The hydrological cycle and ocean circulation of the Maritime Continent in the mid-Pliocene: results from PlioMIP2

Xin Ren[1], Daniel J. Lunt[1], Erica Hendy[1], Anna von der Heydt[2], Ayako Abe-Ouchi[3],
Bette Otto-Bliesner[4], Charles J. R. Williams[1,5], Christian Stepanek[6], Chuncheng Guo[7],
Deepak Chandan[8], Gerrit Lohmann[6], Julia C. Tindall[9], Linda E. Sohl[10,11], Mark A. Chandler[10,11],
Masa Kageyama[12], Michiel L. J. Baatsen[2], Ning Tan[13,14], Qiong Zhang[15], Ran Feng[16], Wing-Le Chan[3],
W. Richard Peltier[8], Xiangyu Li[17,18,19], Youichi Kamae[20], Zhongshi Zhang[7,18], and Alan M. Haywood[9]

[1]School of Geographical Sciences, University of Bristol, Bristol, UK
[2]Institute for Marine and Atmospheric research Utrecht (IMAU), Department of Physics, Utrecht University, Utrecht, the Netherlands
[3]Atmosphere and Ocean Research Institute, The University of Tokyo, Kashiwa, Japan
[4]Climate and Global Change Dynamics lab, National Center for Atmospheric research, USA
[5]NCAS, Department of Meteorology, University of Reading, Reading, UK
[6]Alfred-Wegener-Institut – Helmholtz-Zentrum für Polar and Meeresforschung (AWI), Bremerhaven, Germany
[7]NORCE Norwegian Research Centre, Bjerknes Centre for Climate Research, Bergen, Norway
[8]Department of Physics, University of Toronto, Toronto, Canada
[9]School of Earth and Environment, University of Leeds, Woodhouse Lane, Leeds, West Yorkshire, UK
[10]Center for Climate Systems Research at Columbia University, New York, NY, USA
[11]NASA Goddard Institute for Space Studies, New York, NY, USA)
[12]LSCE/IPSL – Laboratoire des Sciences du Climat et de l'Environnement, UMR8212, CEA-CNRS-UVSQ – CE Saclay, L'Orme des Merisiers, Gif-sur-Yvette Cedex, France
[13]Key Laboratory of Cenozoic Geology and Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China
[14]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France
[15]Department of Physical Geography and Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden
[16]Department of Earth Sciences, College of Liberal Arts and Sciences, University of Connecticut, Storrs, USA
[17]Centre for Severe Weather and Climate and Hydro-geological HazardsWuhan, China
[18]Department of Atmospheric Science, School of Environmental Studies, China University of Geosciences, Wuhan, China
[19]Climate Change Research Center, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
[20]Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

Correspondence: Xin Ren (xinxin.ren@bristol.ac.uk)

Abstract. The Maritime Continent (MC) forms the western boundary of the tropical Pacific Ocean, and relatively small changes in this region can impact the climate locally and remotely. In the mid-Pliocene (from 3.264 to 3.025 million years before present), atmospheric CO_2 concentrations were $\sim$ 400 ppm, and the subaerial Sunda and Sahul shelves made the land-sea distribution of the MC different to today. Topographic changes and elevated levels of CO_2, combined with other forcings,

5   are therefore expected to have driven a substantial climate signal in the MC region at this time. By using the results from the Pliocene Model Intercomparison Project phase 2 (PlioMIP2) we study the mean climatic features of the MC in the mid-Pliocene and changes in Indonesian Throughflow (ITF) with respect to preindustrial. Results show a warmer and wetter mid-Pliocene

climate of the MC and lower sea surface salinity in the surrounding ocean compared with preindustrial. Furthermore, we quantify the volume transfer through the ITF; although the ITF may be expected to be hindered by the subaerial shelves, 10

10   out of 15 models show an increased volume transport compared with preindustrial.

In order to avoid undue influence from closely-related models that are present in the PlioMIP2 ensemble, we introduce a new metric - the multi-cluster mean (MCM), based on cluster analysis of the individual models. We study the effect that the choice of MCM versus the more traditional analysis of multi-model mean (MMM) and individual models has on the discrepancy between model results and reconstructed proxy data. The clusters reveal spatial signals that are not captured by the MMM, so

15   that the MCM provides us with a new way to explore the results from model ensemble that include similar models.

1   Introduction

The Maritime Continent (MC; 10°S–20°N and 90°E–150°E), which consists of more than 22,000 islands, lies in the warmest ocean region, the West Pacific Warm Pool (WPWP) (Yoneyama and Zhang, 2020; RAMAGE, 1968). The near-surface air temperature in most areas of the MC is higher than 27°C throughout the year (Li et al., 2018a). High sea surface temperature

20   leads to substantial moisture flux into the atmosphere. High levels of moisture and energy characterize the unstable climate in this region and make it "the world's strongest atmospheric convection center" (Yoneyama and Zhang, 2020). Based on satellite data it has been established that the MC is the region that produces the largest amount of precipitation in the world (Adler et al., 2017; Yamanaka et al., 2018). A large amount of rainfall releases large quantities of latent heat, that is an important driver of global atmospheric circulation, into the atmosphere (1 mm/y equals to 0.08 W/m^2 (Trenberth et al., 2009; Yamanaka

25   et al., 2018)). According to these characteristics, the MC has been recognized as the "boiler box" of the global climate system (RAMAGE, 1968; Neale and Slingo, 2003).

The MC also impacts on the oceanic hydrological system in that it links the Pacific Ocean and the Indian Ocean via the Indonesian Throughflow (ITF), which is a key component of global ocean circulation. In general, the ITF transports warm and comparably fresh water from the Pacific Ocean into the Indian Ocean. Dramatic tectonic processes and sea level changes shaped

30   this gateway that plays an important role in influencing local climate and that of other regions of the world (Gordon, 2005; Tillinger, 2011; Yoneyama and Zhang, 2020). According to the International Nusantara Stratifcation and Transport Program (INSTANT), from 2004 to 2006 (Gordon et al., 2010; Sprintall et al., 2009) the ITF transported about 15 Sv of water from Pacific Ocean into the Indian Ocean; the heat flux, also known as transport-weighted temperature, of the total ITF export is 17.6 °C during these three years (Sprintall et al., 2009).

35   The MC acts as a source of sensible heat for the atmosphere over land, and the geographical configuration of the MC redistributes sea surface temperatures (SSTs) in the western Pacific and the Indian Ocean; both the effect of sensible heat and the geography of the MC are crucial for the onset of the Walker Circulation (Dayem et al., 2007). The variation of the Walker Circulation leads to various kinds of periodic phenomena that modulate the climate of the MC, including the El Niño–Southern Oscillation (ENSO), the Indian Ocean Dipole (IOD) and the Pacific Decadal Oscillation (PDO). The MC forms the western

40   boundary of the ENSO phenomenon. On the interannual scale, ENSO and IOD play important roles in modulating rainfall

and the ITF transport by affecting zonal winds (Wang, 2019). During El Niño years, the easterly trade winds in the Pacific are weakened. As a result, rainfall is decreased and the ITF is weakened (Zhang et al., 2016; Yamanaka et al., 2018) with a peak-to-trough amplitude of about 5 Sv (Feng et al., 2018). During La Niña years, the situation is reversed. The positive phase of IOD normally occurs during the autumn of an El Niño developing year, which offsets the effects of El Niño on the ITF (Wang,

45 2019; Sprintall and Révelard, 2014). On decadal timescales, the PDO plays a dominant role in regulating the ITF (Wang, 2019; Li et al., 2018b). Similar to the effect of ENSO, in the warm phase of the PDO easterly trade winds become weaker, causing the ITF transport to become weaker, and vice versa (Wang, 2019). A modelling study from Tan et al. (2022) highlighted the importance of the geometry of the MC, and found that the shallow opening of the MC can trigger an active Pacific meridional overturning circulation and enhance upwelling strength along the equator from the Central to Eastern Pacific.

50   It is possible to obtain an "atlas" of the future warming world from climate model simulations (e.g. Iturbide et al. (2021); Gutiérrez et al. (2021)). However, there is no direct evidence for us to validate how well climate models simulate the future. The mid-Pliocene is an interval in the mid-Piacenzian between 3.264 and 3.025 million years (Ma) before Present (BP). According to the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6) Working Group I (WGI) Technical Summary (Arias et al., 2021), the mid-Pliocene is *very likely* 2.5°C to 4.0°C warmer than the preindustrial, and the

55 concentration of CO_2 during this period is *very likely* in the range from 360 ppmv to 420 ppmv, which is substantially higher than the preindustrial period (280 ppmv). In terms of temperature, and geography, the Earth System of the mid-Pliocene is similar to projections for the end of the 21st century (Dowsett et al., 2016). Consequently, the mid-Pliocene is considered one of the most recent analogues to the future warm conditions on a broad scale perspective (Salzmann et al., 2009; Burke et al., 2018).

60   During the Pliocene Epoch, the Indonesian Gateway reorganized, and the ITF was restricted progressively (Karas et al., 2009; Auer et al., 2019). This process contributed to variations of climate throughout the mid-Pliocene. For example, during the mid to late Piacenzian (3.3 Ma –2.6 Ma BP), the reorganization of the ITF led to a strengthening of the South Asian Summer Monsoon and acted as a precursor to east African aridification (Cane and Molnar, 2001; Sarathchandraprasad et al., 2021). By using model results, we can explore the mid-Pliocene climate in this region. This effort is supported by geological records that

65 provide a variety of evidence towards reconstructing the climate over the MC during the mid-Pliocene. For example, deep-sea deposit samples of the mid-Pliocene can be utilized to reconstruct sea surface temperature (Wara et al., 2005; Zhang et al., 2014; Fedorov et al., 2015; O'Brien et al., 2014; McClymont et al., 2020). By comparing results of simulations and geological records, we can validate climate models and gain confidence in projections of future climate if the models perform well for past warm climates.

70   The mid-Pliocene has been broadly studied; previous works by the Pliocene Research Interpretation and Synoptic Mapping (PRISM; e.g., Dowsett et al. (2013, 2016)) and the Pliocene Model Intercomparison Project (PlioMIP; e.g, Haywood et al. (2010, 2016a, 2020)) provide us with a large amount of paleoclimate reconstructions and simulations of environmental and climatic information. The second phase of PlioMIP (PlioMIP2), that this study relates to, is briefly described in section 2 of this paper. Large-scale climate features of the mid-Pliocene as derived from PlioMIP1 and PlioMIP2 are described in Haywood

75 et al. (2013) and Haywood et al. (2020): On a global scale simulated annual averaged surface air temperature is elevated,

with models suggesting an anomaly above the respective conditions of the preindustrial range between 1.84 and 3.6°C based
on PlioMIP1 coupled atmosphere–ocean climate models only and between 1.7 and 5.2°C based on PlioMIP2; annual mean
total precipitation rates are higher by 0.09-0.18 mm d^{-1} (PlioMIP1 coupled atmosphere–ocean climate models only) and
0.07-0.37 mm d^{-1} (PlioMIP2) in the mid-Pliocene. All the models from PlioMIP1 and PlioMIP2 simulated a clear polar

80 amplification of the warming in the mid-Pliocene and show that the warming magnitude in the MC is less than the global
average. Regarding precipitation changes in the region, the PlioMIP1 ensemble shows an increase over most parts of the MC,
especially in the tropics (Fig. 3b in Haywood et al. (2013)). In contrast to PlioMIP1, PlioMIP2 shows spatially heterogeneous
changes in precipitation over the MC (Fig. 5b in Haywood et al. (2020)). Moreover, there is an increase in precipitation minus
evaporation over exposed continents across this region (Fig. 1 in Feng et al. (2022)). PlioMIP1 studies have shown that tropical

85 overturning circulation slows down in the mid-Pliocene due to the changes of sea surface temperature, and that the ascending
branch of the Walker Circulation over the MC weakens (Corvec and Fletcher, 2017). In terms of climate variability, ENSO
exists in the modelled mid-Pliocene, but the amplitude of ENSO is weaker than in the preindustrial. This inference is consistent
from PlioMIP1 (Brierley, 2015) to PlioMIP2 (Oldeman et al., 2021). However, the mean zonal SST gradient does not decrease
consistently across models in PlioMIP2 (Oldeman et al., 2021). Although the MC is important for regional and large-scale

90 climate, there are relatively few studies of the mid-Pliocene focusing on the MC (e.g. Smith et al. (2020)), especially studies
that use the state-of-the-art models. In order to explore this issue further, in this paper we investigate the climate of the MC in
the mid-Pliocene by employing PlioMIP2 simulations.

We study large-scale climate patterns (sea surface temperature, the net water flux at the surface (precipitation minus evapo-
ration), salinity at the sea surface and wind stress) of the MC and investigate ocean circulation and strength of the ITF in the

95 mid-Pliocene. Moreover, we assess the ability of the PlioMIP2 models to simulate the mid-Pliocene climate of the MC by cross-
validating simulations and models with proxy reconstructions and reanalysis data. We investigate differences in predictions by
individual models with a hierarchical clustering method.

The rest of the paper is organized as follows. Section 2 describes the PlioMIP2 models, experiment design and simulation
data that are relevant to this study. Section 3 addresses the following 4 main question:

100 – Q1. How do PlioMIP2 models perform in simulating the mid-Pliocene and preindustrial climate of the MC? (Section
     3.1);

   – Q2. How were climate (sea surface temperature, net water flux into the surface, salinity at sea surface and wind stress)
     shaped in the mid-Pliocene on the MC in the PlioMIP2 simulations? (Section 3.2);

   – Q3. Did the volume transport via the ITF intensify in the mid-Pliocene, and what were the characteristics of this through-
105   flow? (Section 3.3);

   – Q4. How can we best emphasize results from individual models that may otherwise disappear in the multimodel ensemble
     mean (MMM), and avoid duplication of biases shared by model from the same family in the MMM? (Section 3.4)

Finally, we discuss drivers for changes of the ITF and of the climate of the MC (Section 4.1); we summarize performance of individual models and of the MMM in simulating mid-Pliocene and pre-industrial climate (Section 4.2); and we address
110   similarities and differences between results by individual models (Section 4.3).

2 PlioMIP2

In this section, we describe the PlioMIP2 models, experiments and the PlioMIP2 project. In section 2.1, we introduce the PlioMIP2 models and the development of models which participated in the PlioMIP phase 1 and phase 2. Then, the two core experiments of PlioMIP2 adopted in this study are described in terms of boundary conditions (section 2.2).

115   ### 2.1 PlioMIP2 models

PlioMIP (Haywood et al., 2011) is a project aimed at studying climate and environments of the late Pliocene. PlioMIP is part of the Palaeoclimate Modelling Intercomparison Project (PMIP). In consideration of the development of models and proxy data in the past years, this project has progressed to the next phase (PlioMIP2) (Haywood et al., 2016b, 2020). In the second phase, there are more models (17 models in comparison to 11 models that participated in PlioMIP phase 1). Most of the newly
120   participating models are descendants of existing models. For example, CCSM4 participated in PlioMIP1. The models derived from it, CCSM4-Utr, CCSM4-UoT, CESM1.2 and CESM2, participated in the second phase; IPSLCM5A from the Institute Pierre-Simon Laplace (IPSL) participated in PlioMIP, and then joined PlioMIP2 with IPSLCM5A2 and IPSL-CM6A-LR; NorESM-L, which has been developed from CCSM4, participated in the first phase, and then participated in the second phase with NorESM1-F. The atmosphere-only model HadAM3 participated in PlioMIP1 as a component of the UK Hadley Cen-
125   tre atmosphere-ocean general circulation model HadCM3, in PlioMIP2 it contributed via the "CMIP6-class" UK Met Office HadGEM3-GC31-LL (Williams et al., 2021), as well as via HadCM3 (Hunter et al., 2019). Refer to Table 1 in Haywood et al. (2020) which shows the components used in the PlioMIIP2 models. In terms of the atmospheric component and oceanic component, new model versions were developed in that the parameterization changed (CCSM4-Utr and CCSM4-UoT (Peltier and Vettoretti, 2014) use a simplified version of ocean mixing scheme from CCSM4 with parameterization changes; NorESM1-F
130   updated the parameterization from NorESM-L), model components were updated (CESM1.2 and CESM2 adopted the atmospheric components of CAM5 and CAM6), or the resolution has been enhanced (IPSL models; NorESM1-F applied the tripolar grid in the ocean component compared with NorESM-L). All these models have run the core mid-Pliocene experiment (Eoi400) and provide the preindustrial experiment (E280) as a control simulation. Table 1 shows details of the 16 PlioMIP2 models and of HadGEM3. For more details on the 16 PlioMIP2 models, see Haywood et al. (2020).

Table 1: Details of the 16 PlioMIP2 models (Haywood et al., 2020) and of the HadGEM3 model used in this study.

Model ID	modelling centre responsible for simulation	Atmosphere component, resolution, and layers	Ocean component, resolution, and layers	equilibrium climate sensitivity (ECS)	PlioMIP2 Publication
CCSM4	National Centre for Atmospheric Research, US	CAM4 FV0.9°×1.25° L26 (Neale et al., 2010)	Parallel Ocean Program version 2 (POP2) G16 (∼ 1°) L60 (Danabasoglu et al., 2012)	3.2	(Feng et al., 2020a)
CCSM4-Utr	Utrecht University, the Netherlands	CAM4 FV2.5°× 1.9° L26	POP2 with parameterization changes G16 (∼ 1°) L60	3.2	(Baatsen et al., 2022)
CCSM4-UoT	University of Toronto, Canada	CAM4 FV0.9°×1.25° L26	POP2 with parameterization changes G16 (∼ 1°) L60	3.2	(Chandan and Peltier, 2017)
CESM1.2	National Centre for Atmospheric Research, US	CAM5 FV0.9°×1.25° L30 (Neale et al., 2010)	POP2 G16 (∼ 1°) L60	4.1	(Feng et al., 2020a)
CESM2	National Centre for Atmospheric Research, US	CAM6 FV0.9°×1.25° L32 (Danabasoglu et al., 2020)	POP2 with update on parameterization and schemes G16 (∼ 1°) L60 (Danabasoglu et al., 2020)	5.3	(Feng et al., 2020a)
COSMOS	Alfred Wagner Institute, Germany	ECHAM5 T31 (3.75°× 3.75°) L19 (Roeckner et al., 2003)	MPI-OM GR30 (3.0°× 1.8°) L40 (Marsland et al., 2003)	4.7	(Stepanek et al., 2020)

EC-Earth3.3	Stockholm University, Sweden	IFS cycle 36r4 T159 (∼1.125°× 1.125°) L62	NEMO3.3 ORAC1 (1°× 1°) L46 (Madec, 2012)	4.3	(Zhang et al., 2021)
GISS-E2-1-G	Goddard Institute for Space Studies, US	2.0°× 2.5° L40 (Kelley et al., 2020)	GISS Ocean v1 1.0°× 1.25° L40 (Kelley et al., 2020)	3.3	n.a.
HadCM3	University of Leeds, UK	2.5°× 3.75° L19	1.25°× 1.25° L20	3.5	(Hunter et al., 2019)
IPSLCM5A	Laboratoire des Sciences du Climat et de l'Environnement, France	LMDZ5 revision 2063 3.75°× 1.875° L39 (Hourdin et al., 2013)	NEMOv3.2 2.0°× 2.0°, 0.5°in the tropics L31 (Madec, 2012)	4.1	(Tan et al., 2020)
IPSLCM5A2	Laboratoire des Sciences du Climat et de l'Environnement, France	LMDZ5 revision 3342 3.75°× 1.875° L39 (Hourdin et al., 2013; Sepulchre et al., 2020)	NEMOv3.6 2.0°× 2.0°, 0.5°in the tropics L31 (Sepulchre et al., 2020)	3.6	(Tan et al., 2020)
IPSL-CM6A-LR	Laboratoire des Sciences du Climat et de l'Environnement, France	LMDZ6A-LR 2.5°× 1.26° L79 (Hourdin et al., 2020; Boucher et al., 2020)	NEMOv3.6 ∼1°, latitudinal refined at 1/3°in the equatorial region L75 (Hourdin et al., 2020)	4.8	(Tan et al., 2020)
MIROC4m	University of Tokyo, Japan	T42 (∼2.8°× 2.8°) L20	1.4°longitude, 0.56°-1.4°latitude L43	3.9	(Chan and Abe-Ouchi, 2020)
MRI-CGCM2.3	University of Tsukuba, Japan	T42 (∼2.8°× 2.8°) L30	2.5°longitude, 2.0 - 0.5°latitude L26	2.8	(Kamae et al., 2016)

NorESM-L	NORCE Norwegian Research Centre, Bjerknes Centre for Climate Research, Norway	CAM4 Oslo version T31 (∼3.75°× 3.75°) L26	Miami Isopycnic Co-ordinate Ocean Model (MICOM) G37 (∼3.0°× 3.0°) L30	3.1	(Li et al., 2020)
NorESM1-F	NORCE Norwegian Research Centre, Bjerknes Centre for Climate Research, Norway	CAM4 Oslo version 1.9°× 2.5° L26	MICOM ∼1.0°× 1.0°L53	2.3	(Li et al., 2020)
HadGEM3	University of Bristol, UK	N96 (1.875°× 1.25°) L85	NEMO 3.6 1.0°× 1.0°L75	5.55	(Williams et al., 2021)
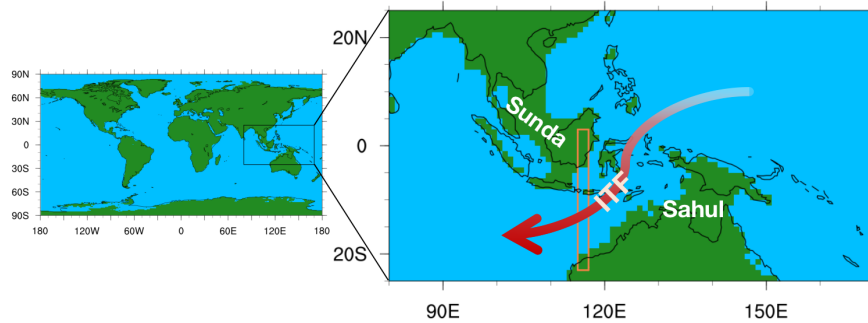
Figure 1. PRISM4 Pliocene land-sea mask. The modern continental outline is indicated with black lines; blue and green shadings indicate ocean and land in the Eoi400 mid-Pliocene experiment. The red rectangle indicates the gateway we defined in this work to calculate the intensity of the Indonesian Throughflow (ITF) in section 3.3. The red arrow denotes the position and direction of the ITF illustratively. The Sunda and Sahul shelves are illustrated in this figure via white labels.

135 ## 2.2 Experimental Design

The global-scale and gridded paleo reconstruction data from PRISM provides data for mid-Pliocene paleo-geography and land ice, etc., to enable preparation of reliable boundary conditions and to carry out climate models simulations of the mid-Pliocene. For PlioMIP phase 1, experiment boundary conditions for the mid-Pliocene have been built based on PRISM3D (Dowsett et al., 2010). In PlioMIP2, new boundary conditions have been adopted, which are derived from the state-of-the-art
140   reconstruction PRISM4 (Dowsett et al., 2016).

In this study we adopt the E280 simulation as a control experiment which is forced with preindustrial conditions, and the Eoi400 simulation as a sensitivity experiment for mid-Pliocene conditions which is forced with all mid-Pliocene boundary conditions; these are topography, ice sheets, soil, vegetation, lakes, land-sea mask and a concentration of 400 ppm of CO_2 in the atmosphere. Simulation naming follows the rules established by Haywood et al. (2016a). The numbers 280 and 400
145   indicate the CO_2 concentration; character o indicates changes to orography, bathymetry, land–sea mask, lakes, and soils from the preindustrial model setup towards mid-Pliocene conditions, character i indicates such changes of ice sheets. Fig. 1 shows the land-sea masks used in the E280 and the Eoi400 experiments indicated by black lines and colours, respectively, and the location of the MC. From this map it is clear that in the mid-Pliocene, several modern ocean gateways were absent, such as the Bering Strait, the connection between Canada and Greenland (Canadian Arctic Archipelago), and some straits in
150   the MC. Details of the experimental design are described by Haywood et al. (2016a). Note that HadGEM3 did not change the land sea mask in the Eoi400 experiment, and therefore for this model the various gateways are the same as in E280. Data from the experiments E280 and Eoi400 used in this study are from the PlioMIP2 dataset, part of which is available at https://esgf-node.llnl.gov/search/cmip6/. In this study, the majority of the results are calculated across the final 100 model years of each simulation. Data has been regridded to a 1° × 1° grid in order to calculate ensemble means; results for individual
155   models are illustrated based on the original grids. The regional averaged value in this paper is calculated from region of 25°S-25°N and 80°E-170°E.

3   Results

In section 3.1, model results are compared to proxy reconstructions and reanalysis data to evaluate model performance for reproducing mid-Pliocene and modern climate. Furthermore, the multimodel mean method (MMM) is evaluated by comparing
160   its results with those derived from individual models. Then in section 3.2, we illustrate the most relevant patterns of mean climate of the MC in mid-Pliocene relative to the preindustrial level in terms of sea surface temperature, precipitation minus evaporation, wind stress and salinity at the ocean surface. Subsequently, we analyse the oceanic circulation in the MC (section 3.3). Lastly, we introduce a new metric - the multi cluster mean (MCM), which takes the weight of models from the different clustering branches into consideration to better illustrate the results from multiple models (section 3.4). Finally, we present
165   comparison between the performance of the multi-cluster mean (MCM), the MMM and individual models.

3.1   Model Skill at Simulating preindustrial and mid-Pliocene MC climate
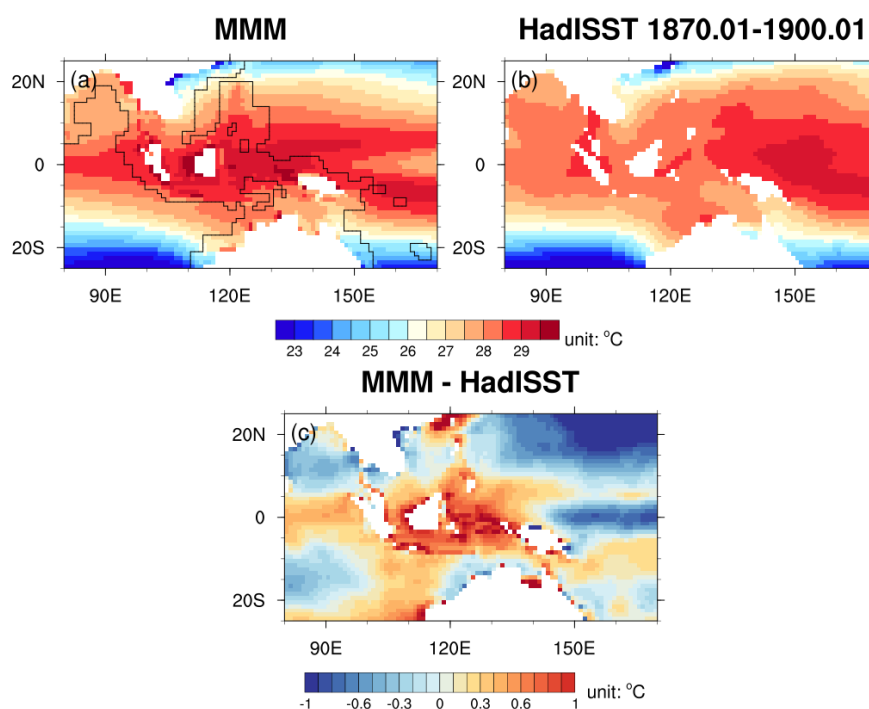
Figure 2. The annual mean sea surface temperature (SST) over the MC from: (a) multimodel ensemble mean (MMM) of the pre-industrial (E280), (b) HadISST1 reanalysed SSTs between January 1870 to January 1900 and (c) their difference. White shading indicates the grid boxes that are land for all the models. In (a), grid boxes outside the black lines are ocean grid boxes for all the models.

In this section, we investigate how well a model performs in reproducing preindustrial climate (section 3.1.1) and mid-Pliocene climate (section 3.1.2). Our aim is to address the following questions: 1. Are the models that perform well in simulating preindustrial climate of the MC also good at simulating mid-Pliocene climate? 2. Does the multimodel ensemble have
170 improved performance for the MC compared to individual models, as is often the case in weather and climate prediction (e.g.Tebaldi and Knutti (2007))?

3.1.1 Comparison with HadISST1

Observational data and reconstructed climatic information from proxy data can offer a way to evaluate simulation results. Here, we adopt the Hadley Centre sea ice and sea surface temperature (SST) reanalysis version 1 (HadISST1) (Rayner, 2003) from
175 January 1870 to January 1900 to evaluate model performance in reproducing preindustrial climate. The HadISST1 dataset is available at https://www.metoffice.gov.uk/hadobs/hadisst/.

Fig. 2 shows the spatial distribution pattern of the preindustrial SSTs over the MC from the HadISST1 reanalysed SSTs and MMM. Despite that models reproduce the SST zonal gradient from the equator to higher latitudes, simulated SSTs are warmer surrounding the tropical islands and cooler in the North Western Pacific Ocean compared to the reanalysis data. By
180 calculating the root mean square error (RMSE) between reanalysed SSTs and model SSTs over the MC, we quantify the model and observation discrepancy for the modern period, which is shown in Fig. S1 and Table S1 in the supplement. The mean discrepancy in the MC between the MMM SSTs and HadISST1 reanalysed SSTs is 0.51°C, which is lower (better) than 13 out of 17 models. By that metric, the MMM better reproduces preindustrial climate in this region than most of the PlioMIP2 models. The MMM metric also shows better agreement with observations than individual simulations in many other studies
185 confer, for example, the study by Williams et al. (2022) in the Deep-Time Model Intercomparison Project (DeepMIP), where the MMM is the best estimate of the actual precipitation in terms of spatial patterns relative to individual DeepMIP models.

In terms of individual models, CCSM4 shows the lowest discrepancy of any individual model (0.39°C); IPSLCM5A and HadCM3 show the highest discrepancies (1.29°C). In order to evaluate the simulated SSTs' spatial patterns, the discrepancy between the MMM SSTs and the HadISST1 reanalysed SSTs is also calculated after bias adjustment (i.e. removing the mean
190 SST over the MC). The values are as shown in Table S1 in the supplement.
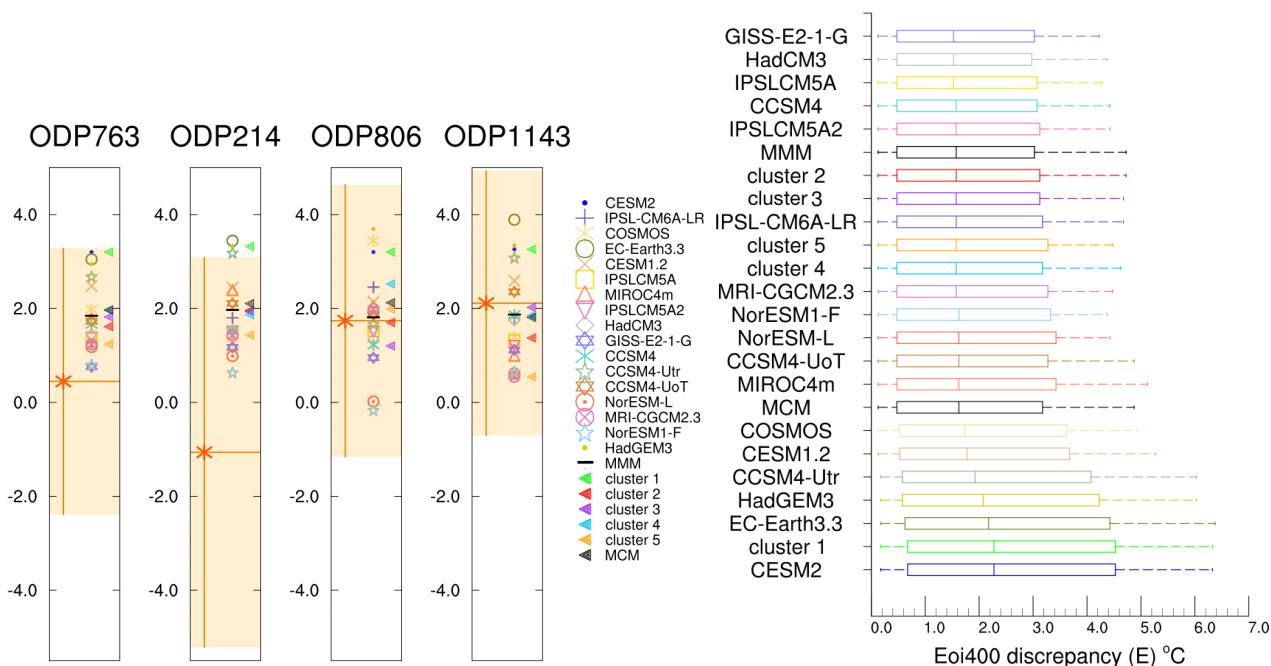
Figure 3. Discrepancy between modelled and reconstructed SST anomalies (SSTA) in the Pliocene. (a) Comparison between simulated and reconstructed SSTAs for 4 sites in the MC. The location of these 4 sites can be seen in Fig. 5a. The orange vertical bars and shading indicate the double standard errors ($\pm 2\sigma$) of the reconstructed SSTs. The orange horizontal bars and the asterisk indicate the reconstructed SSTAs of the site, individually. The short black horizontal bars among models' results indicate the MMM of the simulated SSTAs. (b) Box-plot of model and proxy discrepancies over the MC for every model. Vertical lines inside the boxes are the median values (E) of the possibility distribution function of the discrepancies derived via Monte-Carlo simulations, which is explained in section 3.1.2. Lower and upper box boundaries indicate the interval where the model and proxy discrepancies are *likely* to be (probability range from 66% to 100%). Lower and upper error lines indicate the interval where the model and proxy discrepancies are *most likely* to be (probability range from 90% to 100%). Models are ordered by their value of E. Also shown are the E values for the 5 clusters and the MCM.

3.1.2 Comparison with Pliocene proxies

For the mid-Pliocene, multi-proxy reconstructed SST anomalies (SSTA) data compiled by McClymont et al. (2020) are available for a narrow time slice centred on 3.205 Ma BP. This data (McClymont et al., 2020) is compared with the simulated mid-Pliocene SSTA in this study. Data and the full details for reconstructed SSTA are available at https://doi.pangaea.de/10.
1594/PANGAEA.911847McClymont et al. (2020). There are 4 sites with available data over the MC, as shown in Fig. 5a. The SSTA for ODP1143 is reconstructed with the alkenone-derived (BAYSPLINE) calibration (Tierney and Tingley, 2018); the SSTA for ODP763, ODP214 and ODP806 are reconstructed with the Mg/Ca-derived (BAYMAG) calibration (Tierney et al., 2019). Fig. 3a shows reconstructed and modelled SSTAs for every site individually. Model results for each site are provided as the average of the 9 grid boxes centred on the site to reduce regional bias. For the site ODP763, which is located close to the shore of the MC, and for the site ODP214, which is located in the Indian Ocean, the models all simulate SSTAs that are warmer than the reconstructed data, especially for the BAYMAG-based SSTA estimate at site ODP214, the proxy data show that the Indian Ocean is cooler in the mid-Pliocene than in the preindustrial, but all the models show a warmer SSTA. We note that there is a persistent mismatch between Mg/Ca-derived (BAYMAG) and alkenone-derived (BAYSPLINE) SSTAs. McClymont et al. (2020) show that generally and with few exceptions, Mg/Ca-derived SSTAs are lower than their alkenone-based counterparts. In particular at lower latitudes there is a substantial mismatch, where Mg/Ca even suggests a negative SSTAs mid-Pliocene versus preindustrial, that is not supported by alkenones. Furthermore, studies such as Zhang et al. (2014) and Smith et al. (2020) that reconstructed SSTAs from organic geochemical proxies also suggest that the SSTA in site ODP806 and region near the ITF outlet could potentially be warmer. In the models' results, simulated SSTAs over a broader region of the Indian Ocean surrounding the site ODP214 are all higher in the Eoi400 simulation than the E280 experiment. There are large standard errors (σ) for the proxy reconstructed SSTAs, with $\pm 2\sigma$ shown with the orange shading in Fig. 3a. Most of the models are within 2σ of the reconstructed SSTAs uncertainties. For the site ODP806 in the Pacific warm pool, the MMM SSTA (1.81°C) is very close to the reconstructed SSTAs (1.74°C); for the sites ODP763, ODP806, and ODP1143, all the models' results are within the uncertainties of reconstructions. This implies that, in terms of SSTA, the models perform well at simulating the mid-Pliocene SSTAs over one site of the Western Pacific Warm Pool (WPWP) and the South China Sea, and also have skill in the Indian Ocean close to the coast of Australia. The exception is site ODP214, where models have less skill in reproducing the SSTA of this site. A few of models suggest comparably warm mid-Pliocene SSTA that are outside the uncertainty range of the reconstruction. Yet, we note that relative cold SSTA is derived from the Mg/Ca-reconstruction at this site.

In order to quantify the discrepancies between proxy reconstructed and simulated SSTAs from individual models, we apply Monte-Carlo simulation (Kageyama et al., 2021). The results are shown in Fig. 3b, sorted in ascending order of model and proxy discrepancy median value E. Since there are uncertainties for proxy data, we generate an array containing 1000 random SSTAs using a uniform range distribution derived from the reconstructed SSTAs and $\pm 2\sigma$ for every site. Uncertainties of the simulated SSTAs are derived over the last 100 years of annual mean SSTAs. By subtracting the reconstructed SSTAs arrays from the 100 simulated SSTAs for every site separately, and combining the arrays for the 4 sites, we derive arrays of model and proxy discrepancy for individual models which contains 400,000 (1000×100×4) SSTAs. A box plot illustrating *likely* and *very*
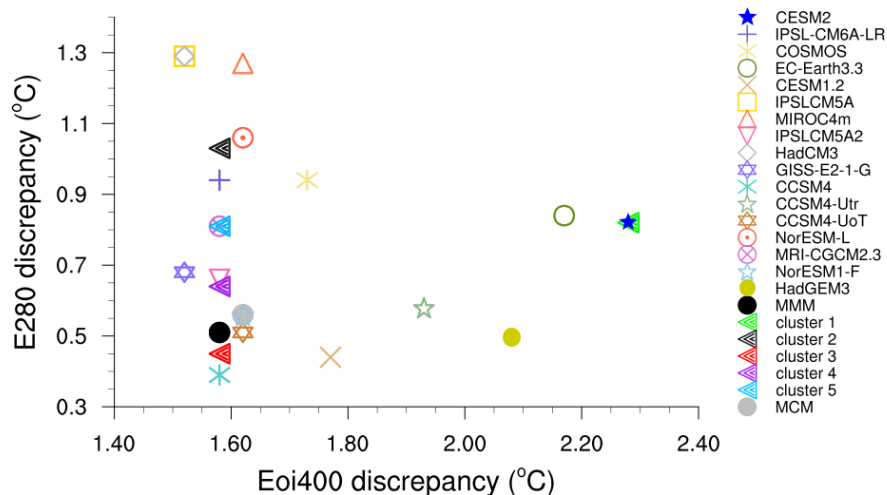
Figure 4. Comparison of the model and data discrepancies between the E280 and Eoi400 simulations. Discrepancy values are as shown in Table S1. The unit is °C.

225 *likely* range of the probability density distribution function of the model and proxy absolute discrepancy is shown for every

model in Fig. 3b. The vertical lines inside the boxes are the medians of the probability density distribution function, which

means that there is a probability of 50% that model and proxy discrepancy are less than this value. Here we use letter E to

denote discrepancy. The box in Fig. 3b shows the range of the *likely* model and proxy discrepancy; the dash error bar shows the

range of the *very likely* model and proxy discrepancy. The likelihood definition adopted here is that from the IPCC Guidance

230 Note (Mastrandrea et al., 2010) *likely* means the probability of the outcome ranges from 66% to 100%; *very likely* means the

probability of the outcome ranges from 90% to 100%. By using this method, we take uncertainties of proxy data and model

simulation into account. From Fig. 3b we find that the E values between proxy and models range from 1.52°C to 2.28°C. As

the majority of the model SSTAs show a warmer bias than the reconstructed SSTAs (Fig. 3a), the models showing relatively

less regional warming also show a lower discrepancy value E with proxy SSTAs over the MC than the models that produce

235 higher warming (Fig. S3). The PlioMIP2 models' MMM shows a discrepancy value E of 1.58°C, which is lower than that of

most of the PlioMIP2 individual models.

In order to explore whether those models, that perform well in simulating preindustrial SSTs over the MC, are also good at simulating the mid-Pliocene SSTAs over the MC, the discrepancies of models and observed SSTs on the one hand and reconstructed SSTAs on the other hand for both experiments are shown in Fig. 4. In terms of simulating the climate of the MC, there is no clear correlation between model performance in preindustrial climate and in mid-Pliocene climate, implying that models which reproduce preindustrial climate of the MC well are not always good at simulating mid-Pliocene climate in accordance with the proxy reconstructions. However, from Fig. 4, models seems to fall into two groups: models where the discrepancy in the Eoi400 experiment appear decoupled from those in the E280 experiment, and models where regional biases are linked between the E280 and Eoi400 simulations (including CESM2, EC-EARTH3.3, HadGEM3, NorESM1-F and CESM1.2). Models in the second group are also appeared to be relatively warm in the Eoi400 simulation (Fig. 6). In terms of simulating the global SSTA, the comparison of the model and data discrepancies on global scale between the E280 and Eoi400 simulations is shown in Fig. S4, which shows a better correlation than the regional SST simulation.

3.2 Mean Climate Features

In this section, we describe the modelled mean climatic characteristics of the mid-Pliocene relative to the preindustrial over the MC in terms of SST, wind stress, precipitation minus evaporation (P-E), and salinity at the ocean surface (SOS).
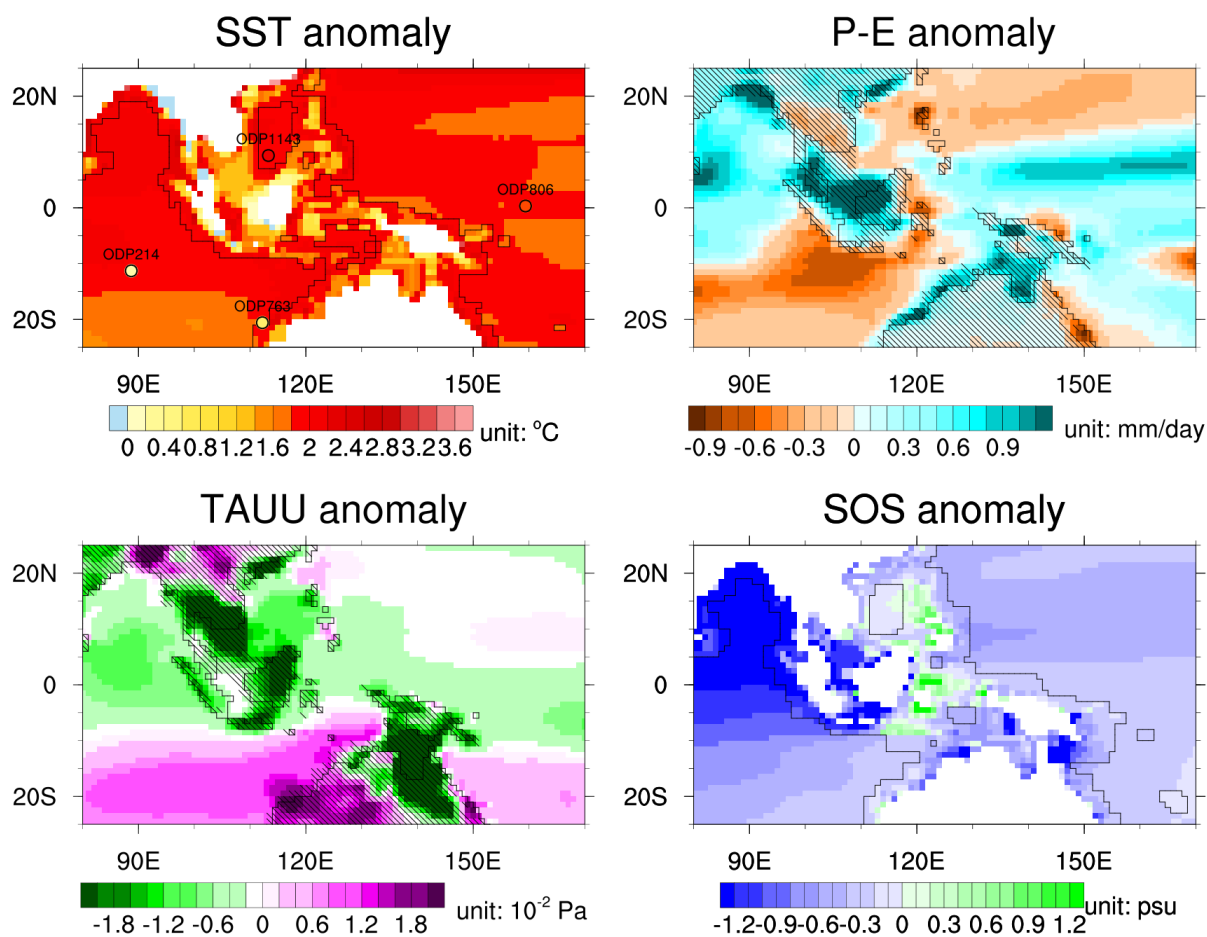
Figure 5. The MMM anomalies over the MC in the Eoi400 simulation relative to the E280 simulation for (a) sea surface temperature (SST), (b) precipitation minus evaporation (P-E), (c) zonal surface wind stress (TAUU) and (d) salinity at the ocean surface (SOS). In (a), circles mark the location of the SSTA proxies, their colour represents the reconstructed SSTAs, and the name of each site is indicated on the top of each circle. White shading and black lines in (a) and (d) are as in Fig. 2. In (b) and (c), the hatchings indicate land grid boxes in the mid-Pliocene experiment; the black lines indicate the continental outlines of the preindustrial experiment. In (c), green shading indicates westward anomaly and purple shading indicates eastward anomaly.

The MMM shows a regional averaged warm SSTA of 1.88°C over the MC, the spatial pattern of which is as shown in Fig. 5a: the Western Pacific Warm Pool and the southern sector of the Indian Ocean in the MC warm less compared with other regions. SSTAs warm most over the North Indian Ocean in the MC region. Oceans surrounding the land masses of the MC also show an apparent warming, but this is not a robust signal because the models' land sea masks differ in the region of the

255   MC. Models showing the greatest regional average warmings in the MC also show the largest global averaged warmings (Fig. S2). The linear regression coefficient between regional (the MC) and global annual SSTAs is ˜1.06, implying that the SSTAs over the MC change at a similar magnitude as global SSTAs. The SSTA spatial pattern over the MC from individual models are shown in Fig. S5.

As shown in Fig. 6, models consistently show warmer SSTs over the MC in the mid-Pliocene relative to the preindus-

260   trial; regional averaged SSTA range from 0.43°C produced by NorESM1-F to 3.33°C produced by HadGEM3. HadGEM3, EC-Earth3.3 and CESM2 all produce SSTAs exceeding 3°C, which represents the 3 largest warm anomalies in the region among the PlioMIP2 models; 3 of the lowest SSTAs, which are all less than 1°C, come from NorESM1-F, NorESM-L, and MRI-CGCM2.3, which are also the models with the lowest equilibrium climate sensitivity (ECS) among PlioMIP2 models (Haywood et al., 2020) as shown in Fig. 6. As there is a relationship between models' ECS and their global temperature

265   anomalies (Eoi400 minus E280) in the study of Haywood et al. (2020), models' ECS correlates with SSTAs on the MC, but the linearity is not exactly linearly. For example, GISS-E2-1-G shows an SSTA less than 1°C, but its ECS is relatively high. CCSM4-UoT and CCSM4-Utr have a modest ECS but show high global near-surface air temperature anomaly (Haywood et al., 2020) as well as a relatively high SSTA in mid-Pliocene over the MC.

The warmer SSTs not only influence the atmospheric circulation in the Pliocene but also bring higher evaporation due to

270   higher water vapour holding capacity of warmer air (the Clausius-Clapeyron equation) and more atmospheric moisture, which contributes to the change of simulated precipitation. As an essential part of the hydrological cycle, freshwater budget over land and ocean (precipitation minus evaporation) influences the oceanic salinity and shapes the climate of the MC. Fig. 5 illustrates the MMM anomalies of P-E, zonal wind stress (TAUU) and SOS over the MC in the mid-Pliocene relative to the preindustrial.

For the MMM P-E anomalies, if we take the MC as a whole, the freshwater budget (P-E) is more positive in the mid-

275   Pliocene over the MC relative to the preindustrial (Fig. 5b); 13 out of 17 models, and the MMM, consistently show an increase in P-E (Fig. S6). As shown in Fig. 5b, it can be seen that unlike the SSTs, P-E does not change consistently over the MC but varies throughout the MC. P-E increases over the Western Pacific Ocean and the North Eastern Indian Ocean, but decreases evidently over the west oceanic sector of the MC gateway. From the hatching and black lines in Fig. 5b, the difference in the distribution of land and sea in the Eoi400 and the E280 simulations is evident. P-E increases over most lands of the MC, and

280   also increases generally over grid cells which are ocean in the preindustrial but land in the mid-Pliocene. The spatial patterns of P-E anomalies from individual models are quite different (Fig. S6). In the case of some models, the simulated P-E anomaly is nearly opposite to each other, such as for COSMOS and EC-Earth3.3. Even the same family of models can show different results (e.g., CCSM-Utr and CCSM-UoT).

A large amount of precipitation releases a large amount of latent heat to the atmosphere, which fuels the atmospheric

285   circulation. In order to explore the changes of atmospheric circulation over the MC, we analyse the zonal surface wind stress
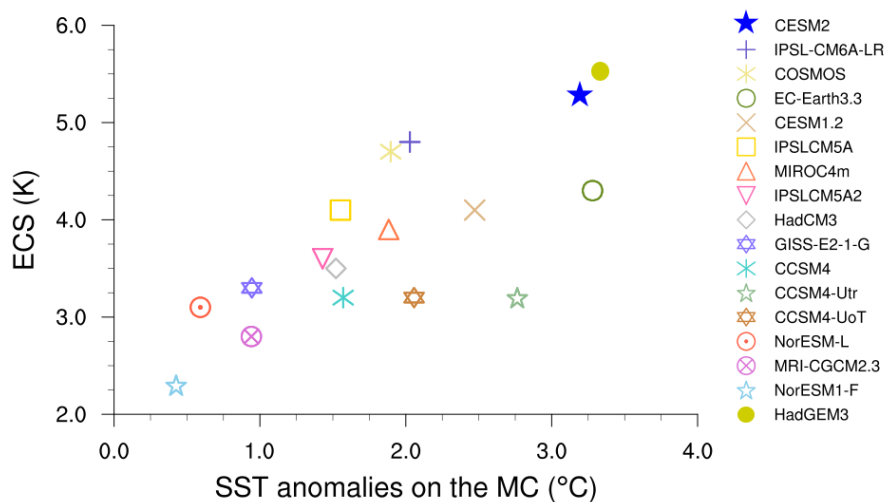
Figure 6. SST anomalies over the MC versus equilibrium climate sensitivity (ECS) from all the PlioMIP2 models.

(TAUU) anomalies, as shown in Fig. 5c. Green shading indicates a westward anomaly. There are westward anomalies over the North Western Pacific Ocean and eastward anomalies over the ocean to the west of Australia. This means that the tropical trade wind has been enhanced on the side of the North Western Pacific Warm Pool. This signal can be seen from half of available PlioMIP2 models (Fig. S7). There are extreme values over some regions which belonged to a land mass during the

290 mid-Pliocene but have turned into ocean in modern times. The extreme values found there might result from the change of land-sea distribution that also impacts on the prevalence of surface roughness. The latter impacts atmospheric flow more over land than over the ocean.

As a result of the changes in the hydrological system, SOS changes in the seas surrounding the MC. Fig. 5d illustrates the spatial distribution of SOS anomalies. We find a decrease of SOS in most of the ocean regions belonging to the MC. This pattern is especially pronounced in the northern part of the Indian Ocean, where one can find SOS to decrease by more than 1.2 PSU. For the Pacific Ocean sector and the southern Indian Ocean sector of the MC, SOS decreases by less than 1 PSU. Overall, we find that seawater becomes fresher at the surface around the MC in the mid-Pliocene in comparison to the preindustrial. This finding is consistent in 14 out of 16 PlioMIP2 models and in the MMM. Spatial patterns of SOS anomalies for individual models are shown for reference in Fig. S8.

3.3 Ocean Currents

Temperature and density differences are drivers of the ITF, which is an important process for regulating the climate of the MC. As the seawater characteristics of Pacific Ocean and Indian Ocean are so different, oceanic flow through the MC does not only transport volume, but also creates a net-transport of salinity and heat between these two oceans. This process can directly influence both local and remote climate. Therefore, we quantify and investigate the changes of oceanic flow through the MC in the mid-Pliocene. In this section, we address the following questions: Is there more fresh and warm water injected into the Indian Ocean from the Pacific Ocean in the mid-Pliocene relative to the preindustrial? Where do the anomalies originate from? Do the anomalies result from the surface ocean, or rather from processes and characteristics that are shaped in deeper layers of the sea?
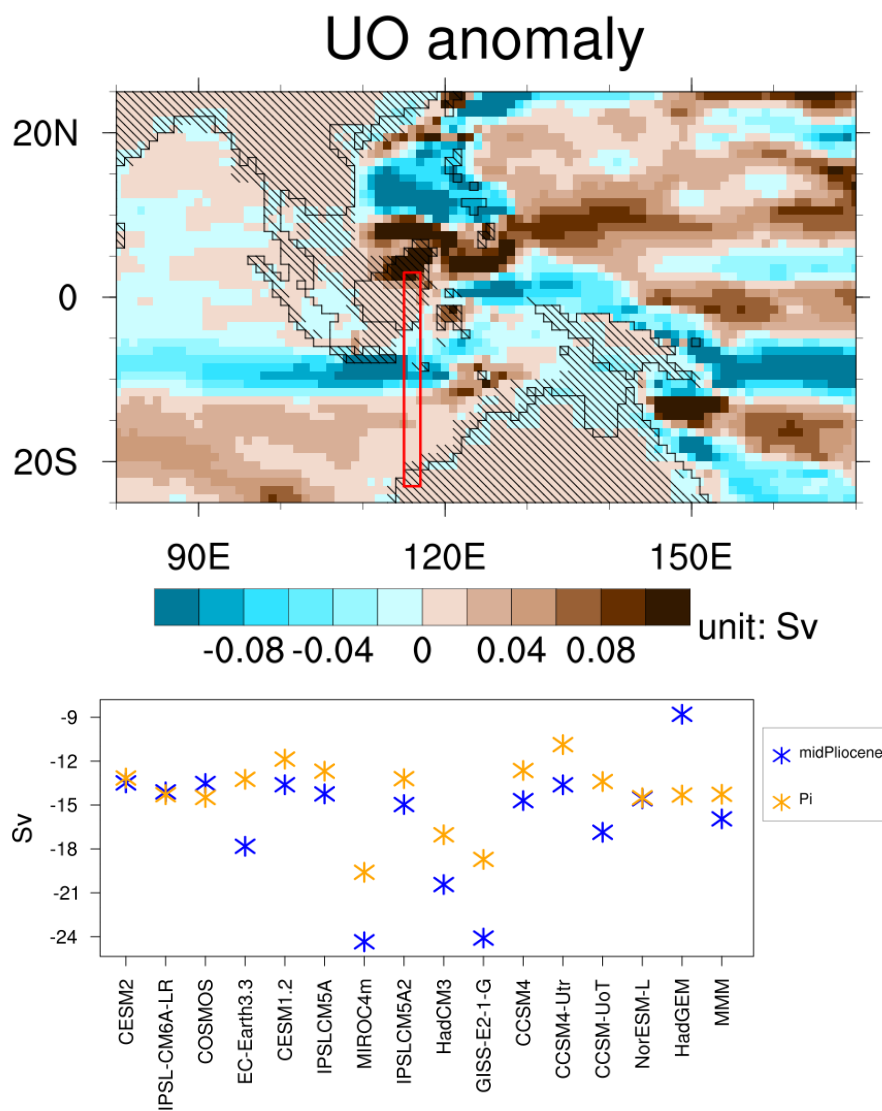
Figure 7. (a) The MMM zonal sea water velocity (UO) anomaly in the Eoi400 experiment relative to the E280 experiment. Blue shading indicates westward anomaly and brown shading indicates eastward anomaly. The red rectangle is the gateway defined to calculate the ocean volume transport intensity in (b). The hatchings indicate land grid boxes in the mid-Pliocene experiment; the black lines indicate the continental outlines of the preindustrial experiment. (b) The integrated ocean volume transport intensity through the gateway. Negative values indicate westward transport.

We study the oceanic flow with the available data. During the calculation of ocean currents, we find that there are some
310 disagreements between results derived from regridded data and analyses performed on individual models' native ocean model
grids. Therefore, to reduce errors and improve the accuracy, we use native grid data to calculate the oceanic flow intensity.
Fig. 1 shows the gateways (red rectangle), and differences in the land sea distribution between the E280 and the Eoi400
simulations. In the modelled mid-Pliocene, some regions, which are islands in the preindustrial, are still a peninsula that is
connected to the Asian mainland in some models (CESM2, EC-Earth3.3, CESM1.2, MIROC4m, HadCM3, GISS-E2-1-G,
315 CCSM4, CCSM4-Utr and CCSM-UoT). In these models except for HadGEM3 which adopted the same land-sea mask of the
MC in both experiments, the Malacca Strait, and some other small straits in the northern part of the MC, are closed as shown
in Fig. 1. Except for this gateway, the strait between New Guinea and Australia is also closed. Therefore, the Sunda and Sahul
shelves caused some straits of the modern MC, that are present in the E280 experiment, to be absent in the Eoi400 experiment.
However, the volume transport through the Malacca Strait is very small compared with the volume transport through the main
320 ITF passage. Therefore, we only take the main part of throughflow, that occurs through the gateway showed in Fig. 1, into
account.

The spatial pattern of the vertical integrated zonal velocity (UO) anomaly is shown in Fig. 7a. There is a westward UO
anomaly along the northeastern coast of the New Guinea, which favours the ocean mass transport of the ITF. For the oceanic
velocity through the gateway, the westward water transport has been strengthened over the north, but weakened over the south.
325 Models and modern observation both show a strong flow through the gateway between MC and Australia. According to the
observation between 2004 and 2006 (Sprintall et al., 2009), the ocean volume transport via the Indonesian Throughflow to the
Indian Ocean is 15 Sv. By integrating the water volume transport through this gateway vertically, we derive the total seawater
transport volume through the MC, which is shown in Fig. 7b. Based on the model results, the flow through the MC ranges
from 10.9 Sv (CCSM4-Utr) to 19.6 Sv (MIROC4m) westward in the preindustrial and from 8.8 Sv (HadGEM3) to 24.3 Sv
330 (MIROC4m) westward in the mid-Pliocene. The result of the E280 experiment from every model shows the same transport
direction and a magnitude comparable to observations. Compared to the preindustrial, the strength of the flow shows a clear
increase in 10 out of 15 models, ranging from 12.2% (IPSLCM5A) to 34.6% (EC-Earth3.3). Exceptions are three very small
changes simulated by CESM2 (2.5%), NorESM-L (0.5%) and IPSL-CM6A-LR (-1.3%), and two reductions that are observed
in the simulation by COSMOS (-6.6%) and HadGEM3 (-38.5%). As mentioned before, the gateway geometry in HadGEM3 is
335 the same in both E280 and Eoi400 experiments, which may lead to a result different from those of the other models. Therefore,
all but three of the PlioMIP2 models show agreement on an increase in ocean volume transport through the MC in the mid-
Pliocene in comparison to the preindustrial. The volume transport of 10 out of 15 models is appreciably increased (NorESM1-F
and MRI-CGCM2.3 are absent). Three of 15 models show very small changes in volume transport. There are only three models
(IPSL-CM6A-LR, COSMOS and HadGEM3) that disagrees with the rest of the ensemble on the sign of transport change, and
340 only two of these predict that volume transport through the gateway changes in an appreciable manner.
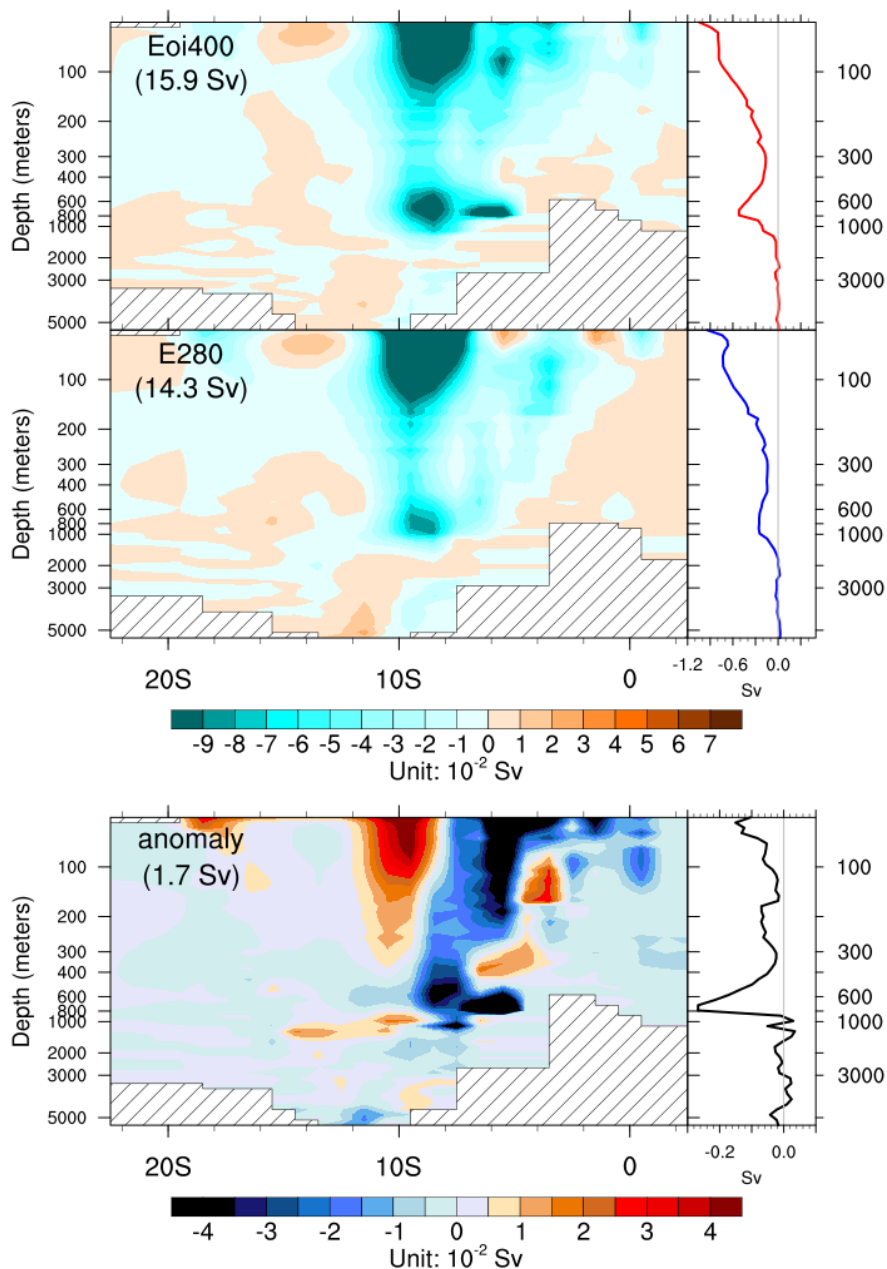
Figure 8. The MMM profiles of the ocean current through the MC in the mid-Pliocene (top panel), in the preindustrial (middle panel) and the anomalies of the mid-Pliocene relative to preindustrial (bottom panel). The graphs attached on the right show the horizontally averaged ocean volume transport. Values in brackets indicate the total water volume transport through the gateway. Warm shading indicates direction from west to east. Cold shading indicates direction from east to west. The first colour bar is for absolute values in top panel and middle panel. The colour bar on the bottom is for anomaly values in the bottom panel.

In order to explore which depth of the ocean determines the changes of throughflow, we investigate the vertical structure of water volume transport through the gateway, the MMM of which is shown in Fig. 8. In the MMM result, the strongest westward water transport is via the current located around 10°S. In general, the directions of water transport in the mid-Pliocene stay the same as for the preindustrial. This encompasses the westward transport through the main ITF passage as well as the relatively

345   weak eastward transport through the south branch of the gateway simulated by some models (slightly yellow shading in Fig. 8). The strong westward flow counterbalances the weak eastward flow, and there is net-transport of fresh and warm water from the Pacific Ocean to the Indian Ocean. In both experiments, water mass is mainly transported in this gateway in the upper ocean above 1000 meters depth, which is consistent from the results of the MMM and individual models (Fig. S9).

The profiles of the throughflow anomalies are shown in the bottom panel of Fig. 8. From the difference between Eoi400

350   experiment and E280 experiment in Fig. 8, it can be seen that the meridional integrated ITF changes largely above 1000 meters depth, especially strongly at the surface and depth of around 800 meters, without water transport direction reverse. In terms of the structure of the ITF anomaly, there is a negative anomaly (stronger flow) in the northern part of the ITF and a positive anomaly (weaker flow) in the flow between 12°S and 8°S. This finding suggests that the ITF developed closer to the northern boundary of the gateway in the mid-Pliocene than it is located during the preindustrial at the layers above 1000 meters.

355   ### 3.4 Clustering

So far we have studied the climatic features of the MC by using the multimodel ensemble method, which has often been employed when analysing results from multiple models. However, we have shown that, although models show agreement on warming and wetter climate over the MC in the mid-Pliocene, the spatial patterns of climate anomalies vary between models (e.g. Fig. S5). These differences might counterbalance each other to some extent, and therefore some important climate signals

360   from individual models will be eliminated when using the MMM. Moreover, some of the employed models belong to a group of general circulation models that share model components or may be similar in other respects. Consequently, similar signals from the same model family may be amplified in the ensemble of multimodel results. In order to investigate differences between models and show the signals which might disappear in the MMM, we use clustering analyses in this section. Thereafter, we introduce a new metric, the multi cluster mean (MCM).

365   In the beginning of this section, we demonstrate the method of classifying models into groups based on the spatial features of SSTAs and precipitation anomalies over the MC from individual models by applying pattern correlation and the hierarchical clustering method, which are derived for example from Sierra et al. (2017) and Knutti et al. (2013a). After averaging the results from multiple clusters, we obtain the MCM results. Then we illustrate the spatial anomaly pattern of every group and of the MCM, and compare them with the MMM. After that, we investigate the performance of cluster groups and of the

370   MCM at the example of the mid-Pliocene and the preindustrial simulation, and investigate if these show less discrepancy with observation/reconstruction than the MMM.

Fig. 9a and Fig. 10a illustrate the procedures of clustering. The first step is to find the two models with the smallest distance and join them into one group. On each subsequent step, the two closest groups or models are merged into a larger group. The distance between two models is the pattern correlation between the multiannual mean SSTAs or precipitation anomalies of
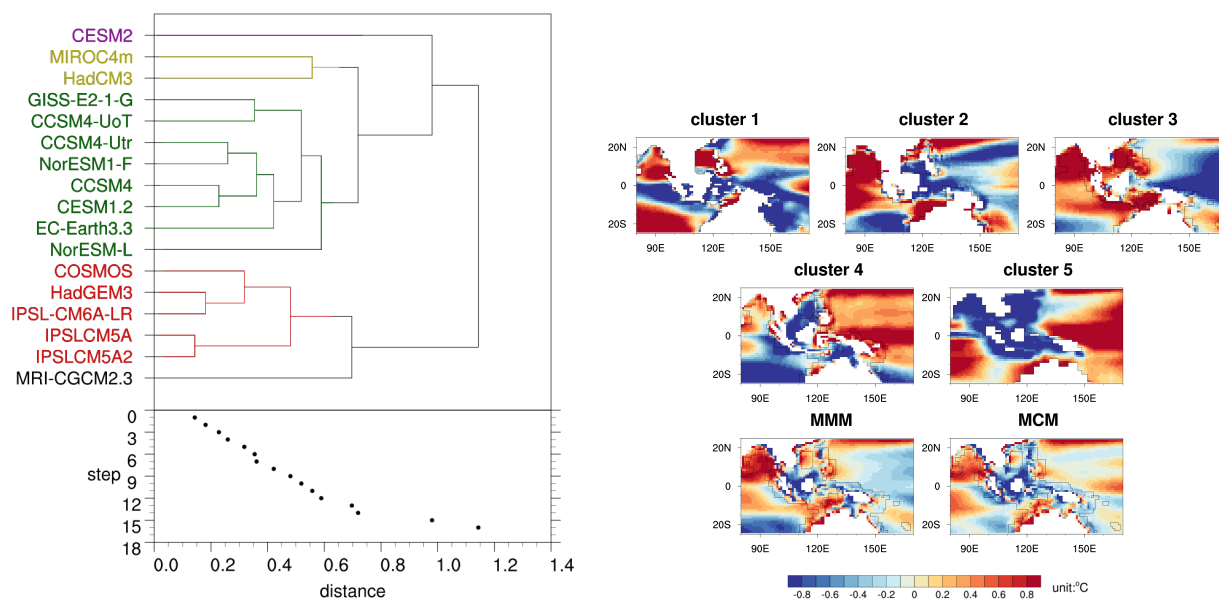
Figure 9. (a) Dendrogram of the SSTA clustering analysis based on the regional SSTAs over the MC from 17 PlioMIP2 models. Below the dendrogram the clustering step is shown as a function of distance between groups. (b) The SSTA patterns of the composite for different clusters, multi-model mean (MMM) and the multi cluster mean (MCM) after removing the regional mean SSTA. Cluster 1: CESM2; Cluster 2: MIROC4m and HadCM3; Cluster 3: GISS-E2-1-G, CCSM4-UoT, CCSM4-Utr, NorESM1-F, CCSM4, CESM1.2, EC-Earth3.3 and NorESM-L; Cluster 4: COSMOS, HadGEM3, IPSL-CM6A-LR, IPSLCM5A and IPSLCM5A2; Cluster 5:MRI-CGCM2.3. White shading indicates the grid boxes that are land for all the models of the cluster. The grid boxes outside the black lines are ocean grid boxes for all the models of the cluster.

375   these two models. The x-axes of the Fig. 9a and the Fig. 10a indicate the distance of every two joined groups at every step. The pattern correlation measures the similarity of the spatial patterns between two models (Taylor, 2001), which is the Pearson Correlation Coefficient of the same variable of two different maps from two models. Here, for SSTAs we only use grid boxes which are ocean in all the models. Here we choose average-linkage method (Wilks, 2011) to merge groups, which means the distances between every group is the averaged distance between individual members of the two groups.

380   In the bottom panel of Fig. 9a, there is a large jump after clustering steps 12, suggesting an appropriate stopping point with 5 groups, the group members of which are shown in Fig. 9a. CESM2 shows an SSTA spatial pattern that different from the other PlioMIP2 models, that it forms cluster 1 on its own. The same is true for MRI-CGCM2.3, but this model could be merged into another cluster if we moved one step forward. Cluster 2 contains MIROC4m and HadCM3. Cluster 3 contains the largest number of models, with 8 members, including almost all the CCSM series model (except for CESM2), NorESM series

385   model, GISS-E2-1-G and EC-Earth3.3. IPSLCM5A2 is an updated version of IPSLCM5A in terms of updated components and technical characteristics (Sepulchre et al., 2020). This is clearly shown in the clustering results, with IPSLCM5A and IP-SLCM5A2 being the two closest models. All the 3 IPSLCM models have been classified into cluster 4 together with COSMOS and HadGEM3. Fig. 9b illustrates the ensemble SSTAs pattern of every cluster after removing the regional mean SSTA. Cluster 4 and cluster 5 show a relatively warmer pattern on the WPWP, which is opposite to results present in cluster 3. For the Indian

390   Ocean, cluster 4 shows a relative weak warm pattern, which means the Indian Ocean warms less than the region, especially the WPWP. Spatial distributions of SSTA in the Indian Ocean are similar for clusters 3 and 2, warmer in the north but less warm in the south, which is opposite to results from cluster 5. Since models are equally weighted in the ensemble, and cluster 3 contains the largest number of PlioMIP2 models, it dominates the MMM SSTA pattern (Fig. 9b). The MCM is the mean of these 5 clusters - in other words, it is the multimodel mean with weights. The MCM can avoid signals being overweighted

395   from the same model family and demonstrate signals which disappear in the MMM. For example, the centre of the WPWP shows a warmer anomaly, a result consistent for all clusters except for cluster 3. Comparison between warming of the WPWP and the eastern Indian Ocean on different clusters indicates 3 possible SSTA patterns: the eastern Indian Ocean warms more than the WPWP (cluster 2, cluster 3 and the MMM); the eastern Indian Ocean warms less than the WPWP (cluster 4); there is no apparent difference in warming between these two regions (cluster 1, cluster 5 and the MCM).
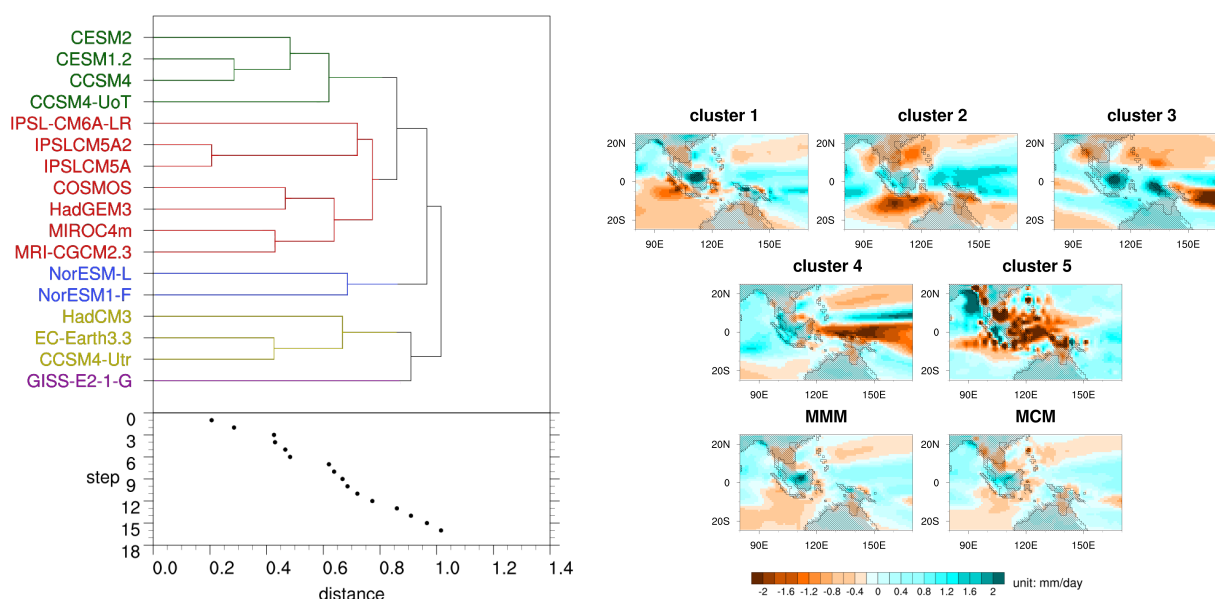
Figure 10. The same as Fig. 9 but for precipitation. Cluster 1: CESM2, CESM1.2, CCSM4 and CCSM4-UoT ; Cluster 2: IPSL-CM6A-LR, IPSLCM5A2, IPSLCM5A, COSMOS, HadGEM3, MIROC4m and MRI-CGCM2.3 ; Cluster 3: NorESM-L and NorESM1-F ; Cluster 4: HadCM3, EC-Earth3.3 and CCSM4-Utr ; Cluster 5: GISS-E2-1-G. The hatchings and the black lines are the same as Fig. 5.
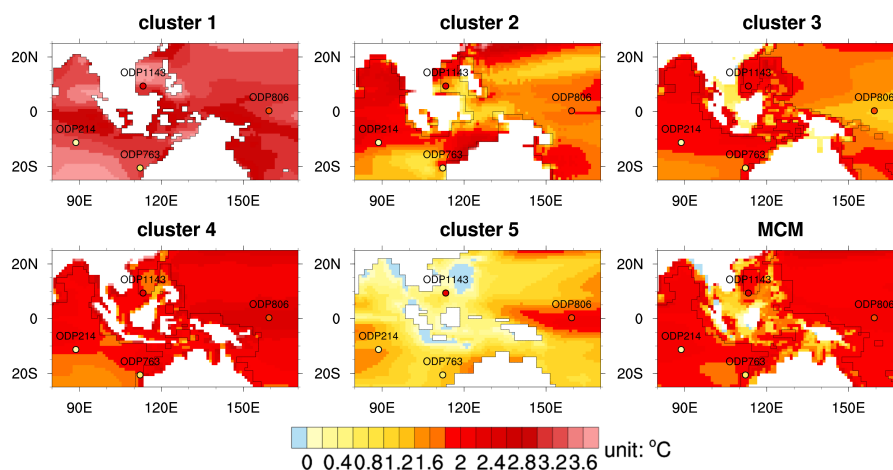
Figure 11. SSTAs over the MC in the mid-Pliocene (Eoi400) relative to the preindustrial (E280) from the SSTA-based clusters and the multi cluster mean (MCM). Cluster 1: CESM2; Cluster 2: MIROC4m and HadCM3; Cluster 3: GISS-E2-1-G, CCSM4-UoT, CCSM4-Utr, NorESM1-F, CCSM4, CESM1.2, EC-Earth3.3 and NorESM-L; Cluster 4: COSMOS, HadGEM3, IPSL-CM6A-LR, IPSLCM5A and IPSLCM5A2; Cluster 5:MRI-CGCM2.3. Units: °C. The MMM SSTAs is shown in Fig. 5a. Note that in this figure, regional averaged SSTA has not been removed as in Fig. 9b. White shading and black lines are the same as in Fig. 9b

400   In addition to the cluster based on SSTAs, we also sort models based on the precipitation anomaly spatial patterns, as shown in Fig. 10. Similar to SSTA analyses, we find a jump after step 12. Therefore, we classify the models into 5 groups (Fig. 10a). Four of the CCSM series models, CESM2, CESM1.2, CCSM4 and CCSM-UoT, form cluster 1. The IPSLCM-type models are members of the cluster 2 together with COSMOS, HadGEM3, MIROC4m and MRI-CGCM2.3. This cluster is the biggest one with 7 models. Similar to clustering based on SSTA, IPSLCM5A and IPSLCM5A2 are still the two most similar models; they are also the closest group for IPSL-CM6A-LR. NorESM-L and NorESM1-F form cluster 3. The remaining model in the CCSM
405   series, CCSM4-Utr, joins EC-Earth3.3 and HadCM3 as cluster 4. For precipitation, cluster 5 (only GISS-E2-1-G) itself show a decrease in precipitation over the centre of the MC. For the WPWP, cluster 1 and cluster 2 show relatively wetter anomalies; cluster 3 and cluster 4 show basically drier anomalies with spatial heterogeneity. Cluster 5, which only contains GISS-E2-1-G, shows a relative drier climate in the northwest but wetter condition in the rest of the WPWP. All clusters except cluster 5 show
410   a relative dry anomaly over the north of the WPWP. Similar to SSTA-based clustering, the cluster with the largest number of models dominates the spatial pattern of the MMM and leads some signals of other clusters to disappear. Overall, compared with the eastern Indian Ocean, the WPWP is relatively wetter (cluster 1, cluster 2, the MMM and the MCM), drier (cluster 3 and cluster 4), or receives similar amount of precipitation (cluster 5). Compared Fig. 9b with Fig. 10b, the pattern of the MCM is more similar to the pattern of the MMM in precipitation than in the SSTA.

415   Fig. 11 compares the 5 clusters with the reconstructed SSTA data. In order to quantify model clusters' performance in simulating the mid-Pliocene SSTAs, the SSTA-based clusters are validated in the same way as individual models and as the MMM, as shown in Fig. 3 and Fig. 4. The performance of the MCM is similar to the MMM for every site. However, in contrast to the MMM, the MCM can take the model weight into account to avoid duplication of model signals from the same family, and also retain the climate signals we might lose in the multimodel ensemble. An example is the warm anomaly signal in the
420   centre of the Western Pacific Ocean.

4   Discussion

4.1   Physical Mechanisms

The MC is a region of the atmospheric-oceanic coupling and exchange, slight changes here will affect climate afar (Gordon, 2005; Yoneyama and Zhang, 2020). Due to the different land distribution, topography, and CO_2 concentration in the mid-
425   Pliocene compared to today, we find that characteristics of mid-Pliocene climate of the MC, such as warmer SSTs, higher freshwater budget, lower SOS, strengthened surface zonal wind in WPWP and increased volume transport through the MC differ in comparison to the preindustrial.

     In terms of atmospheric circulation, PlioMIP1 suggests weakening of the Walker Circulation in the mid-Pliocene (Corvec and Fletcher, 2017); in PlioMIP2 this circulation change proved to be highly model dependent (Han et al., 2021). However, a
430   robust feature across PlioMIP2 models is the westward shift of the Pacific Walker Circulation (Han et al., 2021). Moreover, the Northern Hemisphere warms more than the Southern Hemisphere in the mid-Pliocene (Haywood et al., 2020). As a result of the variation of interhemispheric meridional heat fluxes and energetic constraints for the Intertropical Convergence Zone (ITCZ) position, the Southern Hemisphere Subtropical High pressure system is intensified, and the western Pacific easterly trade is enhanced correspondingly (Pontes et al., 2021), which is consistent with the strengthened zonal wind stress we find in
435   Fig. 5c.

     Warmer SST in the mid-Pliocene results in higher atmospheric moisture content (Allen and Ingram, 2002). The warm West Pacific provides substantial water vapour that contributes to increased precipitation over the MC as part of the "wet-gets-wetter, dry-gets-drier" response, the spatial precipitation anomaly spatial pattern forms as shown in Fig. 5b (Feng et al., 2022).

     In the mid-Pliocene, the subaerial Sunda and Sahul shelves inhibit volume transport between western Pacific Ocean and
440   Indian Ocean to some extent. However, the strengthened westward zonal wind stress over the Western Tropical Pacific Ocean changes sea level and thermocline via Kelvin and Rossby waves (Wang, 2019; Wijffels and Meyers, 2004) and is therefore favourable for water transport through the MC (Wainwright et al., 2008). This is consistent with the intensified ITF found in our study (Fig. 7b). In addition, over the northwest coast of Australia, the modelled wind stress shows an eastward anomaly (Fig. 5c). From the MMM zonal ocean velocity anomaly, there is also an eastward anomaly on the southern part of the ITF (Fig.
445   7a), maybe as a result from the zonal wind stress (Fig. 5c). Based on observations, the variations of trade wind in the Pacific Ocean can explain changes of the ITF for the past decades (Feng et al., 2018). However, under global warming, the weakening of the ITF is more associated with changes of the overturning circulation in the Pacific basin: the zonal volume convergence

via southern Australia decreases. Then upwelling increases to compensate for upper layer water, but transport into the deep
Pacific basin at the latitude of the southern tip of Australia will decrease, ultimately reducing the ITF (Fig. 10 in Sen Gupta
450   et al. (2016)). The mechanisms of the changes of ITF in the mid-Pliocene remains a question to be studied in future work.

Precipitation minus evaporation is the largest freshwater component entering the ocean (Byrne et al., 2016). Increased fresh
water flux provides a way to explain the general decrease of the SOS over the MC, especially the decreased SOS in the Northern
Indian Ocean. The changes of topography also play an important role. The closure of the gateway between New Guinea and
Australia can change the water source from the saline south Pacific to the less saline north Pacific (Cane and Molnar, 2001),
455   which can cause the Northern Indian Ocean SOS to decrease significantly. Furthermore, the higher precipitation and wetter
climate over the Sunda shelves, India, south slope of the Tibetan Plateau and the surrounding regions (as shown in Fig. 5b of
Haywood et al. (2020) and Fig. 1b of Feng et al. (2022)) may increase the local run-off, which can also inject more fresh water
into the ocean and then dilute salinity.

Large scale flow patterns on the other hand seem to be simulated unambiguously across the PlioMIP2 model ensemble. In
460   particular, models in general agree on increased westward ocean volume transport through the MC during the mid-Pliocene.

4.2   Summary of the PlioMIP2 Model Performance

While MC SST are pronouncedly warmer in the mid-Pliocene, it seems that within the PlioMIP2 model ensemble MC climate
is to some extent decoupled from model ECS (Fig. 6). On the other hand, we find that models with a large ECS also produce
rather warm mid-Pliocene conditions around the MC (Fig. 6). When using models to predict climate of the MC for different
465   levels of carbon dioxide or changes in earth surface condition, it is important to note that the models that best reproduce a
recent regional climate of the MC as quantified by a comparison to modern observations are not necessarily the same models
as those that produce the best agreement between simulated and reconstructed mid-Pliocene SSTAs (Fig. 4). But at a global
scale, there is a correlation between model performance on simulating mid-Pliocene and modern climate (Fig. S4). From that
point of view, the mid-Pliocene proves to be a real test bed for the skill of models for climate states that strongly differ from
470   the recent one.

Largely, we find agreement between proxy reconstructions of SSTA on the one hand and mid-Pliocene climate simulations
on the other hand. At ODP Site 214 some of the models fall out of the uncertainty range of the reconstruction at the warm end.
Yet, we note that ODP214 indeed appears somewhat cool in comparison to nearby SSTA reconstructions in the region. This
leaves us with the possibility that the models are not able to capture climate mechanisms that shaped the mid-Pliocene climate
475   recorded at site ODP214. An alternative explanation could be that the discord at this site is to be sought in the proxy data
itself or in its interpretation. We note that the model-data comparison is hampered by large uncertainty ranges of reconstructed
SSTAs and limited availability of suitable SSTA reconstructions in the region. This highlights that more studies are necessary
that increase availability of MC SSTAs during the mid-Pliocene. On the other hand, we find that the simulated climate in
PlioMIP2 has much smaller uncertainty ranges than the reconstructed climate. This is seen from the fact that the spread of
480   simulated SSTAs across the PlioMIP2 model ensemble is much smaller than the uncertainty bars attributed to reconstructions

(Fig. 3a). This hints that the proxy reconstructions at the MC are shaped by aspects of regional climate variability that the models fail to capture.

4.3 Model Hierarchy and Clustering

The multimodel ensemble mean (MMM) has frequently been adopted to illustrate models' results. We found the MMM pro-
485 duces less discrepancy than more than half of the PlioMIP2 individual models in both mid-Pliocene climate simulation and preindustrial climate simulation (Fig. 4). In the case of model independence, the model biases will be partially cancelled, which results in the MMM outperforming individual models (Solomon et al., 2007; Knutti et al., 2010). But some models, such as models adopted in Coupled Model Intercomparison Projects phase 6 (CMIP6) are not all independent, they might have similar biases (Abramowitz et al., 2019). Boé (2018) quantified the relationship between CMIP5 model components replication and
490 proximity of their results, and found there is a clear relationship at both global and regional scales. Model clustering is a way to study proximity of models results' spatial patterns and to weight models to reduce reduplication of similar biases from the same model family.

In PlioMIP2, not all the models are independent, there are models that developed from the same source, with shared codes or using different generations of the same components as shown in Table 1. A study from Ho et al. (2016) adopted outputs
495 from 41 CMIP5 models and clustered models according to the concept of "model genealogy" Knutti et al. (2013b), and they found difference between the weighted and unweighted multi-model ensemble mean, but, the general conclusions from both methods did not change. In this study, we cluster models into groups according to the proximity of their results in simulating SSTAs (section 3.4). Although some models used the same ocean component, they turned out to be in different SSTA-based cluster groups, such as CESM2. Descendants of CCSM4, CCSM4-Utr, CCSM4-UoT, CESM1.2, CESM2, NorESM-L and
500 NorESM1-F, are all in cluster 3, except for CESM2 (Fig. 9a). Compared to the other CCSM4 models, CESM2 has updated the POP ocean component not only with respect to the parameterization but also uses new schemes (Danabasoglu et al., 2020). Moreover, CCSM4, CESM1.2 and CESM2 have very different atmospheric components, in the study of Pacific Walker Circulation, CCSM4 showed strengthened Walker Circulation, while both CESM1.2 and CESM2 showed weakened Walker Circulation (Feng et al., 2020b). Nearly all the models that include the NEMO ocean component, except for EC-Earth3.3,
505 these are EC-Earth3.3, IPSLCM5A, IPSLCM5A2, IPSL-CM6A-LR and HadGEM3, appear in the SSTA-based cluster 4 (Fig. 9a). In the analysis of El Niño variability from PlioMIP2 (Oldeman et al., 2021), the 'La Niña-like' mean state model group and the 'El Niño-like' mean state group are not the same as the SSTA-based clusters found in our study. This implies that clustering depends on the region and on the studied physical quantity. In terms of atmospheric component, some models used the same atmosphere component but haven't been clustered into the same group based on precipitation. For example, CCSM4,
510 CCSM4-UoT, CESM1.2 and CESM2 are all in cluster 1, but CCSM4-Utr is not in this group. The two closest models in both SSTA-based and precipitation-based clusters are IPSLCM5A and IPSLCM5A2. They adopted the same atmospheric compo-
nent and oceanic component, as well as vertical and horizontal resolutions. Compared with the IPSLCM5A, IPSLCM5A2 updated the versions of components and retuned the cloud radiative forcing (Tan et al., 2020), but both models still produce a similar climate signal over the MC. From the discussion above, we conclude that the model configuration can provide to us a

515 reference to cluster models, but even models of the same model family may still produce different climatic signals depending
on the analysis region or the studied climate characteristic. These results suggest that there is not always a clear trace of model
family apparent via the clustering.

Moreover, the multimodel mean will lead to the loss of climatic signal (Knutti et al., 2010). In our study, there are noticeable
variations between spatial anomalies patterns from individual models. By illustrating models' results in cluster groups, climatic
520 signal loss can be reduced, for example the warm anomaly signal in the centre of the Western Pacific Ocean as shown in Fig. 9b.

5   Conclusions

The subaerial Sunda and Sahul shelves and relative high atmospheric CO_2 concentration, combined with other forcings, lead to
a different climate of the MC in the mid-Pliocene than in the preindustrial. In line with the global climate, all the 17 PlioMIP2
525 models (including HadGEM3) show higher SSTs in the mid-Pliocene MC, ranging from +0.43°C to +3.33°C, with a MMM
warming by 1.88°C; 13 out of 17 models show fresh water flux (precipitation minus evaporation) increased in the MC, ranging
from 0.04 mm/day to 0.50 mm/day. In terms of atmospheric circulation, the easterlies over the western tropical Pacific Ocean
show enhanced intensity in the MMM; there is anomalous eastward wind stress over the ocean near the Western Australia.
Seawater salinity affects the thermohaline circulation; 14 out of 16 models (MRI-CGCM2.3 is absent here) show a lower SOS
530 in the MC .

Even though the topography of the MC in the mid-Pliocene acts as a barrier for volume transport between the Pacific Ocean
and the Indian Ocean compared to the preindustrial, 10 out of 15 models show the ITF intensity increasing by 12.2% to 34.6%.
Three models do not show apparent changes in intensity. Only HadGEM3 and COSMOS show a decrease in intensity of the
ITF, the amplitude of westward flow is here lowered by 38.5% and 6.6%, respectively. In terms of spatial difference, the MMM
535 shows that the ITF is stronger in the north but weaker in the south in the mid-Pliocene relative to the preindustrial. Moreover,
the water mass is mainly transported through the ITF in the ocean above 1000 meters depth in both the preindustrial and the
mid-Pliocene. The changes in the volume transport of ITF also mainly occur in the ocean above 1000 meters depth.

The models' performance has also been quantified in this work by comparing the discrepancies between model results and
reanalysis data on the one hand and reconstructed data on the other hand. The results imply that models, which reproduce mod-
540 ern climate well, are not always good at simulating the mid-Pliocene climate of the MC according to the proxy reconstructions.
But at the global scale, there is a correlation between models' performance on simulating the mid-Pliocene climate and modern
climate (see Fig. S5 in appendix).

The comparison between individual models and the MMM suggests that the MMM could reproduce the preindustrial SST
of the reanalysis better than 13 out of 17 models, and produce less discrepancy with reconstructed SSTAs than 12 out of 17
545 models.

There is a big difference in the spatial distribution of anomalies between models. Because similar models will amplify similar
signals, and because some climatic signals from individual models will get lost in the MMM, a new metric MCM has been

introduced. It takes model families weight into account, provides a new perspective on ensemble multiple models' results and demonstrates different climate states with clusters. The evaluations of the MCM show that the MCM also outperforms 12 out of 17 models in simulating preindustrial climate and shows less or equal discrepancy with reconstructed mid-Pliocene SSTAs than 10 out of 17 models, which suggests that the MCM is a plausible method to show multimodel results.

Code and data availability. The PlioMIP2 model data used in this work is available from the PlioMIP2 database upon request from Alan M. Haywood (a.m.haywoodß@leeds.ac.uk). PlioMIP2 data from CESM2, EC- Earth3.3, NorESM1-F, IPSLCM6A, GISS2.1G and HadGEM3 can be obtained from the Earth System Grid Federation (https://esgf-node.llnl.gov/search/cmip6/). The boundary conditions in PlioMIP2 experimental design can be downloaded from https://geology.er.usgs.gov/egpsc/prism/7.2_pliomip2_data.html. Most of the analyses in this paper are processed with NCAR Command Language (NCL), some scripts used have been shared in https://github.com/XinRenn/Xin_MM-cluster.

Author contributions. XR carried out the data process, wrote the majority of the manuscript, and led the paper. DJL and EH provided assistance during the whole work. DJL and CS contribute to some writing. AH, AAO, BOB, CJRW, CS, CG, DC, GL, JCT, LES, MAC, MK, MLJB, NT, QZ, RF, WLC, WRP, XL, YK, ZZ and AMH conducted the simulations and provided data. All authors provided comments and contributed to this paper.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This research is funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813360 4DREEF and the NERC SWEET grant (grant no. NE/P01903X/1). Wing-Le Chan and Ayako Abe-Ouchi acknowledge funding from JSPS (KAKENHI grant no. 17H06104 and MEXT KAKENHI grant no. 17H06323) and computational resources from the Earth Simulator at JAMSTEC, Yokohama, Japan. Development of GISS-E2.1 was supported by the NASA Modeling, Analysis, and Prediction (MAP) Program. CMIP6 simulations with GISS-E2.1 were made possible by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center. Charles J. R. Williams acknowledges the financial support of the UK Natural Environment Research Council (NERC)-funded SWEET project (research grant no. NE/P01903X/1), as well as funding from the European Research Council (under the European Union's Seventh Framework Programme (FP/2007-868 2013) (ERC grant agreement no. 340923 (TGRES)). GL and CS acknowledge institutional funding at AWI via the research program PACES-II of the Helmholtz Association. CS acknowledges funding via the Helmholtz Climate Initiative REKLIM. MB and AvdH acknowledge support by the program of the Netherlands Earth System Science Centre (NESSC), financially supported by the Ministry of Education, Culture and Science (OCW). Xiangyu Li acknowledge funding from National Natural Science Foundation of China (42005042 and 42275047).

References

Abramowitz, G., Herger, N., Gutmann, E., Hammerling, D., Knutti, R., Leduc, M., Lorenz, R., Pincus, R., and Schmidt, G. A.: ESD Reviews: Model dependence in multi-model climate ensembles: Weighting, sub-selection and out-of-sample testing, Earth System Dynamics, 10, 91–105, https://doi.org/10.5194/ESD-10-91-2019, 2019.

580 Adler, R. F., Gu, G., Sapiano, M., Wang, J.-J., and Huffman, G. J.: Global Precipitation: Means, Variations and Trends During the Satellite Era (1979–2014), Surveys in Geophysics, 38, 679–699, https://doi.org/10.1007/s10712-017-9416-4, 2017.

Allen, M. R. and Ingram, W. J.: Constraints on future changes in climate and the hydrologic cycle, Nature, 419, 224–232, 2002.

Arias, P., Bellouin, N., Coppola, E., Jones, R., Krinner, G., Marotzke, J., Naik, V., Palmer, M., Plattner, G.-K., Rogelj, J., Rojas, M., Sillmann, J., Storelvmo, T., Thorne, P., Trewin, B., Rao, K. A., Adhikary, B., Allan, R., Armour, K., Bala, G., Barimalala, R., Berge, S., and Zickfeld,

585 K.: Technical Summary., in: Climate Change 2021: The Physical Science Basis.Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., p. 3949, Cambridge University Press, https://doi.org/10.1017/9781009157896.002, 2021.

Auer, G., De Vleeschouwer, D., Smith, R. A., Bogus, K., Groeneveld, J., Grunert, P., Castañeda, I. S., Petrick, B., Christensen, B., Fulthorpe,

590 C., Gallagher, S. J., and Henderiks, J.: Timing and Pacing of Indonesian Throughflow Restriction and Its Connection to Late Pliocene Climate Shifts, Paleoceanography and Paleoclimatology, 34, 635–657, https://doi.org/10.1029/2018PA003512, 2019.

Baatsen, M. L. J., von der Heydt, A. S., Kliphuis, M. A., Oldeman, A. M., and Weiffenbach, J. E.: Warm mid-Pliocene conditions without high climate sensitivity: the CCSM4-Utrecht (CESM 1.0.5) contribution to the PlioMIP2, Climate of the Past, 18, 657–679, https://doi.org/10.5194/CP-18-657-2022, 2022.

595 Boé, J.: Interdependency in Multimodel Climate Projections: Component Replication and Result Similarity, Geophysical Research Letters, 45, 2771–2779, https://doi.org/10.1002/2017GL076829, 2018.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J. L., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M. A.,

600 Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J. Y., Guenet, B., Guez, Lionel, E., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J. B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A. K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., and Vuichard, N.: Presentation and Evaluation of the IPSL-CM6A-LR

605 Climate Model, Journal of Advances in Modeling Earth Systems, 12, e2019MS002 010, https://doi.org/10.1029/2019MS002010, 2020.

Brierley, C. M.: Interannual climate variability seen in the Pliocene Model Intercomparison Project, Climate of the Past, 11, 605–618, https://doi.org/10.5194/cp-11-605-2015, 2015.

Burke, K. D., Williams, J. W., Chandler, M. A., Haywood, A. M., Lunt, D. J., and Otto-Bliesner, B. L.: Pliocene and Eocene provide best analogs for near-future climates, Proceedings of the National Academy of Sciences, 115, 13 288–13 293,

610 https://doi.org/10.1073/pnas.1809600115, 2018.

Byrne, D., Garcia-Soto, C., Hamilton, G., Leuliette, E., Campos, E., Durack, P. J., Manzella, G. M., Tadokoro, K., Schmitt, R. W., Arkin, P., Bryden, H., Nurse, L., Milliman, J., Inniss, L., and Bernal, P.: Chapter 4. The Ocean's Role in the Hydrological Cycle, chap. 4, pp. 91 – 104, Cambridge University Press, Cambridge, UK, and New York, NY, USA, 582 pp., 2016.

Cane, M. A. and Molnar, P.: Closing of the Indonesian seaway as a precursor to east African aridification around 3–4 million years ago, Nature, 411, 157–162, https://doi.org/10.1038/35075500, 2001.

Chan, W.-L. and Abe-Ouchi, A.: Pliocene Model Intercomparison Project (PlioMIP2) simulations using the Model for Interdisciplinary Research on Climate (MIROC4m), Climate of the Past, 16, 1523–1545, https://doi.org/10.5194/cp-16-1523-2020, 2020.

Chandan, D. and Peltier, W. R.: Regional and global climate for the mid-Pliocene using the University of Toronto version of CCSM4 and PlioMIP2 boundary conditions, Climate of the Past, 13, 919–942, https://doi.org/10.5194/cp-13-919-2017, 2017.

Corvec, S. and Fletcher, C. G.: Changes to the tropical circulation in the mid-Pliocene and their implications for future climate, Climate of the Past, 13, 135–147, https://doi.org/10.5194/cp-13-135-2017, 2017.

Danabasoglu, G., Bates, S. C., Briegleb, B. P., Jayne, S. R., Jochum, M., Large, W. G., Peacock, S., and Yeager, S. G.: The CCSM4 Ocean Component, Journal of Climate, 25, 1361–1389, https://doi.org/10.1175/JCLI-D-11-00091.1, 2012.

Danabasoglu, G., Lamarque, J. F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., van Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The Community Earth System Model Version 2 (CESM2), Journal of Advances in Modeling Earth Systems, 12, e2019MS001 916, https://doi.org/10.1029/2019MS001916, 2020.

Dayem, K. E., Noone, D. C., and Molnar, P.: Tropical western Pacific warm pool and maritime continent precipitation rates and their contrasting relationships with the Walker Circulation, Journal of Geophysical Research, 112, D06 101, https://doi.org/10.1029/2006JD007870, 2007.

Dowsett, H., Robinson, M., Haywood, A., Salzmann, U., Hill, D., Sohl, L., Chandler, M., Williams, M., Foley, K., and Stoll, D.: The PRISM3D paleoenvironmental reconstruction, Stratigraphy, 2010.

Dowsett, H., Dolan, A., Rowley, D., Moucha, R., Forte, A. M., Mitrovica, J. X., Pound, M., Salzmann, U., Robinson, M., Chandler, M., Foley, K., and Haywood, A.: The PRISM4 (mid-Piacenzian) paleoenvironmental reconstruction, Climate of the Past, 12, 1519–1538, https://doi.org/10.5194/cp-12-1519-2016, 2016.

Dowsett, H. J., Robinson, M. M., Stoll, D. K., Foley, K. M., Johnson, A. L. A., Williams, M., and Riesselman, C. R.: The PRISM (Pliocene palaeoclimate) reconstruction: time for a paradigm shift, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371, 20120 524, https://doi.org/10.1098/rsta.2012.0524, 2013.

Fedorov, A. V., Burls, N. J., Lawrence, K. T., and Peterson, L. C.: Tightly linked zonal and meridional sea surface temperature gradients over the past five million years, Nature Geoscience 2015 8:12, 8, 975–980, https://doi.org/10.1038/ngeo2577, 2015.

Feng, M., Zhang, N., Liu, Q., and Wijffels, S.: The Indonesian throughflow, its variability and centennial change, https://doi.org/10.1186/s40562-018-0102-2, 2018.

Feng, R., Otto-Bliesner, B. L., Brady, E. C., and Rosenbloom, N.: Increased Climate Response and Earth System Sensitivity From CCSM4 to CESM2 in Mid-Pliocene Simulations, Journal of Advances in Modeling Earth Systems, 12, e2019MS002 033, https://doi.org/10.1029/2019MS002033, 2020a.

Feng, R., Otto-Bliesner, B. L., Brady, E. C., and Rosenbloom, N.: Increased Climate Response and Earth System Sensitivity From CCSM4 to
CESM2 in Mid-Pliocene Simulations, Journal of Advances in Modeling Earth Systems, 12, 0–2, https://doi.org/10.1029/2019MS002033,
650   2020b.

Feng, R., Bhattacharya, T., Otto-Bliesner, B. L., Brady, E. C., Haywood, A. M., Tindall, J. C., Hunter, S. J., Abe-Ouchi, A., Chan, W. L.,
Kageyama, M., Contoux, C., Guo, C., Li, X., Lohmann, G., Stepanek, C., Tan, N., Zhang, Q., Zhang, Z., Han, Z., Williams, C. J., Lunt,
D. J., Dowsett, H. J., Chandan, D., and Peltier, W. R.: Past terrestrial hydroclimate sensitivity controlled by Earth system feedbacks, Nature
Communications 2022 13:1, 13, 1–11, https://doi.org/10.1038/s41467-022-28814-7, 2022.

655 Gordon, A., Sprintall, J., Van Aken, H., Susanto, D., Wijffels, S., Molcard, R., Ffield, A., Pranowo, W., and Wirasantosa, S.: The In-
donesian throughflow during 2004–2006 as observed by the INSTANT program, Dynamics of Atmospheres and Oceans, 50, 115–128,
https://doi.org/10.1016/j.dynatmoce.2009.12.002, 2010.

Gordon, A. L.: Oceanography of the Indonesian seas and their throughflow, Oceanography, 18, 15–27,
https://doi.org/10.5670/oceanog.2005.01, 2005.

660 Gutiérrez, J.M., R. J., G.T. Narisma, L.M. Alves, M. Amjad, I.V. Gorodetskaya, M. Grose, N.A.B. Klutse, S. Krakovska, J. Li, D. Martínez-
Castro, L.O. Mearns, S.H. Mernild, T. Ngo-Duc, B. van den Hurk, and Yoon, J.-H.: Ocean, Cryosphere and Sea Level Change. Contribution
of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, in: Climate Change 2021: The
Physical Science Basis, p. 3949, Megan Kirchmeier-Young, https://doi.org/10.1017/9781009157896.021, 2021.

Han, Z., Zhang, Q., Li, Q., Feng, R., Haywood, A. M., Tindall, J. C., Hunter, S. J., Otto-Bliesner, B. L., Brady, E. C., Rosenbloom, N.,
665 Zhang, Z., Li, X., Guo, C., Nisancioglu, K. H., Stepanek, C., Lohmann, G., Sohl, L. E., Chandler, M. A., Tan, N., Ramstein, G., Baatsen,
M. L., Von Der Heydt, A. S., Chandan, D., Peltier, W. R., Williams, C. J., Lunt, D. J., Cheng, J., Wen, Q., and Burls, N. J.: Evaluating the
large-scale hydrological cycle response within the Pliocene Model Intercomparison Project Phase 2 (PlioMIP2) ensemble, Climate of the
Past, 17, 2537–2558, https://doi.org/10.5194/cp-17-2537-2021, 2021.

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,
670 N., Salzmann, U., and Sohl, L. E.: Pliocene Model Intercomparison Project (PlioMIP): experimental design and boundary conditions
(Experiment 1), Geoscientific Model Development, 3, 227–242, https://doi.org/10.5194/gmd-3-227-2010, 2010.

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

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

Haywood, A. M., Dowsett, H. J., and Dolan, A. M.: Integrating geological archives and climate models for the mid-Pliocene warm period,
680 Nature Communications, 6, 1–14, https://doi.org/10.1038/ncomms10646, 2016a.

Haywood, A. M., Dowsett, H. J., Dolan, A. M., Rowley, D., Abe-Ouchi, A., Otto-Bliesner, B., Chandler, M. A., Hunter, S. J., Lunt, D. J.,
Pound, M., and Salzmann, U.: The Pliocene Model Intercomparison Project (PlioMIP) Phase 2: scientific objectives and experimental
design, Climate of the Past, 12, 663–675, https://doi.org/10.5194/cp-12-663-2016, 2016b.

Haywood, A. M., Tindall, J. C., Dowsett, H. J., Dolan, A. M., Foley, K. M., Hunter, S. J., Hill, D. J., Chan, W. L., Abe-Ouchi, A., Stepanek,
685 C., Lohmann, G., Chandan, D., Richard Peltier, W., Tan, N., Contoux, C., Ramstein, G., Li, X., Zhang, Z., Guo, C., Nisancioglu, K. H.,

Zhang, Q., Li, Q., Kamae, Y., Chandler, M. A., Sohl, L. E., Otto-Bliesner, B. L., Feng, R., Brady, E. C., Von Der Heydt, A. S., Baatsen, M. L., and Lunt, D. J.: The Pliocene Model Intercomparison Project Phase 2: Large-scale climate features and climate sensitivity, Climate of the Past, 16, 2095–2123, https://doi.org/10.5194/cp-16-2095-2020, 2020.

Ho, J. T., Thompson, J. R., and Brierley, C.: Projections of hydrology in the Tocantins-Araguaia Basin, Brazil: uncertainty assessment using the CMIP5 ensemble, https://doi.org/10.1080/02626667.2015.1057513, 61, 551–567, https://doi.org/10.1080/02626667.2015.1057513, 2016.

Hourdin, F., Grandpeix, J. Y., Rio, C., Bony, S., Jam, A., Cheruy, F., Rochetin, N., Fairhead, L., Idelkadi, A., Musat, I., Dufresne, J. L., Lahellec, A., Lefebvre, M. P., and Roehrig, R.: LMDZ5B: The atmospheric component of the IPSL climate model with revisited parameterizations for clouds and convection, Climate Dynamics, 40, 2193–2222, https://doi.org/10.1007/S00382-012-1343-Y/FIGURES/23, 2013.

Hourdin, F., Rio, C., Grandpeix, J. Y., Madeleine, J. B., Cheruy, F., Rochetin, N., Jam, A., Musat, I., Idelkadi, A., Fairhead, L., Foujols, M. A., Mellul, L., Traore, A. K., Dufresne, J. L., Boucher, O., Lefebvre, M. P., Millour, E., Vignon, E., Jouhaud, J., Diallo, F. B., Lott, F., Gastineau, G., Caubel, A., Meurdesoif, Y., and Ghattas, J.: LMDZ6A: The Atmospheric Component of the IPSL Climate Model With Improved and Better Tuned Physics, Journal of Advances in Modeling Earth Systems, 12, e2019MS001 892, https://doi.org/10.1029/2019MS001892, 2020.

Hunter, S. J., Haywood, A. M., Dolan, A. M., and Tindall, J. C.: The HadCM3 contribution to PlioMIP phase 2, Climate of the Past, 15, 1691–1713, https://doi.org/10.5194/cp-15-1691-2019, 2019.

Iturbide, M., Fernández, J., Gutiérrez, J. M., Bedia, J., Cimadevilla, E., Díez-Sierra, J., Manzanas, R., Casanueva, A., Baño-Medina, J., Milovac, J., Herrera, S., Cofiño, A. S., San Martín, D., García-Díez, M., Hauser, M., Huard, D., and Yelekci, Ö.: Repository supporting the implementation of FAIR principles in the IPCC-WGI Atlas, https://doi.org/10.5281/ZENODO.5171760, 2021.

Kageyama, M., Harrison, S. P., Kapsch, M.-L., Lofverstrom, M., Lora, J. M., Mikolajewicz, U., Sherriff-Tadano, S., Vadsaria, T., Abe-Ouchi, A., Bouttes, N., Chandan, D., Gregoire, L. J., Ivanovic, R. F., Izumi, K., LeGrande, A. N., Lhardy, F., Lohmann, G., Morozova, P. A., Ohgaito, R., Paul, A., Peltier, W. R., Poulsen, C. J., Quiquet, A., Roche, D. M., Shi, X., Tierney, J. E., Valdes, P. J., Volodin, E., and Zhu, J.: The PMIP4 Last Glacial Maximum experiments: preliminary results and comparison with the PMIP3 simulations, Climate of the Past, 17, 1065–1089, https://doi.org/10.5194/cp-17-1065-2021, 2021.

Kamae, Y., Yoshida, K., and Ueda, H.: Sensitivity of Pliocene climate simulations in MRI-CGCM2.3 to respective boundary conditions, Climate of the Past, 12, 1619–1634, https://doi.org/10.5194/cp-12-1619-2016, 2016.

Karas, C., Nürnberg, D., Gupta, A. K., Tiedemann, R., Mohan, K., and Bickert, T.: Mid-Pliocene climate change amplified by a switch in Indonesian subsurface throughflow, Nature Geoscience, 2, 434–438, https://doi.org/10.1038/ngeo520, 2009.

Kelley, M., Schmidt, G. A., Nazarenko, L. S., Bauer, S. E., Ruedy, R., Russell, G. L., Ackerman, A. S., Aleinov, I., Bauer, M., Bleck, R., Canuto, V., Cesana, G., Cheng, Y., Clune, T. L., Cook, B. I., Cruz, C. A., Del Genio, A. D., Elsaesser, G. S., Faluvegi, G., Kiang, N. Y., Kim, D., Lacis, A. A., Leboissetier, A., LeGrande, A. N., Lo, K. K., Marshall, J., Matthews, E. E., McDermid, S., Mezuman, K., Miller, R. L., Murray, L. T., Oinas, V., Orbe, C., García-Pando, C. P., Perlwitz, J. P., Puma, M. J., Rind, D., Romanou, A., Shindell, D. T., Sun, S., Tausnev, N., Tsigaridis, K., Tselioudis, G., Weng, E., Wu, J., and Yao, M. S.: GISS-E2.1: Configurations and Climatology, Journal of Advances in Modeling Earth Systems, 12, e2019MS002 025, https://doi.org/10.1029/2019MS002025, 2020.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. A.: Challenges in Combining Projections from Multiple Climate Models, Journal of Climate, 23, 2739–2758, https://doi.org/10.1175/2009JCLI3361.1, 2010.

Knutti, R., Masson, D., and Gettelman, A.: Climate model genealogy: Generation CMIP5 and how we got there, Geophysical Research Letters, 40, 1194–1199, https://doi.org/10.1002/GRL.50256, 2013a.

725 Knutti, R., Masson, D., and Gettelman, A.: Climate model genealogy: Generation CMIP5 and how we got there, Geophysical Research Letters, 40, 1194–1199, https://doi.org/10.1002/GRL.50256, 2013b.

Li, M., Gordon, A. L., Wei, J., Gruenburg, L. K., and Jiang, G.: Multi-decadal timeseries of the Indonesian throughflow, Dynamics of Atmospheres and Oceans, 81, 84–95, https://doi.org/10.1016/j.dynatmoce.2018.02.001, 2018a.

Li, X., Guo, C., Zhang, Z., Otterå, O. H., and Zhang, R.: PlioMIP2 simulations with NorESM-L and NorESM1-F, pp. 183–197, 2020.

730 Li, Y., Yang, S., Deng, Y., Hu, X., and Cai, M.: A process-level attribution of the annual cycle of surface temperature over the Maritime Continent, Climate Dynamics, 51, 2759–2772, https://doi.org/10.1007/s00382-017-4043-9, 2018b.

Madec, G.: NEMO ocean engine, vol. 27, http://www.nemo-ocean.eu/About-NEMO/Reference-manualshttp://eprints.soton.ac.uk/64324/, 2012.

Marsland, S. J., Haak, H., Jungclaus, J. H., Latif, M., and Röske, F.: The Max-Planck-Institute global ocean/sea ice model with orthogonal
735 curvilinear coordinates, Ocean Modelling, 5, 91–127, https://doi.org/10.1016/S1463-5003(02)00015-X, 2003.

Mastrandrea, M. D., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K. L., Frame, D. J., Held, H., Kriegler, E., Mach, K. J., Matschoss, P. R., Plattner, G.-K., Yohe, G. W., and Zwiers, F. W.: Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties, Tech. rep., http://www.ipcc.ch, 2010.

McClymont, E. L., Ford, H. L., Ho, S. L., Tindall, J. C., Haywood, A. M., Alonso-Garcia, M., Bailey, I., Berke, M. A., Littler, K., Patterson,
740 M. O., Petrick, B., Peterse, F., Ravelo, A. C., Risebrobakken, B., De Schepper, S., Swann, G. E. A., Thirumalai, K., Tierney, J. E., van der Weijst, C., White, S., Abe-Ouchi, A., Baatsen, M. L. J., Brady, E. C., Chan, W.-L., Chandan, D., Feng, R., Guo, C., von der Heydt, A. S., Hunter, S., Li, X., Lohmann, G., Nisancioglu, K. H., Otto-Bliesner, B. L., Peltier, W. R., Stepanek, C., and Zhang, Z.: Lessons from a high CO2 world: an ocean view from ∼ 3 million years ago, Climate of the Past, 16, 1599–1615, https://doi.org/10.5194/cp-16-1599-2020, 2020.

745 Neale, R. and Slingo, J.: The Maritime Continent and its role in the global climate: A GCM study, Journal of Climate, 16, 834–848, https://doi.org/10.1175/1520-0442(2003)016<0834:TMCAIR>2.0.CO;2, 2003.

Neale, R. B., Richter, J. H., Conley, A. J., Park, S., Lauritzen, P. H., Gettelman, A., and Williamson, D. L.: Description of the NCAR Community Atmosphere Model (CAM 4.0), Tech. Rep. April, https://www.researchgate.net/publication/224017879_Description_of_the_ Community_Atmosphere_Model_CAM_40, 2010.

750 O'Brien, C. L., Foster, G. L., Martínez-Botí, M. A., Abell, R., Rae, J. W., and Pancost, R. D.: High sea surface temperatures in tropical warm pools during the Pliocene, Nature Geoscience 2014 7:8, 7, 606–611, https://doi.org/10.1038/ngeo2194, 2014.

Oldeman, A. M., Baatsen, M. L. J., von der Heydt, A. S., Dijkstra, H. A., Tindall, J. C., Abe-Ouchi, A., Booth, A. R., Brady, E. C., Chan, W.-L., Chandan, D., Chandler, M. A., Contoux, C., Feng, R., Guo, C., Haywood, A. M., Hunter, S. J., Kamae, Y., Li, Q., Li, X., Lohmann, G., Lunt, D. J., Nisancioglu, K. H., Otto-Bliesner, B. L., Peltier, W. R., Pontes, G. M., Ramstein, G., Sohl, L. E., Stepanek, C., Tan, N.,
755 Zhang, Q., Zhang, Z., Wainer, I., and Williams, C. J. R.: Reduced El Niño variability in the mid-Pliocene according to the PlioMIP2 ensemble, Climate of the Past, 17, 2427–2450, https://doi.org/10.5194/cp-17-2427-2021, 2021.

Peltier, W. R. and Vettoretti, G.: Dansgaard-Oeschger oscillations predicted in a comprehensive model of glacial climate: A "kicked" salt oscillator in the Atlantic, Geophysical Research Letters, 41, 7306–7313, https://doi.org/10.1002/2014GL061413, 2014.

Pontes, G., Taschetto, A., Seng Gupta, A., Santoso, A., Wainer, I., Haywood, A., Chan, W.-L., Abe-Ouchi, A., Stepanek, C., Lohmann,
760 G., Hunter, S. J., Tindall, J. C., Chandler, M. A., Sohl, L. E., Peltier, W. R., Chandan, D., Kamae, Y., Nisancioglu, K. H., Zhang, Z.,

Contoux, C., Tan, N., Zhang, Q., Otto-Bliesner, B. L., Brady, E. C., Feng, R., von der Heydt, A. S., Baatsen, M. L. J., and Olde-
man, A. M.: Northward ITCZ shift drives reduced ENSO activity in the Mid-Pliocene Warm Period, Nature Geoscience, Preprint,
https://doi.org/10.21203/RS.3.RS-402220/V1, 2021.

RAMAGE, C. S.: ROLE OF A TROPICAL "MARITIME CONTINENT" IN THE ATMOSPHERIC CIRCULATION 1, Monthly Weather
765   Review, 96, 365–370, https://doi.org/10.1175/1520-0493(1968)096<0365:ROATMC>2.0.CO;2, 1968.

Rayner, N. A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, Journal
of Geophysical Research, 108, 4407, https://doi.org/10.1029/2002JD002670, 2003.

Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kirchner, I., Kornblueh, L., Rhodin, A.,
Schlese, U., Schulzweida, U., and Tompkins, A.: The atmospheric general circulation model ECHAM5: Part 1: Model description, MPI
770   Report, pp. 1–140, www.mpimet.mpg.dehttp://en.scientificcommons.org/8586047, 2003.

Salzmann, U., Haywood, A., and Lunt, D.: 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: Mathematical, Physical and Engineering
Sciences, 367, 189–204, https://doi.org/10.1098/rsta.2008.0200, 2009.

Sarathchandraprasad, T., Tiwari, M., and Behera, P.: South Asian Summer Monsoon precipitation variability during late Pliocene: Role of In-
775   donesian Throughflow, Palaeogeography, Palaeoclimatology, Palaeoecology, 574, 110 447, https://doi.org/10.1016/j.palaeo.2021.110447,
2021.

Sen Gupta, A., McGregor, S., Van Sebille, E., Ganachaud, A., Brown, J. N., and Santoso, A.: Future changes to the Indonesian Through-
flow and Pacific circulation: The differing role of wind and deep circulation changes, Geophysical Research Letters, 43, 1669–1678,
https://doi.org/10.1002/2016GL067757, 2016.

780   Sepulchre, P., Caubel, A., Ladant, J.-B., Bopp, L., Boucher, O., Braconnot, P., Brockmann, P., Cozic, A., Donnadieu, Y., Dufresne,
J.-L., Estella-Perez, V., Ethé, C., Fluteau, F., Foujols, M.-A., Gastineau, G., Ghattas, J., Hauglustaine, D., Hourdin, F., Kageyama,
M., Khodri, M., Marti, O., Meurdesoif, Y., Mignot, J., Sarr, A.-C., Servonnat, J., Swingedouw, D., Szopa, S., and Tardif, D.: IPSL-
CM5A2 – an Earth system model designed for multi-millennial climate simulations, Geoscientific Model Development, 13, 3011–3053,
https://doi.org/10.5194/gmd-13-3011-2020, 2020.

785   Sierra, J. P., Arias, P. A., Vieira, S. C., and Agudelo, J.: How well do CMIP5 models simulate the low-level jet in western Colombia?, Climate
Dynamics 2017 51:5, 51, 2247–2265, https://doi.org/10.1007/S00382-017-4010-5, 2017.

Smith, R. A., Castañeda, I. S., Groeneveld, J., De Vleeschouwer, D., Henderiks, J., Christensen, B. A., Renema, W., Auer, G., Bogus, K.,
Gallagher, S. J., and Fulthorpe, C. S.: Retracted: Plio-Pleistocene Indonesian Throughflow Variability Drove Eastern Indian Ocean Sea
Surface Temperatures, Paleoceanography and Paleoclimatology, 35, e2020PA003 872, https://doi.org/10.1029/2020PA003872, 2020.

790   Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tignor, M., and Miller, H.: Contribution of Working Group I to the
Fourth Assessment Report of the Intergovernmental Panel on Climate Change, p. 996, Cambridge University Press, Cambridge, United
Kingdom and New York, NY, USA, https://www.ipcc.ch/report/ar4/wg1/http://www.ipcc.ch/publications_and_data/ar4/wg1/en/ch5s5-es.
html, 2007.

Sprintall, J. and Révelard, A.: The Indonesian Throughflow response to Indo-Pacific climate variability, Journal of Geophysical Research:
795   Oceans, 119, 1161–1175, https://doi.org/10.1002/2013JC009533, 2014.

Sprintall, J., Wijffels, S. E., Molcard, R., and Jaya, I.: Direct estimates of the indonesian throughflow entering the indian ocean: 2004-2006,
Journal of Geophysical Research: Oceans, 114, 7001, https://doi.org/10.1029/2008JC005257, 2009.

Stepanek, C., Samakinwa, E., Knorr, G., and Lohmann, G.: Contribution of the coupled atmosphere–ocean–sea ice–vegetation model COS-MOS to the PlioMIP2, Climate of the Past, 16, 2275–2323, https://doi.org/10.5194/cp-16-2275-2020, 2020.

800 Tan, N., Contoux, C., Ramstein, G., Sun, Y., Dumas, C., Sepulchre, P., and Guo, Z.: Modeling a modern-like pCO2 warm period (Marine Isotope Stage KM5c) with two versions of an Institut Pierre Simon Laplace atmosphere–ocean coupled general circulation model, Climate of the Past, 16, 1–16, https://doi.org/10.5194/cp-16-1-2020, 2020.

Tan, N., Zhang, Z. S., Guo, Z. T., Guo, C. C., Zhang, Z. J., He, Z. L., and Ramstein, G.: Recognizing the Role of Tropical Seaways in Modulating the Pacific Circulation, Geophysical Research Letters, 49, e2022GL099 674, https://doi.org/10.1029/2022GL099674, 2022.

805 Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research: Atmospheres, 106, 7183–7192, https://doi.org/10.1029/2000JD900719, 2001.

Tebaldi, C. and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365, 2053–2075, https://doi.org/10.1098/RSTA.2007.2076, 2007.

Tierney, J. E. and Tingley, M. P.: BAYSPLINE: A New Calibration for the Alkenone Paleothermometer, Paleoceanography and Paleoclima-810 tology, 33, 281–301, https://doi.org/10.1002/2017PA003201, 2018.

Tierney, J. E., Malevich, S. B., Gray, W., Vetter, L., and Thirumalai, K.: Bayesian Calibration of the Mg/Ca Paleothermometer in Planktic Foraminifera, Paleoceanography and Paleoclimatology, 34, 2005–2030, https://doi.org/10.1029/2019PA003744, 2019.

Tillinger, D.: Physical oceanography of the present day Indonesian Throughflow, Geological Society Special Publication, 355, 267–281, https://doi.org/10.1144/SP355.13, 2011.

815 Trenberth, K. E., Fasullo, J. T., and Kiehl, J.: Earth's global energy budget, Bulletin of the American Meteorological Society, 90, 311–323, https://doi.org/10.1175/2008BAMS2634.1, 2009.

Wainwright, L., Meyers, G., Wijffels, S., and Pigot, L.: Change in the Indonesian Throughflow with the climatic shift of 1976/77, Geophysical Research Letters, 35, https://doi.org/10.1029/2007GL031911, 2008.

Wang, C.: Three-ocean interactions and climate variability: a review and perspective, Climate Dynamics, 53, 5119–5136, 820 https://doi.org/10.1007/s00382-019-04930-x, 2019.

Wara, M. W., Ravelo, A. C., and Delaney, M. L.: Permanent El Niño-Like Conditions During the Pliocene Warm Period, Science, 309, 758–761, https://doi.org/10.1126/science.1112596, 2005.

Wijffels, S. and Meyers, G.: An intersection of oceanic waveguides: Variability in the Indonesian throughflow region, Journal of Physical Oceanography, 34, 1232–1253, https://doi.org/10.1175/1520-0485(2004)034<1232:AIOOWV>2.0.CO;2, 2004.

825 Wilks, D. S.: Statistical methods in the atmospheric sciences, vol. 16, Elsevier/Academic Press,, Amsterdam ;, 2011.

Williams, C. J. R., Sellar, A. A., Ren, X., Haywood, A. M., Hopcroft, P., Hunter, S. J., Roberts, W. H. G., Smith, R. S., Stone, E. J., Tindall, J. C., and Lunt, D. J.: Simulation of the mid-Pliocene Warm Period using HadGEM3: experimental design and results from model–model and model–data comparison, Climate of the Past, 17, 2139–2163, https://doi.org/10.5194/cp-17-2139-2021, 2021.

Williams, C. J. R., Lunt, D. J., Salzmann, U., Reichgelt, T., Inglis, G. N., Greenwood, D. R., Chan, W.-L., Abe-Ouchi, A., Donnadieu, 830 Y., Hutchinson, D. K., de Boer, A. M., Ladant, J.-B., Morozova, P. A., Niezgodzki, I., Knorr, G., Steinig, S., Zhang, Z., Zhu, J., Huber, M., Otto-Bliesner, B. L., and R Williams, C. J.: African Hydroclimate During the Early Eocene From the DeepMIP Simulations, Paleoceanography and Paleoclimatology, 37, e2022PA004 419, https://doi.org/10.1029/2022PA004419, 2022.

Yamanaka, M. D., Ogino, S.-Y., Wu, P.-M., Jun-Ichi, H., Mori, S., Matsumoto, J., and Syamsudin, F.: Maritime continent coastlines controlling Earth's climate, Progress in Earth and Planetary Science, 5, 21, https://doi.org/10.1186/s40645-018-0174-9, 2018.

835   Yoneyama, K. and Zhang, C.: Years of the Maritime Continent, Geophysical Research Letters, 47, https://doi.org/10.1029/2020GL087182, 2020.

Zhang, Q., Berntell, E., Axelsson, J., Chen, J., Han, Z., De Nooijer, W., Lu, Z., Li, Q., Zhang, Q., Wyser, K., and Yang, S.: Simulating the mid-holocene, last interglacial and mid-pliocene climate with ec-earth3-lr, Geoscientific Model Development, 14, 1147–1169, https://doi.org/10.5194/GMD-14-1147-2021, 2021.

840   Zhang, T., Yang, S., Jiang, X., and Zhao, P.: Seasonal-interannual variation and prediction of wet and dry season rainfall over the maritime continent: Roles of ENSO and monsoon circulation, Journal of Climate, 29, 3675, https://doi.org/10.1175/JCLI-D-15-0222.1, 2016.

Zhang, Y. G., Pagani, M., and Liu, Z.: A 12-Million-Year Temperature History of the Tropical Pacific Ocean, Science, 344, 84–87, https://doi.org/10.1126/science.1246172, 2014.