For both PPEs and MMEs the response of a model to the change in the boundary conditions will vary and therefore future boundary conditions need to use a suite of models to assess the changes they cause.

5.4.3. What is the role of boundary condition uncertainty in mPWP data-model mismatches?

As discussed above, the changes in the physical boundary conditions between PRISM2 and PRISM3D have exerted an influence both through direct effects and indirect effects on the simulated climate of the mPWP. As a result these changes have led to improvements in the data-model comparisons to three sets of palaeo-data. The *"PMIP Triangle"* (Haywood et al.,. 2013a) outlines the three main causes of data-model mismatches in palaeoclimate studies, modelling, boundary condition and data uncertainties. The change from PRISM2 to PRISM3D and the improvements that have resulted reflect the importance of reducing boundary condition uncertainty. These physical boundary condition changes have played a noticeable role in improving the modelling of the mPWP. However, a range of other boundary condition uncertainties exist with respect to the mPWP. Other potential physical boundary condition uncertainty the recent developments to the reconstruction of Pliocene soils and lakes (Contoux et al., 2013; Pound et al., 2013) and future developments in reconstructing ice sheets and sea level. Similarly, as discussed in Chapter 4, forcing boundary conditions such as the greenhouse gases and orbital forcings (Haywood et al., 2013b; Prescott et al., 2014) could influence the performance of mPWP simulations.

The development of an mPWP soil map and lake distribution acts as one of the future developments to the PRISM3D physical boundary conditions. Pound et al. (2013) highlighted however that primarily these developments had regional influences, a conclusion supported by work from Contoux et al. (2013) looking at megalake Chad for the mPWP and the Holocene. However, the location of these changes led to some important changes to the simulation of the mPWP such as along the southern border of the Sahara desert or through western North America, regions presently simulated as too dry compared to the palaeo-data.

The development of the Pliocene time slices (Haywood et al., 2013b; Prescott et al., 2014) has shown that the selection of orbital parameters can exert strong forcing upon the simulation of high latitude temperatures greater than any changes observed in mean annual temperatures. Both ensembles produced for this investigation have utilised a modern orbital forcing solution. Haywood et al. (2013b) investigating the role of orbital forcing on the modelling of the mPWP highlighted that choosing specific orbits from within the mPWP had an impact on the regional responses of the model. Choosing orbital conditions with a top of atmosphere radiative forcing the same as modern, but with modification of the orbital parameters (obliquity, eccentricity, and precession) could result in regional changes such as 2 to 4°C warming through the North Atlantic. The investigations into the orbital boundary conditions are indicating that they will also improve the reconciliation of data and models for the mPWP.

However, reducing the boundary condition uncertainty through improved representation of the physical characteristics or forcing characteristics of the mPWP is not synonymous with improvements in the modelling. Simulations across the range of potential values for atmospheric CO_2 through the mPWP using the ensemble (Chapter 4) have show that values less than 400 ppmv lead to weaker data-model comparisons, with data-model mismatches increased through high latitudes. The simulations forced with lower concentrations of CO_2 are unable to generate the high latitude warming compared to the standard value of 400 ppmv.

In summary, the improved representation of the physical boundary conditions from PRISM2 to PRISM3D has resulted in stronger simulations of the mPWP. Additionally, future improvements in the boundary conditions such as the inclusion of soils and lakes and the development of specific orbital time slices through the mPWP are displaying the potential to lead to further beneficial changes in the mPWP simulations. However, should improvements in proxies for atmospheric CO_2 indicate a lower than 400 ppmv value for mPWP CO_2 then the ensemble has indicated that this would reduce the strength of mPWP simulations. Should atmospheric CO_2 be lower than 350 ppmv, it will be increasingly hard to generate enough high latitude warming to further improve data-model mismatches in the North Atlantic. It is evident that boundary condition uncertainties have and will continue to impact on the data-model mismatches for the mPWP. Reducing the uncertainty around ice sheets, orography and vegetation in the change from PRISM2 to PRISM3D boundary conditions has resulted in improved simulation of the mPWP and DMC rankings for this PPE.

In circumstances with improved boundary conditions weakening the simulation of the mPWP, the community is left with the decision of whether to include these boundary conditions or not. If the aim of a modelling study is to reduce the uncertainty in data-model mismatches, then it is vital that the experimental design includes the best boundary conditions, even if this leads to weakened simulations with respect to the palaeo-data. To not minimise one region of uncertainty reduces the ability of a study to investigate other regions of uncertainty. By reducing boundary condition uncertainties, potential causes of data-model mismatches can be eliminated and the results highlight regions which require further work and development. Continuing to reduce boundary condition uncertainty, surrounding orbital forcing, greenhouse gases, ice sheets and the land surface will reduce the role of boundary condition uncertainty within the PMIP Triangle and increase the understanding of simulating warmer than modern climates, even if the change results in a weaker simulation in comparison to the palaeo-data. It is unlikely that boundary condition uncertainties can be used to resolve all data-model mismatches for the mPWP, however they offer potential to reduce the role that they play in future data-model comparisons for the mPWP.

5.5. Conclusions

A large number of changes have occurred between the PRISM2 (Dowsett et al., 1999) and PRISM3D (Dowsett 2007; Dowsett et al., 2010a) palaeo-environmental reconstructions. However, prior to this work, no assessment of the performance of climate models forced with these new physical boundary condition changes has been undertaken.

The change to the PRISM3D boundary conditions from PRISM2 resulted in both direct and indirect effects on the simulations of the mPWP. Every member of the ensemble responded with an increase in temperature of 0.42°C and increase in precipitation of 0.024 mm/day. However, the interaction between the perturbed parameters within the ensemble members resulted in a range of temperatures (-0.53 to 0.62°C) and precipitation (-0.023 to 0.094 mm/day) across the ensemble. As a result the PRISM3D ensemble members were warmer than their PRISM2 equivalents (except member B) and wetter (except members J & K). For the data-model comparisons the warmer climate in the PRISM3D ensemble members resulted in improved high latitude data-model comparisons. However, the PRISM3D members were also warmer through tropical latitudes weakening the DMC through this region compared to the PRISM2 ensemble. Overall, the high latitude improvements outweighed the weakening in the tropics and PRISM3D members ranked higher that their PRISM2 equivalents for the

SST and SAT DMCs. For the vegetation biomes, the PRISM3D ensemble members improved high latitude DMCs with increased forest biomes compared to PRISM2 members. However, the impact of warmer and drier tropical climates across most of the PRISM3D members compared to PRISM2 resulted in the PRISM3D members performing poorer or the same as their PRISM2 equivalents. Only members H, M, N & the Standard improved the vegetation biome DMC compared to their PRISM2 equivalents.

Overall, it has been possible to determine that the PRISM3D boundary conditions result in an improvement in the simulation of the mPWP. Despite some good arguments to the contrary, it is evident that mPWP modelling studies should switch to using the PRISM3D boundary conditions. However, analysis of the results has indicated that it is important to assess future developments in boundary conditions using ensembles of models, to fully analyse the magnitude and distribution of both direct and indirect effects from these changes. Finally, to fully understand the role of boundary condition uncertainty on data-model mismatches, it is important that when new developments to boundary conditions are developed, they are included, regardless of whether the improve the data-model mismatches or not. It is only through reducing boundary condition uncertainties that an assessment of where uncertainties in models and palaeo-data are influencing the understanding and simulation of the mid-Pliocene Warm Period.

Chapter 6: Conclusions

The preceding chapters have undertaken the testing of a perturbed physics ensemble to investigate parameter and boundary condition uncertainty in model simulations of the mid-Pliocene Warm Period (mPWP: 3.264 to 3.025 Ma BP). Two full ensembles were undertaken each forced with different physical boundary conditions, based on palaeo-environmental reconstructions by the US Geological Survey as part of the Pliocene Research Interpretations and Synoptic Mapping (PRISM) project. Additionally a sub-ensemble was created from the main ensemble to assess the impact of changing the concentration of atmospheric CO_2 from 400 ppmv to either 300 or 350 ppmv. The ensembles were compared to palaeo-data (data-model comparisons – DMCs) in the form of the PRISM3D Mean Annual Sea Surface Temperature (MASST) dataset (Dowsett et al., 2010b, 2012), a terrestrial surface air temperature dataset (Salzmann et al., 2013) and a reconstruction of Piacenzian vegetation biomes (Salzmann et al., 2008). Rankings were produced based the use of Root Mean Square Error (RMSE) and Cohen's Kappa Statistic (Cohen, 1960) for the DMCs. A final DMC was undertaken using zonal mean annual SSTs for the ensemble members and the palaeo-data to provide an alternative comparison to the site-by-site comparisons used previously.

6.1. Significance of this Research

The research described represents a detailed investigation of the effect of parameter and boundary condition uncertainty on model simulations for the mPWP. The perturbed physics ensembles presented are the first investigation of these two forms of uncertainty in a warmer than modern palaeoclimate setting with a close to modern continental configuration. It is also the first PPE to be tested against a range of palaeo-data covering sea surface temperatures, terrestrial surface air temperatures and vegetation biome reconstructions. The work represents a thorough investigation of the impact of parameter uncertainty on existing data-model mismatches for the mPWP.

6.2. Summary of Main Findings

The aim of the thesis has been to use the PPE to investigate the roles of model parameter and boundary condition uncertainty in modelling simulations of the mPWP. Initial results based on an initial ensemble using end members of the full ensemble identified that the ensemble member with higher Charney sensitivitys than the Standard version of HadCM3 improved the comparison to the PRISM SST data, although the Standard version produced

a better representation of the mPWP vegetation biome data (Chapter 3 & Pope et al., 2011).

From these initial results, Chapters 4 & 5 were written based on the full PPEs. Chapter 4 assessed the intra-model performance of the ensemble members, with the final aim of determining whether any members of the ensemble produced a better representation of the mPWP climate than the Standard. Chapter 5 utilised the PPE to assess the impact on mPWP modelling studies due to the developments in the physical boundary conditions used to simulate the mPWP.

The existing data-model mismatch for the mPWP indicates that in both SST and SAT data-model comparisons, model simulations generate appropriate warming in the tropics, but struggle to replicate the warming observed in the high latitudes, especially through the North Atlantic. Ensemble members with Charney sensitivities lower than the Standard simulation, improved tropical data-model comparisons, but weakened the high latitude DMCs. In the ensemble members with Charney sensitivities higher than the Standard, these ensemble members reduced the high latitude data-model mismatch, but weakened the tropical data-model comparison. The results in Chapter 4 highlighted that for the PPE members, the balance between increasing high latitude warming without overheating the tropical temperatures was crucial to improving the DMCs. Three members of the full PPE produced a better simulation of the mPWP climate based on both the SST and SAT data-model comparisons. These members, M (4.54°C), N (4.62°C) & P (5.40°C) had Charney sensitivities higher than the Standard version, and also at or slightly above the upper boundary of the IPCC likely range (1.5 to 4.5°C – IPCC, 2013). However, the ensemble member with the highest Charney sensitivity, member Q (7.11°C) simulated a weaker mPWP climate. Member Q produced the largest reduction in the high latitude data-model mismatch, but unlike members M, N & P, generated a large mismatch through the tropics, which results in a larger RMSE score and a simulation judged to be weaker. The benefits at high latitudes of the warming in member Q were undone by the weakening of the tropical DMC. An idealised data-model comparison, which combined the best site-by-site DMC value from across the ensemble and included uncertainty estimates on the palaeo-data still failed to fully resolve the data-model mismatch.

Zonal mean data-model comparisons produced in Chapter 4 highlighted these weaknesses, but also indicated that the ensemble range brackets the warming seen in the mid-Pliocene. However, none of the ensemble members were able to replicate the spatial distribution of the warming. The inability of the PPE to simulate this pattern of warming, could be down to a number of variables linked to the experimental design, structural uncertainty, poorly

constrained boundary conditions or unaccounted for data uncertainty. However, at present it is not possible to determine the roles of each of these possible factors.

Finally, Chapter 4 used a sub-ensemble to investigate the uncertainty range in mPWP CO_2. The full ensembles are forced with a value of 400 ppmv, which is towards the top end of the range of values for the mPWP. Based on the results from the sub-ensembles, it would be significantly harder to resolve high latitude data-model mismatches should improvements in the palaeo-proxies for CO_2 indicate that mPWP concentrations were lower than 350 ppmv.

The results in Chapter 5 indicated that the PRISM3D boundary conditions represented an improved set of model boundary conditions compared to PRISM2. The result held across a range of ensemble members and for the ensemble Standard. The effect of the boundary condition changes was both direct and indirect, with effects that occurred across all ensemble members with equal magnitude and others that varied as a result of interactions with the model parameters. For the change from PRISM2 to PRISM3D boundary conditions, there was a warming of 0.42°C and an increase in precipitation of 0.024 mm/day as direct effects. However, there was then indirect effects ranging from -0.53°C to 0.62°C additional warming (or cooling) and changes in precipitation by -0.023 to 0.094 mm/day. As a result, all PRISM3D ensemble members were warmer (bar ensemble member B) and wetter (bar members J & K) than their PRISM2 equivalents.

Across the SST and SAT DMCs this warmer and wetter climate improved the PRISM3D data-model comparisons compared to their PRISM2 equivalents, with the exception of member Q in the SSTs, which, due to increased tropical warming in the PRISM3D ensemble weakened its DMC. For the vegetation biome DMC, the balance of warming and precipitation changes was very important, with a less clear signal in the patterns observed between the two ensembles. Only members, H, M, N & the Standard improved the vegetation biome DMC in PRISM3D compared to PRISM2. For a number of ensemble members, improved high latitude biomes, were weakened by simulating a greater extent of dry biomes in the tropics, with tropical shrublands replacing tropical grasslands and forests. However, despite this, the PRISM3D ensemble members outperformed their PRISM2 equivalents when all three DMCs were combined to produce a final ranking, resulting in the conclusion that the PRISM3D boundary conditions represent an improved set of boundary conditions.

Owing to the importance of boundary condition uncertainty within the PMIP Triangle (Haywood et al., 2013a), combined with upcoming changes to the way the mPWP is simulated and the palaeo-data is produced (focussing on the Pliocene time slices –

Haywood et al., 2013b; Prescott et al., 2014), it was concluded that based on these results, it is important to recommend that the Pliocene modelling community move away from mPWP simulations using the PRISM2 boundary conditions. It was also concluded, that to offer the best chance to further reduce the role of boundary condition uncertainty in data-model mismatches, it is crucial to integrate any improvement to the geological boundary conditions, even if it weakens the DMCs. To not exclude one region of modelling uncertainty (say by using less accurate boundary conditions) would make it far harder to reduce other regions of uncertainty in resolving data-model mismatches.

6.3. Developments & Future Work

The key developments to the work in this thesis revolve around two of the four possible explanations given in Chapter 4 for the ensemble to be able to generate the magnitude but not the spatial distribution of Pliocene warming:

- The parameter space is not fully sampled
- Poorly constrained or misrepresented boundary conditions

Chapter 4 discussed these different factors and how they could be involved in the data-model mismatch. The issue of structural uncertainty has been removed as it is addressed through other projects such as the Pliocene Modelling Intercomparison Project (PlioMIP). Similarly, unaccounted for palaeo-data uncertainty, whilst a potential major source of data-model mismatch, will be investigated and estimated independent of modelling studies. Any developments to this PPE should focus primarily on further reducing the parameter component of modelling uncertainty in the PMIP Triangle. This could be achieved through the use of other PPEs such as those based around ocean parameters or the carbon cycle, but it is likely the greatest gains will be made by focussing on other methods to design the atmospheric PPE. For improved analysis of the impact of the perturbed parameters, it is best to focus on smaller parameter sets, targeting specific regions of model weakness. By perturbed 32 parameters simultaneously, it has not been possible to assign the impact of individual or groups of perturbations to model results. In summary, this thesis investigated one section of parameter uncertainty, which has produced interesting and novel results for Pliocene modelling. However, the results have also posed further questions about parameter uncertainty which the PPE can not answer in its present configuration. Alternative PPE designs offer another method for further investigating the uncertainty and building on the results presented here.

The other noted reason for the remaining data-model mismatch is due to the boundary condition uncertainty. Poorly constrained boundary conditions are a focus in many palaeo-climate studies at present, with work investigating soils, lakes, rivers, greenhouse gases and orbital forcings (i.e. Contoux et al., 2013; Pound et al., 2013; Prescott et al., 2014) during a range of palaeoclimate periods, particularly the Pliocene. As discussed in Chapter 5, it is important to assess the effect of both direct and indirect impacts of boundary condition changes. Whilst this can also be achieved by utilising a multi-model ensemble, the ability to run a PPE *"in house"* makes it a simple and effective way to investigate changes in boundary conditions across a range of Charney sensitivities. Additionally, the technical challenges of translating new boundary conditions onto model grids is limited to only one grid configuration, unlike an MME which will feature a range of different grid configurations. The results in this thesis highlighted that some ensemble members may be weakened in comparison to palaeo-data as a response to changes in boundary conditions.

With respect to improving the data-model mismatch, then the combination of these PPE results with the new lakes and soils boundary conditions could result in a useful improvement particularly to vegetation biome DMCs. These changes could benefit the members of the PPE that simulated too dry a climate at tropical latitudes as these boundary condition changes have been shown to increase precipitation, such as around tropical mega-lakes like Lake Chad (Contoux et al., 2013; Pound et al., 2013). In summary, the PPE provides a simple platform to produce an ensemble of variable sensitivity model simulations to test the impact of boundary condition changes. As a PPE can be undertaken in one institution without having to initiate a large scale project, it makes a PPE used in this thesis the strongest tool for future Pliocene modelling of new boundary conditions. Meanwhile, boundary condition changes could result in improvements to some data-model mismatches, especially with respect to members of this ensemble and tropical biome DMCs through changes to soils and lakes.

Bibliography

Abel, S.J. & Shipway, B.J. 2007. A comparison of cloud-resolving model simulations of trade wind cumulus with aircraft observations taken during RICO. Quarterly Journal of the Royal Meteorological Society. 133 (624). 781–794. doi:10.1002/qj.55

Ackerley, D., Highwood, E.J. & Frame, D.J. 2009. Quantifying the effects of perturbing the physics of an interactive sulphur scheme using an ensemble of GCMs on the climateprediction.net platform. Journal of Geophysical Research. 114. D01203. doi:10.1029/2008JD010532

Ackert, Jr., R.P. & Kurz, M. D. 2004. Age and uplift rates of Sirius Group sediments in the Dominion Range, Antarctica, from surface exposure dating and geomorphology. Global & Planetary Change. 42. 207–225.

Allen, M.R. 1999. Do-it-yourself climate prediction. Nature. 401. 627.

Allen, M.R., Stott, P.A., Mitchell, J.F.B., Schnur, R. & Delworth, T.L. 2000. Quantifying the uncertainty in forecasts of anthropogenic climate change. Nature. 407. 617-620.

Allen, M.R. & Stainforth, D.A. 2002. Towards objective probabilistic climate forecasting. Nature. 419. 228.

Alley, R.B., Clark, P.U., Huybrechts, P. & Joughin, I. 2005. Ice sheet and sea level changes. Science. 310 (5747). pp.456-460

Annan, J.D., Lunt, D.J., Hargreaves, J.C. & Valdes, P.J. 2005a. Parameter estimation in an atmospheric GCM using the Ensemble Kalman Filter. Nonlinear Processes in Geophysics. 12. 363-371.

Annan, J.D., Hargreaves, J.C., Ohgaito, R., Abe-Ouchi, A. & Emori, S. 2005b. Efficiently constraining climate sensitivity with paleoclimate simulations. SOLA. 1. 181-184.

Annan, J.D. & Hargreaves, J.C. 2010. Reliability of the CMIP3 ensemble. Geophysical Research Letters. 37. L02703.

Annan, J.D. & Hargreaves, J.C. 2012. Identification of climatic state with limited proxy data. Climate of the Past. 8. 1141-1151.

Ashworth, A.C., Lewis, A.R., Marchant, D.R., Askin, R.A., Cantrill, D.J., Francis, J.E., Leng, M.J., Newton, A.E., Raine, J.I., Williams, M. & Wolfe, A.P. 2007. The Neogene biota of the Trans-Antarctic mountains. In: Online Proceedings of the ISAES. Cooper, A.K. & Raymond C.R. USGS Open-File Report 2007–1047. Extended Abstract 071. 4.

Balmaseda, M.A., Trenberth, K.E. & Kallen, E. 2013. Distinctive climate signals in reanalysis of global ocean heat content. Geophysical Research Letters. 40.

Bardsley, W.E. 1984. Note on the forms of precipitation totals. Journal of Hydrology. 73 (1-2) 187-191.

Barnett, D.N., Brown, S.J., Murphy, J.M., Sexton, D.M.H. & Webb, M.J. 2006. Quantifying uncertainty in changes in extreme event frequency in response to doubled CO_2 using a large ensemble of GCM simulations. Climate Dynamics. 26. 489-511.

Bartoli, G., Hönisch, B. & Zeebe, R.E. 2011. Atmospheric CO_2 decline during the Pliocene intensification of Northern Hemisphere glaciations. Paleoceanography. 26. PA4213. doi:10.1029/2010PA002055

Beer, J., Mende, W. & Stellmacher, R. 2000. The role of the Sun in climate forcing. Quaternary Science Reviews. 19. pp.403-415

Beerling, D.J. & Royer, D.L. 2011. Convergent Cenozoic CO_2 history. Nature Geoscience. 4. 418-420.

Beesley, J. A. & Moritz, R.E. 1999. Toward an explanation of the annual cycle of cloudiness over the Arctic Ocean. Journal of Climate. 12. 395–415. doi.10.1175/1520-0442(1999)012<0395:TAEOTA>2.0.CO;2

Berger, A.L. 1978. Long-term variations of caloric insolation resulting from the Earth's orbital elements. Quaternary Research. 9. 139-167

Betts, R.A., Cox, P.M., Lee, S.E. & Woodward, F.I. 1997. Contrasting physiological and structural vegetation feedbacks in climate change simulations. Nature. 387. 796-799

Betts, R.A., Cox, P.M., Collins, M., Harris, P.P., Huntingford, C. & Jones, C.D. 2004. The role of ecosystem-atmosphere interactions in simulated Amazonian precipitation decrease and forest dieback under global climate warming. Theoretical and Applied Climatology. 78. 157-175

Bice, K.L., Scotese, C.R., Seidov, D. & Barron, E.J. 2000. Quantifying the role of geographic changes in Cenozoic ocean heat transport using uncoupled atmosphere and ocean models. Earth & Planetary Science Letters. 161. 295–310.

Blair, R.C. 1981. A reaction to 'Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance'. Review of Educational Research. 51. 499-507.

Blenkinsop, S., Zhao, Y., Quinn, J., Berryman, F., Thornes, J., Baker, C. & Fowler, H.J. 2012. Downscaling future wind hazard for SE London using the UKCP09 regional climate model ensemble. Climate Research. 53. 141-156. doi:10.3354/cr01091.

Bonham, S.G., Haywood, A.M., Lunt, D.J., Collins, M. & Salzmann, U. 2009. El Nino southern oscillation, Pliocene climate and equifinality. Philosophical Transactions of the Royal Society A. 367. 127-156.

Booth, B.B.B. &. Jones, C.D. 2011. Terrestrial response of QUMPC ensemble. Hadley Centre Technical Note No. 89.

Booth, B.B.B, Jones, C.D., Collins, M., Totterdell, I.J., Cox, P.M., Sitch, S., Huntingford, C., Betts, R.A., Harris, G.R. & Lloyd, J. 2012. High sensitivity of future global warming to land carbon cycle processes. Environmental Research Letters. 7.

Bowen, G. J., Beerling, D.J., Koch, P.L., Zachos, J.C. & Quattlebaum, T. 2004. A humid climate state during the Palaeocene/Eocene Thermal Maximum. Nature. 432. 495–499. doi:10.1038/nature03115.

Box, G.E.P. & Draper, N.R. 1987. Empirical Model Building and Response Surfaces. John Wiley & Sons. New York, NY.

Braconnot, P., Otto-Bliesner, B., Harrison, S., Joussaume, S., Peterchmitt, J-Y., Abe-Ouchi, A., Crucifix, M., Driesschart, E., Fichefet, Th., Hewitt, C.D., Kageyama, M., Kitoh, A., Laine, A., Loutre, M-F., Marti, O., Merkel, U., Ramstein, G., Valdes, P., Weber, A.L., Yu, Y. & Zhao, Y. 2007. Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum – Part 1: Experiments and large-scale features. Climate of the Past. 3 (2). 261-277

Braconnot, P., Harrison, S.P., Kageyama, M., Bartlein, P.J., Masson-Delmotte, V., Abe-Ouchi, A., Otto-Bliesner, B. & Zhao, Y. 2012. Evaluation of climate models using palaeoclimatic data. Nature Climate Change. 2. 417-424. doi:10.1038/nclimate1456

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

Brierley, C.M., Fedorov, A.V., Liu, Z.H., Herbert, T.D., Lawrence, K.T., LaRiviere, J.P. 2009. Greatly expanded tropical warm pool and weakened Hadley circulation in the early Pliocene. Science. 323. 1714-1718. doi:10.1126/science.1167625.

Brierley, C.M., Collins, M. & Thorpe, A.J. 2010. The impact of perturbations to ocean model parameters on climate and climate change in a coupled model. Climate Dynamics. 34. 325-343.

Brierley, C.M. & Fedorov, A.V. 2010. Relative importance of meridional and zonal sea surface temperature gradients for the onset of the ice ages and Pliocene-Pleistocene climate evolution, Paleoceanography. 25. PA2214. doi:10.1029/2009PA001809.

Briskin, M. & Berggren., W.A. 1975. Pleistocene stratigraphy and quantitative paleo-oceanography of tropical North Atlantic core. Micropaleontology. Special Editions. 167-198.

Brown, J., Collins, M., Tudhope, A.W. & Toniazzo, T., 2008. Modelling mid-Holocene tropical climate and ENSO variability: Towards constraining predictions of future change with palaeo-data. Climate Dynamics. 30. 19-36.

Bruch, A.A., Uhl, D. & Mosbrugger, V. 2007. Miocene climate in Europe. Patterns and evolution: a first synthesis of NECLIME. Palaeogeography, Palaeoclimatology, Palaeoecology. 253. 1–7.

Bryan, F.O., Danabasoglu, G., Nakashiki, N., Yoshida, Y., Kim, D.H., Tsutsui, J. 2006. Response of the North Atlantic thermohaline circulation and ventilation to increasing carbon dioxide in CCSM3. Journal of Climate. 19. 2382–2397. doi:10.1175/JCLI3757.1

Buckley, L., Waaser, S., MacLean, H. & Fox, R. 2011. Does including physiology improve species distribution model predictions of responses to recent climate change? Ecology. 92. 2214–2221.

Chandler, M., Rind, D. & Thompson, R. 1994. Joint investigations of the middle Pliocene climate II: GISS GCM Northern Hemisphere results. Global & Planetary Change. 9. 197-219.

Chandler, M., Dowsett, H. & Haywood, A.M. 2008. The PRISM model/data cooperative: mid-Pliocene data-model comparisons. PAGES News. 16(2).

Charney, J. G. 1979. Carbon dioxide and climate: A scientific assessment. National Academy of Science. 22 pp

Chervin, R.M. & Schneider, S.H. 1976. On determining the statistical significance of climate experiments with general circulation models. Journal of Atmospheric Science. 33. 405-412.

Clark, P.U., Alley, R.B. & Pollard, D. 1999. Northern Hemisphere ice-sheet influences on global climate change. Science. 286. 1104-1111. doi:10.1126/science.286.5442.1104

Clark, P.U., Archer, D., Pollard, D., Blum, J.D., Rial, J.A., Brovkin, V., Mix, A.C., Pisias, N.G. & Roy, M. 2006. The Middle Pleistocene Transition: characteristics, mechanisms, and implications for long term changes in atmospheric pCO_2. Quaternary Science Reviews. doi: 10.1016/j.quascirev.2006.07.008.

Clark, R.T., Murphy, J.M. & Brown, S.J. 2010. Do global warming targets limit heatwave risk? Geophysical Research Letters. 37. L17703.

CLIMAP Project Members. 1981. Seasonal Reconstructions of the Earths surface at the Last Glacial Maximum. Geological Society of America Map and Chart Series. MC-36.

CLIMAP Project Members. 1984. The Last Inter-glacial ocean. Quaternary Research. 21. 123-224.

Cohen, J. 1960. A coefficient of agreement for nominal scales. Educational & Psychological Measurement. 2. 37–46.

Collins, M. 2007. Ensembles and probabilities: A new era in the prediction of climate change. Philosophical Transactions of the Royal Society A. 365. 1957-1970.

Collins, M., Tett, S.F.B. & Cooper, C. 2001. The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. Climate Dynamics. 17. 61–81. doi:10.1007/s003820000094.

Collins, M., Booth, B.B.B., Harris, G.R., Murphy, J.M., Sexton, D.M.H. & Webb, M.J. 2006. Towards quantifying uncertainty in transient climate change. Climate Dynamics. 27. 127-147.

Collins, M., Brierley, C.M., MacVean, M., Booth, B.B.B. & Harris, G.R., 2007. The sensitivity of the rate of transient climate change to ocean physics perturbations. Journal of Climate. 20. 2315-2320.

Collins, M., Booth, B.B.B., Bhaskaran, B., Harris, G.R., Murphy, J.M., Sexton, D.M.H. & Webb, M.J. 2011. Climate model errors, feedbacks and forcings: A comparison of perturbed physics and multi model ensembles. Climate Dynamics. 36. 9-10. 1737-1766

Collins, M., Chandler, R.E., Cox, P.M., Huthnance, J.M., Rougier, J. & Stephenson, D.B. 2012 Quantifying future climate change. Nature Climate Change. 2 (6). 403-409.

Collins, W.D., Ramaswamy, V., Schwarzkopf, M.D., Sun, Y., Portmann, R.W., Fu, Q., Casanova, S.E.B., Dufresne, J-L., Fillmore, D.W., Forster, P.M.D., Galin, V.Y., Gohar, L.K., Ingram, W.J., Kratz, D.P., Lefebvre, M.P., Li, J., Marquet, P., Oinas, V., Tsushima, Y., Uchiyama, T. & Zhong, W.Y. 2006. Radiative forcing by well-mixed greenhouse gases: Estimates from climate models in the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4). Journal of Geophysical Research. 111. D14317. doi:10.1029/2005JD006713

Contoux, C., Jost, A., Ramstein, G., Sepulchre, P., Krinner, G. & Schuster, M. 2013. Megalake Chad impact on climate and vegetation during the late Pliocene and the mid-Holocene. Climate of the Past. 9. 1417-1430. doi:10.5194/cp-9-1417-2013

Covey, C., Brandon, S., Bremer, P.T., Domyancie, D., Garaizar, X., Johannesson, G., Klein, R., Klein, S.A., Lucas, D.D., Tannahill, J. & Zhang, Y. 2011. Quantifying the uncertainties in climate predictions. Technical Report. Lawrence Livermore National Laboratory (LLNL), Livermore, California.

Cox, P.M., Betts, R.A., Bunton, C.B., Essery, R.L.H., Rowntree, P.R. & Smith, J. 1999. The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. Climate Dynamics. 15. 183-203

Cox, P.M., Betts, R.A., Collins, M., Harris, P.P., Huntingford C. & Jones, C.D. 2004. Amazonian forest dieback under climate-carbon cycle projections for the 21st century Theoretical and Applied Climatology. 78 (1-3). 137-156. doi:10.1007/s00704-004-0049-4

Coxall, H.K., Wilson, P.A., Pälike, H., Lear, C.H. & Backman, J. 2005. Rapid stepwise onset of Antarctic glaciation and deeper calcite compensation in the Pacific Ocean. Nature. 433. 53-57. doi:10.1038/nature03135

Cronin, T.M., Dwyer, G.S., Baker, P.A., Rodriguez-Lazaro, J. & Briggs Jr., W.M. 1996. Deep-sea ostracode shell chemistry (Mg/Ca ratios) and late Quaternary Arctic Ocean history. 117–134. In: Andrews, J.T., Austin, W.E.N., Bergsten, H., Jennings, A.E. (Eds.). Late Quaternary Palaeoceanography of the North Atlantic Margins. Special Publication. 111. Geological Society, London.

Cronin, T.M., Dowsett, H.J., Dwyer, G.S., Baker, P.A. & Chandler, M.A. 2005. Mid-Pliocene deep sea bottom water temperatures based on Ostracode Mg/Ca ratios. Marine Micropaleontology. 54. 249-261.

Crossley, J.F. & Roberts, D.L. 1995. The thermodynamic/dynamic sea ice model. Unified Model Documentation Paper 45, version 1.

Crowley, T.J. 1991. Are there any satisfactory geologic analogs for a future greenhouse warming? Journal of Climate. 3. 1282–1292. doi:10.1175/1520-0442(1990)003<1282:ATASGA>2.0.CO;2

Crucifix, M. 2006. Does the Last Glacial Maximum constrain climate sensitivity? Geophysical Research Letters. 33. L18701. doi:10.1029/2006GL027137.

Currie, B.S., Rowley, D.B. & Tabor, N.J. 2005. Middle Miocene paleoaltimetry of Southern Tibet: implications for the role of mantle thickening and delamination in the Himalayan orogeny. Geology. 33. 181–184.

Curry, J.A., Schramm, J.L. & Ebert, E.E. 1993. Impact of clouds on the surface radiation balance of the Arctic Ocean. Meteorology and Atmospheric Physics. 51. 197-217.

Curry, J.A., Schramm, J.L. & Ebert, E.E. 1995. Sea ice-albedo climate feedback mechanism. Journal of Climate. 8. 240-247.

Curry, J.A., Rossow, W.B., Randall, D. & Schramm, J.L. 1996. Overview of Arctic cloud and radiation characteristics. Journal of Climate. 9. 1731-1764.

Cusack, S., Slingo, A., Edwards, J.M. & Wild, M. 1998. The radiative impact of a single aerosol climatology on the Hadley Centre atmospheric GCM. Quarterly Journal of the Royal Meteorological Society. 124. 2517-2526.

DeConto, R.M. & Pollard, D. 2003. Rapid Cenozoic glaciation of Antarctica induced by declining atmospheric CO_2. Nature. 421. 245-249.

DeConto, R.M., Pollard, D., Wilson, P.A., Palike, H., Lear, C.H. & Pagani, M. 2008. Thresholds for Cenozoic bipolar glaciation. Nature. 455. 652–656.

Dekens, P.S., Ravelo, A.C., McCarthy, M.D. & Edwards, C.A. 2008. A 5 million year comparison of Mg/Ca and alkenone paleothermometers. Geochemistry, Geophysics, Geosystems. 9. Q10001. doi: 10.1029/2007GC001931.

Demicco, R.V. 2004. Modelling sea-floor spreading rates through time. Geology. 32(6). 485–488.

Des Marais, D.L. & Juenger, T. E. 2010. Pleiotropy, plasticity, and the evolution of plant abiotic stress tolerance. Annals of the New York Academy of Sciences. 1206. 56-79.

DeWeaver, E.T. 2007. Uncertainty in climate model projections of Arctic sea ice decline: An evaluation relevant to polar bears. USGS Alaska Science Center, Anchorage, Administrative Report.

DeWeaver, E.T., Hunke, E.C. & Holland M.M. 2008. Sensitivity of Arctic sea ice thickness to intermodel variations in the surface energy budget. In DeWeaver, E.T., Bitz, C.M. & Tremblay L.B. (Eds.) Arctic sea ice decline: Observations, projections, mechanisms, and implications. American Geophysical Union, Washington, D.C. doi:10.1029/180GM05

Dijkstra, H.A. & Neelin, J.D. 1999. Imperfections of the Thermohaline Circulation: Multiple equalibria and flux correction. Journal of Climate. 12. 1382-1392.

Dolan, A.M., Haywood, A.M., Hill, D.J., Dowsett, H.J., Hunter, S.J., Lunt, D.J. & Pickering, S.J. 2011. Sensitivity of Pliocene ice sheets to orbital forcing. Palaeogeography, Palaeoclimatology, Palaoecology. 309 (1-2). 98-110. doi:10.1016/j.palaeo.2011.03.030

Dolan, A.M., Koenig, S.J., Hill, D.J., Haywood, A.M. & DeConto, R.M. 2011. Pliocene Ice Sheet Modelling Intercomparison Project (PLISMIP) - experimental design. Geoscientific Model Development. 5. 963-974. doi: 10.5194/gmd-5-963-2012

Dowsett, H.J. 1991. The development of a long-range foraminifer transfer function and application to Late Pleistocene North Atlantic climatic extremes. Paleoceanography. 6 (2). 259-273.

Dowsett, H.J. 2007. The PRISM palaeoclimate reconstruction and Pliocene sea surface temperature. 459-480. In Williams, M., Haywood, A.M., Gregory, F.J. & Schmidt, D.N. (Eds.) Deep-time perspectives on climate change: Marrying the signal from computer models and biological proxies. The Micropalaeontological Society.. Special Publications. The Geological Society, London. 251-312.

Dowsett, H.J. & Cronin, T.M. 1990. High eustatic sea level during the Middle Pliocene: Evidence from the southern U.S. Atlantic coastal plain. Geology. 18. 435-438.

Dowsett, H.J. & Poore, R.Z. 1990. A new planktic foraminifer transfer function for estimating Pliocene through Holocene sea surface temperatures. Marine Micropaleontology. 16. 1–23

Dowsett, H.J. & Poore, R.Z. 1991. Pliocene sea surface temperatures of the North Atlantic ocean at 3.0 Ma. Quaternary Science Reviews. 10. 189-204.

Dowsett, H.J. & Robinson, M.M. 1998. Application of the Modern Analog Technique (MAT) of sea surface temperature estimation to middle Pliocene North Pacific planktonic foraminifer assemblages. Palaeontologia Electronica 1. http://palaeo-electronica.org/1998_1/dowsett/issue1.htm

Dowsett, H.J. & Robinson, M.M. 2006. Stratigraphic framework for Pliocene paleoclimate reconstruction: the correlation conundrum. Stratigraphy. 3. 53-64.

Dowsett, H.J., Cronin, T.M., Poore, R.Z., Thompson, R.S., Whatley, R.C. & Wood, A.D. 1992. Micropalentological evidence for increase meridional heat transport in the North Atlantic Ocean during the Pliocene. Science. 258. 1133-1135.

Dowsett, H.J., Thompson, R., Barron, J., Cronin, T., Fleming, F., Ishman, S., Poore, R., Willard, D. & Holtz, T. 1994. Joint investigations of the middle Pliocene climate I: PRISM palaeoenvironmental reconstructions. Globlal & Planetary Change. 9. 169-195.

Dowsett, H.J., Barron, J. & Poore, R. 1996. Middle Pliocene sea surface temperatures: A global reconstruction. Marine Micropaleontology. 27. 13-25.

Dowsett, H.J., Barron, J.A., Poore, R.Z., Thompson, R.S., Cronin, T.M., Ishman, S.E. & Williams, D.A. 1999. Pliocene paleoenvironmental reconstruction: PRISM2. Edited. 99-535. U.S. Geological Survey.

Dowsett, H.J., Chandler, M.A., Cronin, T.M. & Dwyer, G.S. 2005. Middle Pliocene sea surface temperature variability. Paleoceanography. 20.

Dowsett, H.J., Robinson, M.M., Dwyer, G.S., Chandler, M.A., & Cronin, T.M. 2006. PRISM3 DOT1 Atlantic Basin Reconstruction. U.S. Geological Survey Data Series. 189.

Dowsett, H.J., Robinson, M.M. & Foley, K.M. 2009a. Pliocene Three-Dimensional Global Ocean Temperature Reconstruction. Climate of the Past. 5. 769-783.

Dowsett, H.J., Chandler, M. & Robinson, M.M. 2009b. Surface Temperatures of the Mid Pliocene North Atlantic Ocean: Implications for Future Climate. Philosophical Transactions of the Royal Society A. 367. 16.

Dowsett, H.J., Robinson, M.M., Haywood, A.M., Salzmann, U., Hill, D.J., Sohl, L., Chandler, M., Williams, M., Foley, K. & Stoll, D.K. 2010a. The PRISM3D paleoenvironmental reconstruction. Stratigraphy. 7 (2-3). 123-140.

Dowsett, H.J., Robinson, M.M., Stoll, D.K. & Foley, K.M. 2010b Mid-Pliocene mean annual sea surface temperature analysis for data-model comparisons. Stratigraphy. 7 (2-3). 189-198.

Dowsett, H.J., Haywood, A.M., Valdes, P.J., Robinson, M.M., Lunt, D.J., Hill, D.J., Stoll, D.K. & Foley, K.M. 2011. Sea surface temperatures of the mid-Piacenzian warm period: A comparison of PRISM3 and HadCM3. Palaeogeography, Palaeoclimatology, Palaeoecology. 309 (1-2). 83-91.

Dowsett, H.J., Robinson, M.M., Haywood, A.M., Hill, D.J., Dolan, A.M., Stoll, D.K., Chan, W-L., Abe-Ouchie, A., Chandler, M.A., Rosenbloom, N.A., Otto-Bliesner, B.L., Bragg, F.J., Lunt, D.J., Foley, K.M. & Riesselman, C.R. 2012. Assessing confidence in Pliocene sea surface temperatures to evaluate predictive models. Nature Climate Change. 2. 365-371.

Dowsett, H., Foley, K., Stoll, D., Chandler, M., Sohl, L., Bentsen, M., Otto-Bliesner, B., Bragg, F., Chan, W-L., Contoux, C., Dolan, A., Haywood, A., Jonas, J., Jost, A., Kamae, Y., Lohmann, G., Lunt, D., Nisancioglu, K., Abe-Ouchi, A., Ramstein, G., Riesselman, C., Robinson, M., Rosenbloom, N., Salzmann, U., Stepanek, C., Strother, S., Ueda, H., Yan, Q. & Zhang, Z. 2013. Sea surface temperature of the mid-Piacenzian ocean: A data-model comparison. Nature Scientific Reports. 3. doi:10.1038/srep02013

Dunkley-Jones, T., Ridgwell, A., Lunt, D.J., Maslin, M.A., Schmidt, D.N. & Valdes, P.J. 2010. A Palaeogene perspective on climate sensitivity and methane hydrate instability. Philosophical Transactions of the Royal Society A. 368. 2395–2415. doi:10.1098/rsta.2010.0053

Dutton, J.F. & Barron, E.J. 1997. Miocene to present vegetation changes: A possible piece of the Cenozoic puzzle. Geology. 25. 39–41.

Dwyer, G.S. & Chandler, M.A. 2009. Mid-Pliocene sea level and continental ice volume based on coupled benthic Mg/Ca palaeotemperatures and oxygen isotopes. Philosophical Transactions of the Royal Society A. 367. 157-168. doi:10.1098/rsta.2008.0222.

Edwards, M. 1992. Global gridded elevation and bathymetry. National Geophysical Data Center. NOAA.

Edwards, J.M. & Slingo., A. 1996. Studies with a flexible new radiation code. I: Choosing a configuration for a large scale model. Quarterly Journal of the Royal Meteorological Society. 122. 689-719.

Ekart, D.D., Cerling, T.E., Montanez, I.P. & Tabor, N.J. 1999. A 400 million year carbon isotope record of pedogenic carbonate: implications for paleoatmospheric carbon dioxide. American Journal of Science. 299. 805–827. doi:10.2475/ajs.299.10.805

Etourneau, J., Martinez, P., Blanz, T., Schneider, R. 2009. Pliocene-Pleistocene variability of upwelling activity, productivity and nutrient cycling in the Benguela region. Geology. 37. 871–874.

Fedorov, A.V., Brierley, C.M. & Emanuel, K. 2010. Tropical cyclones and permanent El Niño in the early Pliocene epoch. Nature. 463. 1066-1070. doi:10.1038/nature08831

Fedorov, A.V., Brierley, C.M., Lawrence, K.T., Liu, Z., Dekens, P.S. & Ravelo, A.C. 2013. Patterns and mechanisms of early Pliocene warmth. Nature. 496. 43–49. doi:10.1038/nature12003

Filippelli, G.M. & Flores, J.A. 2009. From the warm Pliocene to the cold Pleistocene: A tale of two oceans. Geology. 37. 959-960. doi:10.1130/focus102009.1

Fischer, E.M., Lawrence, D.M. & Sanderson, B.M. 2011. Quantifying uncertainties in projections of extremes – a perturbed land surface experiment. Climate Dynamics 37. 1381-1398.

Flower, B.P. & Kennett, J.P. 1994. The middle Miocene climatic transition: East Antarctic ice sheet development, deep ocean circulation and global carbon cycling. Palaeogeography, Palaeoclimatology, Palaeoecology. 108. 537–555.

Foley, J.A., Kutzbach, J.E., Coe, M.T. & Levis, S. 1994. Feedbacks between climate and boreal forests during the Holocene epoch. Nature. 371. 52-54

Foster, G.L. 2008. Seawater pH, p CO_2 and [CO_{2-3}] variations in the Caribbean Sea over the last 130 kyr: A boron isotope and B/Ca study of planktic foraminifera. Earth & Planetary Science Letters. 271 (1-4). 254-266. doi:10.1016/j.epsl.2008.04.015

Foster, G.L., Lunt, D.J. & Parrish, R.R. 2010. Mountain uplift and the glaciation of North America – a sensitivity study. Climate of the Past. 6 (5). 707-717. doi:10.5194/cp-6-707-2010

Fowler H., Cooley D., Sain S. & Thurston M. 2010. Detecting change in UK extreme precipitation using results from the climateprediction.net BBC climate change experiment. Extremes. doi:10.1007/s10687-010-0101-y.

Frame, D.J., Aina, T., Christensen, C.M., Faull, N.E., Knight, S.H.E., Piani, C., Rosier, S.M., Yamazaki, K., Yamazaki, Y. & Allen, M.R. 2009. The climateprediction.net BBC climate change experiment: Design of the coupled model ensemble. Philosophical Transactions of the Royal Society A. 367. 855-870.

Fröhlich, C. 2006. Solar irradiance variability since 1978: Revision of the PMOD composite during solar cycle 21. Space Science Reviews. 125. 53–65. doi: 10.1007/s11214-006-9046-5.

Ganopolski, A., Petoukhov, V., Rahmstorf, S., Brovkin, V., Claussen, M., Eliseev, A. & Kubatzki, C. 2001. CLIMBER-2: a climate system model of intermediate complexity. Part II: Model sensitivity. Climate Dynamics. 17. 735–751.

Ganopolski, A., Calov, R., Robinson, A. & Willeit, M. 2011. Analysis of the role of climate forcings and feedbacks for the Mid-Pliocene climate using a set of transient simulations from an Earth System Model. Abstract PP13B-1824 presented at 2011 Fall Meeting, AGU, San Francisco, California. 5-9th Dec.

Gent, P.R., Danabasoglu, G., Donner, L.J., Holland, M.M., Hunke, E.C., Jayne, S.R., Lawrence, D.M., Neale, R.B., Rasch, P.J., Vertenstein, M., Worley, P.H., Yang, Z.-L. & Zhang, M. 2011. The Community Climate System Model Version 4. Journal of Climate. 24. 4973–4991. doi:10.1175/2011JCLI4083.1

Gladenkov, A., Oleinik, A., Marincovich, L. & Barinov, K. 2002. A refined age for the earliest opening of Bering Strait. Palaeogeography, Palaeoclimatology, Palaeoecology. 183 (3-4). 321-328.

Goodchild, M.F., Guoqing, S. & Shiren. Y. 1992. Development and test of an error model for categorical data. International Journal of Geographical Information Systems. 6 (2). 87-103.

Gordon, C., Cooper, C., Senior, C.A., Banks, H., Gregory, J.M., Johns, T.C., Mitchell, J.F.B. & Wood, R.A. 2000. The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. Climate Dynamics. 16. 147-168

Gradstein, F.M. & Ogg., J.G. 2009. The geologic timescale. Oxford University Press. Oxford.

Gradstein, F.M., Ogg, J.G., Schmitz, M. & Ogg, G. 2012. The geologic timescale 2012. Elsevier. 1176pp.

Gregoire, L.J., Valdes, P.J., Payne, A.J. & Kahana, R. 2011. Optimal tuning of a GCM using modern and glacial constraints. Climate Dynamics. 37. 705-719. doi:10.1007/s00382-010-0934-8

Gregory, D. & Rowntree, P.R. 1990. A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependant closure. Monthly Weather Review. 118. 1483-1506.

Gregory, J.M. & Forster, P.M. 2008. Transient climate response estimated from radiative forcing and observed temperature change, Journal of Geophysical. Research. 113. D23105. doi:10.1029/2008JD010405.

Gregory, J.M., Browne, O.J.H., Payne, A.J., Ridley, J.K. & Rutt, I.C. 2012. Modelling large-scale ice-sheet–climate interactions following glacial inception. Climate of the Past. 8. 1565-1580. doi:10.5194/cp-8-1565-2012.

Gregory-Wodzicki, K.M. 2000. Uplift history of the Central and Northern Andes: A review, Geological Society of America Bulletin. 112. 1091–1105.

Haney, R.L. 1971. Surface thermal boundary condition for ocean circulation models. Journal of Physical Oceanography. 1. 241–248. doi:10.1175/1520-485(1971)001<0241:STBCFO>2.0.CO;2

Hansen, J., Sato, M., Ruedy, R., Lo, K., Lea, D.W. & Medina-Elizade, M (2006), Global temperature change. Proceedings of the National Academy of Sciences. 103 (39). 14288-14293.

Hargreaves, J.C. & Annan, J.D. 2009. On the importance of paleoclimate modelling for improving predictions of future climate change. Climate of the Past. 5. 803-814.

Harris, N. 2006. The elevation history of the Tibetan Plateau and its implication for the Asian monsoon. Palaeogeography, Palaeoclimatology, Palaeoecology. 241. 4–15.

Harrison, S.P. & Prentice, I.C. 2003. Climate and CO_2 controls on global vegetation distribution at the Last Glacial Maximum: Analysis based on palaeovegetation data, biome modelling and palaeoclimate simulations. Global Change Biology. 9. 983-1004.

Harrison, T.M. & Yin, A. 2004 Timing and processes of Himalayan and Tibetan uplift. Himalayan Journal of Science. 2. 152–153.

Hasumi, H. 2002. Sensitivity of the global thermohaline circulation to interbasin freshwater transport by the atmosphere and the Bering Strait throughflow. Journal of Climate. 15. 2516-2526.

Hawkins, E. & Sutton, R.T. 2009. The potential to narrow uncertainty in regional climate predictions. Bulletin of the American Meteorological Society. 90 (8). 1095-1107. doi:10.1175/2009BAMS2607.1

Hawkins, E. & Sutton, R.T. 2011. The potential to narrow uncertainty in projections of regional precipitation change. Climate Dynamics. 37 (1-2). 407-418. doi:10.1007/s00382-010-0810-6

Haxeltine, A. & Prentice, I.C. 1997. BIOME3: An equilibrium terrestrial biosphere model based on ecophysiological constrains, resource availability and competition among plant function types. Global Biogeochemical Cycles. 10. 693-709

Haywood, A.M. & Valdes, P.J. 2004. Modelling Pliocene warmth: contribution of atmosphere, oceans and cryosphere. Earth & Planetary Science Letters. 218. 363-377.

Haywood, A.M. & Valdes, P.J. 2006. Vegetation cover in a warmer world simulated using a dynamic global vegetation model for the mid-Pliocene. Palaeogeography, Palaeoclimatology, Palaeoecology. 237. 412-427.

Haywood, A.M., Valdes, P.J. & Sellwood, B.W. 2000a. Global scale palaeoclimate reconstruction of the middle Pliocene climate using the UKMO GCM: Initial results. Global & Planetary Change. 25. 239-256.

Haywood, A.M., Sellwood, B.W. & Valdes, P.J. 2000b. Regional warming: Pliocene (3 Ma) palaeoclimate of Europe and the Mediterranean. Geology. 28 (12). 1063-1066.

Haywood, A.M., Valdes, P.J., Sellwood, B.W., Kaplan, J.O. & Dowsett, H.J. 2001. Modelling Middle Pliocene warm climates of the USA. Palaeontologia Electronica. 4. 21.

Haywood, A.M., Valdes, P.J. & Sellwood, B.W. 2002a. Magnitude of climate variability during the Middle Pliocene warmth: A palaeoclimate modelling study. Palaeogeography, Palaeoclimatology, Palaeoecology. 188. 1-24.

Haywood, A.M., Valdes, P.J., Sellwood, B.W. & Kaplan, J.O. 2002b, Antarctic climate during the middle Pliocene: Model sensitivity to ice sheet variation. Palaeogeography, Palaeoclimatology, Palaeoecology. 182. 93-115.

Haywood, A.M., Valdes, P.J., Francis, J.E. & Sellwood, B.W. 2002c. Global Middle Pliocene biome reconstruction: A data/model synthesis. Geochemistry, Geophysics, Geosystems. 3 (12). 1072. doi:10.1029/2002GC000358

Haywood, A.M., Dekens, P., Ravelo, A.C. & Williams, M. 2005. Warmer tropics during the mid-Pliocene? Evidence from alkenone paleothermometry and a fully coupled ocean-atmosphere GCM. Geochemistry, Geophysics, Geosystems. 6. Q03010. doi: 10.1029/2004GC000799

Haywood, A.M., Valdes, P.J. & Peck, V.L. 2007a. A permanent El Nino like state during the Pliocene. Paleoceanography. 22.

Haywood, A.M., Valdes, P.J., Hill, D.J. & Williams, M. 2007b. The mid-Pliocene warm period: A test bed for integrating data and models. 443-458. In Williams, M., Haywood, A.M., Gregory, F.J. & Schmidt, D.N (Eds.) Deep-time perspectives on climate change: Marrying the signal from computer models and biological proxies. The Micropalaeontological Society. Special Publications. The Geological Society, London. 251-312.

Haywood, A.M., Dowsett, H.J., Valdes, P.J., Lunt, D.J., Francis, J.E. & Sellwood, B.W. 2009a. Introduction: Pliocene climate, processes and problems, Philosophical Transactions of the Royal Society A. 367. 3-17.

Haywood, A.M., Chandler, M.A., Valdes, P.J., Salzmann, U., Lunt, D.J. & Dowsett, H.J. 2009b. Comparison of mid-Pliocene climate predictions produced by the HadAM3 and GCMAM3 general circulation models. Global & Planetary Change. 66. 208-224.

Haywood, A.M., Dowsett, H.J., Otto-Bliesner, B., Chandler, M.A., Dolan, A.M., Hill, D.J., Lunt, D.J., Robinson, M.M., Rosenbloom, N., Salzmann, U. & Sohl, L.E. 2010. Pliocene Model Intercomparison Project (PlioMIP): experimental design and boundary conditions (Experiment 1). Geoscientific Model Development. 3. 227-242.

Haywood, A.M., Dowsett, H.J., Robinson, M.M., Stoll, D.K., Dolan, A.M., Lunt, D.J., Otto-Bliesner, B. & Chandler, M.A. 2011a. Pliocene Model Intercomparison Project (PlioMIP): experimental design and boundary conditions (Experiment 2). Geoscientific Model Development. 4. 571-577. doi:10.5194/gmd-4-571-2011

Haywood, A.M., Ridgwell, A., Lunt, D.J., Hill, D.J., Pound, M.J., Dowsett, H.J., Dolan, A.M., Francis, J.E. & Williams, M. 2011b. Are there pre-Quaternary geological analogues for a future greenhouse gas-induced global warming? Philosophical Transactions of the Royal Society A. 369. 933-956.

Haywood, A.M., Hill, D.J., Dolan, A.M., Pickering, S.J., Otto-Bliesner, B.L., Rosenbloom, N.A., Bragg, F., Lunt, D.J., Chan, W-L., Abe-Ouchi, A., Chandler, M.A., Sohl, L., Contoux, C., Ramstein, G., Jost, A., Dowsett, H.J., Kamae, Y., Ueda, H., Lohmann, G., Stepanek, C., Salzmann, U., Yan, Q. & Zhang, Z. 2013a. Large-scale features of Pliocene climate: Results from the Pliocene Model Intercomparison Project. Climate of the Past. 9. 191-209. doi:10.5194/cp-9-191-2013

Haywood, A.M., Dolan, A.M., Pickering, S.J., Dowsett, H.J., McClymont, E.L., Prescott, C.L., Salzmann, U., Hill, D.J., Hunter, S.J., Lunt, D.J., Pope, J.O. & Valdes, P.J. 2013b. On the identification of a Pliocene time slice for data-model comparison. Philosophical Transactions of the Royal Society A. 371 (2001). 20120515. doi.org/10.1098/rsta.2012.0515

Henrot, A-J., François, L., Favre, E., Butzin, M., Ouberdous, M. & Munhoven, G. 2010. Effects of CO_2, continental distribution, topography and vegetation changes on the climate at the Middle Miocene: a model study. Climate of the Past. 6. 675-694. doi:10.5194/cp-6-675-2010.

Herold, N., Seton, M., Muller, R.D., You, Y. & Huber, M. 2008. Middle Miocene tectonic boundary conditions for use in climate models. Geochemistry, Geophysics, Geosystems. 9 (10). Q10009. doi:10.1029/2008GC002046.

Herold, N., Huber, M., Müller, R.D. & Seton, M. 2012. Modelling the Miocene climatic optimum ocean circulation. Paleoceanography. 27. PA1209.

Hewitt, C.D., Stouffer, R.J., Broccoli, A.J., Mitchell, J.F.B. & Valdes, P.J. 2002. The effect of ocean dynamics in a coupled GCM simulation of the Last Glacial Maximum. Climate Dynamics. 20. 203-218

Hill, D.J. 2009. Modelling Earth's cryosphere during Peak Pliocene Warmth. PhD Thesis. University of Bristol/British Antarctic Survey. 334.

Hill, D.J., Haywood, A.M., Hindmarsh, R.C.A. & Valdes, P.J. 2007. Characterising ice sheets during the mid-Pliocene. 539-562. In Williams, M., Haywood, A.M., Gregory, F.J. & Schmidt, D.N (Eds.) Deep-time perspectives on climate change: Marrying the signal from computer models and biological proxies. The Micropalaeontological Society. Special Publications. The Geological Society, London.

Hill, D.J., Csank, A.Z., Dolan, A.M. & Lunt, D.J. 2011. Pliocene climate variability: Northern Annular Mode in models and tree-ring data. Palaeogeography, Palaeoclimatology, Palaoecology. 309 (1-2). 118-127.

Hill, D.J., Haywood, A.M., Lunt, D.J., Hunter, S.J., Bragg, F.J., Contoux, C., Stepanek, C., Sohl, L., Rosenbloom, N.A., Chan, W-L., Kamae, Y., Zhang, Z., Abe-Ouchi, A., Chandler, M.A., Jost, A., Lohmann, G., Otto-Bliesner, B.L., Ramstein, G. & Ueda, H. 2014. Evaluating the dominant components of warming in Pliocene climate simulations. Climate of the Past. 10. 79-90. doi:10.5194/cp-10-79-2014.

Hodson, D.L.R., Keeley, S.P.E., West, A., Ridley, J., Hawkins, E. & Hewitt, H.T. 2013. Identifying uncertainties in Arctic climate change projections. Climate Dynamics. 40 (11-12). 2849-2865. doi: 10.1007/s00382-012-1512-z

Holland, M.M. & Bitz, C.M. 2003. Polar amplification of climate change in coupled models. Climate Dynamics. 21. 221-232

Hollis, C.J., Taylor, K.W.R., Handley, L., Pancost, R.D., Huber, M., Creech, J.B., Hines, B.R., Crouch, E.M., Morgans, E.G., Crampton, J.S., Gibbs, S., Pearson, P.N. & Zachos, J.C. 2012. Early Paleogene temperature history of the Southwest Pacific Ocean: Reconciling proxies and models. Earth & Planetary Science Letters. 349-350. 53-66

Howell, F.W., Haywood, A.M., Francis, J., Wade, B., Pickering, S.J. & Dowsett, H.J. 2013. The sea ice albedo effect: Improving polar amplification in model simulations of the mid-Pliocene. Geophysical Research Abstracts. 15. GU2013-929.

Howell, F.W., Haywood, A.M., Dolan, A.M., Dowsett, H.J., Francis, J.E., Hill, D.J., Pickering, S.J., Pope, J.O., Salzmann, U. & Wade, B.S. 2014. Can uncertainties in sea ice albedo reconcile patterns of data-model discord for the Pliocene and 20th/21st centuries? Geophysical Research. Letters. 41. doi:10.1002/2013GL058872.

Hu, A.X., Meehl, G.A., Washington, W.M. & Dai, A. 2004. Response of the Atlantic thermohaline circulation to increased atmospheric CO_2 in a coupled model. Journal of Climate. 17. 4267–4279.

Huber, M. & Sloan, L.C. 2001. Heat transport, deep waters, and thermal gradients: Coupled simulation of an Eocene greenhouse climate. Geophysical Research Letters. 28 (18). 3481-3484.

Huber, M. & Caballero, R. 2011. The early Eocene equable climate problem revisited. Climate of the Past. 7. 603-633. doi:10.5194/cp-7-603-2011

Hutson, W.H. 1979. The Agulhas current during the Late Pleistocene: Analysis of modern faunal analogs. Science. 207. 64-66.

Huebener, H., Cubasch, U., Langematz, U., Spangehl, T., Nierhorster, F., Fast, I. & Kunze, M. 2007. Ensemble climate simulations using a fully coupled ocean-troposphere-stratosphere general circulation model. Philosophical Transactions of The Royal Society A. 365. 2089–2101. doi:10.1098/rsta.2007.2078

Hutton, 1788. Theory of the Earth; or an investigation of the laws observable in the composition, dissolution, and restoration of land upon the Globe. Transactions of the Royal Society of Edinburgh, 1 (2). 209–304.

Hyndman, R.J. & Koehler, A.B. 2006. Another look at measures of forecast accuracy. International Journal of Forecasting. 22 (4). 679-688.

Imbrie, J. & Kip, N.G. 1971. A new micropaleontological method for quantitative paleoclimatology: Applied to a Late Pleistocene Caribbean core. 71-181. In: Turekian, K.K. (Ed.). The Late Cenozoic glacial ages. Yale University Press, New Haven, CT.

IPCC, 2000. Special report on emissions scenarios. Cambridge University Press, Cambridge & New York.

IPCC. 2001. Climate Change 2001: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK & New York.

IPCC. 2007. Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, 906. Cambridge University Press, Cambridge, UK & New York.

IPCC. 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.

Irvine, P.J., Gregoire, L.J., Lunt, D.J. & Valdes, P.J. 2013. An efficient method to generate a perturbed parameter ensemble of a fully coupled AOGCM without flux-adjustment. Geoscientific Model Development. 6. 1447-1462. doi:10.5194/gmd-6-1447-2013

Ivanovic, R.F., Valdes, P.J., Flecker, R., Gregoire, L.J. & Gutjahr, M. 2013. The parameterisation of Mediterranean-Atlantic water exchange in the Hadley Centre model HadCM3 and its effect on modelled North Atlantic climate. Ocean Modelling. 62. 11-16.

Ivanovic, R.F., Valdes, P.J., Flecker, R. & Gutjahr, M. 2014a. Modelling global-scale climate impacts of the late Miocene Messinian Salinity Crisis, Climate of the Past, 10, pp.607-622. doi: 10.5194/cp-10-607-2014

Ivanovic, R.F., Valdes, P.J., Gregoire, L.J., Flecker, R. & Gutjahr, M. 2014b. Sensitivity of modern climate to the prescence, strength and salinity of Mediterranean-Atlantic exchange in a global circulation model. Climate Dynamics. 42 (3-4). 859-877.

Jackson, C.S., Sen, M.K., Huerta, G., Deng Y. & Bowman, K. 2008. Error reduction and convergence in climate prediction. Journal of Climate 21. 6698-6709.

Jackson, L.C., Vellinga, M. & Harris, G.R. 2012. The sensitivity of the meridional overturning circulation to modelling uncertainty in a perturbed physics ensemble without flux adjustment. Climate Dynamics. 39 (1-2). 277-285. doi:10.1007/s00382-011-1110-5.

Jackson, L. & Vellinga, M. 2013. Multidecadal to centennial variability of the AMOC: HadCM3 and a perturbed physics ensemble. Journal of Climate. 26. 2390–2407. doi:10.1175/JCLI-D-11-00601.1

Jakob, C. & Klein, S.A. 2000. A parameterisation of the effects of cloud and precipitation overlap for use in general circulation models. Quarterly Journal of the Royal Meteorological Society. 126. 2525-2544.

Jenness, J. & Wynne, J.J. 2005. Cohen's Kappa and classification table metrics 2.0: an ArcView 3x extension for accuracy assessment of spatially explicit models. U.S. Geological Survey Open-File Report OF 2005-1363. 1-86.

Jimenez-Moreno, G., Fauquette, S. & Suc, J.-P. 2008. Vegetation, climate and palaeoaltitude reconstructions of the eastern Alps during the Miocene based on pollen records from Austria, central Europe. Journal of Biogeography. 35. 1638–1649. doi:10.1111/j.1365-2699.2008.01911.x

Johns, T.C. 1996. A description of the second Hadley Centre Coupled Model (HadCM2). Climate Research Technical Note No. 71. Meteorological Office, Bracknell, UK. 19pp.

Johns, T.C., Carnell, R.E., Crossley, J.F., Gregory, J.M., Mitchell, J.F.B., Senior, C.A., Tett, S.F.B. & Wood, R.A. 1997. The second Hadley Centre coupled ocean atmosphere GCM: model description, spin up and validation. Climate Dynamics. 132. 103-134.

Johns, T.C., Gregory, J.M., Ingram, W.J., Johnson, C.E., Jones, A., Lowe, J.A., Mitchell, J.F.B., Roberts, D.L., Sexton, D.M.H., Stevenson, D.S., Tett, S.F.B. & Woodage, M.J. 2003. Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 Model under updated emissions scenarios. Climate Dynamics. 20. 583-612.

Jones, S.M., White, N.J. & Maclennan, J.C. 2002. V-shaped ridges around Iceland: implications for spatial and temporal patterns of mantle convection. Geochemistry, Geophysics, Geosystems, 3. doi:101029/2002GC000361

Jones, C., Gregory, J., Thorpe, R., Cox, P., Murphy, J., Sexton, D. & Valdes, P. 2005. Systematic optimisation and climate simulation of FAMOUS, a fast version of HadCM3. Climate Dynamics. 25 (2–3). 189–204.

Joshi, M.M., Gregory, J.M., Webb, M.J., Sexton, D.M.H. & Johns, T.C. 2008. Mechanisms for the land/sea warming contrast exhibited by simulations of climate change. Climate Dynamics. 30. 455-465.

Joshi, M.M., Webb, M.J., Maycock, A.C. & Collins, M. 2010. Stratospheric water vapour and high climate sensitivity in a version of the HadSM3 climate model. Atmospheric Chemistry & Physics. 10. 7161-7167.

K-1 model developers. 2004. K-1 coupled model (MIROC) description. K-1 technical report. 1. Hasumi, H. & Emori, S. (eds.). Center for Climate System Research, University of Tokyo. 34pp. http://www.ccsr.u-tokyo.ac.jp/kyosei/hasumi/MIROC/tech-repo.pdf

Kaplan, J.O. 2001. Geophysical applications of vegetation modelling. PhD Thesis, Lund University.

Kaplan, J.O. & New, M. 2006. Arctic climate change with a 2°C global warming: Timing, climate patterns and vegetation change. Climatic Change. 79. 213-241

Karas, C., Nurnberg, D., Gupta, A.K., Tiedemann, R., Mohan, K. & Bicket, T. 2009. Mid-Pliocene climate change amplified by a switch in Indonesian subsurface through flow. Nature Geoscience. 2 (6). 434-438.

Karas, C., Nurnberg, D., Tiedemann, R. & Garbe-Schonberg, D. 2010. Pliocene climate change of the Southwest Pacific and the impact of ocean gateways. Earth & Planetary Science Letters. doi:10.1016/j.epsl.2010.10.028

Katz, M.E., Cramer, B.S., Toggweiler, J.R., Esmay, G., Liu, C., Miller, K.G., Rosenthal, Y., Wade, B.S. & Wright, J.D. 2011. Impact of Antarctic Circumpolar Current development on late Paleogene ocean structure. Science. 332. 1076-1079. doi:10.1126/science.1202122

Keeling, C.D., Piper, S.C., Bacastow, R.B., Wahlen, M., Whorf, T.P., Heimann, M. & Meijer, H.A. 2005. Atmospheric CO_2 and $^{13}CO_2$ exchange with the terrestrial biosphere and oceans from 1978 to 2000: Observations and carbon cycle implications. 83-113. In Ehleringer, J.R., Cerling, T.E. & Dearing, M.D (Eds.). A History of Atmospheric CO_2 and its effects on Plants, Animals, and Ecosystems. Springer Verlag, New York.

Kennett, J.P. 1977. Cenozoic evolution of Antarctic glaciation, the circum-Antarctic Ocean, and their impact on global paleoceanography. Journal of Geophysical Research. 87. 3843–3860.

Kennett, J.P. & Stott, L.D. 1991. Abrupt deep-sea warming, palaeoceanographic changes and benthic extinctions at the end of the Paleocene. Nature. 353. 225–229. doi:10.1038/353225a0

Kleman, J. & Borgstrom, I. 1996. Reconstruction of palaeo-ice sheets: The use of geomorphological data. Earth Surface Processes and Landforms. 21. 893–909. doi:10.1002/(SICI)1096-9837

Klocke, D., Pincus, R. & Quaas, J. 2011. On constraining estimates of climate sensitivity with present-day observations through model weighting. Journal of Climate. 24. 6092-6099. doi:10.1175/2011JCLI4193.1

Knutson, T.R., Delworth, T.L., Dixon, K.W. & Stouffer. 1999. Model assessment of regional surface temperature trends (1949-1997). Journal of Geophysical Research. 104. 30981-30996.

Knutson, T.R., Delworth, T.L., Dixon, K.W., Held, I.M., Lu, J., Ramaswamy, V. Schwarzkopf, M.D., Stenchikov, G. & Stouffer, R.J. 2006. Assessment of twentieth-century regional surface temperature trends using the GFDL CM2 coupled models. Journal of Climate. 19. 1624–1651. doi: http://dx.doi.org/10.1175/JCLI3709.1

Knutti, R., Meehl, G.A., Allen, M.R. & Stainforth, D.A. 2006. Constraining climate sensitivity from the seasonal cycle in surface temperature. Journal of Climate, 19 (17), 4224-4233

Knutti, R. 2010. The end of model democracy? Climatic Change. 102. 395–404. doi:10.1007/s10584-010-9800-2.

Knutti, R., Masson, D. & Gettleman, A. 2013. Climate model genealogy: Generation CMIP5 and how we got there. Geophysical Research Letters. 40. 1194-1199. doi:10.1002/grl.50256.

Kohler, P., Bintanja, R., Fischer, H., Joos, F., Knutti, R., Lohmann, G. & Masson-Delmotte, V. 2010. What caused Earth's temperature variation during the last 800,000 years? Data based evidence on radiative forcing and constraints on climate sensitivity. Quaternary Science Reviews. 29 (1-2). 129-145. DOI: 10.1016/j.quascirev.2009.09.026

Kopp, G. & Lean, J.L. 2011. A new, lower value of total solar irradiance: Evidence and climate significance, Geophysical Research Letters. 38. L01706. doi:10.1029/2010GL045777.

Krantz, D.E. 1991. A chronology of Pliocene sea-level fluctuations: The U.S. middle Atlantic Coastal plain record. Quaternary Science Reviews. 10. 163-174.

Krinner, G., Lézine, A.-M., Braconnot, P., Sepulchre, P., Ramstein, G., Grenier, C. & Gouttevin, I. 2012. A reassessment of lake and wetland feedbacks on the North African Holocene climate. Geophysical Research Letters. 39. L07701. doi:10.1029/2012GL050992

Krebs, U., Park, W. & Schneider, B. 2011. Pliocene aridification of Australia caused by tectonically induced weakening of the Indonesian throughflow. Palaeogeography, Palaeoclimatology, Palaeoecology. 309 (1-2). 111-117.

Kuhlemann, J., Dunkl, I., Brugel, A., Spiegel, C. & Frisch, W. 2006. From source terrains of the Eastern Alps to the Molasse Basin: Detrial record of non-steady-state exhumation. Tectonophysics. 413. 301–316.

Kurschner, W.M., Van der Burgh, J., Visscher, H. & Dilcher, D.L. 1996. Oak leaves as biosensors of late Neogene and early Pleistocene paleoatmospheric CO_2 concentrations. Marine Micropaleontology. 27. 299–231. doi:10.1016/0377-8398(95)00067-4

Kurschner, W.M., Kvacek, Z. & Dilcher, D.L. 2008. The impact of Miocene atmospheric carbon dioxide fluctuations on climate and the evolution of terrestrial ecosystems. Proceedings of the National Academy of Science. 105. 449–453. doi:10.1073/pnas.0708588105

Kutzbach J. & Liu, Z. 1997. Response of the African Monsoon to Orbital Forcing and Ocean Feedbacks in the Middle Holocene. Science. 278. 440-443.

Kutzbach, J.E., Guetter, P.J., Ruddiman, W.F. & Prell, W.L. 1989. Sensitivity of climate to Late Cenozoic uplift in Southern Asia and the American west: Numerical experiments. Journal of Geophysical Research. 94 (15). 18. 393-318.

Lambert, S. J. & Boer, G.J. 2001. CMIP1 Evaluation and Intercomparison of Coupled Climate Models. Climate Dynamics. 17. 83-106.

LaRiviere, J.P., Ravelo, A.C., Crimmins, A., Dekens, P.S., Ford, H.L., Lyle, M. & Wara, M.W. 2012. Late Miocene decoupling of oceanic warmth and atmospheric carbon dioxide forcing. Nature. 486.

Laskar, J., Robutel, P., Joutel, F., Gastineau, M., Correia, A.C.M. & Levard, B. 2004. A long term numerical solution for the insolation quantities of the Earth. Astronomy and Astrophysics. 428. 261-285.

Lawrence, K.T., Liu, Z. & Herbert, T.D. 2006. Evolution of the Eastern Tropical Pacific through Plio-Pleistocene glaciations. Science. 312. 79-83.

Lawrence, K.T., Herbert, T.D., Dekens, P.S. & Ravelo, A.C. 2007. The application of the alkenone organic proxy to the study of Plio-Pleistocene climate. 539-562. In Williams, M., Haywood, A.M., Gregory, F.J. & Schmidt, D.N (Eds.) Deep-time perspectives on climate change: Marrying the signal from computer models and biological proxies. The Micropalaeontological Society. Special Publications. The Geological Society, London.

Lawrence, D.M., Oleson, K.W., Flanner, M.G., Fletcher, C.G., Lawrence, P.J., Levis, S., Swenson, S.C. & Bonan, G. B. 2012. The CCSM4 land simulation, 1850–2005: Assessment of surface climate and new capabilities. Journal of Climate. 25. 2240–2260. doi:10.1175/JCLI-D-11-00103.1

Lean, J. L. & Rind, D.H. 2009. How will Earth's surface temperature change in future decades? Geophysical Research Letters. 36. L15708. doi:10.1029/2009GL038932

Lear, C. 2007. Mg/Ca Palaethemometry: A new window on Cenozoic climate change. In Williams, M., Haywood, A.M., Gregory, F.J. & Schmidt, D.N (Eds.) Deep-time perspectives on climate change: Marrying the signal from computer models and biological proxies. The Micropalaeontological Society. Special Publications. The Geological Society, London.

Lear, C., Elderfield, H. & Wilson, P. 2000. Cenozoic deep-sea temperatures and global ice volumes from Mg/Ca in benthic foraminiferal calcite. Science. 287. 269-272.

Lear, C., Bailey, T., Pearson, P., Coxall, H. & Rosenthal, Y. 2008. Cooling and ice growth across the Eocene-Oligocene transition. Geology. 36. 251-254.

Levitus, S., Antonov, J. & Boyer. 2005. Warming of the world ocean 1955-2003. Geophysical Research Letters. 32. L02604. doi:10.1029/2004GL021592.

Lin, L.I-K. 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics. 45. 255-268.

Lisiecki, L.E. & Raymo, M.E. 2005. A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. Paleoceanography. 20. PA1003.

Liu, Z., Pagani, M. Zinniker, D., DeConto, R., Huber, M., Brinkhuis, H., Shah, S.R., Leckie, R.M. & Pearson, A. 2009. Global cooling during the Eocene-Oligocene climate transition. Science. 323. 1187-1190.

Livezey, R.E. & Chen, W.Y. 1983. Statistical field significance and its determination by Monte Carlo techniques, Monthly Weather Review. 111. 46-59

Lomax, R.G. & Hahs-Vaughn, D.L. 2007. Statistical Concepts: A Second Course. Taylor & Francis, Sussex. UK. ISBN 0-8058-5850-4

Lopez, A., Fung, F., New, M., Watts, G., Weston, A. & Wilby, R. 2009. From climate model ensembles to climate change impacts: a case study of water resource management in the South West of England. Water Resource Research. 45. W08419.

Lopez, A., Smith, L.A., & Suckling, E. In Prep. Pattern scaled climate change scenarios: are these useful for adaptation? Working paper, Grantham Research Institute on Climate Change and the Environment, London, UK

Lorenz, E.N. 1963. Deterministic nonperiodic flow. Journal of Atmospheric Sciences. 20. 130–141. doi: dx.doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

Lowenstein, T.K. & Demicco, R.V. 2006. Elevated Eocene atmospheric CO_2 and its subsequent decline. Science. 313. 1928. doi:10.1126/science.1129555

Lunt, D.J., Valdes, P.J., Haywood, A.M. & Rutt, I.C. 2008a. Closure of the Panama seaway during the Pliocene: Implications for climate and Northern Hemisphere glaciation. Climate Dynamics. 30. 1-18.

Lunt, D.J., Foster, G.L., Haywood, A.M. & Stone, E.J. 2008b, Late Pliocene Greenland glaciation controlled by a decline in atmospheric CO_2 levels. Nature. 454. 1102-1105.

Lunt, D.J., Flecker, R., Valdes, P.J., Salzmann, U., Gladstone, R. & Haywood, A.M. 2008c. A methodology for targeting palaeo proxy data acquisition: A case study for the terrestrial Late Miocene. Earth & Planetary Science Letters. 271. 53-62.

Lunt, D.J., Haywood, A.M., Foster, G.L. & Stone, E.J. 2009. The Arctic cryosphere in the mid-Pliocene and the future. Philosophical Transactions of the Royal Society A. 367. 49-67.

Lunt, D.J., Haywood, A.M., Schmidt, G.A., Salzmann, U., Valdes, P.J. & Dowsett, H.J. 2010. Earth system sensitivity inferred from Pliocene modelling and data. Nature Geoscience. 3. 60-64.

Lunt, D.J., Haywood, A.M., Schmidt, G.A., Salzmann, U., Valdes, P.J., Dowsett, H.J. & Loptson, C.A. 2012a. On the causes of mid-Pliocene warmth and polar amplification. Earth & Planetary Science Letters. 321-322. 128-138. doi:10.1016/j.epsl.2011.12.042

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

Lüthi, D., Le Floch, M., Bereiter, B., Blunier, T., Barnola, J-M, Siegenthaler, U., Raynaud, D., Jouzel, J., Fischer, H., Kawamura, K. & Stocker, T.F. 2008. High-resolution carbon dioxide concentration record 650,000–800,000 years before present. Nature. 453. 379-382. doi:10.1038/nature06949

Lutz, B.P. 2011. Shifts in North Atlantic planktic foraminifer biogeography and subtropical gyre circulation during the mid-Piacenzian warm period. Marine Micropalaeontology. 80 (3-4). 125-149.

Lyell, C. 1830. Principles of Geology. John Murray, Albermarle Street. London.

Marincovich, L. & Gladenkov, A.Y. 1999. Evidence for an early opening of the Bering Strait. Nature. 397 (6715). 149-151.

Markwick, P.J. 2007. The palaeogeographic and palaeoclimatic significance of climate proxies for data-model comparisons. 251-312. In Williams, M., Haywood, A.M., Gregory, F.J. & Schmidt, D.N (Eds.) Deep-time perspectives on climate change: Marrying the signal from computer models and biological proxies. The Micropalaeontological Society. Special Publications. The Geological Society, London.

Marques da Silva, C., Landau, B., Domenech, R. & Martinell, J. 2010. Pliocene Atlantic molluscan assemblages from the Mondego Basin (Portugal): Age and palaeoceanographic implications, Palaeogeography, Palaeoclimatology, Palaeoecology. 285 (3-4). 248-254

Masson, D. & Knutti, R. 2011. Spatial-scale dependence of climate model performance in the CMIP3 ensemble. Journal of Climate. 24 (11). 2680–2692. doi:10.1175/2011JCLI3513.1

Matthiessen, J., Knies, J., Vogt, C. & Stein, R. 2009. Pliocene palaeoceanography of the Arctic Ocean and sub Arctic seas. Philosophical Transactions of the Royal Society A. 367. 21–48.

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Hakk, H., Jungclaus, J., Klocke, D., Matei, D., Mikolajewicz., U., Notz, D., Pincus, R., Schmidt, H. & Tomassini, L. 2012. Tuning the climate of a global model. Journal of Advances in Modelling Earth Systems. 4. M00A01. doi:10.1029/2012MS000154.

May, W. 2008. Climatic changes associated with a global "2°C-stabilization" scenario simulated by the ECHAM5/MPI-OM coupled climate model. Climate Dynamics. 31 (2-3). 283-313.

McMillan, M.E., Heller, P.L. & Wing, S.L. 2006. History and causes of post-Laramide relief in the Rocky Mountain orogenic plateau. Geological Society of America Bulletin. 118 (3-4). 393-405. doi:10.1130/B25712.1

Meehl, G., Boer, G.J., Covey, C., Latif, M. & Stouffer, R.J. 2000. The Coupled Model Intercomparison Project. Bulletin of the American Meteorological Society. 81. 313-318.

Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J.F.B., Stouffer, R.J. & Taylor, E. 2007. The WCRP CMIP3 multi-model dataset: A new era in climate change research. Bulletin of the American Meteorological Society. 88. 1383-1394.

Micheels, A., Bruch, A.A., Uhl, D., Utescher, T. & Mosbrugger, V. 2007. A Late Miocene climate model simulation with ECHAM4/ML and its quantitative validation with terrestrial proxy data. Palaeogeography, Palaeoclimatology, Palaeoecology. 253. 251–270. doi:10.1016/j.palaeo.2007.03.042

Miller, J.R. & Russell, G.L. 2002. Projected impact of climate change on the energy budget of the Arctic Ocean by a global climate model. Journal of Climate. 15. 3028–3042.

Miller, K.G., Wright, J.D., Browning, J.V., Kulpecz, A., Kominz, A., Naish, T.R., Cramer, B.S., Rosenthal, Y., Peltier, W.R. & Sosdian, S. 2012. High tide of the warm Pliocene: Implications of global sea level for Antarctic deglaciation. Geology. 40 (5). 407-410.

Molteni, F. 2003. Atmospheric simulations using a GCM with simplified physical parameterizations I: Model climatology and variability in multi-decadal experiments. Climate Dynamics. 20. 175–191.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Binger, R.L., Harmel, R.D. & Veith, T. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE. 50 (3). 885-900.

Mosbrugger, V. & Utescher, T. 1997. The coexistence approach - a method for quantitative reconstructions of Tertiary terrestrial palaeoclimate data using plant fossils. Palaeogeography, Palaeoclimatology, Palaeoecology. 134. 61-86.

Moucha, R., Forte, A.M., Mitrovica, J.X., Rowley, D.B., Quere, S., Simmons, N.A. & Grand, S.P. 2008. Dynamic topography and long-term sea-level variations: There is no such thing as a stable continental platform: Earth & Planetary Science Letters. 271. 101–108. doi:10.1016/j.epsl.2008.03.056

Murray, J.W. 2001. The niche of benthic foraminifera, critical thresholds and proxies. Marine Micropaleontology. 41. 1-7.

Murphy, J.M., Sexton, D.M.H., Barnett, D.N., Jones, G.S., Webb, M.J., Collins, M. & Stainforth, D.A. 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature. 430. 768-772.

Murphy, J.M., Booth, B.B.B. Collins, M., Harris, G.R., Sexton, D.M.H. & Webb, M.J. 2007. A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. Philosophical Transactions of the Royal Society A. 365. 1993-2028.

Murphy, J.M. Sexton, D., Jenkins, G., Boorman, P., Booth, B., Brown, K., Clark, R., Collins, M., Harris, G. & Kendon, E. 2009a. UKCP09 climate change projections. ISBN 978-1-906360-02-3.

Murphy, J.M., Collins, M., Doblas-Reyes, F. & Palmer, T. 2009b. Development of ensemble prediction systems [Research Theme 1]. In: van der Linden, P. & Mitchell, J.F.B (Eds.). 2009b. ENSEMBLES: Climate change and its impacts: Summary of research and results from the ENSEMBLES project. Met Office Hadley Centre, Fitzroy Road, Exeter EX1 3PB, UK. 160pp

Murphy, J.M., Clark, R., Collins, M., Jackson, C., Rodwell, M., Rougier, J.C., Sanderson, B., Sexton, D.M.H. & Yokohata, T. 2011. Perturbed parameter ensembles as a tool for sampling model uncertainties and making climate projections. In Proceedings of ECMWF Workshop on Model Uncertainty. 20-24th June 2011. 183–208. Available online:http://www.ecmwf.int/publications/library/ecpublications/pdf/workshop/2011/Modeluncertainty/Murphy.pdf.

Murray, J.W. 2001. The niche of benthic foraminifera, critical thresholds and proxies. Marine Micropaleontology. 41 (1-2). 1–7.

Naish, T.R., Woolfe, K.J., Barrett, P.J., Wilson, G.S., Cliff, A., Bohaty, S.M., Bucker, C.J., Claps, M., Davey, F.J., Dunbar, G.B., Dunn, A.G., Fielding, C.R., Florindo, F., Hannah, M.J., Harwood, D.M., Henrys, S.A., Krissek, L.A., Lavelle, M., van der Meer, J.J.M., McIntosh, W.C., Niessen, F., Passchier, S., Powell, R.D., Roberts, A.P., Sagnotti, L., Scherer, R.P., Strong, P.C., Talarico, F., Verosub, K.L., Villa, G., Watkins, D.K., Webb, P-N. & Wonik, T. 2001. Orbitally induced oscillations in the East Antarctic ice sheet at the Oligocene/Miocene boundary. Nature. 413. 719–723.

Naish, T., Powell, R., Levy, R., Wilson, G., Scherer, R., Talarico, F., Krissek, L., Niessen, F., Pompilio, M., Wilson, T., Carter, L., DeConto, R., Huybers, P., McKay, R., Pollard, D., Ross, J., Winter, D., Barrett, P., Browne, G., Cody, R., Cowan, E., Crampton, J., Dunbar, G., Dunbar, N., Florindo, F., Gebhardt, C., Graham, I., Hannah, M., Hansaraj, D., Harwood, D., Helling, D., Henrys, S., Hinnov, L., Kuhn, G., Kyle, P., Läufer, A., Maffioli, P., Magens, D., Mandernack, K., McIntosh, W., Millan, C., Morin, R., Ohneiser, C., Paulsen, T., Persico, D., Raine, I., Reed, J., Riesselman, C., Sagnotti, L., Schmitt, D., Sjunneskog, C., Strong, P., Taviani, M., Vogel, S., Wilch, T. & Williams, T. 2009. Obliquity paced Pliocene West Antarctic ice sheet oscillations. Nature. 458. 322-328.

Neelin, J.D., Bracco, A., Luo, H., McWilliams J.C. & Meyerson, J.E. 2010. Considerations for parameter optimization and sensitivity in climate models. Proceedings of the National Academy of Sciences. 107. 21349-21354.

New, M., Lopez, A., Dessai, S. & Wilny, R. 2007. Challenges in using probabilistic climate change information for impact assessments: An example from the water sector. Philosophical Transactions of the Royal Society A. 365 (1857). 2117-2131

Niehörster, F. & Collins, M. 2009. A comparison of perturbed physics ensembles constructed with structural different models. ENSEMBLES Deliverable Report D1.14 of Working Group I

Ohkouchi, N., Kawamura, K., Kawahata, H., & Okada, H. 1999. Depth ranges of alkenone production in the central Pacific Ocean. Global Biogeochemical Cycles. 13 (2). 695–704. doi:10.1029/1998GB900024

Oreopoulos, L. & Khairoutdinov, M. 2003. Overlap properties of clouds generated by a cloud-resolving model. Journal of Geophysical Research. 108. 4479. doi:10.1029/2002JD003329.

Pagani, M., Zachos, J.C., Freeman, K.H., Tipple, B. & Bohaty, S. 2005. Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. Science. 309. 600–603. doi:10.1126/science.1110063

Pagani, M., Caldeira, K., Archer, D. & Zachos, J.C. 2006a. An ancient carbon mystery. Science. 314. 1556-1557.

Pagani, M., Pedentchouk, N., Huber, M., Sluijs, A., Schouten, S., Brinkhuis, H., Sinninghe-Damste, J.S., Dickens, G.R. & Expedition 302 Scientists. 2006b. Arctic hydrology during global warming at the Palaeocene/Eocene Thermal Maximum. Nature. 442. 671–675. doi:10.1038/nature05043

Pagani, M., Liu, Z. LaRiviere, J. & Ravelo, A.C. 2010. High Earth-system climate sensitivity determined from Pliocene carbon dioxide concentrations. Nature Geoscience. 31. 27-30.

PALAEOSENS Project Members. 2012. Making sense of palaeoclimate sensitivity. Nature. 491. 683-691. doi:10.1038/nature11574

Pälike, H., Lyle, M.W., Nishi, H., Raffi, I., Ridgwell, A., Gamage, K., Klaus, A., Acton, G., Anderson, L., Backman, J.,Baldauf, J., Beltran, C., Bohaty, S.M., Bown, P., Busch, W., Channell, J.E.T., Chun, C.O.J., Delaney, M., Dewangan, P., Dunkley Jones, T., Edgar, K.M., Evans, H., Fitch, P., Foster, G.L., Gussone, N., Hasegawa, H., Hathorne, E.C., Hayashi, H., Herrle, J.O., Holbourn, A., Hovan, S., Hyeong, K., Iijima, K., Ito, T., Kamikuri, S., Kimoto, K., Kuroda, J., Leon-Rodriguez, L., Malinverno, A., Moore Jr, T.C., Murphy, B.H., Murphy, D.P., Nakamura, H., Ogane, K., Ohneiser, C., Richter, C., Robinson, R., Rohling, E.J. Romero, O., Sawada, K., Scher, H., Schneider, L., Sluijs, A., Takata, H., Tian, J., Tsujimoto, A., Wade, B.S., Westerhold, T., Wilkens, R., Williams, T., Wilson, P.A., Yamamoto, Y., Yamamoto, S., Yamazaki, T. & Zeebe, R.E. 2012. A Cenozoic record of the equatorial Pacific carbonate compensation depth. Nature. 488. 609–614. doi:10.1038/nature11360

Panchuk, K., Ridgwell, A. & Kump, L.R. 2008. Sedimentary response to Paleocene–Eocene Thermal Maximum carbon release: A model-data comparison. Geology. 36. 315–318.

Pardaens, A.K., Banks, H.T., Gregory, J.M. & Rowntree, P.R. 2003. Freshwater transports in HadCM3. Climate Dynamics. 21. 177–195.

Pearson, P. & Palmer, M. 2000. Atmospheric carbon dioxide concentrations over the past 60 million years. Nature. 406. 695-699.

Pearson, P.N., Ditchfield, P.W., Singano, J., Harcourt-Brown, K.G., Nicholas, C.J., Olsson, R.K., Shackleton, N.J. & Hall, M.A. 2001. Warm tropical sea surface temperatures in the Late Cretaceous and Eocene epochs. Nature. 413. 481–487.

Pearson, P.N., Foster, G.L. & Wade, B.S. 2009. Atmospheric carbon dioxide through the Eocene-Oligocene climate transition. Nature. 461. 1110-1113. doi:10.1038/nature08447

Petit, J., Jouzel, J., Raynaud, D., Barkov, N., Barnola, J., Basile, I., Bender, M., Chappellaz, J., Davis, M., Delaygue, G., Delmotte, M., Kotlyakov, V.M., Legrand, M., Lipenkov, V.Y., Lorius, C., Pepin, L., Ritz, C., Saltzman, E. & Stievenard, M. 1999. Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. Nature. 399 (6735). 429–436.

Piani, C., Frame, D.J., Stainforth, D.A. & Allen, M.R. 2005. Constraints on climate change from a multi-thousand member ensemble of simulations. Geophysical Research Letters. 32. L23825. doi:10.1029/2005GL024452.

Pollard, D. & DeConto, R.M. 2009. Modelling West Antarctic ice sheet growth and collapse through the past five million years. Nature. 458. 329-332.

Poore, R.Z. 1981. Temporal and spatial distribution of ice-rafted mineral grains in Pliocene sediments of the North Atlantic: Implications for Late Cenozoic climatic history. 505–515. In: Warme, J.E.A. Douglas, R.G., and Winterer, E.L. (Eds.) The Deep Sea Drilling Project: A decade of progress. Society of Economic Paleontologists and Mineralogists Special Publication. 32.

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

Pope, J.O., Collins, M., Haywood, A.M., Dowsett, H.J., Hunter, S.J., Lunt, D.J., Pickering, S.J. & Pound, M.J. 2011. Quantifying Uncertainty in Model Predictions for the Pliocene (Plio-QUMP): Initial results, Palaeogeography, Palaeoclimatology, Palaoecology. 309 (1-2). 128-140. doi:10.1016/j.palaeo.2011.05.004

Pound, M.J., Haywood, A.M., Salzmann, U., Riding, J.B., Lunt, D.J. & Hunter, S.J. 2011. A Tortonian (Late Miocene, 11.61-7.25 Ma) global vegetation reconstruction. Palaeogeography, Palaeoclimatology, Palaeoecology. 300 (1-4). 29-45.

Pound, M.J., Tindall, J., Pickering, S.J., Haywood, A.M., Dowsett, H.J. & Salzmann, U. 2013. Late Pliocene lakes and soils: a data-model comparison for the analysis of climate feedbacks in a warmer world. Climate of the Past. 10 (1). 3175-3207. doi:10.5194/cp-10-167-2014.

Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A. & Solomon, A.M. 1992. A global biome model based on plant physiology and dominance, soil properties and climate. Journal of Biogeography. 19. 117-134.

Prescott, C., Haywood, A., Tindall, J., Dolan, A., Hunter, S., Pope, J. & Pickering, S. 2013. Climate response to changes in orbital forcing around the first Pliocene Time Slice. Geophysical Research Abstracts. 15. EGU2013-8134.

Prescott, C.L., Haywood, A.M., Dolan, A.M., Tindall, J., Hunter, S.J., Pickering, S.J. & Pope, J.O. 2014. Climate variability and Pliocene marine isotope stage KM5c. Earth & Planetary Science Letters. 400. 261-271. doi: 10.1016/j.epsl.2014.05.030

Quade, J., Cerling T.E. & Bowman, J.R. 1989. Development of Asian monsoon revealed by marked ecological shift during the latest Miocene in northern Pakistan. Nature. 342. 163–166. doi:10.1038/342163a0

Ramanathan, V., Barkstrom, B.R. & Harrison, E.F. 1989. Climate and the Earth's radiation budget. Physics Today. 42 (5). 22–32

Randall, D.A. & Wielicki, B.A. 1997. Measurements, Models and Hypotheses in the Atmospheric Sciences. Bulletin of the American Meteorological Society. 78. 399–406. doi: dx.doi.org/10.1175/1520-0477

Raper, S.C.B., Gregory, J.M. & Stouffer, R.J. 2002. The role of climate sensitivity and ocean heat uptake on AOGCM transient temperature response. Journal of Climate. 15. 124–130. doi: 10.1175/1520-0442(2002)015<0124:TROCSA>2.0.CO;2

Raymo, M.E., Grant, B., Horowitz, M. & Rau, G.H. 1996. Mid-Pliocene warmth: Stronger greenhouse and stronger conveyor. Marine Micropaleontology. 27. 313-326.

Raymo, M.E., Hearty, P., DeConto, R., O'Leary, M., Dowsett, H.J., Robinson, M.M. & Mitrovica, J.X. 2009. PLIOMAX: Pliocene maximum sea level project. PAGES News. 172.

Raymo, M.E., Mitrovica, J.X., O'Leary, M.J., DeConto, R.M. & Hearty, P.J. 2011. Departures from eustasy in Pliocene sea-level records. Nature Geoscience. 4. 328-332. doi:10.1038/NGEO1118.

Reed-Sterrett, C., Dekens, P.S., White, L.D. & Aiello, I.W. 2010. Cooling upwelling regions along the California margin during the early Pliocene: evidence for a shoaling thermocline Stratigraphy. 7 (2-3). 141-150.

Reynolds, R.W. & Smith, T.M. 1995. A high-resolution global sea surface temperature climatology. Journal of Climate. 8. 1571-1583.

Roberts, M.J. & Wood, R.A. 1997. Topographic sensitivity studies with a Bryan-Cox type ocean model. Journal of Physical Oceanography. 26. 1495-1527.

Robinson, M.M. 2009. New quantitative evidence of extreme warmth in the Pliocene Arctic. Stratigraphy. 64. 265-275.

Robinson, M.M., Valdes, P.J., Haywood, A.M., Dowsett, H.J., Hill, D.J. & Jones, S.M. 2011. Bathymetric controls on Pliocene North Atlantic and Arctic sea surface temperature and deepwater production. Palaeogeography, Palaeoclimatology, Palaoecology. 309 (1-2). 92-97. doi:10.1016/j.palaeo.2011.01.004

Roeckner, E., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kornblueh, L., Manzini, E., Schlese, U. & Schulzweida, U. 2006. Sensitivity of simulated climate to horizontal and vertical resolution in the ECHAM5 atmosphere model. Journal of Climate. 19. 3771–3791.

Rougier, J.C. 2007. Probabilistic inference for future climate using an ensemble of climate model evaluations. Climatic Change. 81. 247-264.

Rougier, J.C. & Crucifix, M. 2012. Uncertainty in climate science and climate policy. In Lloyd, L. & Winsberg, E. (Eds). Conceptual Issues in Climate Modeling, University of Chicago Press, forthcoming

Rougier J.C. & Sexton, D.M.H. 2007. Inference in ensemble experiments. Philosophical Transactions of the Royal Society A. 365. 2133-2143. doi:10.1098/rsta.2007.2071.

Rougier, J.C, Sexton, D.M.H., Murphy, J.M. & Stainforth, D. 2009. Analysing the climate sensitivity of the HadSM3 climate model using ensembles from different but related experiments. Journal of Climate. 22. 3540-3557.

Rougier, J.C., Edwards, T.L., Collins, M. & Sexton, D.M.H. 2011. Low noise projections of complex simulator output: A useful took when checking for code errors. In Proceedings of ECMWF Workshop on Model Uncertainty. 20-24th June 2011. 183–208. Available online: http://www.ecmwf.int/publications/library/ecpublications/pdf/workshop/2011/Modeluncertainty/Rougier.pdf.

Rowell, D.P. 2012. Sources of uncertainty in future changes in local precipitation. Climate Dynamics. 39 (7-8). 1929-1950. doi:10.1007/s00382-011-1210-2.

Rowlands, D.J., Frame, D.J., Ackerley, D., Aina, T., Booth, B.B.B., Christensen, C., Collins, M., Fauli, N., Forest, C.E., Grandey, B.S., Greyspeerft, E., Highwood, E.J., Ingram, W.J., Knight, S., Lopez, A., Massey, N., McNamara, F., Meinshausen, N., Piani, C., Rosier, S.M., Sanderson, B.M., Smith, L.A., Stone, D.A., Thurston, M., Yamazaki, K., Yamazaki, Y.H. & Allen, M.R. 2012. Broad range of 2050 warming from an observational constrained large climate model ensemble. Nature Geoscience. 5. 356-260.

Ruddiman, W.F. 2003. Orbital insolation, ice volume and greenhouse gases. Quaternary Science Reviews. 22. 1597–1629.

Ruddiman, W.F., McIntyre, A. & Raymo, M. 1986. Paleoenvironmental results from North Atlantic Sites 607 and 609. Initial Reports of the Deep Sea Drilling Project. 94. 855–878.

Sagoo, N., Valdes, P., Flecker, R. & Gregoire, L. 2013. The Early Eocene equable climate problem: Can a perturbed physics climate model ensemble identify possible solutions? Philosophical Transactions of the Royal Society A. 371 (2001). 20130123. doi.org/10.1098/rsta.2013.0123

Salzmann, U., Haywood, A.M., Lunt D.J., Valdes, P.J. & Hill, D.J. 2008. A new global biome reconstruction and data-model comparison for the middle Pliocene. Global Ecology and Biogeography. 17. 432-447. doi:10.1111/j.1466-8238.2008.00381.x.

Salzmann, U., Haywood, A.M., & Lunt, D.J. 2009. The past is a guide to the future? Comparing Middle Pliocene vegetation with predicted biome distributions for the twenty-first century. Philosophical Transactions of the Royal Society A. 367. 189-204. doi:10.1098/rsta.2008.0200.

Salzmann, U., Nelson, A.E., Riding, J.B. & Smellie, J.L. 2011. How likely is a green Antarctic Peninsula during warm Pliocene interglacials? A critical reassessment based on new palynofloras from James Ross Island. Palaeogeography, Palaeoclimatology, Palaeoecology. 309 (1-2). 73-82. doi:10.1016/j.palaeo.2011.01.028.

Salzmann, U., Dolan, A.M., Haywood, A.M., Chan, W-L., Hill, D.J., Abe-Ouchi, A., Otto-Bliesner, B., Bragg, F., Chandler, M.A., Countoux, C., Dowsett, H.J., Jost, A., Kamae, Y., Lohmann, G., Lunt, D.J., Pickering, S.J., Pound, M.J., Ramstein, G., Rosenbloom, N.A., Sohl, L., Stepanek, C., Ueda, H. Zhang, Z. 2013. Data-model discord reveals uncertainties in simulating terrestrial warming of the Pliocene. Nature Climate Change.

Sanderson, B.M. 2007. Feedback mechanisms and constraints on climate sensitivity from a perturbed physics ensemble of general circulation models. PhD Thesis. University of Oxford. 184pp.

Sanderson, B.M. 2011. A multimodel study of parametric uncertainty in predictions of climate response to rising greenhouse gas concentrations. Journal of Climate. 24. 1362-1377.

Sanderson, B.M. 2013. On the estimation of systematic error in regression-based predictions of climate sensitivity. Climatic Change. 118 (3-4). 757-770.

Sanderson, B.M., Piani, C., Ingram, W.J., Stone, D.A. & Allen, M.R. 2008a. Towards constraining climate sensitivity by linear analysis of feedback patterns in thousands of perturbed physics GCM simulations. Climate Dynamics. 30. 175-190.

Sanderson, B.M., Knutti, R., Aina, T., Christensen, C., Faull, N., Frame, D.J., Ingram, W.J., Piani, C., Stainforth, D.A., Stone, D.A. & Allen, M.R. 2008b. Constraints on model response to greenhouse gas forcing and the role of sub grid-scale processes. Journal of Climate. 21. 2384-2400.

Sarnthein, M., Bartoli, G., Prange, M., Schmittner, A., Schneider, B., Weinelt, M., Andersen, N. & Garbe-Schönberg, D. 2009. Mid-Pliocene shifts in ocean overturning circulation and the onset of Quaternary-style climates. Climate of the Past. 5. 269–283.

Schouten, S., Hopmans, E.C., Schefuss, E. & Sinninghe-Dampsté, J.S. 2002. Distributional variations in marine crenarchaeotal membrane lipids: A new tool for reconstructing ancient sea water temperatures? Earth & Planetary Science Letters. 204. 265-274. doi:10.1016/S0012-821X(02)00979-2.

Scroxton, N., Bonham, S.J., Rickaby, R.E.M., Lawrence, S.H.F., Hermoso, M. & Haywood, A.M. 2011. Persistent El Niño-Southern Oscillation variation during the Pliocene Epoch. Paleoceanography. 26. doi:10.1029/2010PA002097

Seager, R., Battisti, D.S., Yin, J., Gordon, N., Naik, N., Clement, A. C. & Cane, M.A. 2002. Is the Gulf Stream responsible for Europe's mild winters. Quarterly Journal of the Royal Meterological Society. 128. 2563–2586.

Seki, O., Foster, G.L., Schmidt, D.N., Mackensen, A., Kawamura, K. & Pancost, R.D. 2010. Alkenone and Boron-based Pliocene pCO_2 records. Earth & Planetary Science Letters. 292 (1-2). 201-211.

Sexton, D.M.H., Murphy, J.M., Collins, M. & Webb, M.J. 2012a. Multivariate probabilistic projections using imperfect climate models. Part I: outline of methodology. Climate Dynamics. 38 (11-12). 2513-2542. doi:10.1007/s00382-011-1208-9

Sexton, D.M.H., Murphy, J.M., Collins, M. & Webb, M.J. 2012b. Multivariate prediction using imperfect climate models. Part II: Using a large number of observations to constrain a prediction. Climate Dynamics. 38 (11-12). 2543-2558.

Shaffer, G. & Bendtsen, J. 1994. Role of the Bering Strait in controlling North Atlantic ocean circulation and climate. Nature. 367. 354–357.

Shellito, C.J., Sloan, L.C. & Huber, M. 2003. Climate model sensitivity to atmospheric CO_2 levels in the Early-Middle Paleogene. Palaeogeography, Palaeoclimatology, Palaeoecology. 193. 113–123.

Shiogama, H., Wanatabe, M., Yoshimori, T., Yokohata, T., Ogura, T., Anna, J., Hargreaves, J., Abe, M., Kamae, Y., O'ishi, R., Nobui, R., Emori, S., Nozawa, A., Abe-Ouchi, A. & Kimoto, M. 2012. Physics parameter uncertainty and observational constrains to climate feedback: An ensemble of coupled atmosphere ocean GCMs without flux corrections. Climate Dynamics. 39. 3041-3056

Shukla, J. 1998. Predictability in the midst of chaos: A scientific basis for climate forecasting. Science. 282. 728. DOI: 10.1126/science.282.5389.728

Shukla, J., Palmer, T. N., Hagedorn, R., Hoskins, B., Kinter, J., Marotzke, J., Miller, M. & Slingo, J. 2010. Toward a new generation of world climate research and computing facilities. Bulletin of the American Meteorological Society. 91. 1407–1412. doi:10.1175/2010BAMS2900.1

Singh, J., Knapp, H.V. & Demissie, M. 2004. Hydrologic modeling of the Iroquois River watershed using HSPF and SWAT.ISWS CR 2004-08. Champaign, Ill. Illinois State Water Survey. Available at: www.sws.uiuc.edu/pubdoc/CR/ISWSCR2004-08.pdf.

Sloan, L.C., Crowley, T.J. & Pollard, D. 1996. Modelling of Middle Pliocene climate with the NCAR GENESIS general circulation model. Marine Micropaleontology. 27. 51-61.

Sluijs, A., Brinkhuis, H., Schouten, S., Bohaty, S.M., John, C.M., Zachos, J.C., Reichart, G-J., Sinninghe-Damsté, J.S., Crouch, E.M. & Dickens, G.R. 2007. Environmental precursors to rapid light carbon injection at the Palaeocene/Eocene boundary. Nature. 450. 1218–1221. doi:10.1038/nature06400

Soden, B.J., Held, I.M., Colman, R., Shell, K.M., Kiehl, J.T., & Shields, C.A. 2008. Quantifying climate feedbacks using radiative kernels. Journal of Climate. 21. pp.3504-3520

Sohl, L.E., Chandler, M.A., Schnunk, R.B., Mankoff, K., Jonas, J.A., Foley, K.M. & Dowsett, H.J. 2009. PRISM3/GISS topographic reconstruction, edited. United States Geological Survey.

Spicer, R.A., Harris, N.B.W., Widdowson, M., Herman, A.B., Guo, S., Valdes, P.J., Wolfe, J.A. & Kelley, S.P. 2003. Constant elevation of southern Tibet over the past 15 million years. Nature. 412. 622–624. doi:10.1038/nature01356

Stainforth, D.A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D.J., Kettleborough, J.A., Knight, S., Martin, A., Murphy, J.M., Piani, C., Sexton, D., Smith, L.A., Spicer, R.A., Thorpe, A.J. & Allen, M.R. 2005. Uncertainty in predictions of the climate response to rising levels of greenhouse gases. Nature. 433. 403-406.

Stott, P.A. & Forest, C.E. 2007. Ensemble climate predictions using climate models and observational constraints. Philosophical Transactions of the Royal Society A. 365. 2029-2052.

Student. 1908. The probable error of a mean. Biometrika. 6. 1-25

Stults, D.Z., Wagner-Cremer, F. & Axsmith, B.J. 2011. Atmospheric paleo-CO_2 estimates based on Taxodium distichum (Cupressaceae) fossils from the Miocene and Pliocene of Eastern North America. Palaeogeography, Palaeoclimatology, Palaeoecology. 309. 327-332.

Tebaldi, C. & Knutti, R. 2007.The use of the multimodel ensemble in probabilistic climate projections. Philosophical Transactions of the Royal Society A. 365. 2053-2075.

Tett, S.F.B., Mineter, M.J., Cartis, C., Rowlands, D.J. & Liu, P. 2013. Can top of the atmosphere radiation measurements constrain climate predictions? Part 1: Tuning. Journal of Climate.

Texier, D., de Noblet, N., Harrison, S.P., Haxeltine, A., Jolly, D., Joussaume, S., Laarif, F., Prentice, I.C. & Tarasov, P. 1997. Quantifying the role of biosphere-atmosphere feedbacks in climate change: coupled model simulation for 6000 years BP and comparison with palaedata for northern Eurasia and Africa. Climate Dynamics. 13. 865-882

Thompson, R.S. & Fleming, R.F. 1996. Middle Pliocene vegetation: reconstructions, paleoclimatic inferences and boundary conditions for climate modelling. Marine Micropaleontology. 27. 27-49.

Tindall, J., Valdes, P.J. & Sime, L. 2008. Stable water isotopes in HadCM3: The isotopic signature of ENSO and the tropical amount effect. Journal of Geophysical Research. 114. D04111

Trenberth, K.E., Stepaniak, D.P. & Caron, J.M. 2000. The global monsoon as seen through the divergent atmospheric circulation. Journal of Climate. 13. 3969-3993.

Tripati, A.K., Roberts, C.D. & Eagle, R.A. 2009. Coupling of CO_2 and ice sheet stability over major climatic transitions of the past 20 million years. Science. 326 (5958). 1394-1397.

Twohy, C.H., Schanot, A.J. & Cooper, W.A. 1997. Measurement of condensed water content in liquid and ice clouds using an airborne counterflow virtual impactor. Journal of Atmospheric & Oceanic Technology. 14. 197-202.

Tziperman, E., Toggweiler, J.R., Feliks, Y. & Bryan, K. 1994. Instability of the thermohaline circulation with respect to mixed boundary conditions: Is it really a problem for realistic models? Journal of Physical Oceanography. 24. 218-232.

Valdes, P.J. 2011. Built for stability. Nature Geoscience. 4. 414-416.

Van de Wal, R.S.W., de Boer, B., Lourens, L.J., Köhler, P. & Bintanja, R. 2011. Reconstruction of a continuous high-resolution CO_2 record over the past 20 million years. Climate of the Past. 7. 1459-1469. doi:10.5194/cp-7-1459-2011

Von der Heydt, A. & Dijkstra, H.A. 2008. The effect of ocean gateways on ocean circulation patterns in the Cenozoic. Global Planetary Change. 62. 132–146.

Watanabe, M., Shiogama, H., Yokohata, T., Kamae, Y., Yoshimori, M., Ogura, T., Annan, J.D., Hargreaves, J.C., Emori, S. & Kimoto, M. 2012. Using a multiphysics ensemble for exploring diversity in cloud–shortwave feedback in GCMs. Journal of Climate. 25. 5416–5431. doi:10.1175/JCLI-D-11-00564.1

Wardlaw, B.R. & Quinn, T.M. 1991. The record of Pliocene sea-level change at Enewetak Atoll. Quaternary Science Reviews. 10. 247–258.

Watterson, I.G. 1996. Non-dimensional measures of climate model performance. International Journal of Climatology. 16(4). 379–391.

Weaver, A.J., Eby, M., Fanning, A.F. & Wiebe, E.C. 1998. Simulated influence of carbon dioxide, orbital forcing and ice sheets on the climate of the last glacial maximum. Nature. 394. 847–853.

Webb, M.J., Senior, C.A., Sexton, D.M.H., Ingram, W.J., Williams, K.D., Ringer, M.A., McAveney, B.J., Colman, R., Soden, B.J., Gudgel, R., Knutson, T., Emori, S., Ogura, T., Tsushima, Y., Andronova, N., Li, B., Musat, I., Bony, S. & Taylor, K.E. 2006. On the contribution of local feedback mechanisms to the range of climate sensitivity in two GCM ensembles. Climate Dynamics. 27. 17-38.

Webb, M.J., Lambert, F.H. & Gregory, J.M. 2013. Origins of differences in climate sensitivity, forcing and feedback in climate models. Climate Dynamics. 40 (3-4). 677-707.

Weigel, A.P., Knutti, R., Liniger, M.A. & Appenzeller, C. 2010. Risks of model weighting in multimodel climate projections. Journal of Climate. 23. 4175–4191. doi:10.1175/2010JCLI3594.1

Weijers, J.W.H., Schouten, S., Slujis, A., Brinkhuis, H. & Sinninghe-Damste, J.S. 2007. Warm Arctic continents during the Palaeocene-Eocene thermal maximum. Earth & Planetary Science Letters. 261. 230–238.

Wigley, T.M.L. & Santer, B.D. 1990. Statistical comparison of spatial fields in model validation, perturbation and predictability experiments. Journal of Geophysical Research. 95.

Willard, D.A. 1997. Plio-Pleistocene pollen assemblages from the Yermak Plateau, Arctic Ocean: ODP Sites 910 and 911. Proceedings of the Ocean Drilling Program, Scientific Results. 151. 297-305.

Willeit, M., Ganopolski, A. & Feulner, G. 2013 On the effect of orbital forcing on mid-Pliocene climate, vegetation and ice sheets, Climate of the Past. 9. 1749-1759. doi:10.5194/cp-9-1749-2013.

Williams, K.D.S., Senior, C.A. & Mitchell, J.F.B. 2001. Transient climate change in the Hadley Centre models: The role of physical processes. Journal of Climate. 14. 2659-2674.

Williams, M., Haywood, A.M., Taylor, S.P., Valdes, P.J., Sellwood, B.W. & Hillenbrand, C.D. 2005. Evaluating the efficacy of planktonic foraminifer calcite $\delta^{18}O$ data for sea surface temperature reconstruction for the Late Miocene. Geobios. 38 (6). 843–863.

Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L. & Yamazaki. 2013. History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. Climate Dynamics. 41 (7-8). 1703-1729.

Wolfe, J.A. 1993. A method of obtaining climatic parameters from leaf assemblages. U.S. Geological Survey Bulletin. 2040. 73

Wolff, E.W., Barbante, C., Becagli, S., Bigler, M., Boutron, C.F., Castellano, E., de Angelis, M., Federer, U., Fischer, H., Funde, F., Hansson, M., Hutterli, M., Jonsell, U., Karlin, T., Kaufmann, P., Lambert, F., Littot, G.C., Mulvaney, R., Röthlisberger, R., Ruth, R., Severi, M., Siggaard-Andersen, M.L., Sime, L.C., Steffensen, J.P., Stocker, T.F., Traversi, R., Twarloh, B., Udisti, R., Wagenbach, D. & Wegner, A. 2010. Changes in environment over the last 800,000 years from chemical analysis of the EPICA Dome C ice core. Quaternary Science Reviews. 29. 285-295.

Xie, S., Xu, K.M., Cederwall., R.T., Bechtold, P., Del Genio, A.D., Klein, S.A., Cripe, D.G., Ghan, S.J., Gregory, D., Lacoballis, S.F., Krueger, S.K., Lohmann, U., Petch, J.C., Randall, D.A., Rotstayn, L.D., Somerville, R.C.J., Sud, Y.C., Von Salzen, K., Walker, G.K., Wolf, A., Yio, J.J., Zhang, G.J. & Zhang, M. 2002. Intercomparison and evaluation of cumulus parameterizations under summertime mid-latitude continental conditions. Quarterly Journal of the Royal Meteorological Society. 128. 1095-1135.

Xu, K.M., Cederwall, R.T., Donner, L.J., Grabowski, W.W., Guichard, F., Johnson, D.E., Khairoutdinov, M., Krueger, S.K., Petch, J.C., Randall, D.A., Seman, C.J., Tao, W-K., Wang, D., Xie, S., Yio, J.J. & Zhang, M.H. 2002. An intercomparison of cloud-resolving models with the atmospheric radiation measurement summer 1997 intensive observation period data. Quarterly Journal of the Royal Meteorological Society. 128. 593–624. doi:10.1256/003590002321042117

Yokohata, T., Webb, M.J., Collins, M., Williams, K.D., Yoshimori, M., Hargreaves, J.C. & Annan, J. 2010. Structural similarities and differences in climate response to CO_2 increase between two perturbed physics ensembles. Journal of Climate. 23. 1392-1410.

Yokohata, T., Annan, J.D., Collins, M., Jackson, C.S., Tobis, M., Webb, M.J. & Hargreaves, J.C. 2012. Reliability of multi-model and structurally different single-model ensembles. Climate Dynamics. 39 (3-4). 599-616. doi:10.1007/s00382-011-1203-1.

Yokohata, T., Annan, J.D., Collins, M., Jackson, C.S., Shiogama, H., Watanabe, M., Emori, S., Yoshimori, M., Abe, M., Webb, M.J. & Hargreaves, J.C. 2013. Reliability and importance of structural diversity of climate model ensembles. Climate Dynamics. doi: 10.1007/s00382-013-1733-9

Yoshimori, M., Hargreaves, J.C., Annan, J.D. & Yokohata, T. 2011. Dependency of feedbacks on forcing and climate state in physics parameter ensembles. Journal of Climate. 24. 6440-6455.

You, Y., Huber, M., Muller, D., Poulsen, C. & Ribbe, J. 2009. Simulation of the Middle Miocene Climate Optimum. Geophysical Research Letters. 36. L04702. doi:10.1029/2008GL036571

Zachos, J.C., Quinn, T.M. & Salamy, K. 1996. High resolution (104 yr) deep-sea foraminiferal stable isotope time series. Paleoceanography. 11. 251-266.

Zachos, J.C., Flower, B.P. & Paul, H. 1997. Orbitally paced climate oscillations across the Oligocene/Miocene boundary. Nature. 388. 567-571.

Zachos, J.C., Pagani, M., Sloan, L., Thomas, E. & Billups, K. 2001. Trends, rhythms and aberrations in global climate 65 Ma to Present. Science. 292. 686.

Zachos, J.C., Rohl, U., Schellenberg, S.A., Sluijs, A., Hodell, D.A., Kelly, D.C., Thomas, E., Nicolo, M., Raffi, I., Lourens, L.J., McCarren, H. & Kroon, D. 2005. Rapid acidification of the ocean during the Paleocene-Eocene thermal maximum. Science. 308 (5728). 1611-1615. doi:10.1126/science.1109004

Zachos, J.C., Dickens, G.R. & Zeebe, R.E. 2008. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. Nature. 451. doi:10.1038/nature.

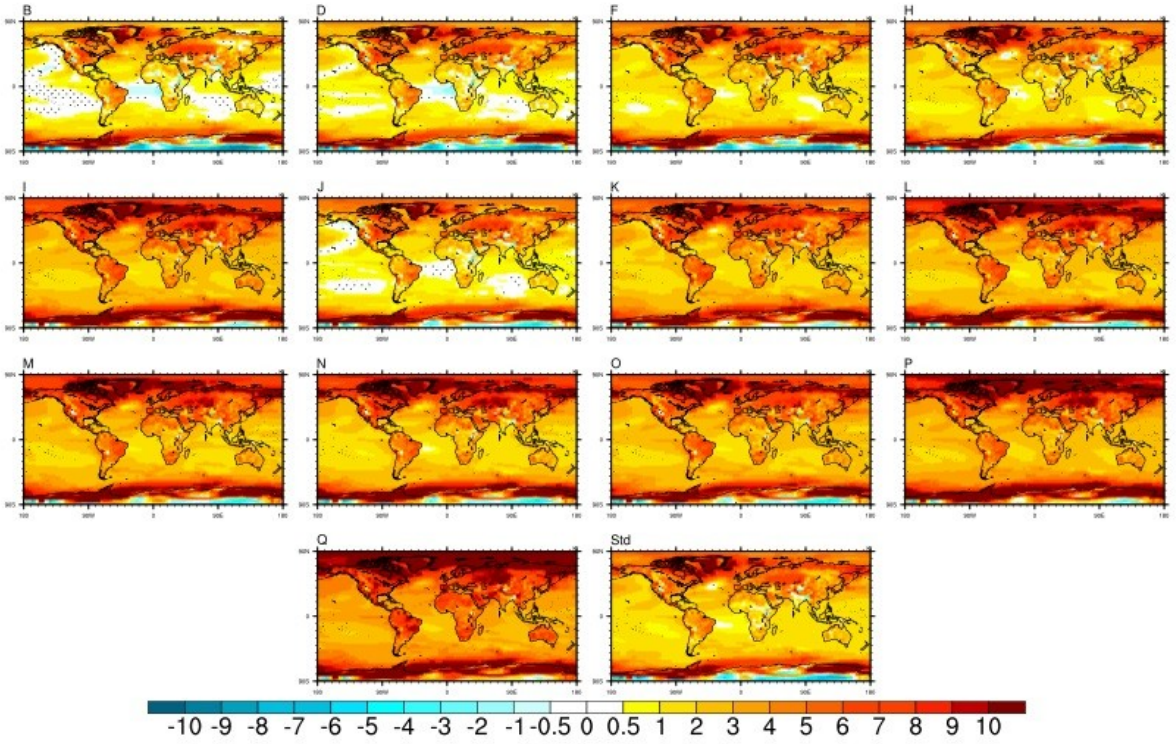
Zeebe, R.E., Zachos, J.C. & Dickens, G.R. 2009. Carbon dioxide forcing alone insufficient to explain Palaeocene–Eocene Thermal Maximum warming. Nature Geoscience. 2. 576-580. doi:doi:10.1038/ngeo578

Zhang, Z.S., Nisancioglu, K., Bentsen, M., Tjiputra, J., Bethke, I., Yan, B., Risebrobakken, B., Andsersson, C. & Jansen, E. 2012. Pre-industrial and mid-Pliocene simulations with NorESM-L. Geoscientific Model Development. 5. 523-533.

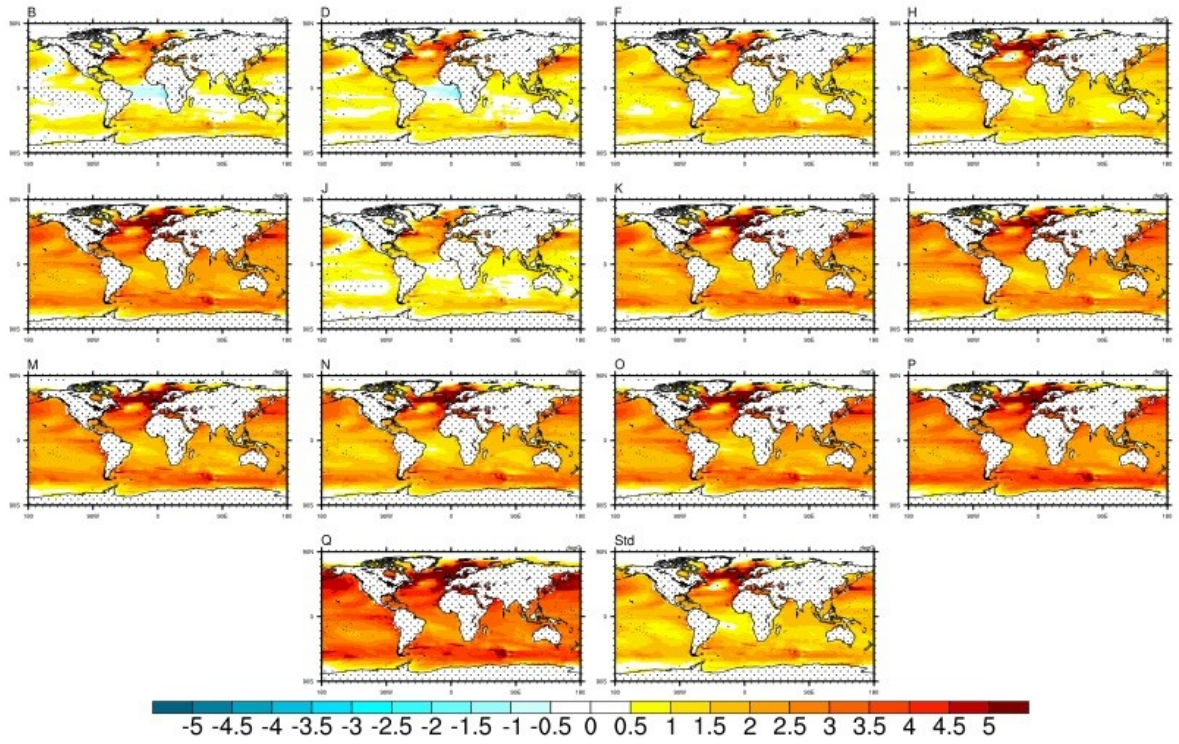
Zhang, Z-S., Nisancioglu, K.H., Chandler, M.A., Haywood, A.M., Otto-Bliesner, B.L., Ramstein, G., Stepanek, C., Abe-Ouchi, A., Chan, W-L., Bragg, F.J., Contoux, C., Dolan, A.M., Hill, D.J., Jost, A., Kamae, Y., Lohmann, G., Lunt, D.J., Rosenbloom, N.A., Sohl, L.E. & Ueda, H. 2013. Mid-Pliocene Atlantic meridional overturning circulation not unlike modern? Climate of the Past Discussions. 9. 1297-1319. doi:10.5194/cpd-9-1297-2013

Zheng, W., Zhang, Z., Chen, L., & Yu, Y. 2013. The mid-Pliocene climate simulated by FGOALS-g2. Geoscientific Model Development. 6. 1127-1135. doi:10.5194/gmd-6-1127-2013

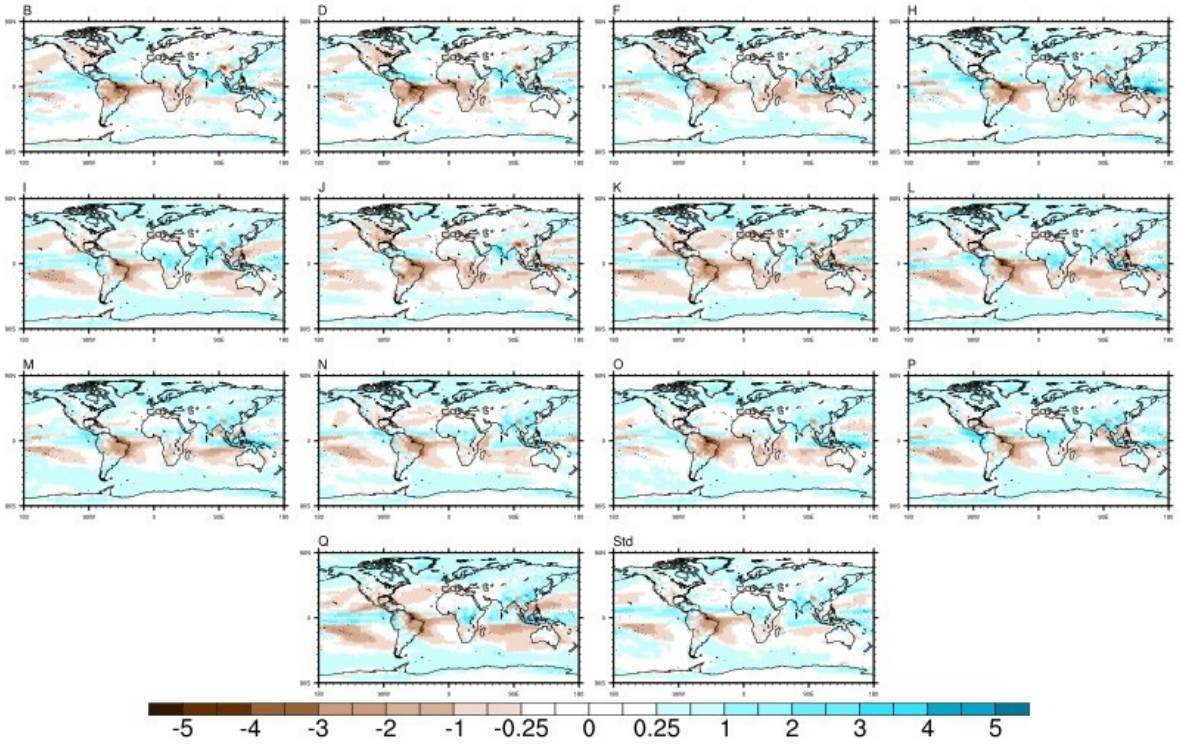
Appendix A



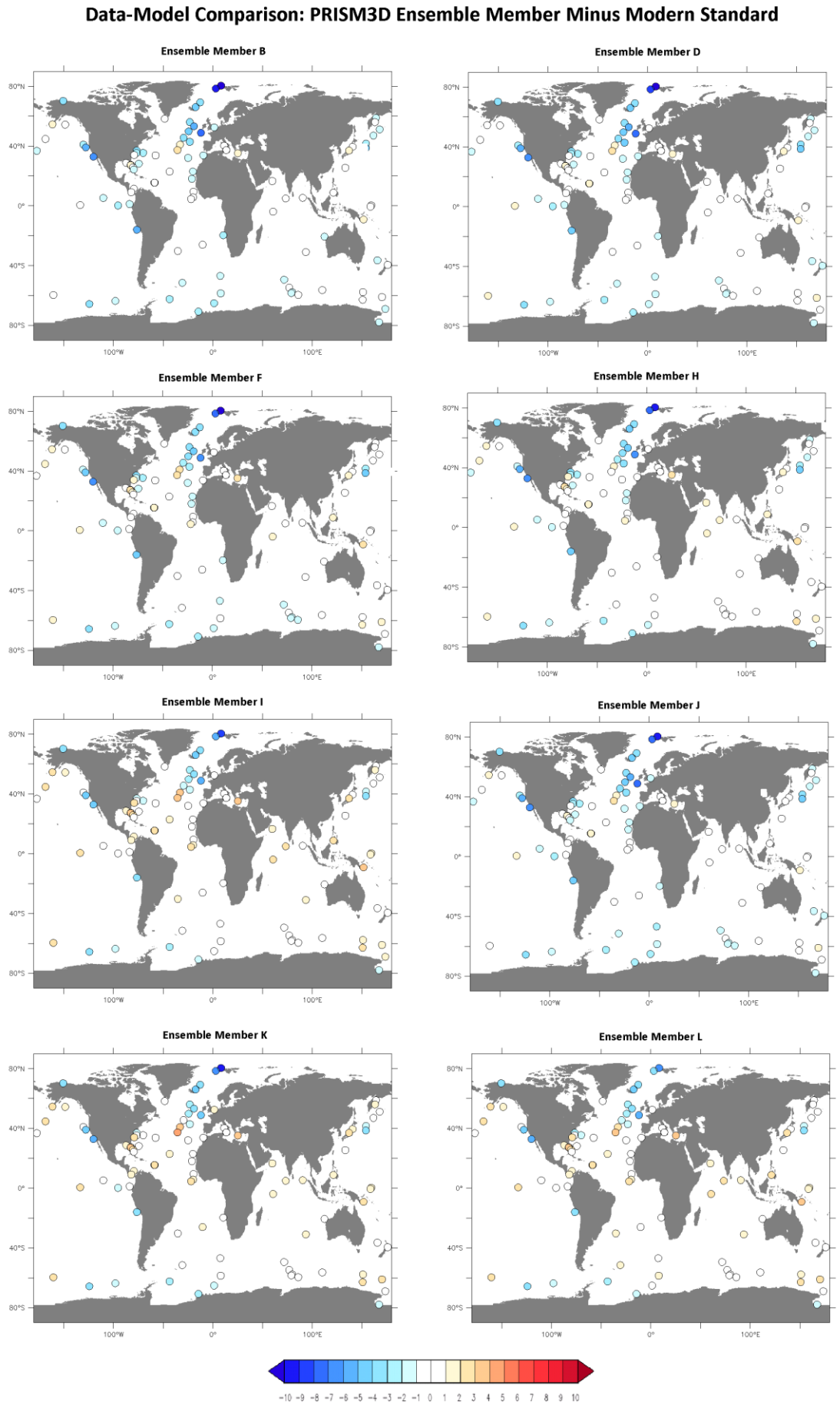
A1: Global annual mean plots for *"Pliocene ensemble member minus modern Standard"* for surface air temperature (SAT - °C) for all 13 ensemble members and the Standard. The coloured outlines represent which grouping the ensemble member belongs to either 'colder than Standard' (blue) or 'warmer than Standard' (red).

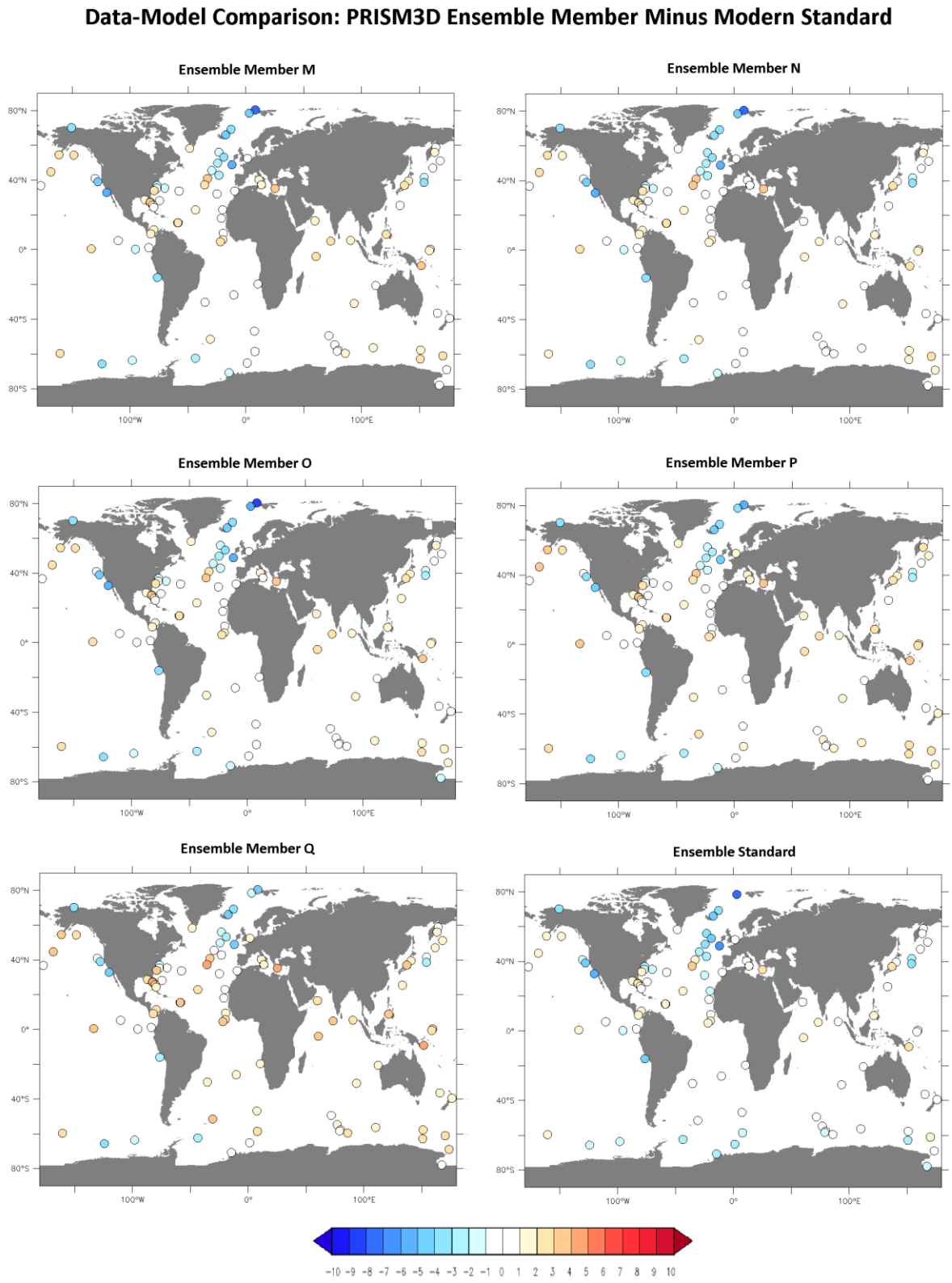


A2: Global annual mean plots for *"Pliocene ensemble member minus modern Standard"* for sea surface temperature (SST - °C) for all 13 ensemble members and the Standard. The coloured outlines represent which grouping the ensemble member belongs to either 'colder than Standard' (blue) or 'warmer than Standard' (red).

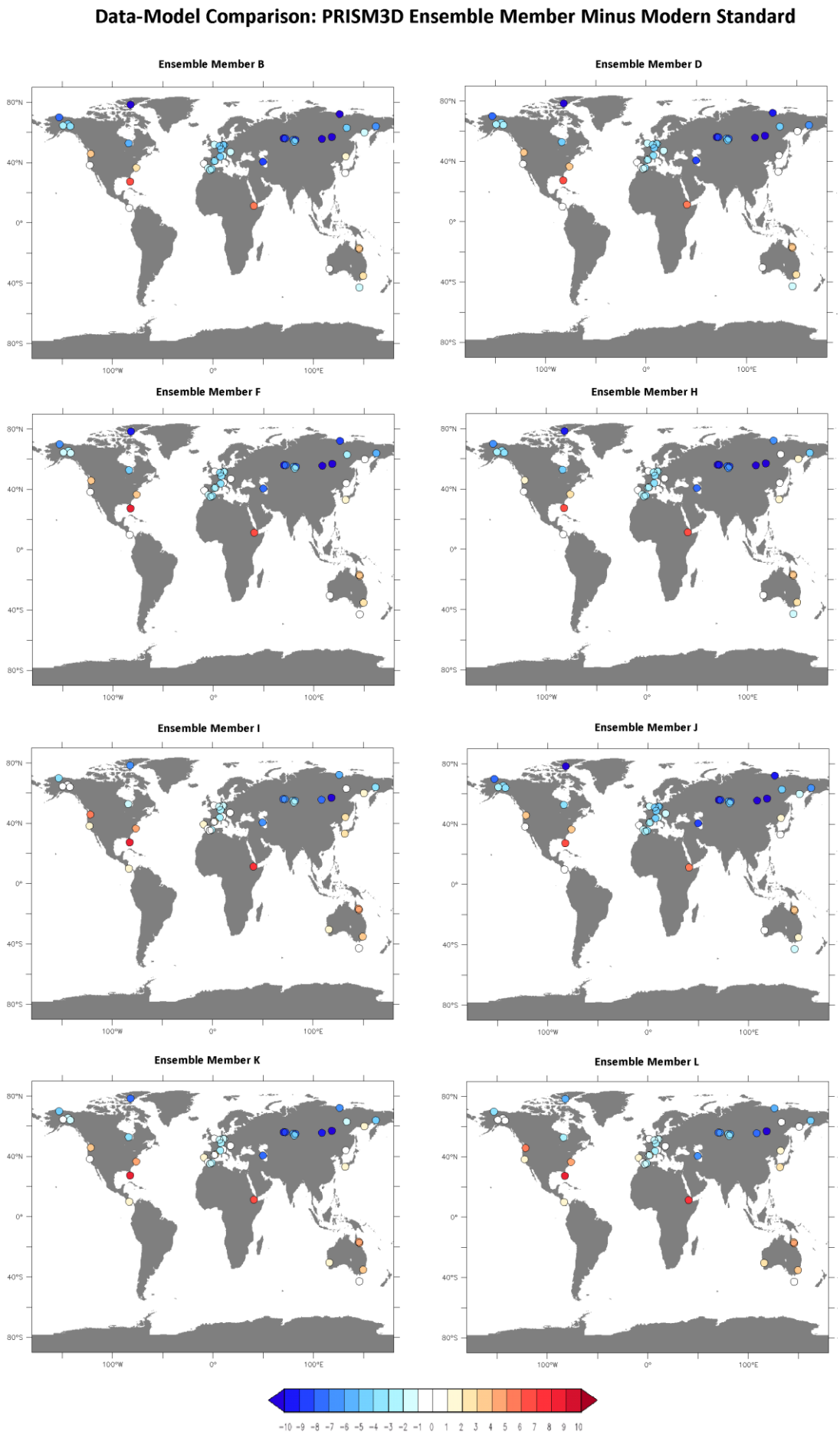


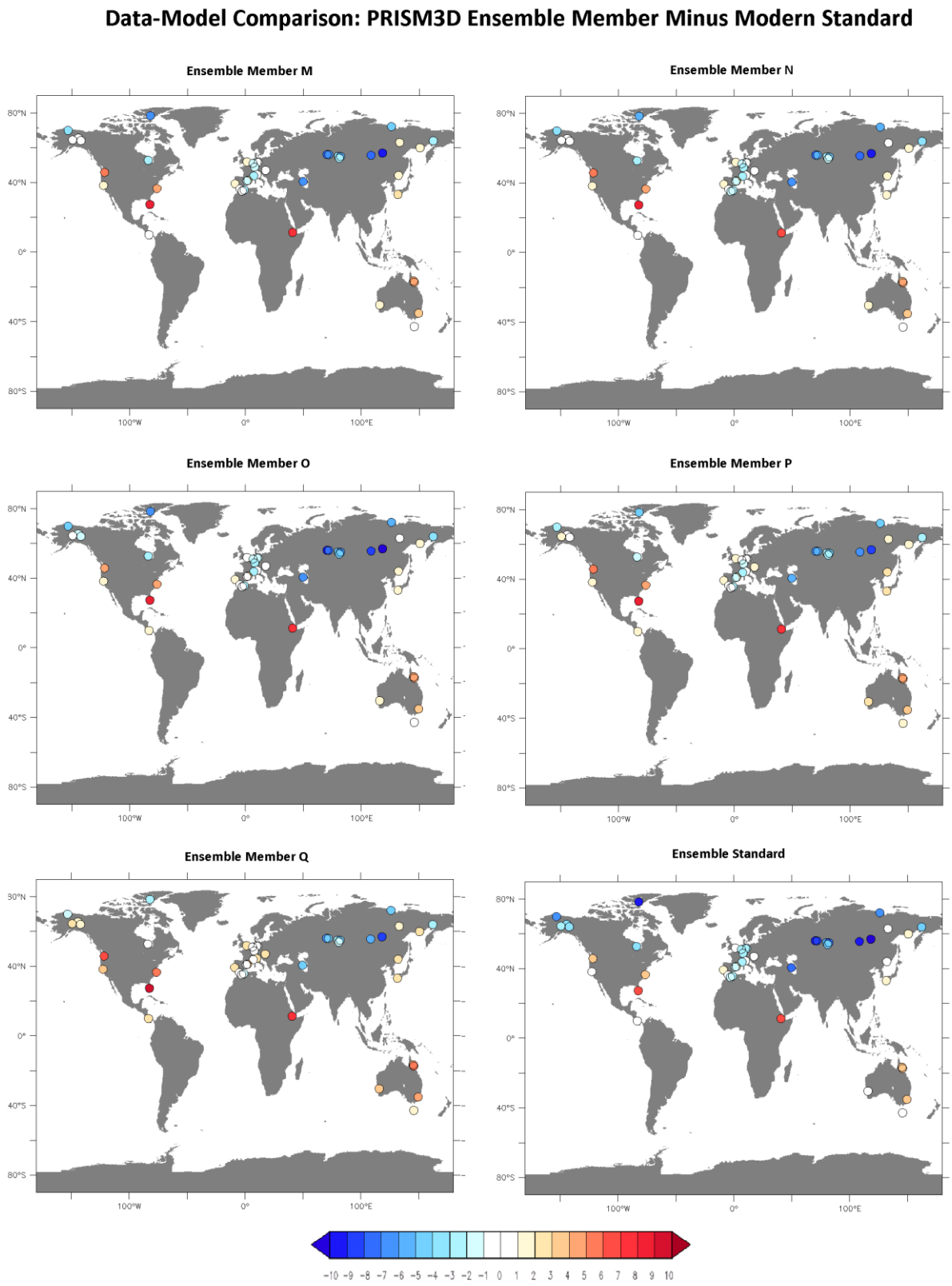
A3: Global annual mean plots for *"Pliocene ensemble member minus modern Standard"* for precipitation (mm/day) for all 13 ensemble members and the Standard. The coloured outlines represent which grouping the ensemble member belongs to either 'colder than Standard' (blue) or 'warmer than Standard' (red).


Data-Model Comparison: PRISM3D Ensemble Member Minus Modern Standard
Ensemble Member B
Ensemble Member D
Ensemble Member F
Ensemble Member H
Ensemble Member I
Ensemble Member J
Ensemble Member K
Ensemble Member L
80°N
40°N
0°
40°S
80°S
100°W
0°
100°E
−10 −9 −8 −7 −6 −5 −4 −3 −2 −1 0 1 2 3 4 5 6 7 8 9 10



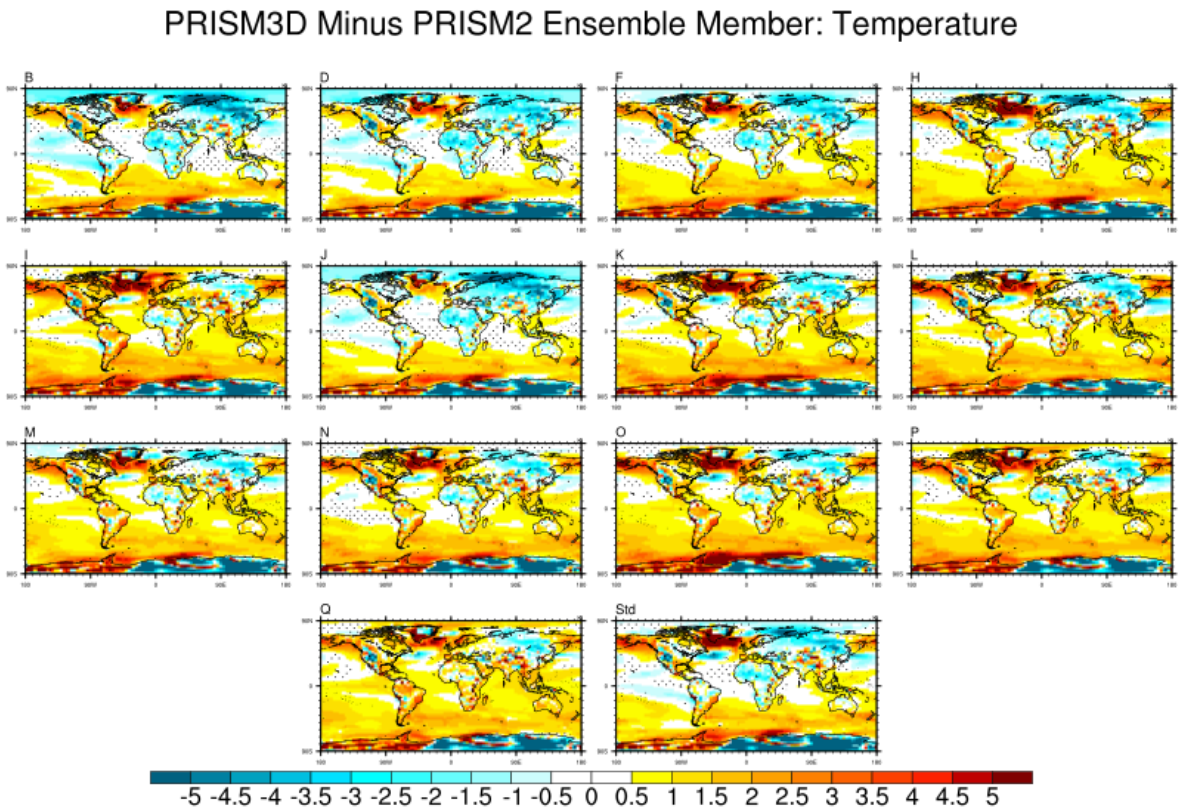
A4: Site-by-site *"Pliocene ensemble member minus modern Standard"* sea surface temperature data-model comparisons for all 13 ensemble members and the Standard


Data-Model Comparison: PRISM3D Ensemble Member Minus Modern Standard
Ensemble Member B
Ensemble Member D
Ensemble Member F
Ensemble Member H
Ensemble Member I
Ensemble Member J
Ensemble Member K
Ensemble Member L
80°N
40°N
0°
40°S
80°S
100°W
0°
100°E
−10 −9 −8 −7 −6 −5 −4 −3 −2 −1 0 1 2 3 4 5 6 7 8 9 10

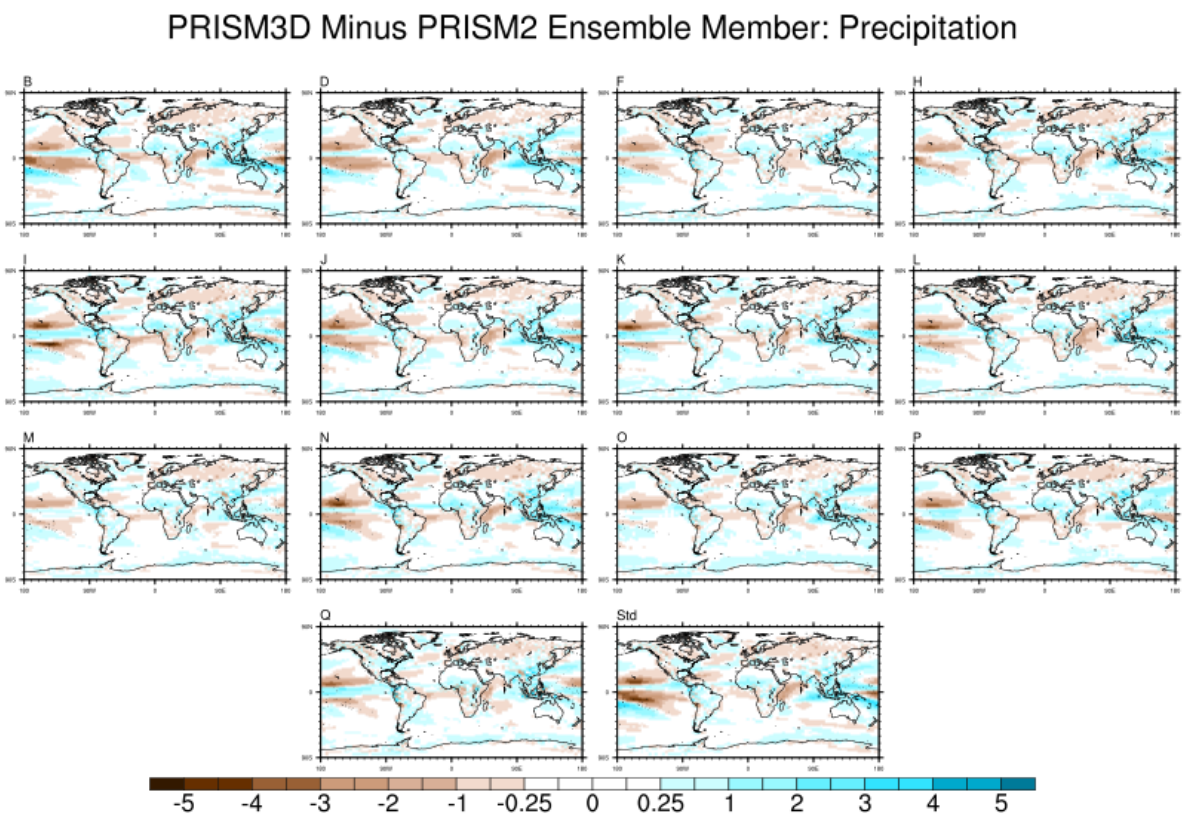


A5: Site-by-site *"Pliocene ensemble member minus modern Standard"* surface air temperature data-model comparisons for all 13 ensemble members and the Standard

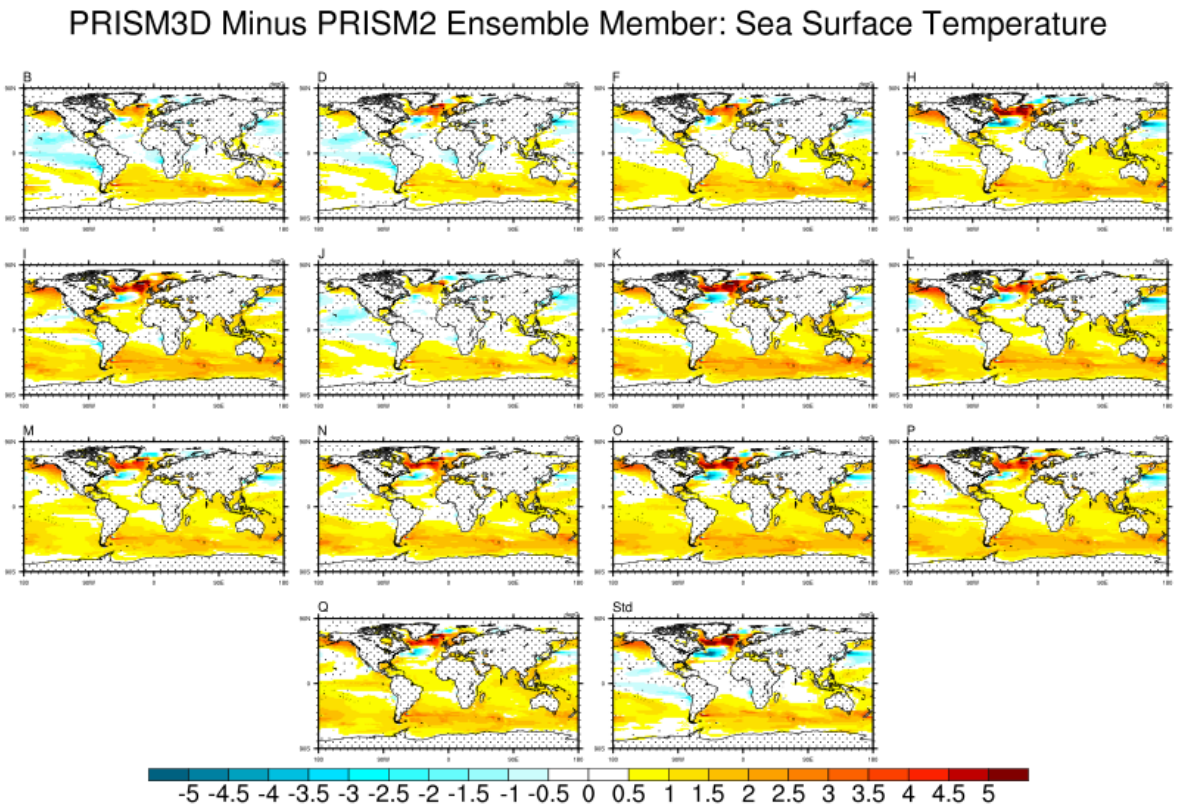
Appendix B



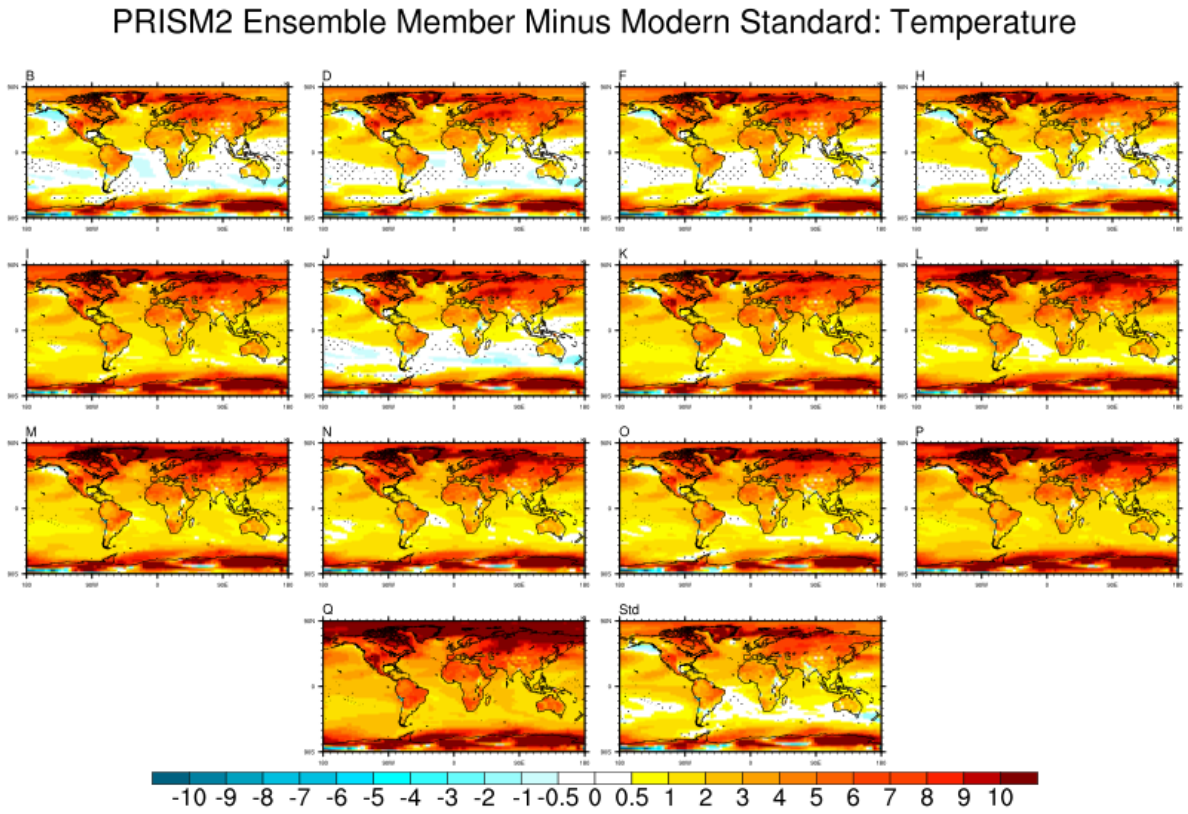
B1 - Global mean annual intra-model comparisons for "PRISM3D minus PRISM2 ensemble member" surface air temperature (SAT - °C) for all 13 members and the Standard



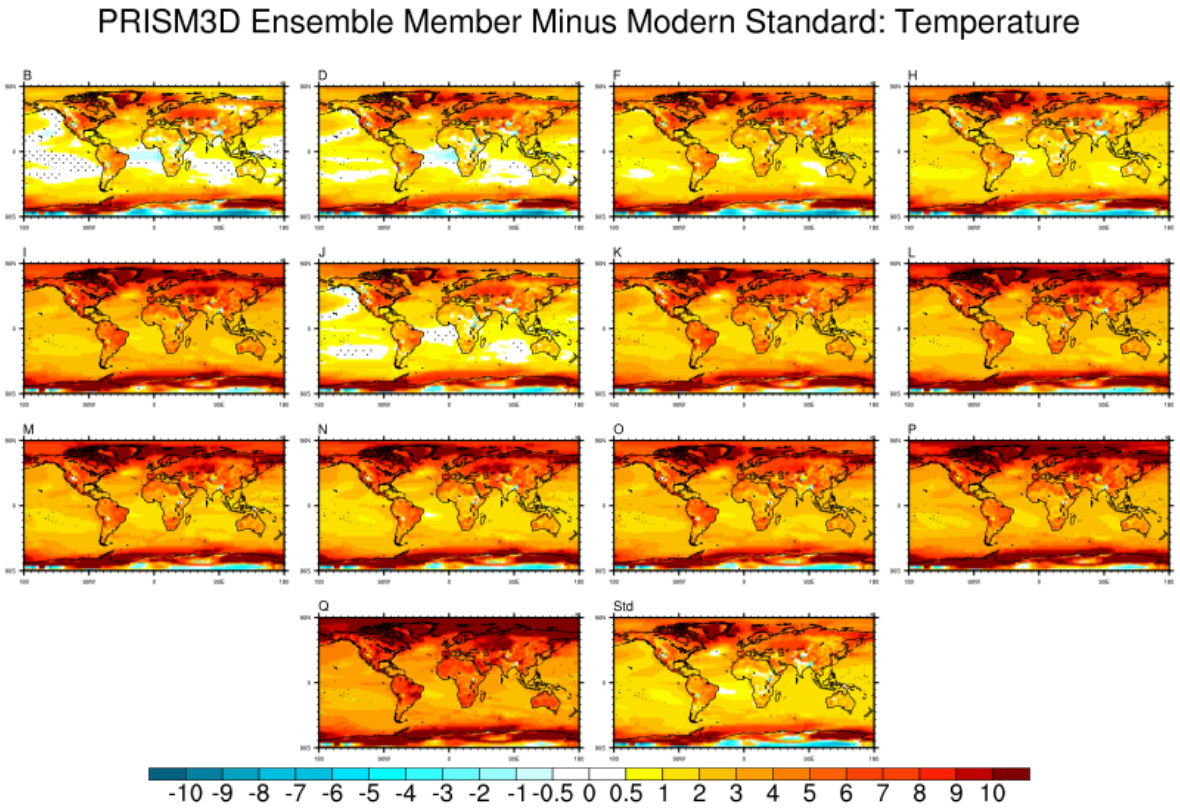
B2 - Global mean annual intra-model comparisons for "PRISM3D minus PRISM2 ensemble member" precipitation (mm/day) for all 13 ensemble members and the Standard



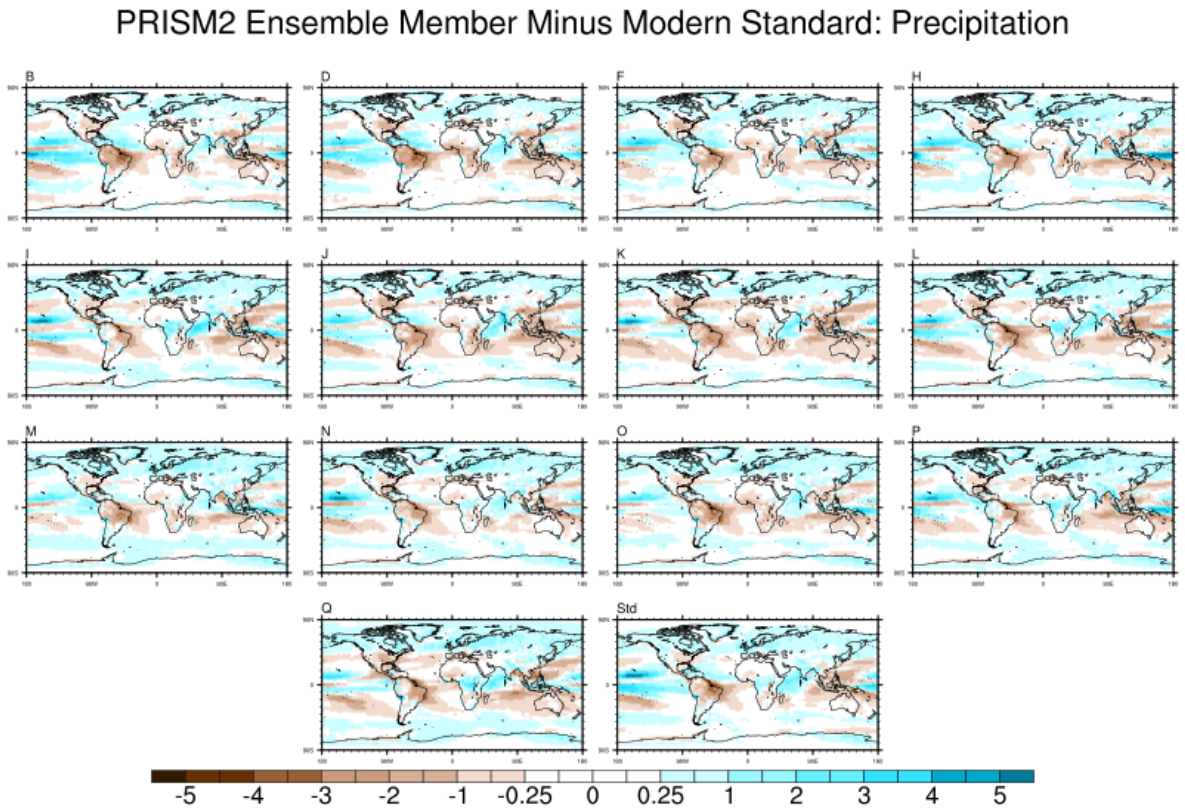
B3 - Global mean annual intra-model comparisons for "PRISM3D minus PRISM2 ensemble member" sea surface temperature (SST - °C) for all 13 ensemble members and the Standard



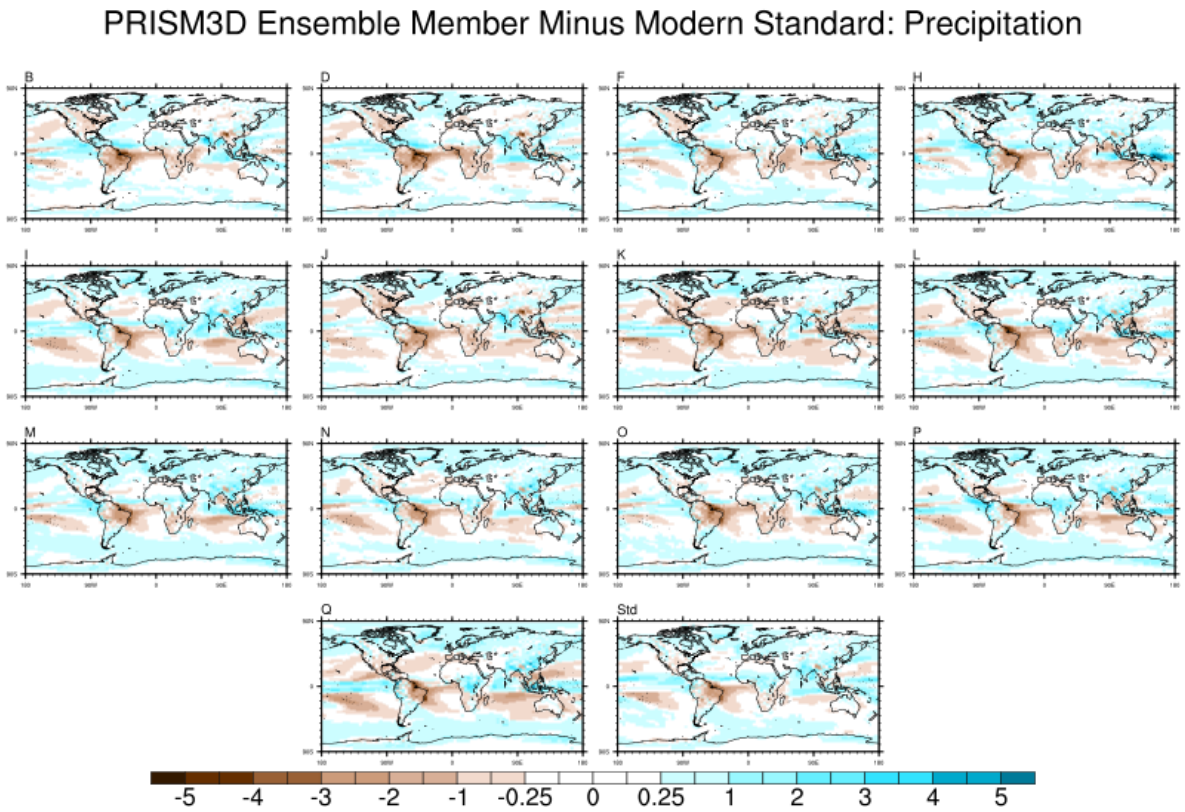
B4 - Global mean annual intra-model comparisons for "PRISM2 ensemble member minus modern Standard" surface air temperature (SAT - °C) for all 13 members and the Standard



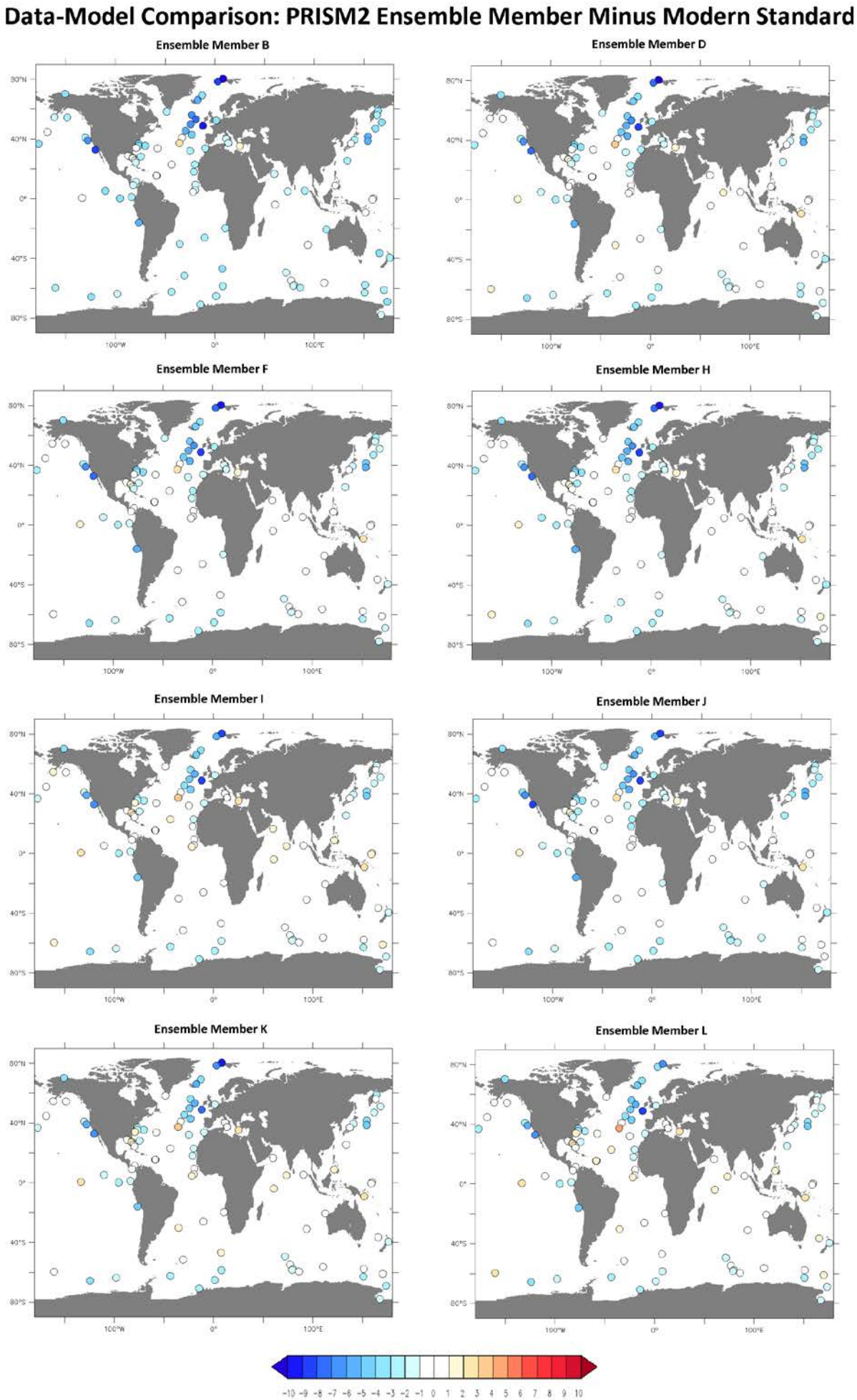
B5 - Global mean annual intra-model comparisons for "PRISM3D ensemble member minus modern Standard" surface air temperature (SAT - °C) for all 13 members and the Standard

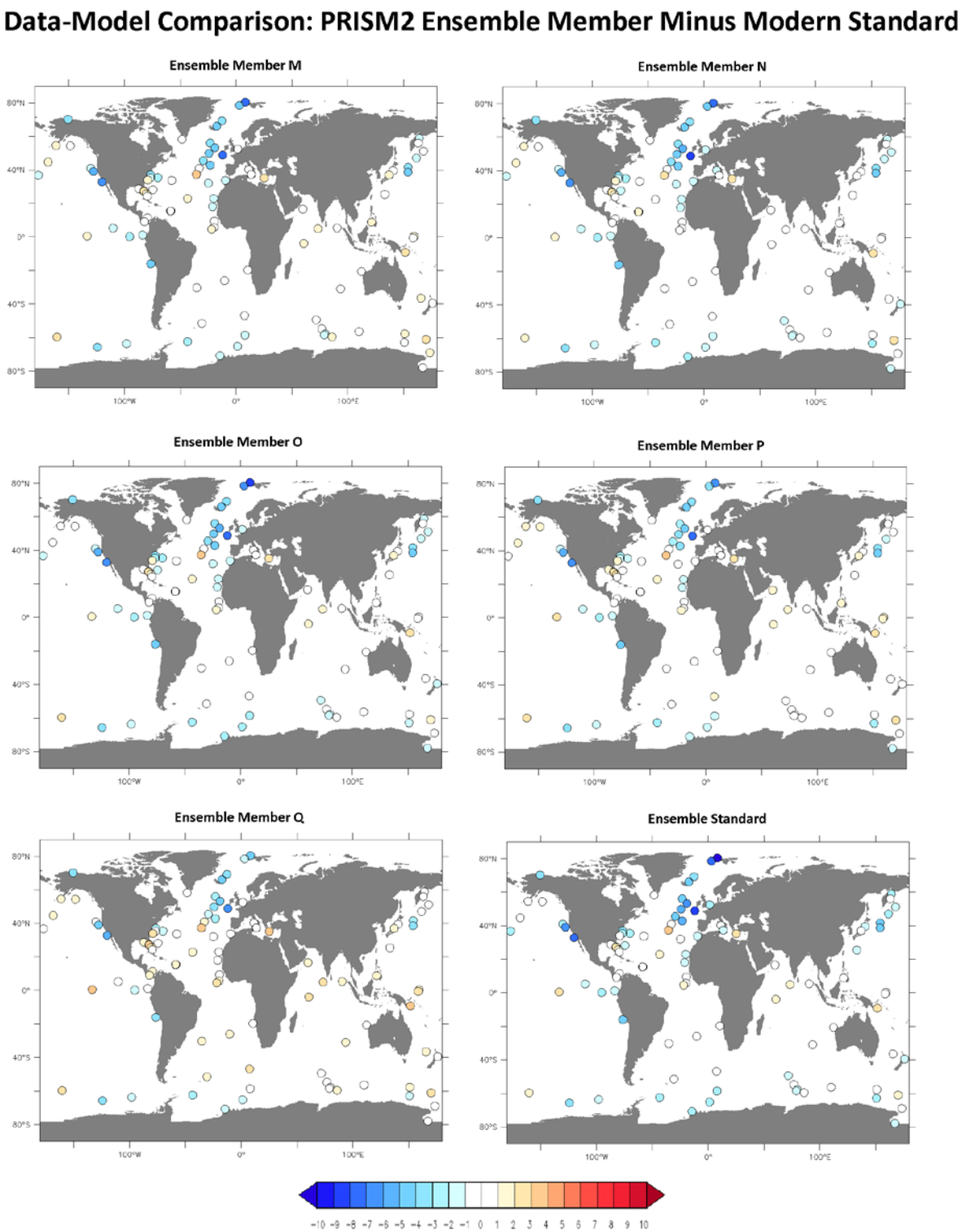


B6 - Global mean annual intra-model comparisons for "PRISM2 ensemble member minus modern Standard" precipitation (mm/day) for all 13 members and the Standard

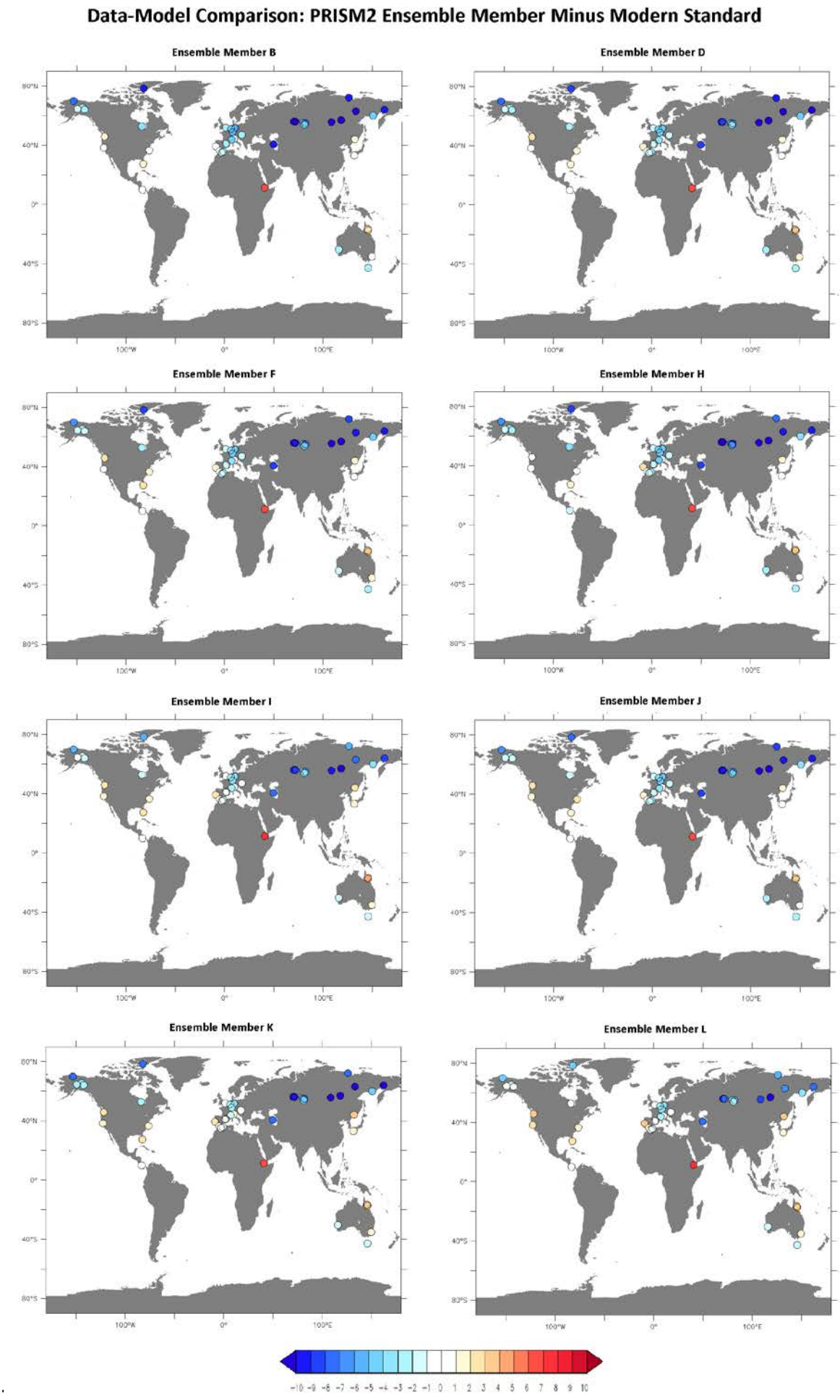


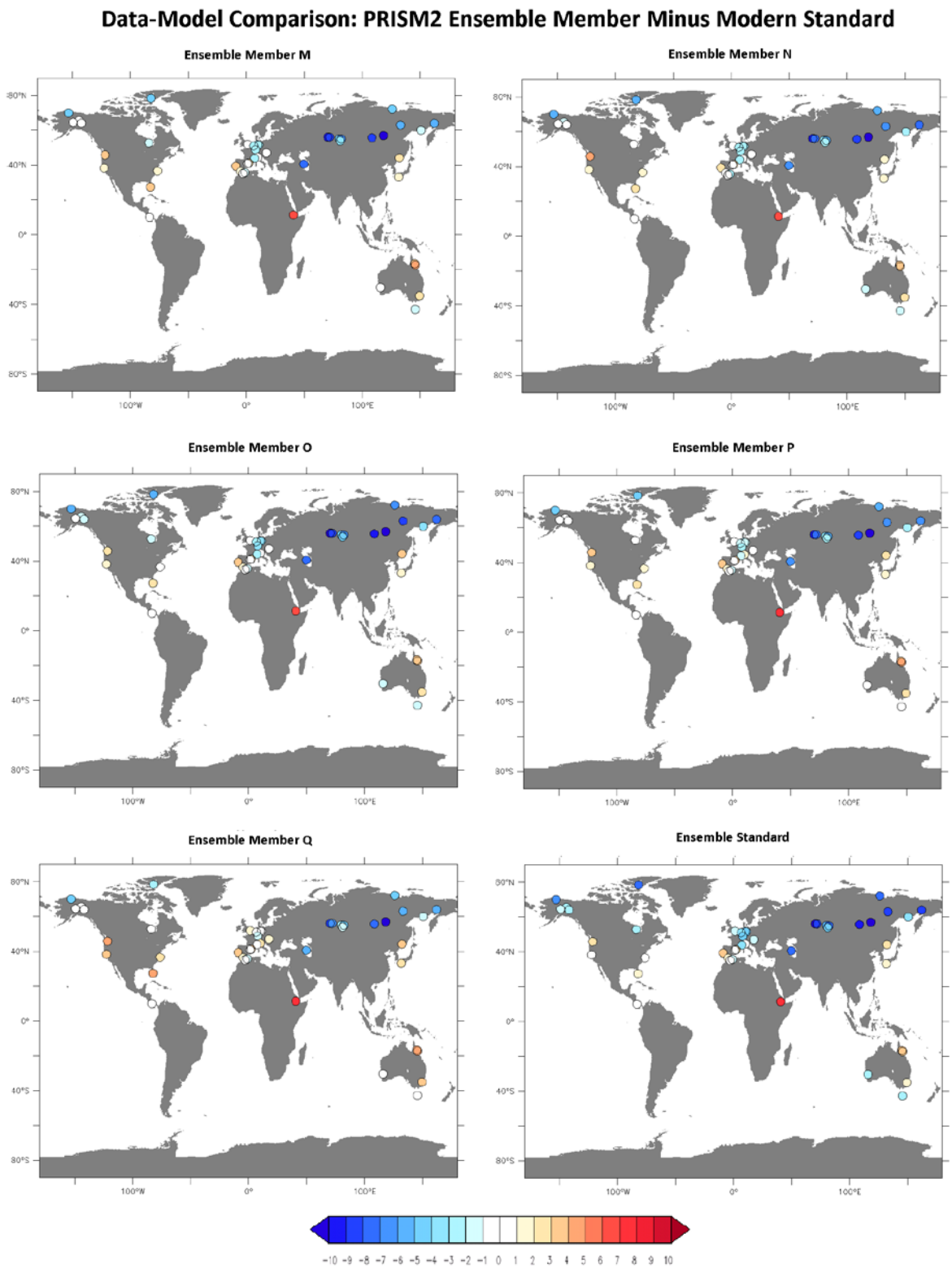
B7 - Global mean annual intra-model comparisons for "PRISM3D ensemble member minus modern Standard" precipitation (mm/day) for all 13 members and the Standard


Data-Model Comparison: PRISM2 Ensemble Member Minus Modern Standard
Ensemble Member B
Ensemble Member D
Ensemble Member F
Ensemble Member H
Ensemble Member I
Ensemble Member J
Ensemble Member K
Ensemble Member L
80°N
40°N
0°
40°S
80°S
100°W
0°
100°E
−10 −9 −8 −7 −6 −5 −4 −3 −2 −1 0 1 2 3 4 5 6 7 8 9 10



B8 - Global distribution of the PRISM2 site-by-site sea surface temperature (SST - °C) data-model comparisons for all 13 ensemble members and the Standard


Data-Model Comparison: PRISM2 Ensemble Member Minus Modern Standard
Ensemble Member B
Ensemble Member D
Ensemble Member F
Ensemble Member H
Ensemble Member I
Ensemble Member J
Ensemble Member K
Ensemble Member L
80°N
40°N
0°
40°S
80°S
100°W
0°
100°E
−10 −9 −8 −7 −6 −5 −4 −3 −2 −1 0 1 2 3 4 5 6 7 8 9 10



B9 - Global distribution of the PRISM2 site-by-site surface air temperature (SAT - °C) data-model comparisons for all 13 ensemble members and the Standard

Palaeogeography, Palaeoclimatology, Palaeoecology 309 (2011) 128–140

Contents lists available at ScienceDirect

Palaeogeography, Palaeoclimatology, Palaeoecology

journal homepage: www.elsevier.com/locate/palaeo

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

James O. Pope [a,*], Matthew Collins [b,c], Alan M. Haywood [a], Harry J. Dowsett [d], Stephen J. Hunter [a], Daniel J. Lunt [e], Steven J. Pickering [a], Matthew J. Pound [a]

[a] School of Earth & Environment, University of Leeds, Leeds, LS2 9JT, UK
[b] Met Office Hadley Centre, Exeter, EX1 3PB, UK
[c] College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, EX4 4QF, UK
[d] United States Geological Survey, National Center, Reston, VA 20192, USA
[e] School of Geographical Sciences, University of Bristol, Bristol, BS8 1SS, UK

ARTICLE INFO

Article history:
Received 10 December 2010
Received in revised form 27 April 2011
Accepted 6 May 2011
Available online 15 May 2011

Keywords:
Perturbed Physics Ensemble
Pliocene
Data/model comparison

ABSTRACT

Examination of the mid-Pliocene Warm Period (mPWP; ~3.3 to 3.0 Ma BP) provides an excellent opportunity to test the ability of climate models to reproduce warm climate states, thereby assessing our confidence in model predictions. To do this it is necessary to relate the uncertainty in model simulations of mPWP climate to uncertainties in projections of future climate change. The uncertainties introduced by the model can be estimated through the use of a Perturbed Physics Ensemble (PPE). Developing on the UK Met Office Quantifying Uncertainty in Model Predictions (QUMP) Project, this paper presents the results from an initial investigation using the end members of a PPE in a fully coupled atmosphere–ocean model (HadCM3) running with appropriate mPWP boundary conditions. Prior work has shown that the unperturbed version of HadCM3 may underestimate mPWP sea surface temperatures at higher latitudes. Initial results indicate that neither the low sensitivity nor the high sensitivity simulations produce unequivocally improved mPWP climatology relative to the standard. Whilst the high sensitivity simulation was able to reconcile up to 6 °C of the data/model mismatch in sea surface temperatures in the high latitudes of the Northern Hemisphere (relative to the standard simulation), it did not produce a better prediction of global vegetation than the standard simulation. Overall the low sensitivity simulation was degraded compared to the standard and high sensitivity simulations in all aspects of the data/model comparison.

The results have shown that a PPE has the potential to explore weaknesses in mPWP modelling simulations which have been identified by geological proxies, but that a 'best fit' simulation will more likely come from a full ensemble in which simulations that contain the strengths of the two end member simulations shown here are combined.

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 (CO_2) (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

* Corresponding author. Tel.: +44 113 343 9085; fax: +44 113 343 6716.
E-mail address: eejop@leeds.ac.uk (J.O. Pope).

0031-0182/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.palaeo.2011.05.004

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 CO_2 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 climate*prediction*.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 CO_2. 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 CO_2 concentration of 400 ppmv 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 Greenland Ice Sheet by 50% and the Antarctic Ice Sheet by 33% volume (Dowsett et al., 1999; Haywood and Valdes, 2004). During the model simulations, the vegetation type and land ice volume and area are held constant. A sea ice model is part of the HadCM3 (Gordon et al., 2000) and was initialised by the PRISM2 sea ice reconstruction (Dowsett et al., 1996). BIOME4 also used the same land–sea mask as the HadCM3 simulations. CO_2 was set in the BIOME4 offline calculations to the same value as was used in all the climate model simulations.

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 CO_2, O_3 and H_2O (Edwards & Slingo, 1996; Gordon et al., 2000; Johns et al., 2003). Background aerosols in the model are determined using prescribed pre-industrial emissions 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 CO_2 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 $1.25° \times 1.25°$ 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–Faeroes–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 input into the model along with data on soil depth, soil texture properties, the absolute minimum temperature and atmospheric CO_2, from which a biome is classified for each grid point (Kaplan, 2001; Salzmann et al., 2008). BIOME4 is programmed with 28 biome classifications, which are determined based on the combination of dominant and sub-dominant plant functional types (PFT). The 12 PFT's cover distinct categories of flora from Arctic to tropical environments (e.g. tropical grassland or cool conifer woodland (Salzmann et al., 2009)). Using the climate inputs and the CO_2 level, the model calculates whether the growth of each PFT could occur within those climatic parameters through the calculation of net primary productivity (NPP). Parameters such as photosynthetic pathway and seasonal fluxes in temperature and precipitation are used along with semi-empirical rules related to balances between forest and grass taxa to determine the potential PFTs at each grid point; the combination of potential PFTs then determines the final biome (Kaplan, 2001). BIOME4 was run at the resolution of HadCM3, enabling the required fields to be taken directly from the HadCM3 outputs without adjusting the resolution. The climatological fields used are averaged for 30 years, the standard for analysing the data from palaeoclimate modelling studies (PMIP2; Braconnot et al., 2007). Then each average month is used to construct 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 simulation (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 climate*prediction*.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 (PRISM0 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 MASST) 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
The perturbed parameter labels, identifiers, values and areas of influence in the three simulations (high sensitivity, standard simulation and low sensitivity).

	Label	Sub-group	Identifier/properties	Low sensitivity	Control	High sensitivity	Area of influence
			Climate sensitivity (°C)	2.19	3.3	7.11	
1	VF1	–	Icefall speed (ms^{-1})	0.93976	1	0.54306	Large scale cloud
2	CT	–	Cloud droplet to rain conversion rate ($\times 10^{-4}\,s^{-1}$)	1.63E-04	1.00E-04	3.50E-04	Large scale cloud
3	RHCrit	–	Threshold of relative humidity for cloud formation	0.84077	0.7	0.7	Large scale cloud
4	CW	Land	Cloud droplet to rain conversion rate ($\times 10^{-4}\,kg\,m^{-3}$)	1.75E-04	2.00E-04	1.41E-04	Large scale cloud
		Sea	Cloud droplet to rain conversion rate ($\times 10^{-4}\,kg\,m^{-3}$)	4.25E-05	5.00E-05	3.23E-05	Large scale cloud
5	MinSIA	–	Minimum sea ice albedo — dependence of SIA on temperature	0.50666	0.5	0.60104	Sea ice
6	ice_tr	–	Sea ice minimum temperature	9.5243	10	3.836	Sea ice
7	Ent	–	Entrainment rate coefficient	2.89934	3	2.37726	Convection
8	Icesize	–	Ice particle size (μm)	30.761	30	28.099	Radiation
9	Cape	–	CAPE closure — intensity of convective mass flux	Off	Off	Off	Convection
10	flux_g0	–	Boundary layer flux profile	16.9231	10	6.9039	Boundary layer
11	Charnoc	–	Charnock constant — roughness lengths and surface fluxes over sea	0.0128	1.20E-02	0.0134	Boundary layer
12	soillev	–	Number of soil levels accessed for evapotranspiration	2	4	4	Land surface
13	lambda	–	Asymptotic neutral mixing length parameter	0.3356	0.15	0.135	Boundary layer
14	cnv_rl	–	Free convective roughness over sea	4.14E-03	1.30E-03	1.67E-03	Boundary layer
15	oi_diff	–	Ocean ice diffusion	3.67E-04	3.75E-04	3.73E-04	Sea ice
16	dyndel	–	Order of diffusion operator	4	6	6	Dynamics
17	dyndiff	–	Diffusion *e*-folding time	8.513	12	8.5	Dynamics
18	eacfbl	–	Cloud fraction at saturation boundary layer	0.79598	0.5	0.51262	Large scale cloud
19	eacftrp	–	Cloud fraction at saturation free troposphere value	0.64799	0.5	0.50631	Large scale cloud
20	k_gwd	–	Surface gravity wavelength	1.98E + 04	2.00E + 04	1.67E + 04	Dynamics
21	k_lee	–	Trapped lee wave constant	2.97E + 05	3.00E + 05	2.50E + 05	Dynamics
22	gw_lev	–	Starting level for gravity wave drag	3	3	3	Dynamics
23	s_sph	sw	Non-spherical ice particles shortwave radiation properties	2	2	7	Radiation
		lw	Non-spherical ice particles shortwave radiation properties	1	1	7	Radiation
24	c_sph	sw	Non-spherical ice particles longwave radiation properties	3	3	7	Radiation
		lw	Non-spherical ice particles longwave radiation properties	1	1	7	Radiation
25	rhparam	–	Flow dependant RHcrit	0	0	1	Large scale cloud
26	vertcld	–	Vertical gradient of cloud water	1	0	0	Large scale cloud
27	canopy	–	Surface canopy energy exchange	0	0	1	Land surface
28	cnv_upd	–	Convective anvils updraft factor	0.51016	Off	Off	Convection
29	anvil	–	Convective anvils shape factor: radiative properties of convective cloud	2.59138	Off	–	Convection
30	sto_res	–	Dependence of stomatal conductance on CO2	0	1	1	Land surface
31	f_rough	–	Forest roughness lengths	0	0	3	Land surface
32	sw_absn	–	Shortwave water vapour continuum absorption	1	0	0	Radiation

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 (P_e) from the proportion of observed agreement (P_o) and this result is normalised through dividing it by the maximum possible difference ($1-P_e$) to generate the Kappa Statistic:

$$k = (P_e - P_o) / (1 - P_e).$$

The P_e 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
The temperature and precipitation values for 'Pliocene minus modern' anomalies for the high sensitivity, low sensitivity and standard simulations and for the Pliocene 'high sensitivity minus standard' and Pliocene 'low sensitivity minus standard' simulations.

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

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 MASST dataset by the PRISM group. Two comparisons were plotted, with the standard simulations for the

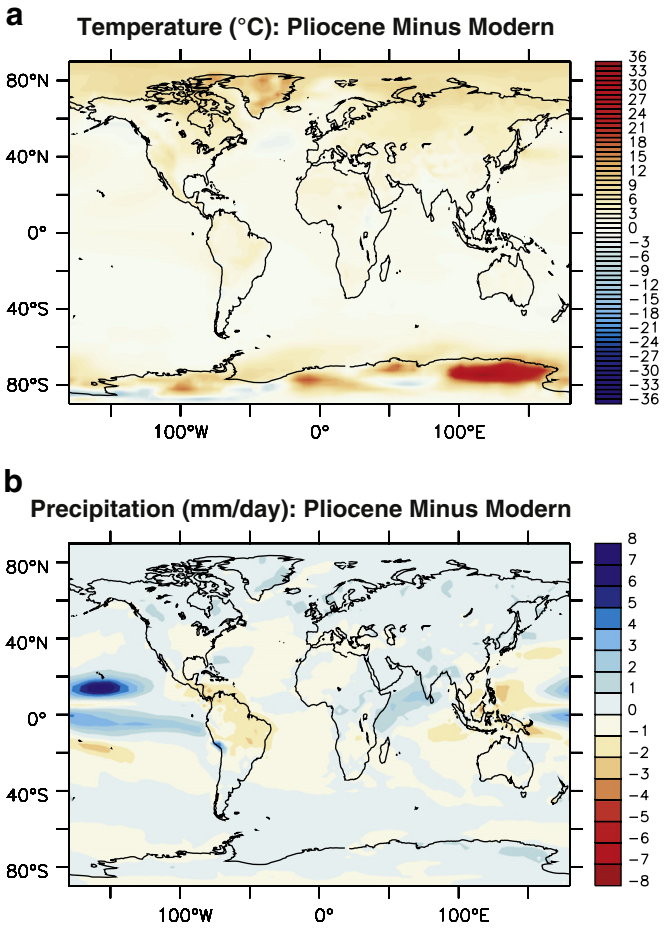


Fig. 1. Pliocene minus modern anomalies for the standard simulations for temperature (1a) and precipitation (1b).

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 CO_2 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

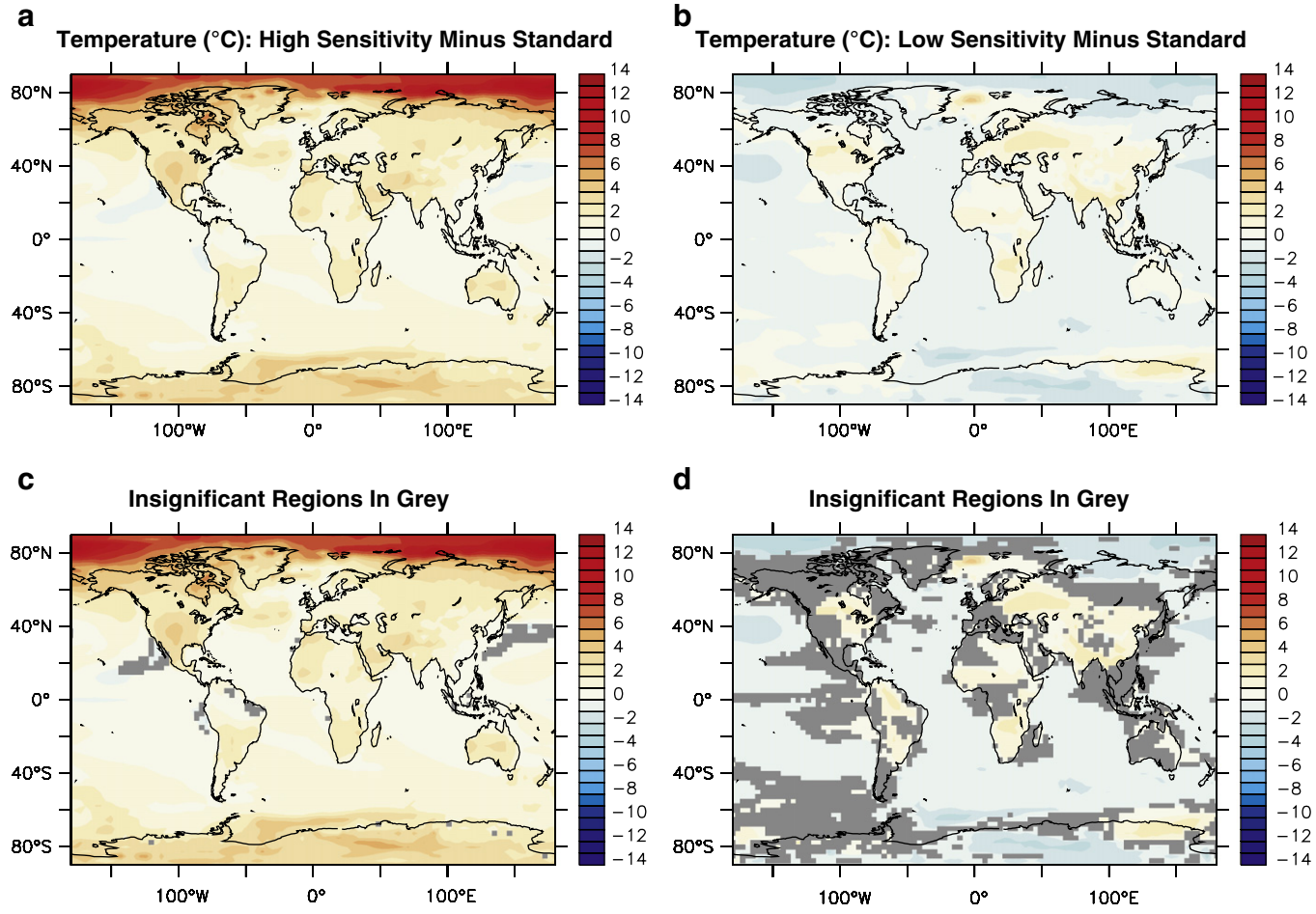


Fig. 2. Mean annual temperature anomalies for 'high sensitivity minus standard' (2a) and 'low sensitivity minus standard' (2b) in °C. Student's t-test was applied to the anomaly plots and the insignificant regions, which are plotted in grey and overlain over the 'high minus standard' anomaly (2c) and the 'low minus standard' anomaly (2d).

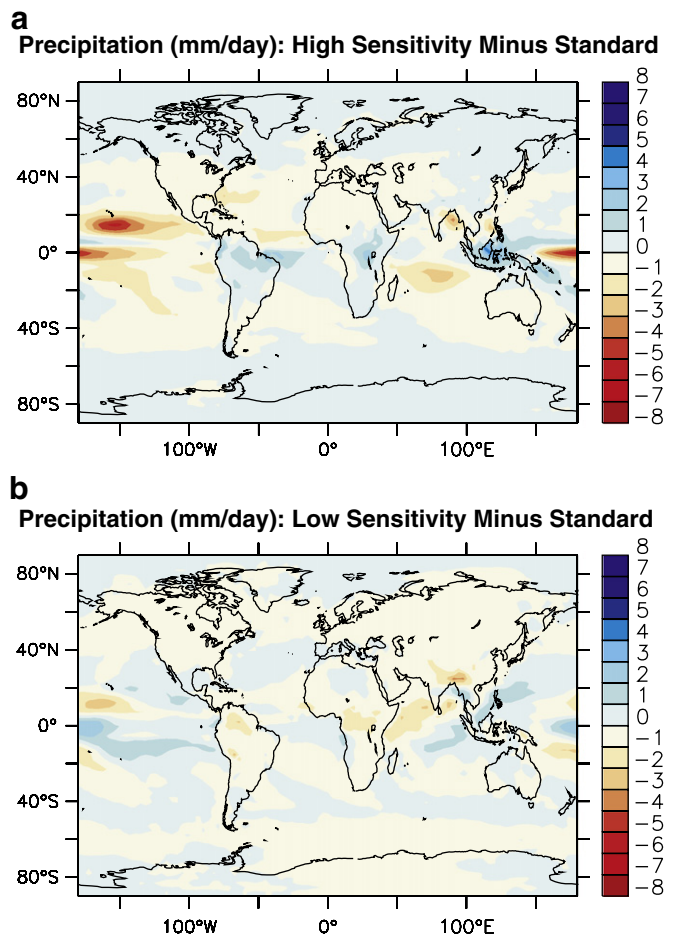


Fig. 3. Precipitation anomalies for 'high sensitivity minus standard' (3a) and 'low sensitivity minus standard' (3b) in mm/day.

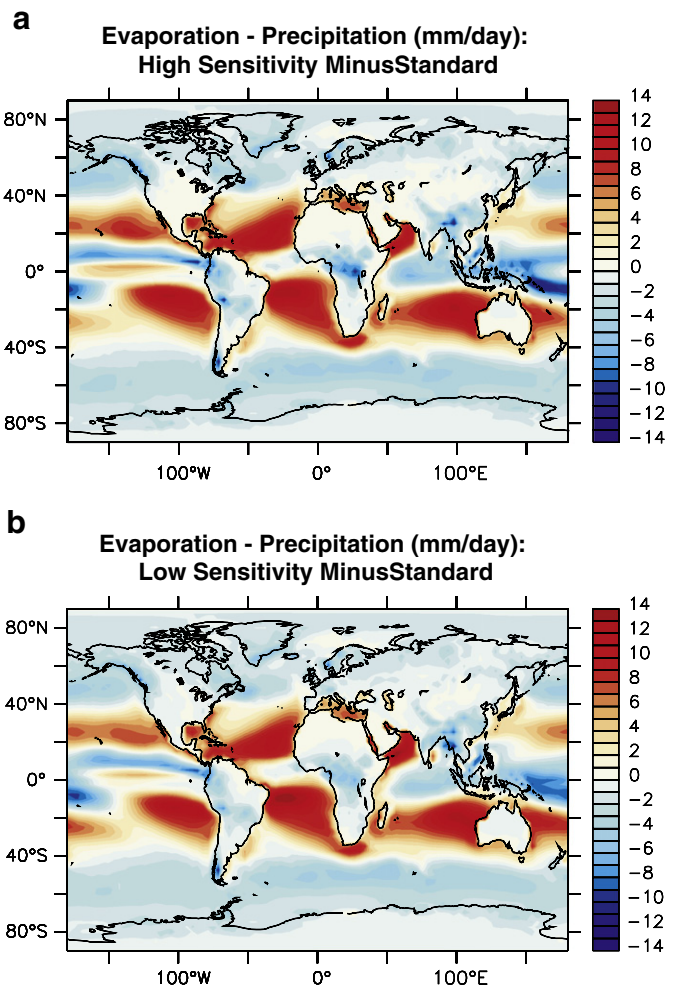


Fig. 4. Evaporation–precipitation anomalies for 'high sensitivity minus standard' (4a) and 'low sensitivity minus standard' (4b) in mm/day of precipitation with negative numbers indicating that there has been net precipitation and positive numbers indicating that there has been net evaporation. These anomalies were calculated using [Evaporation (high/low) minus Precipitation (high/low)] + [Evaporation (standard) minus Precipitation (standard)].

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

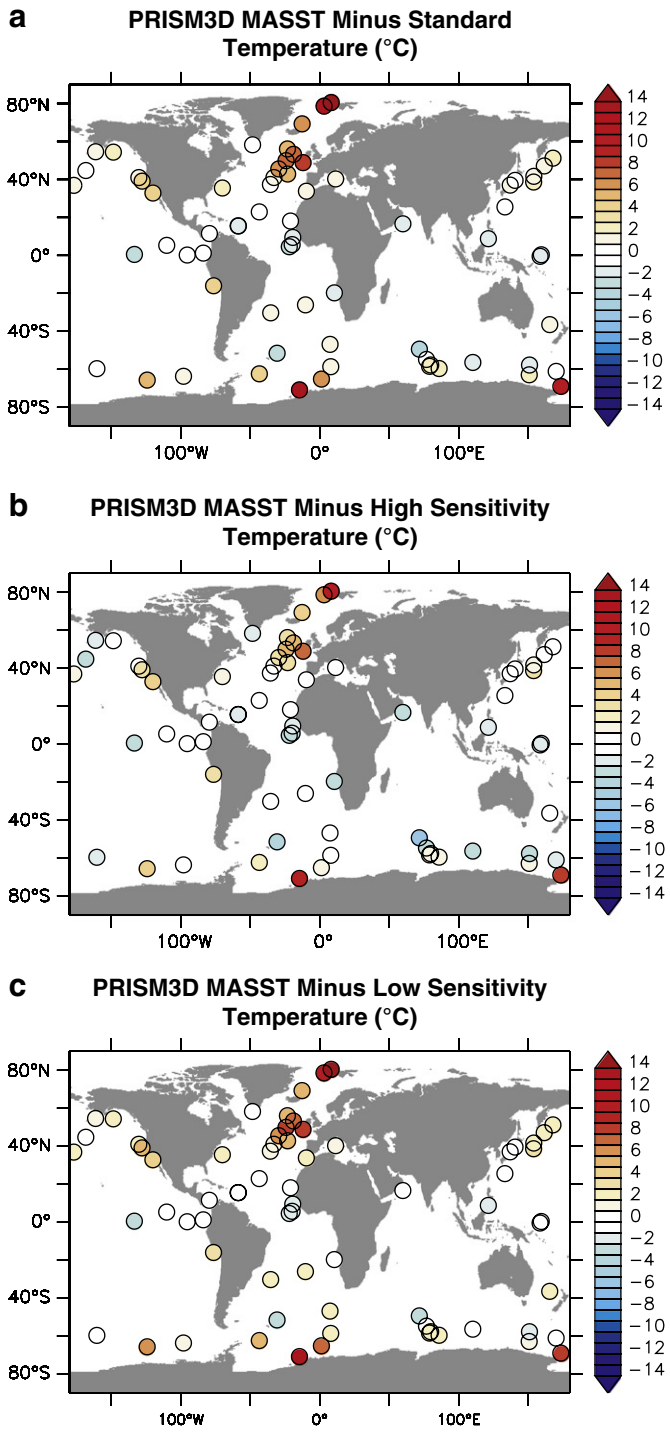


Fig. 5. Data/model comparison using the PRISM3D Mean Annual Sea Surface Temperature (MASST) dataset. PRISM3D MASST minus standard (5a), PRISM3D MASST minus high sensitivity simulation (5b) and PRISM3D MASST minus low sensitivity simulation (5c) in °C. Root mean square errors (RMSE) were calculated for each comparison as (a) 4.37, (b) 3.25 and (c) 4.38.

types are not being compared to a proxy directly, the weakness is not a serious concern for this paper, but will be addressed in future developments of the Plio-QUMP project.

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 data/model 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 similarly in the data/model comparison. The standard simulation data/model comparison shows that the model over-estimates 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 skilful 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

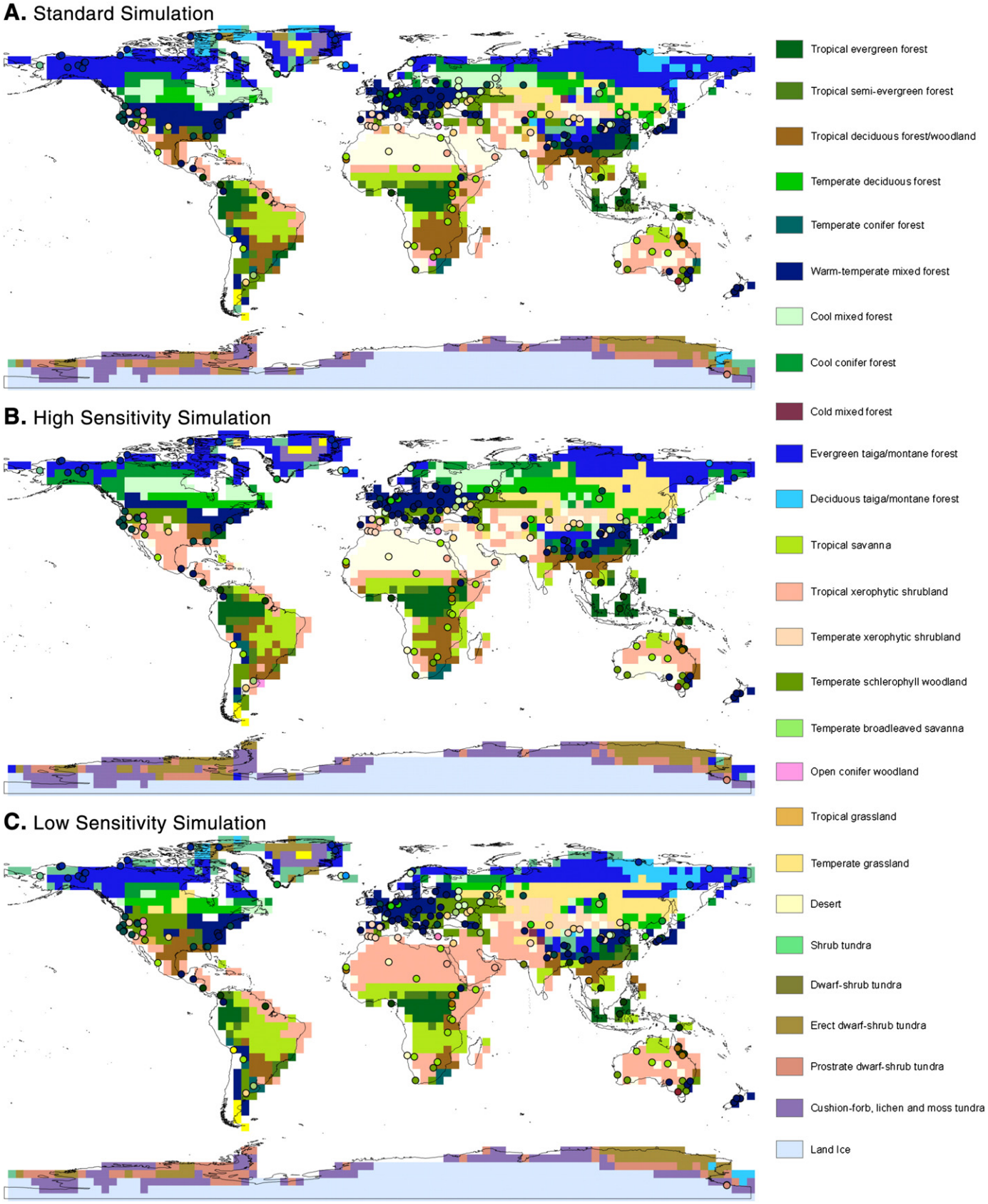


Fig. 6. BIOME4 outputs for the standard simulation (6A), high sensitivity simulation (6B) and the low sensitivity simulation (6C).

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 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 mPWP, 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 mPWP 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 palaeobotanical 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

The Natural Environment Research Council (NERC) is acknowledged for its support to J.O.P, M.J.P. D.J.L and A.M.H. A.M.H. also acknowledges the Leverhulme Trust for the award of a Philip Leverhulme Prize (2008). H.J.D acknowledges the US Department of Interior and USGS Office of Global Change. D.J.L. is currently a NERC/RCUK research fellow, and acknowledges their support. We also wish to acknowledge the PRISM group for use of their palaeoenvironmental reconstructions and Ulrich Salzmann for his palaeobotanical reconstruction. The three simulations were run at the UK Met Office where M.C. was supported by the Met Office DECC/Defra contract. We wish to thank the two anonymous reviewers for their comments and suggestions which helped improve the quality of this paper.

References

Ackert Jr., R.P., Kurz, M.D., 2004. Age and uplift rates of Sirius Group sediments in the Dominion Range, Antarctica, from surface exposure dating and geomorphology. Global and Planetary Change 42, 207–225.

Ashworth, A.C., Lewis, A.R., Marchant, D.R., Askin, R.A., Cantrill, D.J., Francis, J.E., Leng, M.J., Newton, A.E., Raine, J.I., Williams, M., Wolfe, A.P., 2007. The Neogene biota of the Transantarctic Mountains. Online Proceedings of the ISAES (http://pubs.usgs.gov/of/2007/1047/). In: Cooper, A.K., Raymond, C.R., et al. (Eds.), USGS Open-File Report 2007–1047, Extended Abstract 071, p. 4.

Braconnot, P., Otto-Bliesner, B., Harrison, S., Joussaume, S., Peterchmitt, J.-Y., Abe-Ouchi, A., Crucifix, M., Driesschart, E., Fichefet, Th., Hewitt, C.D., Kageyama, M., Kitoh, A., Laine, A., Loutre, M.-F., Marti, O., Merkel, U., Ramstein, G., Valdes, P., Weber, A.L., Yu, Y., Zhao, Y., 2007. Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum — part 1: experiments and large-scale features. Climate of the Past. 3 (2), 261–277.

Brierley, C.M., Collins, M., Thorpe, A.J., 2010. The impact of perturbations to ocean model parameters on climate and climate change in a coupled model. Climate Dynamics 34, 325–343.

Brown, J., Collins, M., Tudhope, A.W., Toniazzo, T., 2008. Modelling mid-Holocene tropical climate and ENSO variability: towards constraining predictions of future change with palaeo-data. Climate Dynamics. 30, 19–36.

Charney, J.G., 1979. Carbon dioxide and climate: a scientific assessment. National Academy of Science 22 pp.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 2, 37–46.

Collins, M., Booth, B.B.B., Harris, G.R., Murphy, J.M., Sexton, D.M.H., Webb, M.J., 2006. Towards quantifying uncertainty in transient climate change. Climate Dynamics 27, 127–147.

Collins, M., Brierley, C.M., MacVean, M., Booth, B.B.B., Harris, G.R., 2007. The sensitivity of the rate of transient climate change to ocean physics perturbations. Journal of Climate 20, 2315–2320.

Collins, M., Booth, B.B.B., Bhaskaran, B., Harris, G.R., Murphy, J.M., Sexton, D.M.H., Webb, M.J., 2010. Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi model ensembles. Climate Dynamics. doi:10.1007/s00382-010-0808-0.

Congalton, R.G., Green, K., 1999. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. Lewis Publishers, Boca Raton, FL. 137 pp.

Dowsett, H.J., 2007. The PRISM palaeoclimate reconstruction and Pliocene sea surface temperature. 459–480 pp. The Micropalaeontological Society, Special Publications. The Geological Society, London.

Dowsett, H., Thompson, R., Barron, J., Cronin, T., Fleming, F., Ishman, S., Poore, R., Willard, D., Holtz, T., 1994. Joint investigations of the middle Pliocene climate I: PRISM palaeoenvironmental reconstructions. Globlal & Planetary Change 9, 169–195.

Dowsett, H.J., Barron, J., Poore, R., 1996. Middle Pliocene sea surface temperatures: a global reconstruction. Marine Micropaleontology 27, 13–25.

Dowsett, H.J., Barron, J.A., Poore, R.Z., Thompson, R.S., Cronin, T.M., Ishman, S.E., Williams, D.A., 1999. Pliocene Paleoenvironmental Reconstruction: PRISM2. edited. U.S.Geological Survey, pp. 99–535.

Dowsett, H.J., Chandler, M.A., Cronin, T.M., Dwyer, G.S., 2005. Middle Pliocene sea surface temperature variability. Paleoceanpgraphy 20.

Dowsett, H.J., Robinson, M.M., Haywood, A.M., Salzmann, U., Hill, D.J., Sohl, L., Chandler, M., Williams, M., Foley, K., Stoll, D.K., 2010a. The PRISM3D paleoenvironmental reconstruction. Stratigraphy 7 (2–3), 123–140.

Dowsett, H.J., Robinson, M.M., Stoll, D.K., Foley, K.M., 2010b. Mid-Pliocene mean annual sea surface temperature analysis for data-model comparisons. Stratigraphy. 7 (2–3), 189–198.

Dowsett, H.J., Haywood, A.M., Valdes, P.J., Robinson, M.M., Lunt, D.J., Hill, D.J., Stoll, D.K., Foley, K.M., 2011. Sea surface temperatures of the mid-Piacenzian warm period: a comparison of PRISM3 and HadCM3. Palaeogeogr. Palaeoclimatol. Palaeoecol. 309, 83–91 (this issue).

Edwards, J.M., Slingo, A., 1996. Studies with a flexible new radiation code. I: Choosing a configuration for a large scale model. Quarterly Journal of the Royal Meteorological Society 122, 689–719.

Gordon, C., Cooper, C., Senior, C.A., Banks, H., Gregory, J.M., Johns, T.C., Mitchell, J.F.B., Wood, R.A., 2000. The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. Climate Dynamics 16, 147–168.

Gregory, D., Rowntree, P.R., 1990. A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependant closure. Monthly Weather Review 118, 1483–1506.

Harrison, S.P., Prentice, I.C., 2003. Climate and CO_2 controls on global vegetation distribution at the Last Glacial Maximum: analysis based on palaeovegetation data, biome modelling and palaeoclimate simulations. Global Change Biology 9, 983–1004.

Haywood, A.M., Valdes, P.J., 2004. Modelling Pliocene warmth: contribution of atmosphere, oceans and cryosphere. Earth and Planetary Science Letters 218, 363–377.

Haywood, A.M., Valdes, P.J., Sellwood, B.W., 2000. Global Scale palaeoclimate reconstruction of the middle Pliocene climate using the UKMO GCM: initial results. Global and Planetary Change 25, 239–256.

Haywood, A.M., Valdes, P.J., Peck, V.L., 2007. A permanent El Nino like state during the Pliocene. Paleoceanography 22.

Haywood, A.M., Chandler, M.A., Valdes, P.J., Salzmann, U., Lunt, D.J., Dowsett, H.J., 2009a. Comparison of mid-Pliocene climate predictions produced by the HadAM3 and GCMAM3 general circulation models. Global and Planetary Change 66, 208–224.

Haywood, A.M., Dowsett, H.J., Otto-Bliesner, B., Chandler, M.A., Dolan, A.M., Hill, D.J., Lunt, D.J., Robinson, M.M., Rosenbloom, N., Salzmann, U., Sohl, L.E., 2009b. Pliocene Model Intercomparison Project PlioMIP: experimental design and boundary conditions experiment 1. Geoscientific Model Development Discussions 2, 1215–1244.

Hegerl, G.C., Crowley, T.J., Hyde, W.T., Frame, D.J., 2006. Climate sensitivity constrained by temperature reconstructions over the past seven centuries. Nature 440, 1029–1032.

Hill, D.J., Csank, A.Z., Dolan, A.M., Lunt, D.J., 2011. Pliocene climate variability: Northern Annular Mode in models and tree-ring data. Palaeogeogr. Palaeoclimatol. Palaeoecol. 309, 118–127 (this issue).

IPCC (Ed.), 2001. Climate Change 2001: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK & New York.

IPCC (Ed.), 2007. Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK & New York, p. 906.

Jenness, J., Wynne, J.J., 2005. Cohen's Kappa and classification table metrics 2.0: an ArcView 3× extension for accuracy assessment of spatially explicit models. U.S. Geological Survey Open-File Report OF 2005–1363, 1–86.

Johns, T.C., Carnell, R.E., Crossley, J.F., Gregory, J.M., Mitchell, J.F.B., Senior, C.A., Tett, S.F.B., Wood, R.A., 1997. The second Hadley Centre coupled ocean atmosphere GCM : model description, spin up and validation. Climate Dynamics 132, 103–134.

Johns, T.C., Gregory, J.M., Ingram, W.J., Johnson, C.E., Jones, A., Lowe, J.A., Mitchell, J.F.B., Roberts, D.L., Sexton, D.M.H., Stevenson, D.S., Tett, S.F.B., Woodage, M.J., 2003. Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 Model under updated emissions scenarios. Climate Dynamics 20, 583–612.

Kaplan, J.O. 2001. Geophysical applications of vegetation modelling. PhD Thesis, Lund University.

Karas, C., Nurnberg, D., Gupta, A.K., Tiedemann, R., Mohan, K., Bicket, T., 2009. Mid-Pliocene climate change amplified by a switch in Indonesian subsurface through flow. Nature Geoscience 2 (6), 434–438.

Karas, C., Nurnberg, D., Tiedemann, R., Garbe-Schonberg, D., 2010. Pliocene climate change of the Southwest Pacific and the impact of ocean gateways. Earth and Planetary Science Letters 301 (1-2), 117–124.

Lisiecki, L.E., Raymo, M.E., 2005. A Pliocene–Pleistocene stack of 57 globally distributed benthic δ18O records. Paleoceanography 20PA1003.
Lunt, D.J., Valdes, P.J., Haywood, A.M., Rutt, I.C., 2008a. Closure of the Panama seaway during the Pliocene: implications for climate and northern hemisphere Glaciation. Climate Dynamics 30, 1–18.
Lunt, D.J., Foster, G.L., Haywood, A.M., Stone, E.J., 2008b. Late Pliocene Greenland Glaciation controlled by a decline in atmospheric CO2 levels. Nature 454, 1102–1105.
Lunt, D.J., Haywood, A.M., Foster, G.L., Stone, E.J., 2009. The Arctic cryosphere in the mid-Pliocene and the future. Philosophical Transactions of The Royal Society A 367, 49–67.
Lunt, D.J., Haywood, A.M., Schmidt, G.A., Salzmann, U., Valdes, P.J., Dowsett, H.J., 2010. Earth system sensitivity inferred from Pliocene modelling and data. Nature Geoscience 3, 60–64.
Matthews, E., 1985. Prescription of Land-surface Boundary Conditions in GISS GCM II: A Simple Method Based on High-resolution Vegetation Databases NASA Report, TM 86096 20.
Meehl, G., Boer, G.J., Covey, C., Latif, M., Stouffer, R.J., 2000. The Coupled Model Intercomparison Project. Bulletin of the American Meteorological Society 81, 313–318.
Murphy, J.M., Sexton, D.M.H., Barnett, D.N., Jones, G.S., Webb, M.J., Collins, M., Stainforth, D.A., 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430, 768–772.
Murphy, J.M., Booth, B.B.B., Collins, M., Harris, G.R., Sexton, D.M.H., Webb, M.J., 2007. A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. Philosophical Transactions of The Royal Society A 365, 1993–2028.
Pagani, M., Liu, Z., LaRiviere, J., Ravelo, A.C., 2010. High Earth-system climate sensitivity determined from Pliocene Carbon Dioxide concentrations. Nature Geoscience 31, 27–30.
Piani, C., Frame, D.J., Stainforth, D.A., Allen, M.R., 2005. Constraints on climate change from a multi-thousand member ensemble of simulations. Geophysical Research Letters 32, 5.
Pope, V.D., Gallani, M.L., Rowntree, P.R., Stratton, R.A., 2000. The impact of new physical parameterizations in the Hadley Centre climate model: HadAM3. Climate Dynamics 16, 123–146.
Pound, M.J., Haywood, A.M., Salzmann, U., Riding, J.B., Lunt, D.J., Hunter, S.J., 2011. A Tortonian (Late Miocene, 11.61–7.25 Ma) global vegetation reconstruction. Palaeogeography, Palaeoclimatology, Palaeoecology 300 (1–4), 29–45.
Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A., Solomon, A.M., 1992. A global biome model based on plant physiology and dominance, soil properties and climate. Journal of Biogeography 19, 117–134.
Raymo, M.E., Grant, B., Horowitz, M., Rau, G.H., 1996. Mid-Pliocene warmth: stronger greenhouse and stronger conveyor. Marine Micropaleontology 27, 313–326.
Raymo, M.E., Hearty, P., De Conto, R., O'Leary, M., Dowsett, H.J., Robinson, M.M., Mitrovica, J.X., 2009. PLIOMAX: Pliocene maximum sea level project. PAGES News, 172.
Reynolds, R.W., Smith, T.M., 1995. A high-resolution global sea surface temperature climatology. Journal of Climate 8, 1571–1583.
Roberts, M.J., Wood., R.A., 1997. Topographic sensitivity studies with a Bryan–Cox type ocean model. Journal of Physical Oceanography 26, 1495–1527.
Robinson, M.M., 2009. New quantitative evidence of extreme warmth in the Pliocene Arctic. Stratigraphy 64, 265–275.
Robinson, M.M., Valdes, P.J., Haywood, A.M., Dowsett, H.J., Hill, D.J., Jones, S.M., 2011. Bathymetric controls on Pliocene north Atlantic and Arctic Sea surface temperature and deepwater production. Palaeogeogr. Palaeoclimatol. Palaeoecol. 309, 92–97 (this issue).
Rougier, J., Sexton, D.M.H., Murphy, J.M., Stainforth, D., 2009. Analysing the Climate Sensitivity of the HadSM3 Climate Model Using Ensembles from Different but Related Experiments. Journal of Climate 22, 3540–3557.
Salzmann, U., Haywood, A.M., Lunt, D.J., Valdes, P.J., Hill, D.J., 2008. A new global biome reconstruction and data-model comparison for the middle Pliocene. Global Ecology and Biogeography 17, 432–447.
Salzmann, U., Haywood, A.M., Lunt, D.J., 2009. The past is a guide to the future? Comparing middle Pliocene vegetation with predicted biome distributions for the twenty-first century. Philosophical Transactions of The Royal Society A 367, 189–204.
Sanderson, B.M., Piani, C., Ingram, W.J., Stone, D.A., Allen, M.R., 2008a. Towards constraining climate sensitivity by linear analysis of feedback patterns in thousands of perturbed physics GCM simulations. Climate Dynamics 30, 175–190.
Sanderson, B.M., Knutti, R., Aina, T., Christensen, C., Faull, N., Frame, D.J., Ingram, W.J., Piani, C., Stainforth, D.A., Stone, D.A., Allen, M.R., 2008b. Constraints on model response to greenhouse gas forcing and the role of sub grid-scale processes. Journal of Climate 21, 2384–2400.
Seki, O., Foster, G.L., Schmidt, D.N., Mackensen, A., Kawamura, K., Pancost, R.D., 2010. Alkenone and Boron-based Pliocene pCO_2 records. Earth and Planetary Science Letters 2921–2, 201–211.
Stainforth, D.A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D.J., Kettleborough, J.A., Knight, S., Martin, A., Murphy, J.M., Piani, C., Sexton, D., Smith, L.A., Spicer, R.A., Thorpe, A.J., Allen, M.R., 2005. Uncertainty in predictions of the climate response to rising levels of greenhouse gases. Nature 433, 403–406.
Stott, P.A., Forest, C.E., 2007. Ensemble climate predictions using climate models and observational constraints. Philosophical Transactions of The Royal Society A 365, 2029–2052.
Tebaldi, C., Knutti.,, R., 2007. The use of the multimodel ensemble in probabilistic climate projections. Philosophical Transactions of The Royal Society A 365, 2053–2075.
Tziperman, E., Toggweiler, J.R., Feliks, Y., Bryan, K., 1994. Instability of the thermohaline circulation with respect to mixed boundary conditions: is it really a problem for realistic models? Journal of Physical Oceanography 24, 218–232.
Webb, M.J., Senior, C.A., Sexton, D.M.H., Ingram, W.J., Williams, K.D., Ringer, M.A., McAveney, B.J., Colman, R., Soden, B.J., Gudgel, R., Knutson, T., Emori, S., Ogura, T., Tsushima, Y., Andronova, N., Li, B., Musat, I., Bony, S., Taylor, K.E., 2006. On the contribution of local feedback mechanisms to the range of climate sensitivity in two GCM ensembles. Climate Dynamics 27, 17–38.
Williams, K.D.S., C.A., Mitchell, J.F.B., 2001. Transient climate change in the Hadley Centre models: the role of physical processes. Journal of Climate 14, 2659–2674.