

My co-authors and I would like to thank the reviewers for their reviews of the paper. We have addressed the comments as detailed below:

Reviewer 1

This version of the paper is much improved over the previous versions in that it now has almost become a scientifically sound description of the PlioMIP ensemble. I'm hence happy to have this paper published subject to the authors addressing the following remaining issues:

1. More details are needed regarding the calculation of sea-ice thickness. It remains unclear whether the actual (floe) thickness is analyzed, or the so-called equivalent thickness that gives mean sea-ice thickness including the contribution of the open water part of the grid cell. Since the authors describe an impact of sea-ice concentration on thickness in the caption of figure 7, I assume that the equivalent thickness is used. However, this thickness is not related to observations of ice thickness (which give actual floe thickness), and is not related to any of the existing studies examining the ice-thickness patterns in CMIP simulations. Hence, either the authors change their analysis to an analysis of floe thickness (by dividing equivalent thickness by concentration), or the reference to existing studies becomes impossible to interpret since these are all based on an analysis of floe thickness. This important point was made already in my first review of this paper, I believe.

Thickness is floe thickness in the models. This is now clarified in section 2.2. The Figure 7 caption was supposed to make clear that the low (or zero) concentration will also mean low thickness (i.e. very little ice left). That part of the caption has been removed to clear up any uncertainty.

2. On a related note, how is spatial mean ice thickness calculated? How is open water considered in the averaging?

Cells with no ice are not included in the averaging. The calculation looks at the average thickness (floe thickness) of all cells north of 80N containing ice. Cells appropriately weighted for size.

3. Why does RMSE of ice thickness not have any unit throughout the entire paper?

Changed to m throughout

4. Around line 135: A greater relative change of summer sea ice does not imply an enhanced annual cycle. Only a greater absolute change does.

Absolute change is also greater, as stated earlier in the paragraph. However, the sentence referring to enhanced seasonal cycle has been removed due to changes made based on other comments.

5. There is still quite a bit of unnecessarily vague/repetitive language throughout the paper, e.g. l. 151: "The ensemble mean extent amplitude of the mid-Pliocene simulations is by 20 % greater than the pre-industrial ensemble mean amplitude, further indication of the enhanced seasonal sea ice extent cycle in the mid-Pliocene simulations." I simply don't get why, an in-

creased amplitude is only an indication of an increased amplitude.

Reference to ‘further indication of enhanced seasonal sea ice extent’ has been removed.

e.g., l.210 "Sea ice volume appears to show a stronger relationship with SAT, with a correlation coefficient of -0.76 versus -0.18" Why is this only apparently a stronger relationship?

Changed ‘appears to show’ to ‘show’

e.g., l. 289: CCSM4 has a change of 4.1 % in the extent amplitude, so I don't understand why this model only "appears" to not have a substantial change in the annual cycle

Changed to ‘there is no substantial change’

Please also check the logical flow within parts of the paper. For example, why does the fact that MRI-CGCM historical simulations have a larger extent than observed show that that model "consistently simulates Arctic sea ice extent larger than the ensemble mean". To me, this only shows that it has a larger extent than observed.

Added ‘and over 18% greater than the ensemble mean (for CMIP5 simulations),’ in this sentence, so is now consistent.

6. l.295: Why does the fact that pre-industrial ice is very thick in summer in MRI imply that there is a greater reduction in the maximum extent than in the minimum extent? I would expect the ice to then be even thicker during winter, hindering a loss of extent also then.

This part has been removed in the rearranging of the section to the results area from discussion.

7. l.305: Why is there disagreement "particularly in summer"? From Figure 6 it seems that absolute model spread is even larger in winter, which is surprising given the geometry of the Arctic Ocean with its rather narrow passages at the edge of winter sea ice.

‘Particularly in summer’ removed

8. l.470: Please give more details as to which lengths of simulations are required for such analysis? From table 1 it seems that some of the models that were not analyzed have even longer runs than the three models listed here.

The issue was not with the length of runs, but the availability of the timeseries for the data. Most models only provided the averaged values. The analysed models had full timeseries for the runs performed. To avoid potential confusion, the ‘sufficiently long time series’ section has been removed.

9. Figure 4:a "jet" colorbar is very hard to read for color-blind readers. In addition, for the lower panel a color scale that diverges from 0 seems useful, or, indeed, a color scale that has no negative values since these do not seem to exist in your analysis.

Changed to parula colorbar, and lower panel changed to go from 0 to 1, rather than -1 to 1

10. *Related to Figure 4: Details should be given on how RMSE and spatial correlation were calculated.*

Added to methods section.

11. *Much of sections 4.1 and 4.2 up to line 305 do not read like a discussion, but instead like a continuation of the results section. It might be a good idea to indeed incorporate those subsections into section 3.*

Lines 225-275 moved to results in a ‘Comparison to CMIP5 simulations’ section of the pre-industrial results. 277-304 have been moved to the Pliocene extent section of results. Some text has been edited or removed so that there is no repetition.

Reviewer 2

The authors made a large effort and substantially improved the manuscript. The main problem of this study is still, that as long as it is unclear how ice conditions in the mid-Pliocene really were no new insights on model performance and reliability, compared to existing studies, can be presented. However, despite this problem, I find that the study now provides a sufficiently large number of interesting results to be publishable after responding to a number of minor comments.

1. *Line 7-8: Why is this in contrast to proxy-data? If I understood right (see also introduction lines 27ff), it depends on the proxy data if we assume ice-free summer in Pliocene or not. In this way, the model uncertainties are in line with the uncertainties from the different proxies.*

This part of the abstract changed to ‘Half of the PlioMIP models simulate ice-free conditions in the mid-Pliocene. This spread amongst the ensemble are in line with the uncertainties amongst proxy reconstructions for mid-Pliocene sea-ice extent.’

2. *Line 24-25: I guess you mean RCP8.5?*

Changed to ‘Under RCP 4.5, many models predict seasonally sea ice-free conditions in the Arctic by the end of the 21st century, with some projections suggesting an ice free Arctic by 2030 under RCP 8.5’. This is in agreement with data shown in given references.

3. *Lines 34-36: I totally agree, these are key questions but unfortunately this article is not really investigating these questions.*

Ok, this sentence removed

4. *Line 59: add ?in? before Haywood et al. (2011)*

Done

5. *Line 113: ‘in Figure 4’ instead of ‘by Figure 4’*

Done

6. *Figure 4: What do you show? Annual means, FMA, ASO? Please clarify in the figure caption. Do you calculate the RMSE only for north of 80N? Please clarify in the figure caption.*

‘Mean annual’ added before Arctic sea ice thickness in first sentence in caption. Ice-covered regions north of 60N also added.

7. *Figure 4/lines 112ff: It seems that figure 4 is showing the RMSE from one specific model with respect to another specific model (respect to the ensemble mean). Please clarify this in the figure caption. To avoid misunderstanding you might want to call it RMSD (root mean square deviation) and not RMSE because it is not really an error you are calculating but the deviation or difference between two models (if I understood correctly).*

Changed to RMSD

8. *Line 132: Here a line or two on a possible relation between ice extent in pre-industrial and in Pliocene would be nice. Table 2 shows that those 3 models with highest ice extent in the pre-industrial period show also the highest extent in the mid-Pliocene period. However, the reductions of ice extent between mid-Pliocene and pre-industrial periods seem to be independent of the ice extent in the pre-industrial period. The ice extent in MIROC4m and MRI-CGCM is almost the same in the PI-period but MIROC4m simulates a twice as large reduction in the mid-Pliocene period.*

This is covered in section 3.4

9. *Lines 138/139: This is only true for MIROC4m and NorESM. The other 6 models show almost the same annual cycle in pre-industrial and Pliocene climates.*

Line stating entire ensemble characterized by change in annual cycle removed.

10. *Line 145: I do not really like this ‘overlap’. Maybe better: ‘This indicates the large spread/ uncertainty in the representation of sea ice extent in the models.’*

Changed as advised

11. *Lines 138-139 and Lines 150-156: These discussions should be merged. And as stated earlier, only 2 models show a substantially reduced annual cycle, the others more or less the same. These two models are thus also responsible for the increase by 20% in the ensemble mean. Why are these two models so outstanding in this respect?*

Discussions have been merged, and references to the mid-Pliocene cycle being enhanced have been removed, save for the line stating the ensemble mean cycle is increased by 20%. This is within the section stating that not all models see this pattern.

12. *Line 173: ‘overall no more or less similar’ sounds awkward. Maybe better: ‘but the differences between thickness patterns are comparable’*

Changed as advised

13. Line 183: *'have' instead of 'had'*

Done

14. Line 191: *delete 'on the other hand'*

Done

15. Line 193: *you could add that SD is much smaller in mid-Pliocene in summer because little ice is left in almost all models.*

SD is actually higher in mid-Pliocene summer than pre-industrial summer

16. Line 194: *delete one 'of'*

Done

17. Lines 261-264: *These sentences are not clear, please reformulate. What is indeed interesting (and maybe this is meant by lines 261-264) is that CCSM4 and MRI-CGCM show a higher annual ice extent in mid-Pliocene than in the recent past – CMIP5 (1979-2005) simulations. Further, HadCM3 simulates a larger ice extent in the 1979-2005 period than in the pre-industrial period. Here, it might be useful to look at the ice area as well and not only at the ice extent.*

Changed to *'In both PlioMIP simulations, compared to the ensemble mean CCSM4 simulates a greater mean and HadCM3 simulates a smaller mean annual extent. In CMIP5, the reverse is true'*

Looking at ice area rather than extent does not make any significant difference to the analysis

18. *Another interesting addition to the comparison with CMIP5 would be to compare the mid-Pliocene ice conditions to the ice conditions in a future period with similar greenhouse-gas concentrations in CMIP5-simulations.*

This has not been done as the comparable figures have not been published, unlike the historical results in Shu et al. [2015]

19. Line 307ff: *I absolutely agree. But given the fact that proxies obviously allow for a variety of conclusions on the sea ice conditions in mid-Pliocene, I wonder if it is possible to realistically reconstruct mid-Pliocene sea ice? This touches also the main problem of this study. As long as it is unclear how ice conditions in the mid-Pliocene really were, we do not get a lot of new insights on model performance or reliability of models.*

Whether it is possible to realistically reconstruct mid-Pliocene sea ice is a challenge for those working in the proxy data fields. Before that happens, it is useful to have a documentation of the different spread of model simulations of mid-Pliocene sea ice, which also helps to highlight

the need for further study of proxies.

20. Lines 368ff: Are you sure NorESM and CCSM really use exactly the same values for their albedo? Even if NorESM and CCSM use the same ice model, this does not necessarily mean that the albedo is exactly the same. Often, the albedo is used for model tuning.

It is not stated anywhere that the albedo values are the same, just that the scheme is the same. In fact, the point made is that despite the scheme being the same, there are large differences in the effects of sea ice albedo in the models? simulation of Arctic sea ice.

21. Figure 12: I appreciate that you show the wind patterns and it is fine to only show selected models with quite some differences in the wind patterns (and the other models in the supplement) but I wonder why you chose different models for Pliocene and preindustrial. It would be more straightforward to show the same models to understand differences between the atmospheric circulations and their possible influence on sea ice distribution in Pliocene and pre-industrial.

Changed the panel to both pre-industrial and mid-Pliocene plots for IPSL and MIROC to have consistency.

List of changes

Abstract

- One sentence changed relating to proxy data interpretations of mid-Pliocene sea ice

Introduction

- Minor language and structure changes

Methods

- No COSMOS in CMIP5 removed, explanation of similar model that exists
- Explanation of root mean square deviation and spatial pattern correlation calculations
- Ice thickness stated to be floe thickness

Results

- Root mean square error changed to root mean square deviations, units (m) added
- CMIP5 comparison moved to results
- Remaining parts of section 4.1 moved to section 3.3.1
- Some sentences removed to avoid repetition as a result of sections being moved
- Minor language and structural changes

Discussion

- Description of Figure 12 altered
- Reference to HadCM3 being ‘best’ Pliocene sea ice simulation removed
- Minor language and structural changes

Conclusions

- Reference to HadCM3 being ‘best’ Pliocene sea ice simulation removed

Tables and Figures

- Table 2 – Numbers added to COSMOS CMIP5, mean values adjusted accordingly
- Figure 4 – Colour bar and scales changed
- Figure 7 – Caption edited
- Figure 12 – Solely MIROC and IPSL plots shown, caption updated accordingly

Bibliography

Q. Shu, Z. Song, and F. Qiao. Assessment of sea ice simulations in the cmip5 models. *The Cryosphere*, 9(1):399–409, 2015. doi: 10.5194/tc-9-399-2015.

Arctic sea ice simulation in the PlioMIP ensemble

F.W. Howell¹, A.M. Haywood¹, B.L. Otto-Bliesner², F. Bragg³, W.-L. Chan⁴,
M.A. Chandler⁵, C. Contoux⁶, Y. Kamae⁷, A. Abe-Ouchi^{4,8}, N.A. Rosenbloom²,
C. Stepanek⁹, and Z. Zhang¹⁰

¹School of Earth and Environment, University of Leeds, Leeds, UK

²National Center for Atmospheric Research, Boulder, Colorado, USA

³School of Geographical Sciences, University of Bristol, Bristol, UK

⁴Atmosphere and Ocean Research Institute, University of Tokyo, Kashiwa, Japan

⁵Columbia University - NASA/GISS-E2-R, New York, NY, USA

⁶Aix-Marseille Université, CNRS, IRD, CEREGE UM34, Aix en Provence, France

⁷Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

⁸Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan

⁹Alfred Wegener Institute - Helmholtz Centre for Polar and Marine Research, Bremerhaven,
Germany

¹⁰Bjerknes Centre for Climate Research, Bergen, Norway

Correspondence to: F.W. Howell (eefwh@leeds.ac.uk)

Abstract.

Eight general circulation models have simulated the mid-Pliocene Warm Period (mid-Pliocene, 3.264 to 3.025 Ma) as part of the Pliocene Modelling Intercomparison Project (PlioMIP). Here, we analyse and compare their simulation of Arctic sea ice for both the pre-industrial and the mid-Pliocene. Mid-Pliocene sea ice thickness and extent is reduced, and the model spread of extent is more than twice the pre-industrial spread in some summer months. ~~As for the proxy record, the simulated predominant sea ice state is ambiguous; half of the models in the ensemble~~ Half of the PlioMIP models simulate ice-free conditions in the mid-Pliocene ~~summer, in contrast to proxy data evidence that suggests the possibility of perennial sea ice.~~ This spread amongst the ensemble is in line with the uncertainties amongst proxy reconstructions for mid-Pliocene sea-ice extent. Correlations between mid-Pliocene Arctic temperatures and sea ice extents are almost twice as strong as the equivalent correlations for the pre-industrial simulations. The need for more comprehensive sea ice proxy data is highlighted, in order to better compare model performances.

1 Introduction

15 The mid-Pliocene warm period (mid-Pliocene), spanning 3.264 to 3.025 Myr ago (Dowsett et al., 2010) was a period exhibiting episodes of global warmth, with estimates of an increase of 2 to 3°C in global mean temperatures in comparison to the pre-industrial period (Haywood et al., 2013). The mid-Pliocene is the most recent period of earth history that is thought to have atmospheric CO₂ concentrations resembling those seen in the 21st century, with concentrations estimated to be

20 between 365 and 415 ppm (e.g. Pagani et al. (2010); Seki et al. (2010)). Therefore, this time period is a useful interval in which to study the dynamics and characteristics of sea ice in a warmer world.

September 2012 saw Arctic sea ice fall to a minimum extent of $3.4 \times 10^6 \text{ km}^2$, a reduction of $4.2 \times 10^6 \text{ km}^2$ since the beginning of satellite observations in 1979 (Parkinson and Comiso, 2013; Zhang et al., 2013a). ~~The Arctic is widely predicted to become seasonally ice-free before the~~ Under
25 RCP 4.5, many models predict seasonally sea ice-free conditions in the Arctic by the end of the 21st century (~~under RCP 4.5~~) (e.g. Stroeve et al. (2012); Massonnet et al. (2012)), with some projections suggesting an ice free Arctic by 2030 under RCP 8.5 (Wang and Overland, 2012), whilst other studies (e.g. Boé et al. (2009)) suggest a later date for the disappearance of summer Arctic sea ice.

There is debate concerning whether the Arctic sea ice in the mid-Pliocene was seasonal or peren-
30 nial. Darby (2008) suggests that the presence of iron grains in marine sediments extracted from the Arctic Coring Expedition (ACEX) core, located on the Lomonosov Ridge (87.5°N , 138.3°W), shows that there was year round coverage of sea ice at this location, whilst there are indications from ostracode assemblages and ice rafted debris sediments as far north as Meighen Island (approx. 80°N) that Pliocene Arctic sea ice was seasonal (Cronin et al., 1993; Moran et al., 2006; Polyak et al., 2010).
35 The prospect of the Arctic becoming ice-free in summer in the future increases the importance of the investigation of past climates which may have had seasonal Arctic sea ice. ~~Of particular interest is an understanding of the processes and sensitivities of Arctic sea ice under such conditions and of the general impact of reduced summer Arctic sea ice on climate.~~

Whilst many studies have focused on the simulation of Arctic sea ice for present and future climate
40 by a variety of modelling groups (e.g. Arzel et al. (2006), Parkinson et al. (2006), Stroeve et al. (2007), Johnson et al. (2007), Holland and Stroeve (2011), Stroeve et al. (2012), Johnson et al. (2012), Blanchard-Wrigglesworth and Bitz (2014), Stroeve et al. (2014), Shu et al. (2015)), there has been little focus on the simulation of past sea ice conditions by an ensemble of models, particularly for climates with warmer than modern temperatures and reduced Arctic sea ice cover. Berger et al.
45 (2013) looks at the response of sea ice to insolation changes in simulations of mid-Holocene climate by PMIP2 and PMIP3 models, which shows that all the models simulate a modest reduction in summer sea ice extent in the mid-Holocene compared to the pre-industrial control (mean difference is lower than the difference in the mean observational Arctic sea ice extents for 1980-1989 and 2000-2009), but in the winter approximately half simulate a more extensive mid-Holocene sea ice
50 cover.

The Pliocene Modelling Intercomparison Project (PlioMIP) is a multi-model experiment which compares the output of different models' simulations of the mid-Pliocene, as well as pre-industrial simulations, each following a standard experimental design, set out in Haywood et al. (2010, 2011) (further details in section 2.1). In this study we analyse the simulation of Arctic sea ice in each of the
55 participating models in PlioMIP Experiment 2 (see Table 1), focusing on both the pre-industrial and

mid-Pliocene outputs. We quantify the variability of sea ice extent and thickness in both simulations, and present an overview of some of the important mechanisms influencing the simulation of sea ice.

2 Methods

2.1 PlioMIP experimental design

60 Two experimental designs for the PlioMIP simulations are described, Experiment 1 in Haywood et al. (2010) and Experiment 2 [in](#) Haywood et al. (2011). Experiment 1 used atmosphere only GCMs (AGCMs), whilst Experiment 2 used coupled atmosphere-ocean GCMs (AOGCMs). Both experimental designs describe the model set-up for pre-industrial and mid-Pliocene simulations. The PRISM3D reconstruction provides the boundary conditions for the mid-Pliocene simulations, which
65 in Experiment 1 also includes the prescribed SSTs and sea ice extents. SST reconstruction utilises a multi-proxy approach, based on faunal analysis, alkenone unsaturation index palaeothermometry, and foraminiferal Mg/Ca ratios Dowsett et al. (2010). Maximum sea ice extent in the mid-Pliocene is set as equal to modern sea ice extent minimum, with sea-ice free conditions for the mid-Pliocene minimum extent (Haywood et al., 2010). These boundary conditions are based on inferences from
70 the SST reconstruction, and evidence from diatoms and sedimentological data (Dowsett et al., 2010). In both Experiment 1 and Experiment 2, atmospheric CO₂ is 405 ppm, and a modern orbital configuration is used.

In Table 1, details of the eight models which ran PlioMIP Experiment 2 simulations are summarised. With the exception of GISS-E2-R, each model was also used for Experiment 1 simulations.
75 Four of the models (CCSM4, GISS-E2-R, HadCM3 and IPSLCM5A) are also represented in the CMIP5 ensemble, the results for which are contrasted with the PlioMIP results. Higher resolution versions of MIROC4m and NorESM-L, and an updated version of MRI-CGCM also ran CMIP5 simulations. ~~COSMOS was not represented in CMIP5, or any related version of it~~ [For COSMOS, results from the model MPI-ESM-LR, which has a higher resolution and an updated version of the
80 ECHAM model in COSMOS, are shown.](#)

2.2 Analysis of results

We focus on the key sea ice metrics of extent (defined as the area of ocean where sea ice concentration is at least 15%), thickness ([floe thickness](#)), and volume. ~~We follow the example of Berger et al. (2013) and examine the mean sea ice thickness~~ [Root mean square deviations \(RMSD\) and spatial pattern correlations
85 \(SPC\) are calculated for mean annual sea ice thicknesses. Analysis of spatial averages of sea ice thickness covers north of 80°N - \(following the example of Berger et al. \(2013\) \), whereas the RMSD and SPC are calculated for ice covered areas north of 60°N. SPC is calculated using Pearson product-moment coefficient of linear correlation.](#)

To understand differences in the models' simulation of sea ice, we quantify correlations between
90 the sea ice metrics and sea surface and surface air temperatures. We also compare the pre-industrial
and mid-Pliocene sea ice extents to establish how closely correlated they are. This enables us to
determine to which degree the mid-Pliocene sea ice cover is influenced by the temperatures and
control simulations.

In our analysis, we define winter as the months February to April (FMA), and summer as the
95 months August to October (ASO). The rationale is that in at least half of the models these are the
three months with the highest and lowest mean sea ice extents respectively. This is in contrast to the
typical seasonal definitions of winter (December to February) and summer (June to August).

3 Results

3.1 Pre-industrial sea ice simulations

100 3.1.1 Sea ice extent

Plots of the mean summer and winter pre-industrial Arctic sea ice concentrations are shown in Figure
1. Across the eight-member ensemble, the multi-model mean annual sea ice extent is $16.17 \times 10^6 \text{ km}^2$
(Table 2), with a winter (FMA) multi-model mean of $20.90 \times 10^6 \text{ km}^2$, and a summer (ASO) multi-
model mean of $10.98 \times 10^6 \text{ km}^2$. The individual models' annual means range from $12.27 \times 10^6 \text{ km}^2$
105 (IPSLCM5A) to $19.85 \times 10^6 \text{ km}^2$ (MIROC4m) (Table 2), and monthly multi-model means range
from a minimum of $10.01 \times 10^6 \text{ km}^2$ (September) to a maximum of $21.24 \times 10^6 \text{ km}^2$ (March, Figure
2). The lowest individual monthly extent is $7.00 \times 10^6 \text{ km}^2$ (HadCM3, September), with the highest
monthly extent produced by MRI-CGCM (March), measuring $27.01 \times 10^6 \text{ km}^2$ (Figure 2).

Figure 2 reveals the differences in the annual sea ice extent cycles across the ensemble. The sea
110 ice extent amplitudes of NorESM-L and IPSLCM5A are 6.39 and $7.36 \times 10^6 \text{ km}^2$ respectively (Table
2). These are the only models in the ensemble with seasonal amplitudes below $10 \times 10^6 \text{ km}^2$. Other
models in the ensemble show a much larger seasonal cycle, in particular GISS-E2-R, MIROC4m and
MRI-CGCM, which have sea ice extent amplitudes of 14.03 , 14.05 , and $15.91 \times 10^6 \text{ km}^2$ respectively
(Table 2). The ensemble mean sea ice extent amplitude is $11.18 \times 10^6 \text{ km}^2$.

115 3.1.2 Sea ice thickness

North of 80°N , the multi-model mean annual thickness is 2.97 m , with a winter multi-model mean of
 3.29 m and a summer multi-model mean of 2.52 m . Across the ensemble, the annual mean thickness
varies from 2.27 m (HadCM3) to 3.81 m (CCSM4). The winter thicknesses range from 2.56 m
(NorESM-L) to 4.01 m (CCSM4), with summer between 1.27 m (GISS-E2-R) and 3.60 m (CCSM4).
120 Plots of mean winter and summer pre-industrial Arctic sea ice thicknesses are shown in Figure 3.

Root mean square errors (RMSE) and spatial pattern correlations (RMSE-RMSDs and SPCs) for mean annual Arctic sea ice thickness are shown by Figure ?? (for ice-covered areas north of 60°N) are shown in Figure 4. MIROC4m has the highest spatial pattern correlation SPC with the ensemble mean (0.93), despite the thickest ice in its simulation being located north of Eastern Siberia, opposite the region of thickest ice in many of the models (see Figure 3). It also has the lowest RMSE-RMSD (0.55 m), marginally lower than COSMOS (0.56 m). MRI-CGCM displays the lowest spatial pattern correlation SPC with the ensemble mean (0.76) and the highest RMSE-RMSD (1.33 m). The lowest spatial pattern correlation SPC between two models is 0.51 (HadCM3 and MRI-CGCM), which have a RMSE-RMSD of 1.83 m, the highest of the ensemble. HadCM3 has a thickness spatial pattern which appears by eye very different to other PlioMIP models, with the thickest ice in a wedge bounded approximately by the 70°N latitude line, and 120°W and 150°E (see Figure 3). However, it has a greater spatial pattern correlation SPC with the ensemble mean than GISS-E2-R or MRI-CGCM, and the RMSE-RMSD between the ensemble mean thickness and HadCM3 is lower than GISS-E2-R or MRI-CGCM when compared to the ensemble mean (Figure ??4).

3.2 Comparison to CMIP5 simulations

Before examining the simulations of Arctic sea ice for the mid-Pliocene, the simulations of pre-industrial sea ice cover by individual models are assessed. A comparison with observed sea ice characteristics is a suitable methodology. Ideally, we would have compared the output of the pre-industrial simulations to observations of sea ice from the same time period. However, the most spatially and temporally comprehensive observations of sea ice originate from satellites. Respective data sets date back only as far as 1979, which is more than 100 years after the time period that the pre-industrial simulations represent.

Whilst there are observations of sea ice characteristics available dating back to the early 20th century, that could have been used for the comparison, most, particularly the earliest, are ship-based observations of ice margins. These observations are only available for the spring and summer months (e.g. Thomsen (1947); Walsh and Chapman (2001)), and the sea ice extent in the remaining months must be estimated by extrapolation. Frequency and location of these observations are determined by shipping patterns, rather than by the scientific need for spatial and temporal coverage.

Due to the differences between the climate states represented by models and the chosen observations, we do not make any direct comparisons. However, all of the PlioMIP models, or related versions, are represented in the CMIP5 ensemble, for which historical simulations exist that can be directly compared to modern observations.

Shu et al. (2015) provides an assessment of the historical simulation of Arctic sea ice by the CMIP5 models for the period 1979-2005. Their results show that for the historical simulations by the PlioMIP models in CMIP5, MRI-CGCM simulates the highest mean annual sea ice extent ($15.01 \times 10^6 \text{ km}^2$), compared to the satellite observational mean of $12.02 \times 10^6 \text{ km}^2$ for the comparable

period (1979-2005). MRI-CGCM simulates the second highest pre-industrial mean annual sea ice extent (just 0.05×10^6 km² less than MIROC4m), and the highest mid-Pliocene mean annual sea ice extent. The CMIP5 historical extent simulated by MRI-CGCM is almost 25% greater than the observational mean, and over 18% greater than the ensemble mean (for CMIP5 simulations), showing MRI-CGCM consistently simulates Arctic sea ice extent larger than the ensemble mean.

In contrast, MIROC4m simulates a pre-industrial mean annual sea ice extent that is similar to the MRI-CGCM PlioMIP simulation, and represents the lowest historical mean annual sea ice extent of the CMIP5 models that are included in the PlioMIP ensemble (10.66×10^6 km², Shu et al. (2015)). The NorESM-M, the higher resolution version of NorESM-L, which simulates both the lowest PlioMIP pre-industrial and mid-Pliocene mean annual sea ice extents, is the CMIP5 model which simulates the closest historical mean annual sea ice extent to the observations (12.01×10^6 km², just 0.01×10^6 km² lower than the observations). As NorESM-L does with the PlioMIP simulations, NorESM-M simulates the lowest sea ice extent amplitude of the PlioMIP models in CMIP5 (Shu et al., 2015).

In addition to the mean annual sea ice extent simulated by each model in the CMIP5 historical and PlioMIP simulations, Table 2 shows the ensemble mean annual extents for these sets of simulations. In both pre-industrial and mid-Pliocene simulations, compared to the ensemble mean CCSM4 simulates a greater mean and HadCM3 simulates a smaller mean annual extent. In the CMIP5 simulations, the reverse is true (see Table 2).

Arctic sea ice thickness in the CMIP5 simulations is analysed in Stroeve et al. (2014). The correlations between the spatial patterns of Arctic sea ice thickness in the simulations (average over the years 1981-2010) and observations from Kwok et al. (2009) are less than 0.4 for all the considered PlioMIP models — with the exception of CCSM4, which has the highest SPC of the entire CMIP5 ensemble. For each PlioMIP model, the spatial patterns of sea ice thickness in the pre-industrial simulation resembles the thickness spatial pattern in that model's CMIP5 simulation, shown in Stroeve et al. (2014). It has been noted that the SPC between different ensemble simulations with the same model is significantly higher than the correlation between one model and the observations, which suggests that poor correlations are more likely explained by biases within the models, rather than by natural variability.

3.3 Pliocene simulations

3.3.1 Sea ice extent

In agreement with enhanced greenhouse forcing each model in the ensemble simulates a smaller sea ice extent in the mid-Pliocene simulation in comparison to the pre-industrial (Figures 1 and 5). The multi-model mean annual extent for the mid-Pliocene simulations is 10.84×10^6 km², a reduction of 5.33×10^6 km² (33.0%) in comparison to the respective multi-model mean of the pre-industrial

simulations. Annual means in the ensemble range from $7.60 \times 10^6 \text{ km}^2$ (NorESM-L), to $15.84 \times 10^6 \text{ km}^2$ (MRI-CGCM) (Table 1).

The lowest multi-model monthly mean extent is $3.15 \times 10^6 \text{ km}^2$ (September), and the highest is
195 $16.59 \times 10^6 \text{ km}^2$ (March). In comparison to the pre-industrial simulation, the lowest multi-model
monthly mean extent is reduced by $6.86 \times 10^6 \text{ km}^2$ (69%). The reduction for the highest monthly
multi-model mean is $4.65 \times 10^6 \text{ km}^2$ (22%). The relative change in the lowest extent is therefore
over three times greater than the relative change in the highest extent. ~~Therefore,~~

MRI-CGCM, CCSM4 and MIROC4m simulate the highest maximum mid-Pliocene sea ice extents
200 in the ensemble. Both CCSM4 and MRI-CGCM also provide the highest two minimum extents, but
MIROC4m is one of the four models that simulates an ice-free Arctic summer. As a result, the
sea ice extent amplitude in MIROC4m in the mid-Pliocene simulations is $\approx 64\%$ greater than the
pre-industrial simulation extent amplitude (Table 2). The ensemble mean extent amplitude of the
mid-Pliocene ~~is characterized by an enhanced seasonal cycle of~~ simulations is by $\approx 20\%$ greater
205 than the pre-industrial ensemble mean amplitude.

Not all of the models, however, show this trend. Table 2 lists the seasonal extent amplitudes
for each model's PlioMIP simulation, in addition to the mean annual sea ice extent, ~~with severely~~
reduced sea ice during boreal summer. Three of the eight models (CCSM4, IPSLCM5A and
MRI-CGCM) simulate mid-Pliocene sea ice extent amplitudes which are smaller than the pre-industrial
210 extent amplitudes. For CCSM4 and IPSLCM5A, the differences in extent amplitude between pre-industrial
and mid-Pliocene are less than 10^6 km^2 , and represent changes of 4.1% and 6.1% respectively, so
there is no substantial change in the annual cycles of both simulations by CCSM4 and IPSLCM5A.
The increase in MRI-CGCM on the other hand is larger ($2.22 \times 10^6 \text{ km}^2$, or 13.9%).

In four of the eight models (COSMOS, GISS-E2-R, MIROC4m and NorESM-L) the mid-Pliocene
215 Arctic Ocean is ice-free at some time during the summer (August – September, Figure 6). In contrast
to this, CCSM4 and MRI-CGCM simulate minimum sea ice extents of $8.90 \times 10^6 \text{ km}^2$ and 8.26×10^6
 km^2 respectively, which both exceed the pre-industrial minimum of HadCM3 ($7.00 \times 10^6 \text{ km}^2$), with
the CCSM4 minimum also exceeding the NorESM-L pre-industrial minimum ($8.34 \times 10^6 \text{ km}^2$).
~~Consequently, there is an overlap in sea ice extents between the mid-Pliocene and pre-industrial~~
220 ~~simulations~~ This indicates the large spread in the representation of sea ice extent in the models.

~~MRI-CGCM, CCSM4 and MIROC4m simulate the highest maximum~~ For those models that
simulate summer sea ice in the mid-Pliocene ~~sea ice extents in the ensemble. Both CCSM4 and~~
the summer sea ice conditions vary strongly. Summer sea ice in HadCM3 is confined to the Arctic
basin, with concentrations that do not exceed 60%, and very low concentrations along all ice edges.
225 The summer sea ice margin in MRI-CGCM ~~also provide the highest two minimum extents, but~~
MIROC4m is one of the four models that simulates an ice-free Arctic summer. As a result, ~~,~~ on
the other hand, extends almost to the southern tip of Greenland, and a large proportion of the sea

ice extent amplitude in MIROC4m in the cover is characterized by concentrations greater than 90% (Figure 5).

230 Four of the five models with larger mid-Pliocene simulations is $\approx 64\%$ greater than the pre-industrial simulation extent amplitude (Table 2). The ensemble mean extent amplitude of the mid-Pliocene simulations is by $\approx 20\%$ greater than the pre-industrial ensemble mean amplitude, further indication of the enhanced seasonal sea ice extent cycle extent amplitudes simulated ice-free conditions for part of the summer in the mid-Pliocene simulations. Not all of the models, however, show this trend.

235 Only five models (the four with . The increase in extent amplitude ranges from a 9.4% increase in COSMOS to a 101.3% increase in NorESM-L. It might be expected that simulating a seasonally ice-free summers and HadCM3) simulate a higher mid-Pliocene sea ice extent amplitude, the remaining three models simulate a (slightly) lower annual cycle Arctic would lead to a decrease in extent amplitude, as the minimum extent has decreased as low as possible, however this is not the case. As

240 Figure 3 shows, the four models with seasonally ice-free mid-Pliocene simulations have the thinnest pre-industrial summer ice, which disappears in the mid-Pliocene simulations (Table 2) summer, whereas much of the winter sea ice has simply thinned, so there is less of a reduction in extent.

3.3.2 Sea ice thickness

Plots of the mean summer and winter mid-Pliocene Arctic sea ice thicknesses are shown in Figure

245 7. The multi-model mean annual sea ice thickness is 1.30 m, which, compared to the pre-industrial simulations, is a reduction of 1.7 m (56%). Across the ensemble, the annual mean thicknesses range from 0.44 m (NorESM-L) to 2.56 m (MRI-CGCM). The multi-model winter mean thickness is 1.77 m, 1.5 m (46%) less than the pre-industrial, whereas the summer multi-model mean thickness drops by 1.8 m (71%) to 0.74 m. Similarly to the sea ice extent, the summer sea ice thickness

250 shows a greater relative decline with respect to pre-industrial than during the winter, although the contrast is not as stark for the thickness. The individual model winter sea ice thicknesses range from 0.79 m (NorESM-L) to 2.78 m (MRI-CGCM), with the summer sea ice thicknesses between 0.3 m (NorESM-L) and 2.24 m (MRI-CGCM).

~~Spatial pattern correlations and RMSEs~~ SPCs and RMSDs between the pre-industrial and mid-

255 Pliocene simulations are shown in Figure ???. All but five of the mid-Pliocene RMSEs-RMSDs are lower than the equivalent RMSE-RMSD for the pre-industrial simulations. This trend is not seen in the ~~spatial pattern correlations~~ SPCs, where just over half (19 out of 36) of the mid-Pliocene correlations are higher than the corresponding pre-industrial correlation. These results show that the differences in thicknesses between the models are lower in the mid-Pliocene simulations, but the

260 ~~differences between~~ thickness patterns are overall no more or less similar comparable. Lower overall RMSEs-RMSDs are likely to be at least part in due to the increase in the area of ice-free ocean, and lower mean thicknesses in the mid-Pliocene simulations compared to the pre-industrial.

GISS-E2-R has the highest ~~correlation-SPC~~ with the ensemble mean (0.90), with NorESM-L the lowest (0.60). NorESM-L has correlations of less than 0.5 with two models, CCSM4 (0.49) and MRI-CGCM (0.27). As with the pre-industrial results, MRI-CGCM has the highest ~~RMSE-RMSD~~ compared to the ensemble of all the simulations (1.05 m), and the ~~RMSE-RMSD~~ of 1.46 m between MRI-CGCM and NorESM-L is the highest between any two models. The highest ~~spatial-pattern correlation-SPC~~ between two models is 0.97, between COSMOS and MIROC4m, which also have the lowest ~~RMSE-RMSD~~, at 0.11 m.

Figure ?? ~~also shows RMSEs and spatial-pattern correlations-4~~ also shows ~~RMSDs and SPCs~~ between each model's pre-industrial and mid-Pliocene runs. All but two models ~~had spatial-pattern correlations~~ ~~have SPCs~~ exceeding 0.9 between the thicknesses of both simulations, with the exceptions being GISS-E2-R (0.81) and NorESM-L (0.56). The ~~spatial-pattern correlation-SPC~~ between the ensemble means is 0.79.

3.4 Variability across the ensemble

The standard deviation (SD) of the monthly ensemble sea ice extents and thicknesses for both the pre-industrial and mid-Pliocene simulations is shown in Figure 8. In each month from December to June, the mid-Pliocene extent SD is lower than the pre-industrial extent SD. During these months, the maximum extent SD in both simulations occurs in February, and SD decreases each month from February to June. In the pre-industrial simulation, extent SD is lowest in July, following which it increases each month until to the February peak. In the mid-Pliocene simulations ~~on the other hand,~~ SD increases after June to July and then August, and reaches maximum SD in October. SD in August and October are greater than in February/March in the mid-Pliocene extent. The annual cycle of ~~of~~ pre-industrial sea ice thickness SD has a minimum in May, and maximum in September. The mid-Pliocene sea ice thickness SD annual cycle follows a similar pattern, with the lowest SD in March, and maximum in July, both two months earlier than the equivalent pre-industrial extremes.

3.5 Correlation of sea ice characteristics in the ensemble

The correlation coefficient between the mean summer sea ice extents of the pre-industrial and mid-Pliocene simulations is 0.47, compared to a correlation coefficient of 0.87 between the mean winter sea ice extents of both time slices (Figure 9 a,b). The models' annual mean sea ice extents for the two climate states show a correlation coefficient of 0.74 (not shown). Sea ice thicknesses simulated by the pre-industrial and mid-Pliocene simulations are strongly correlated in both summer and winter, with correlation coefficients of 0.82 and 0.85 respectively (Figure 9 c,d). Whilst the winter pre-industrial sea ice thickness shows a weak relationship with the mid-Pliocene winter sea ice extent (Figure 9 f), with a correlation coefficient of just 0.30, the relationship between the summer values is stronger, with a correlation coefficient of 0.81 (Figure 9 e). It should be noted that with a sample size of just 8,

only correlation coefficients greater than 0.70 are significant at the 95% level, and only those greater than 0.83 are significant at the 99% level.

The simulated mid-Pliocene sea ice extent and sea ice volume ~~appear to~~ show a stronger relationship with both surface air temperatures (SATs) and sea surface temperatures (SSTs) than the pre-industrial sea ice extent and sea ice volume (Figure 10). The correlation coefficient of the mid-Pliocene mean annual sea ice extent and the SAT, is -0.76, the correlation coefficient of the pre-industrial sea ice extent with SAT is -0.18. For SST the correlation with mid-Pliocene sea ice extents is -0.73, for pre-industrial sea ice extent the correlation coefficient is -0.26. For the summer, the mid-Pliocene sea ice extents have a correlation coefficient of -0.88 with both SAT and SST (not shown). In contrast, the pre-industrial sea ice extents have correlation coefficients of -0.27 (SAT) and -0.32 (SST) respectively (not shown). Mean annual pre-industrial SATs and SSTs have correlations with mean annual pre-industrial sea ice volume of -0.12 and -0.29 respectively. This contrasts to the respective mid-Pliocene correlation coefficients of -0.83 and -0.82. This confirms that the simulated mid-Pliocene sea ice extents and volumes have — independently from the season — a stronger negative correlation with temperatures than the simulated pre-industrial sea ice extents.

4 Discussion

4.1 Pre-industrial simulations

~~Before examining the simulations of Arctic sea ice for the mid-Pliocene, the simulations of pre-industrial sea ice cover by individual models are assessed. A comparison with observed sea ice characteristics is a suitable methodology. Ideally, we would have compared the output of the pre-industrial simulations to observations of sea ice from the same time period. However, the most spatially and temporally comprehensive observations of sea ice originate from satellites. Respective data sets date back only as far as 1979, which is more than 100 years after the time period that the pre-industrial simulations represent.~~

~~Whilst there are observations of sea ice characteristics available dating back to the early 20th century, that could have been used for the comparison, most, particularly the earliest, are ship-based observations of ice margins. These observations are only available for the spring and summer months (e.g. Thomsen (1947); Walsh and Chapman (2001)), and the sea ice extent in the remaining months must be estimated by extrapolation. Frequency and location of these observations are determined by shipping patterns, rather than by the scientific need for spatial and temporal coverage.~~

~~Due to the differences between the climate states represented by models and the chosen observations, we do not make any direct comparisons. However, all of the PlioMIP models, with the exception of COSMOS, are represented in the CMIP5 ensemble, for which historical simulations exist that can be directly compared to modern observations.~~

Shu et al. (2015) provides an assessment of the historical simulation of Arctic sea ice by the CMIP5 models for the period 1979-2005. Their results show that for the historical simulations by the 7 PlioMIP models represented in CMIP5, MRI-CGCM simulates the highest mean annual sea ice extent ($15.01 \times 10^6 \text{ km}^2$), compared to the satellite observational mean of $12.02 \times 10^6 \text{ km}^2$ for the comparable period (1979-2005). MRI-CGCM simulates the second highest PlioMIP pre-industrial mean annual sea ice extent (just $0.05 \times 10^6 \text{ km}^2$ less than MIROC4m), and the highest mid-Pliocene mean annual sea ice extent. The CMIP5 historical extent simulated by MRI-CGCM is almost 25 greater than the observational mean, showing MRI-CGCM consistently simulates Arctic sea ice extent larger than the ensemble mean.

In contrast, MIROC4m simulates a PlioMIP pre-industrial mean annual sea ice extent that is similar to the MRI-CGCM PlioMIP simulation, and represents the lowest historical mean annual sea ice extent of the CMIP5 models that are included in the PlioMIP ensemble ($10.66 \times 10^6 \text{ km}^2$, Shu et al. (2015)). The NorESM-L, which simulates both the lowest PlioMIP pre-industrial and mid-Pliocene mean annual sea ice extents, is the CMIP5 model which simulates the closest historical mean annual sea ice extent to the observations ($12.01 \times 10^6 \text{ km}^2$, just $0.01 \times 10^6 \text{ km}^2$ lower than the observations). As with the PlioMIP pre-industrial simulations, NorESM-L simulates the lowest sea ice extent amplitude of the PlioMIP models in CMIP5 (Shu et al., 2015).

In addition to the mean annual sea ice extent simulated by each model in the CMIP5 historical and PlioMIP simulations, Table 2 shows the ensemble mean annual extents for these sets of simulations. In both PlioMIP simulations, the ensemble mean annual extent is lower than the mean annual extent simulated by CCSM4, and higher than the mean annual extent simulated by HadCM3. However, in the CMIP5 historical simulations, the ensemble mean annual extent is greater than the CCSM4 mean annual extent, and lower than the mean annual extent simulated by HadCM3.

In the following, we also assess the simulated pre-industrial sea ice thickness. The simulation of Arctic sea ice thickness in the CMIP5 simulations is analysed in Stroeve et al. (2014). The correlations between the spatial patterns of Arctic sea ice thickness in the simulations (average over the years 1981-2010) and observations from Kwok et al. (2009) are less than 0.4 for all the considered PlioMIP models — with the exception of CCSM4, which has the highest spatial pattern correlation of the entire CMIP5 ensemble. For each PlioMIP model, the spatial patterns of sea ice thickness in the pre-industrial simulation resembles the thickness spatial pattern in that model's CMIP5 simulation, shown in Stroeve et al. (2014). It has been noted that the spatial pattern correlation between different ensemble simulations with the same model is significantly higher than the correlation between one model and the observations, which suggests that poor correlations are more likely explained by biases within the models, rather than by natural variability.

4.1 Mid-Pliocene simulations

Four models out of the eight-member PlioMIP ensemble (COSMOS, GISS-E2-R, MIROC4m and NorESM-L) simulate almost ice-free conditions in the mid-Pliocene summer. For those models that simulate summer sea ice in the mid-Pliocene the summer sea ice conditions vary strongly. In summer sea ice in HadCM3 is confined to the Arctic basin, with concentrations that do not exceed 60, and very low concentrations along all ice edges. The summer sea ice margin in MRI-CGCM, on the other hand, extends almost to the southern tip of Greenland, and a large proportion of the sea ice cover is characterized by concentrations greater than 90 (Figure 5).

Table 2 lists the seasonal extent amplitudes for each model's PlioMIP simulation, in addition to the mean annual sea ice extent. Three of the eight models (CCSM4, IPSLCM5A and MRI-CGCM) simulate mid-Pliocene sea ice extent amplitudes which are smaller than the pre-industrial extent amplitudes. For CCSM4 and IPSLCM5A, the differences in extent amplitude between pre-industrial and mid-Pliocene are less than 10^6 km², and represent changes of 4.1 and 6.1 respectively, and so there does not appear to have been a substantial change in the annual cycles of both simulations by CCSM4 and IPSLCM5A. The increase in MRI-CGCM on the other hand is larger (2.22×10^6 km², or 13.9). The reduction in sea ice between the extent maxima in the MRI-CGCM simulation is largely due to the loss of lower concentration, thinner sea ice from regions further south than the pre-industrial maximum sea ice margins in other models (see Figures 4 and 5). Much of the pre-industrial sea ice in the summer months in MRI-CGCM is close to 100 concentration and greater than 4 m thick. Consequently, the maximum extent reduced by a greater amount than the minimum extent.

Four of the five models with larger mid-Pliocene extent amplitudes simulated ice-free conditions for part of the summer in the mid-Pliocene. The increase in extent amplitude ranges from a 9.4 increase in COSMOS to a 101.3 increase in NorESM-L. It might be expected that simulating a seasonally ice-free mid-Pliocene sea ice cover would lead to a decrease in extent amplitude, as the minimum extent has decreased as low as possible, however this is not the case. As Figure 3 shows, the four models with seasonally ice-free mid-Pliocene simulations have the thinnest pre-industrial summer ice, which disappears in the mid-Pliocene summer, whereas much of the winter sea ice has simply thinned, so there is less of a reduction in extent.

Given the pronounced disagreement within the ensemble with regard to the nature of mid-Pliocene sea ice particularly in summer, the comparison of the different models' sea ice simulation with a reconstruction of mid-Pliocene Arctic sea ice from proxy data could prove insightful. An independent data set, like a reconstructed palaeo sea ice characteristic, may indicate which models simulate the mid-Pliocene climate more realistically. A reasonable performance of a model in simulating mid-Pliocene sea ice may also improve confidence in its prediction of future sea ice, in particular if its simulation of present day sea ice matches observations closely. If a model simulation matches well with observations/proxy reconstructions for just one climate, this may not necessarily be due to a good model performance — rather, the model may be producing “the right answers for the

wrong reasons”, such as error compensation (Massonnet et al., 2012) . However, a greater degree of confidence could be held in the predictions from a model which produces sea ice simulations that
405 closely match both modern observations in a modern simulation and proxy data-based reconstructions in a mid-Pliocene simulation, as the probability that the model compares well to the data by chance for both is reduced.

Relating proxy data to mid-Pliocene sea ice is, however, subject to limitations due to uncertainty in the proxy itself. Darby (2008) demonstrates evidence for perennial Arctic sea ice in the mid-Pliocene,
410 whilst the presence of IP_{25} , a biomarker proxy for sea ice coverage (Belt and Müller, 2013) in mid-Pliocene sediments, recovered from two boreholes in the Atlantic-Arctic gateway (located at $80.16^{\circ}N, 6.35^{\circ}E$ and $80.28^{\circ}N, 8.17^{\circ}E$, see Figure 11), implies that the maximum sea ice margin during the mid-Pliocene extended southwards beyond these two sites, but the minimum margin did not (Knies et al., 2014) . The locations of these sites are within the maximum mid-Pliocene
415 sea ice margins simulated by all of the PlioMIP models, but also within the minimum sea ice margins simulated by three of the models that simulate summer sea ice (CCSM4, IPSLCM5A and MRI-CGCM) — although the sea ice concentration at these sites is less than 50 in the CCSM4 and IPSLCM5A simulations. The extent of the sea ice minimum in HadCM3 does not reach the location of the sites analysed in Knies et al. (2014) , and so is consistent with the conclusions drawn from
420 proxy data in both the studies by Darby (2008) and Knies et al. (2014) .

A greater spatial coverage of sea ice proxy data, such as that used in Knies et al. (2014) , would improve the analysis of the simulation of sea ice by the PlioMIP models. At the moment, limited data availability does not allow for robust model-proxy comparisons. The sea ice simulated by HadCM3 has the closest agreement with the proxy data indications from Darby (2008) and Knies et al. (2014) ,
425 but greater data coverage may provoke a different conclusion.

4.1 Causes of PlioMIP ensemble variability

4.1.1 Influence of the sea ice models

The sea ice components of each model differ in resolution, representation of sea ice dynamics and thermodynamics, and formulation of various parameterisations, such as sea ice albedo. The key
430 details of each model’s sea ice component are summarised in Table 1. The models CCSM4 and NorESM-L use the same sea ice component, based on CICE4 (Hunke, 2010), although NorESM-L has a coarser model grid in the atmosphere than CCSM4, and furthermore employs a completely different ocean component (Table 1).

The sea ice dynamics of the ensemble members can be categorised into three groups. First,
435 CCSM4, NorESM-L, and MIROC4m, that all use the elastic-viscous-plastic (EVP) rheology of Hunke and Dukowicz (1997). Second, COSMOS, GISS-E2-R, and IPSLCM5A, that are based on viscous-plastic (VP) rheologies (Marsland et al., 2003; Zhang and Rothrock, 2000; Fichefet and

Morales Maqueda, 1999). Third, HadCM3 and MRI-CGCM, that do not consider any type of sea ice rheology, the sea ice following simple free drift dynamics (Cattle and Crossley, 1995; Mellor and
440 Kantha, 1989). In PlioMIP, there does not appear to be any link between the type of dynamics of the sea ice components and the simulated sea ice extents — MRI-CGCM and MIROC4m produce the two highest annual means for pre-industrial whilst having very different sea ice dynamics. The three models that produce the lowest pre-industrial extents, i.e. NorESM-L, IPSLCM5A, and HadCM3, employ different rheologies — EVP, VP and no rheology respectively.

445 Most of the models use a leads parameterisation in their sea ice thermodynamics component, with only CCSM4 and NorESM-L employing explicit melt pond schemes. The models HadCM3 and COSMOS both use the leads parameterisation based on Hibler (1979). The models HadCM3, MIROC4m and MRI-CGCM all utilise the 'zero-layer' model developed by Semtner (1976). Similarly to the considered sea ice dynamics, there is no clear influence of the thermodynamics schemes
450 used in the models on the simulated pre-industrial sea ice extent.

The simulation of Arctic sea ice by means of GCMs has been demonstrated to be very sensitive to the parameterisation of sea ice albedo. This has been observed in the case of variations of albedo in different models (Hodson et al., 2013), and adjusting the parameterisation in one specific model (Howell et al., 2014). Hill et al. (2014) show that clear sky albedo is the dominant factor in high
455 latitude warming in the PlioMIP ensemble. The four models that display the highest warming effect from the clear sky albedo are those four models that simulate an ice-free mid-Pliocene summer (COSMOS, GISS-E2-R, MIROC4m, and NorESM-L). The NorESM-L shows the largest warming due to clear sky albedo, CCSM4 on the other hand shows the smallest clear sky albedo effect. Both NorESM-L and CCSM4 use the same sea ice component, based on CICE4 (Hunke and Lipscomb,
460 2008). This sea ice model employs a shortwave radiative transfer scheme to internally simulate the sea ice albedo, and by that produce a more physically based parameterisation (Holland et al., 2011).

Yet, it appears that the performance of this albedo scheme is very sensitive to differences in other components of the climate models: NorESM-L (that shows a large contribution of clear sky albedo) uses the same atmosphere component as CCSM4 (low contribution of clear sky albedo), albeit at a
465 lower resolution version in the PlioMIP experiment, but it employs a different ocean component, that also has a lower resolution than the ocean component used in CCSM4. The contrast in the contribution of clear sky albedo to high latitude warming between NorESM-L and CCSM4 is reflected in the large difference in their simulations of summer mid-Pliocene sea ice. One cause is certainly the nature of the sea-ice albedo feedback mechanism (Curry et al., 1995). Reduced albedo at high lati-
470 tudes can be both a cause of and a result of a reduced sea ice extent. Models with parameterisations with a lower sea ice albedo minimum have therefore a greater potential to amplify the warming that originates from other sources in simulations of the mid-Pliocene, such as greenhouse gas emissivity. The low sea ice albedo assumed in NorESM-L is a likely explanation for the low sea ice extents it simulates (Figures 2 and 6), both in mid-Pliocene and pre-industrial simulations.

475 Second to NorESM-L, for MIROC4m clear sky albedo has the highest contribution to high latitude
warming. In MIROC4m there is a fixed albedo of 0.5 for bare sea ice, with higher albedo for snow-
covered sea ice, that furthermore varies according to ambient surface air temperature (K-1 Model
Developers, 2004). Of the six models that do not use a radiative transfer scheme to internally simulate
sea ice albedo (those except NorESM-L and CCSM4), only GISS-E2-R has an albedo minimum
480 lower than 0.5. Yet, this model allows the albedo to vary between 0.44 and 0.84 (Schmidt et al.,
2006). All other models also allow the sea ice albedo to vary, and consequently MIROC4m has a
lower overall albedo. This may help to explain the ability of MIROC4m to simulate an ice-free mid-
Pliocene summer, despite simulating one of the highest winter sea ice extents for both pre-industrial
and mid-Pliocene.

485 As the parameterisation of sea ice albedo is kept unchanged between pre-industrial and mid-
Pliocene simulations, differences in the parameterisation between the models should have similar
effects in both simulations. However, if there is a temperature threshold above which the ice-albedo
feedback becomes more dominant in some of the models, then this could explain the different influ-
ence of the sea ice parameterisation on pre-industrial and mid-Pliocene simulations.

490 General circulation models are tuned to best reproduce modern day climate conditions, and pa-
rameterisations are based on modern observations (Hunke, 2010; Mauritsen et al., 2012). When sim-
ulating the climate of time periods with different climate states, such as the mid-Pliocene, models
that are tuned towards present day conditions may be biased in some regions. However, it is disputed
to which extent the adjustment of parameters, such as sea ice albedo, within the limits of observa-
495 tional uncertainties can affect the overall sea ice cover and compensate for other shortcomings in the
model (Eisenman et al., 2007; DeWeaver et al., 2008; Eisenman et al., 2008).

4.1.2 Influence of the control simulation

Massonnet et al. (2012) describe the characteristics of Arctic sea ice simulated by the CMIP5 ensem-
ble for the time period from 1979-2010 as being related in a 'complicated manner' to the simulated
500 future change in September Arctic sea ice extent. Figure 9 demonstrates, based on correlation values,
that some combinations of sea ice characteristics in the pre-industrial and mid-Pliocene simulations
are much stronger related to each other than others. In section 3.2 it was highlighted that the differ-
ences in the PlioMIP models' simulation of sea ice for 1979-2005 in CMIP5 are not consistent with
the differences in pre-industrial or mid-Pliocene simulations in the PlioMIP ensemble.

505 All of the models that simulate thinner pre-industrial summer sea ice than the ensemble mean also
simulate ice-free conditions during the mid-Pliocene summer, with the exception of HadCM3. Hol-
land and Bitz (2003) demonstrate that the thickness of sea ice in control simulations has a stronger
influence on the climate state of the Northern Hemisphere polar region in simulations of future cli-
mates than sea ice extent. Massonnet et al. (2012) find that those CMIP5 models that predict an
510 earlier disappearance of September Arctic sea ice generally have a smaller initial September sea ice

extent. In PlioMIP, mean summer pre-industrial sea ice thicknesses have correlation coefficients of 0.81 and 0.82 with mean summer mid-Pliocene sea ice extents and thicknesses, respectively. Mean summer pre-industrial sea ice extents on the other hand show weaker correlations with mean summer mid-Pliocene sea ice extents and thicknesses, with respective correlation coefficients of 0.47
515 and 0.51. The relatively thin pre-industrial summer sea ice simulated in PlioMIP by COSMOS, GISS-E2-R, MIROC4m and NorESM-L therefore appears to be an important factor for the ability of those models to simulate an ice-free mid-Pliocene summer. An exception is HadCM3, that simulates perennial sea ice in the mid-Pliocene, despite simulating relatively thin (within the PlioMIP ensemble) pre-industrial sea ice.

520 **4.1.3 Influence of atmosphere and ocean on the sea ice simulation**

In the mid-Pliocene simulations, the correlation coefficient between Arctic surface temperatures and simulated sea ice extent is much higher than the corresponding correlation coefficient in the pre-industrial simulations (Figure 10 a,b). Pre-industrial sea ice is thicker than mid-Pliocene sea ice, which could explain the lower sensitivity of the pre-industrial sea ice extent to surface temperatures.
525 However, similar differences in correlation strength between the pre-industrial and mid-Pliocene simulations are also seen for mean sea ice volume (Figure 10, c,d), so there is no strong relationship between warmer pre-industrial simulations and those with less total ice.

In the pre-industrial simulations, much of the ocean north of 60°N is fully covered with sea ice, so all SSTs will be -1.8°C. The uniformity of the SSTs in this region could be a plausible explanation
530 for the weak correlation between the overall Arctic sea ice extents and SSTs north of 60°N in the pre-industrial simulations of the PlioMIP ensemble. The reduced sea ice coverage in the mid-Pliocene simulations, particularly during the summer months, enables on the other hand a greater range of possible SST values. This is potentially the reason for a much stronger correlation with the simulated mid-Pliocene sea ice extents (Figure 10). In the models, the presence of ice in a grid box, even at low
535 concentrations, restricts the warming in the ocean. Larger parts of the ocean are ice-free for longer periods in the year in the mid-Pliocene simulations than in the pre-industrial simulations, meaning longer periods in the mid-Pliocene simulations where the ocean can warm. This will in turn affect the warming of the atmosphere in the models, and so is a possible reason for better correlation between sea ice extent and surface temperatures in the mid-Pliocene simulations.

540 In addition to SATs and SSTs, there are of course other atmospheric and oceanic influences on the simulation of Arctic sea ice. The Atlantic Meridional Overturning Circulation (AMOC) contributes significantly to poleward oceanic heat transport and has been shown to have a strong impact on Arctic sea ice (e.g. Mahajan et al. (2011); Day et al. (2012); Miles et al. (2014)). Zhang et al. (2013b) analyse the simulation of the AMOC in both pre-industrial and mid-Pliocene simulations of
545 the PlioMIP ensemble and find that there is little difference between each model's pre-industrial and mid-Pliocene AMOC simulation. There is no consistent change in northward ocean heat transport,

with half the models simulating a slight (less than 10%) increase, and half the models simulating a slight decrease (less than -15%). Of the models which simulate increased northward ocean heat transport (COSMOS, GISS-E2-R, IPSL-CM5A and MRI-CGCM), only two (COSMOS and GISS-E2-R) simulate an ice-free mid-Pliocene summer. This suggests that the influence of AMOC and northward oceanic heat transport on the ensemble variability of sea ice in the mid-Pliocene simulation of PlioMIP is not the most important factor.

An analysis of multi-decadal variability influence on Arctic sea ice extent in selected CMIP3 simulations (covering 1953-2010) by Day et al. (2012) showed a significant correlation between Arctic sea ice extents and Atlantic Multi-decadal Oscillation (AMO) indices. Kwok (2000) and Parkinson (2008) demonstrate evidence of the North Atlantic Oscillation (NAO) on Arctic sea ice. Table 3 shows annual and decadal correlations between Arctic sea ice extent and AMO and NAO indices for simulations from three PlioMIP models (CCSM4, HadCM3 and NorESM-L), ~~for which sufficiently long time series were available to perform the calculations.~~

All three models show a small but significant (at 90% level) correlation between the pre-industrial annual Arctic sea ice extents and the NAO indices. The correlation coefficients at the decadal time scale are increased for both HadCM3 and NorESM-L, but are not significant for any of the models. None of the correlations between mid-Pliocene Arctic sea ice extents and NAO indices are significant at the 90% level. The correlations between pre-industrial Arctic sea ice extents and AMO indices are all not significant at the 90% level. For the mid-Pliocene simulations, only the correlation between the annual Arctic sea ice extents and AMO indices from the CCSM4 simulations is significant at the 90% level.

There is no significant correlation between decadal sea ice extents and NAO/AMO indices in the three models shown, and so it is unlikely that differences in the mean sea ice extents (representing averages representing between 30 and 200 years worth of climatology) between different models and simulations can be explained by different influences of these variability indices. To more thoroughly investigate this would require much longer timeseries from all the modelling groups, which ~~is~~ are not available. A comprehensive analysis of the relationships between variability indices and sea ice in the PlioMIP simulations is beyond the scope of this paper.

Patterns of ice thicknesses are strongly influenced by the motion of sea ice in the models. In each model, the equations used to determine sea ice motion account for stresses on the ice from surface winds and ocean currents, with the exceptions of HadCM3, which does not take surface winds into account (Gordon et al., 2000), and MRI-CGCM, where the ocean currents are not taken into account in determining ice motion (Mellor and Kantha, 1989).

Figure 12 shows the mean annual 10 m surface winds ~~for the COSMOS~~ and sea ice thicknesses for the IPSL-CM5A and MIROC4m ~~mid-Pliocene simulations, where~~ simulations. In MIROC4m, the dominant wind direction between 90°E and 180°E over the Arctic basin is towards the northern coast of Eastern Siberia, where a build up of thicker ice is present. Similarly, in ~~the~~ IPSL-CM5A

pre-industrial simulation (Figure 12), the dominant wind direction is towards the north of Greenland and the Canadian Arctic Archipelago where the thickest ice is. In the NorESM-L pre-industrial simulation (shown in Figure 12), the thickest ice is present between Greenland and the Canadian Arctic Archipelago. In the corresponding mid-Pliocene simulation, the mean annual 10 m surface winds over this region are weaker, and in a western direction, rather than north-west, towards the Canadian Arctic Archipelago, and so conditions are less conducive for a build up of thicker ice.

585
590 Mean annual 10 m winds and sea ice thicknesses for all simulations (excluding CCSM4, for which 10m-10 m winds are not an output) are included in the supplementary information.

In HadCM3, the ocean surface currents form a vortex in part of the Arctic basin (Beaufort Gyre), where the thickest sea ice is present in both simulations (see Figure 13). Given that the sea ice motion is entirely determined by the surface ocean current, its influence on the spatial pattern of sea ice thickness is clear. If sea ice motion were instead determined by surface wind stresses in addition to the ocean currents (which do not have the same patterns in HadCM3), this should result in a different configuration of sea ice in the Arctic basin, and would likely affect the location of the sea ice margins simulated by the model. Mean annual surface ocean currents and sea ice thicknesses for all simulations are included in the supplementary information.

595

600 Understanding the more precise influences of winds and ocean currents on the modelled sea ice, and the causes of differences between models, as well as different simulations with the same model, would require a far more extensive analysis. Differences in seasonal, as well as annual patterns, alongside atmospheric circulations at higher levels, may be explored in further work.

4.2 Sea ice proxy data

605 Given the large spread within the ensemble with regard to the nature of mid-Pliocene sea ice, the comparison of the different models' sea ice simulation with a reconstruction of mid-Pliocene Arctic sea ice from proxy data could prove insightful. The recent development of organic biomarkers proxies such as IP₂₅ to reconstruct past sea ice presence (e.g. Knies et al. (2014)), may indicate which models simulate the mid-Pliocene climate more realistically. A reasonable performance of a

610 model in simulating mid-Pliocene sea ice may also improve confidence in its prediction of future sea ice, in particular if its simulation of present day sea ice matches observations closely. If a model simulation matches well with observations/proxy reconstructions for just one climate, this may not necessarily be due to a good model performance — rather, the model may be producing “the right answers for the wrong reasons”, such as error compensation (Massonnet et al., 2012) . However, a

615 greater degree of confidence could be held in the predictions from a model which produces sea ice simulations that closely match both modern observations in a modern simulation and proxy data-based reconstructions in a mid-Pliocene simulation, as the probability that the model compares well to the data by chance for both is reduced.

620 Relating proxy data to mid-Pliocene sea ice is, however, subject to limitations due to uncertainty in the proxy itself. Darby (2008) demonstrates evidence for perennial Arctic sea ice in the mid-Pliocene, whilst the presence of IP₂₅, a biomarker proxy for sea ice coverage (Belt and Müller, 2013) in mid-Pliocene sediments, recovered from two boreholes in the Atlantic-Arctic gateway (located at 80.16°N, 6.35°E and 80.28°N, 8.17°E, see Figure 11), implies that the maximum sea ice margin during the mid-Pliocene extended southwards beyond these two sites, but the minimum margin
625 did not (Knies et al., 2014). The locations of these sites are within the maximum mid-Pliocene sea ice margins simulated by all of the PlioMIP models, but also within the minimum sea ice margins simulated by three of the models that simulate summer sea ice (CCSM4, IPSLCM5A and MRI-CGCM) — although the sea ice concentration at these sites is less than 50% in the CCSM4 and IPSLCM5A simulations. The extent of the sea ice minimum in HadCM3 does not reach the location
630 of the sites analysed in Knies et al. (2014), and so is consistent with the conclusions drawn from proxy data in both the studies by Darby (2008) and Knies et al. (2014). A greater spatial coverage of sea ice proxy data, such as that used in Knies et al. (2014), would improve the analysis of the simulation of sea ice by the PlioMIP models. At the moment, limited data availability does not allow for robust model-proxy comparisons.

635 **5 Conclusions**

We have presented a detailed analysis of the simulation of Arctic sea ice in the PlioMIP model ensemble, for both pre-industrial control and mid-Pliocene simulations. The sea ice in the mid-Pliocene simulations is overall less extensive and thinner than the pre-industrial sea ice, with a 33% decrease in mean annual sea ice extent for the ensemble mean, and a 56% reduction in the ensemble
640 mean annual sea ice thickness. The changes in the mid-Pliocene, relative to the pre-industrial, are largest during the summer months, both in absolute and relative terms, and for both sea ice extent and sea ice thickness.

The simulated mid-Pliocene sea ice extents are strongly negatively correlated with the Arctic temperatures. In contrast, there is only a weak correlation between pre-industrial sea ice extents and
645 temperature. Hill et al. (2014) identified clear sky albedo as the dominant driver of high latitude warming in the mid-Pliocene simulations of PlioMIP, particularly in those models that simulate an ice-free mid-Pliocene summer. Sea ice-albedo feedbacks may contribute to the stronger relationship between surface temperatures and sea ice in the mid-Pliocene simulations, as the feedback mechanism enhances the warming that originates from increased greenhouse gas concentrations. The effect
650 of the sea ice-albedo feedback does not appear to be similarly pronounced in the pre-industrial simulations. If it is the case that some models see an enhanced ice-albedo feedback in warmer climates, then this is likely to affect those models' prediction of future Arctic sea ice change.

The HadCM3 is the only model that simulates both perennial mid-Pliocene Arctic sea ice and a minimum sea ice extent that is completely located north of the location of the two sites studied in Knies et al. (2014), located at 80.16°N, 6.35°E and 80.28°N, 8.17°E, where IP₂₅ proxy data indicates the presence of a sea ice margin in the mid-Pliocene. ~~HadCM3 therefore produces the mid-Pliocene simulation that is in best agreement with the proxy inferences from Darby (2008) and Knies et al. (2014), i.e. presence of perennial sea ice and a relatively northern location of summer sea ice during the mid-Pliocene. Yet, it should be noted that the~~ However, this proxy evidence is sparse, ~~with available data~~ originating from just two sites in the same region. ~~Furthermore, the understanding of mid-Pliocene sea ice is still too low to have confidence in this simulation, particularly considering that the HadCM3 CMIP5 simulation is not closest to the observations, and the model simulates an unrealistic sea ice distribution, in part due to the sea ice motion having no influence from the surface winds. Of course, if the~~ If the proxy studies indicating seasonal mid-Pliocene Arctic sea ice (e.g. Cronin et al. (1993); Moran et al. (2006); Polyak et al. (2010)) are correct, then the mid-Pliocene Arctic sea ice in COSMOS, GISS-E2-R, MIROC4m and NorESM-L models concur with the data indication.

Given the limited amount of suitable proxy data, we are currently not able to make firm judgments with respect to a selection of models that simulate a more accurate mid-Pliocene Arctic sea ice cover if compared to the geologic record. The availability of additional proxy data may enable such conclusion in the future, could help to identify strengths and weaknesses in the different models' simulations of sea ice, as well as gauge confidence in their predictions of future sea ice.

However, as discussed in section 4.1.3, there are numerous atmospheric and oceanic factors that influence the simulation of Arctic sea ice. As highlighted by Massonnet et al. (2012), a model can simulate the 'right' results for the wrong reasons, perhaps due to error compensation. This does not mean that the analysis of sea ice simulations for past climates, such as the mid-Pliocene, is not valuable and justified, but that it is important to highlight that the forcings behind the sea ice simulation have to be better understood. Variability modes, such as NAO or AMO, whilst shown to have influence on sea ice extent from an annual viewpoint, do not appear to exert significant influence over the mean sea ice state on a decadal time scale. The models' representation of sea ice motion, and by extension ocean currents and surface winds, are an important influence on the distribution of sea ice, and worthy of a more detailed study. Future studies must particularly aim at quantifying the contribution of the various forcings on the sea ice in warmer climates.

Acknowledgements. F.W. Howell acknowledges NERC for the provision of a doctoral training grant. A.M. Haywood acknowledges that the research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 278636. B. Otto-Bliesner and N. Rosenbloom recognise that NCAR is sponsored by the US National Science Foundation (NSF) and computing resources were provided by the Climate Simulation Lab-

690 oratory at NCAR's Computational and Information Systems Laboratory (CISL), sponsored by the NSF and
other agencies. C. Stepanek acknowledges financial support from the Helmholtz Graduate School for Polar and
Marine Research and from the Helmholtz Climate Initiative REKLIM. Funding for M.A. Chandler has been
provided by NSF Grant ATM0323516 and NASA Grant NNX10AU63A. F. Bragg acknowledges NERC grant
NE/H006273/1. W.-L. Chan and A. Abe-Ouchi acknowledge financial support from the Japan Society for the
695 Promotion of Science and computing resources at the Earth Simulator Center, JAMSTEC. Y. Kamae acknowl-
edges S. Yukimoto, O. Arakawa, and A. Kitoh in Meteorological Research Institute in Japan for providing
source code of the MRI-CGCM model. Z. Zhang would like to thank M. Bentsen, J. Tjiputra, and I. Bethke
from Bjerknæs Center for Climate Research for the contribution to the development of NorESM-L.

References

- Arzel, O., Fichefet, T., and Goosse, H.: Sea ice evolution over the 20th and 21st centuries as simulated by
700 current AOGCMs, *Ocean Modelling*, 12, 401 – 415, doi:10.1016/j.ocemod.2005.08.002, 2006.
- Belt, S. and Müller, J.: The Arctic sea ice biomarker IP₂₅: a review of current understanding, recommendations
for future research and applications in palaeo sea ice reconstructions, *Quaternary Science Reviews*, 79, 9–25,
2013.
- Berger, M., Brandefelt, J., and Nilsson, J.: The sensitivity of the Arctic sea ice to orbitally induced insolation
705 changes: a study of the mid-Holocene Paleoclimate Modelling Intercomparison Project 2 and 3 simulations,
Clim. Past, 9, 969–982, doi:10.5194/cp-9-969-2013, 2013.
- Blanchard-Wrigglesworth, E. and Bitz, C.: Characteristics of Arctic Sea-Ice Thickness Variability in GCMs, *J.*
Climate, 27, 8244–8258, doi:10.1175/JCLI-D-14-00345.1, 2014.
- Boé, J. L., Hall, A., and Qu, X.: September sea-ice cover in the Arctic Ocean projected to vanish by 2100,
710 *Nature Geoscience*, 2, 341–343, doi:10.1038/ngeo467, 2009.
- Bragg, F. J., Lunt, D. J., and Haywood, A. M.: Mid-Pliocene climate modelled using the UK Hadley Centre
Model: PlioMIP Experiments 1 and 2, *Geoscientific Model Development*, 5, 1109–1125, doi:10.5194/gmd-
5-1109-2012, 2012.
- Cattle, H. and Crossley, J.: Modeling Arctic climate change, *Philosophical Transactions of the Royal Society*
715 *A-Mathematical, Physical and Engineering Sciences*, 352, 201–213, doi:10.1098/rsta.1995.0064, 1995.
- Chan, W. L., Abe-Ouchi, A., and Ohgaito, R.: Simulating the mid-Pliocene climate with the MIROC general
circulation model: experimental design and initial results, *Geoscientific Model Development*, 4, 1035–1049,
doi:10.5194/gmd-4-1035-2011, 2011.
- Chandler, M. A., Sohl, L. E., Jonas, J. A., Dowsett, H. J., and Kelley, M.: Simulations of the mid-Pliocene
720 Warm Period using two versions of the NASA/GISS ModelE2-R Coupled Model, *Geoscientific Model De-
velopment*, 6, 517–531, doi:10.5194/gmd-6-517-2013, 2013.
- Contoux, C., Ramstein, G., and Jost, A.: Modelling the mid-Pliocene Warm Period climate with the IPSL
coupled model and its atmospheric component LMDZ5A, *Geoscientific Model Development*, 5, 903–917,
doi:10.5194/gmd-5-903-2012, 2012.
- 725 Cronin, T. M., Whatley, R., Wood, A., Tsukagoshi, A., Ikeya, N., Brouwers, E. M., and Briggs, W. M.: Mi-
crofaunal evidence for elevated Pliocene temperatures in the Arctic ocean, *Paleoceanography*, 8, 161–173,
1993.
- Curry, J. A., Schramm, J. L., and Ebert, E. E.: Sea ice-albedo climate feedback mechanism, *Journal of Climate*,
8, 240–247, doi:10.1175/1520-0442, 1995.
- 730 Darby, D. A.: Arctic perennial ice cover over the last 14 million years, *Paleoceanography*, 23, PA1S07,
doi:10.1029/2007pa001479, 2008.
- Day, J. J., Hargreaves, J. C., Annan, J. D., and Abe-Ouchi, A.: Sources of multi-decadal variability in Arctic
sea ice extent, *Environmental Research Letters*, 7, 034 011, 2012.
- DeWeaver, E., Hunke, E., and Holland, M.: Comment on “On the reliability of simulated Arctic sea ice in
735 global climate models” by I. Eisenman, N. Untersteiner, and J. S. Wettlaufer, *Geophysical Research Letters*,
L04501, doi:10.1029/2007GL031325, 2008.

- Dowsett, H. J., Robinson, M. M., Haywood, A. M., Salzmann, U., Hill, D. J., Sohl, L., Chandler, M. A., Williams, M., Foley, K., and Stoll, D.: The PRISM3D paleoenvironmental reconstruction, *Stratigraphy*, 7, 123–139, 2010.
- 740 Eisenman, I., Untersteiner, N., and Wettlaufer, J. S.: On the reliability of simulated Arctic sea ice in global climate models, *Geophysical Research Letters*, 34, L10 501, doi:10.1029/2007gl029914, 2007.
- Eisenman, I., Untersteiner, N., and Wettlaufer, J. S.: Reply to comment by E. T. DeWeaver et al. on “On the reliability of simulated Arctic sea ice in global climate models”, *Geophysical Research Letters*, 35, n/a–n/a, doi:10.1029/2007GL032173, 104502, 2008.
- 745 Fichefet, T. and Morales Maqueda, M. A.: Modelling the influence of snow accumulation and snow-ice formation on the seasonal cycle of the Antarctic sea-ice cover, *Climate Dynamics*, 15, 251–268, 1999.
- Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., Mitchell, J. F. B., and Wood, R. A.: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments, *Climate Dynamics*, 16, 147–168, 2000.
- 750 Haywood, A. M., Dowsett, H. J., Otto-Bliesner, B., Chandler, M. A., Dolan, A. M., Hill, D. J., Lunt, D. J., Robinson, M. M., Rosenbloom, N., Salzmann, U., and Sohl, L. E.: Pliocene Model Intercomparison Project (PlioMIP): experimental design and boundary conditions (Experiment 1), *Geoscientific Model Development*, 3, 227–242, doi: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. L., and Chandler, M. A.: Pliocene Model Intercomparison Project (PlioMIP): experimental design and boundary conditions (Experiment 2), *Geosci. Model Dev.*, 4, 571–577, doi:10.5194/gmd-4-571-2011, 2011.
- 755 Haywood, A. M., Hill, D. J., Dolan, A. M., Otto-Bliesner, B. L., Bragg, F. J., 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, S. Z.: Large-scale features of Pliocene climate: results from the Pliocene Model Intercomparison Project, *Clim. Past*, 9, 191–209, doi:10.5194/cp-9-191-2013, 2013.
- 760 Hibler, W. D.: A dynamic-thermodynamic sea ice model, *Journal of Physical Oceanography*, 9, 815–846, doi:10.1175/1520-0485, 1979.
- Hill, D. J., Haywood, A. M., Lunt, D. J., Hunter, S. J., Bragg, F. J., Contoux, C., Stepanek, C., Sohl, L., Rosenbloom, N. A., Chan, W. L., Kamae, Y., Zhang, Z., Abe-Ouchi, A., Chandler, M. A., Jost, A., Lohmann, G., Otto-Bliesner, B. L., Ramstein, G., and Ueda, H.: Evaluating the dominant components of warming in Pliocene climate simulations, *Climate of the Past*, 10, 79–90, doi:10.5194/cp-10-79-2014, 2014.
- 765 Hodson, D., Keeley, S., West, A., Ridley, J., Hawkins, E., and Hewitt, H.: Identifying uncertainties in Arctic climate change projections, *Climate Dynamics*, 40, 2849–2865, doi:10.1007/s00382-012-1512-z, 2013.
- 770 Holland, M. M. and Bitz, C. M.: Polar amplification of climate change in coupled models, *Climate Dynamics*, 21, 221–232, doi:10.1007/s00382-003-0332-6, 2003.
- Holland, M. M. and Stroeve, J.: Changing seasonal sea ice predictor relationships in a changing Arctic climate, *Geophysical Research Letters*, 38, n/a–n/a, doi:10.1029/2011GL049303, 118501, 2011.
- 775 Holland, M. M., Bailey, D. A., Briegleb, B. P., Light, B., and Hunke, E. C.: Improved Sea Ice Shortwave Radiation Physics in CCSM4: The Impact of Melt Ponds and Aerosols on Arctic Sea Ice, *Journal of Climate*, 25, 1413–1430, doi:10.1175/jcli-d-11-00078.1, 2011.

- Howell, F. W., Haywood, A. M., Dolan, A. M., Dowsett, H. J., Francis, J. E., Hill, D. J., Pickering, S. J., Pope, J. O., Salzmann, U., and Wade, B. S.: Can uncertainties in sea ice albedo reconcile patterns of data-model discord for the Pliocene and 20th/21st centuries?, *Geophysical Research Letters*, 41, 2011–2018, doi:10.1002/2013gl058872, 2014.
- 780 Hunke, E. C.: Thickness sensitivities in the CICE sea ice model, *Ocean Modelling*, 34, 137–149, 2010.
- Hunke, E. C. and Dukowicz, J. K.: An elastic-viscous-plastic model for sea ice dynamics, *Journal of Physical Oceanography*, 27, 1849–1867, doi:10.1175/1520-0485, 1997.
- Hunke, E. C. and Lipscomb, W. H.: CICE: The Los Alamos sea ice model user’s manual, version 4.0., Tech. Rep. LA-CC-06-012, Los Alamos, New Mexico., p. 76, 2008.
- 785 Johnson, M., Gaffigan, S., Hunke, E., and Gerdes, R.: A comparison of Arctic Ocean sea ice concentration among the coordinated AOMIP model experiments, *Journal of Geophysical Research: Oceans*, 112, C04S11, doi:10.1029/2006JC003690, 2007.
- Johnson, M., Proshutinsky, A., Aksenov, Y., Nguyen, A. T., Lindsay, R., Haas, C., Zhang, J., Diansky, N., Kwok, R., Maslowski, W., Häkkinen, S., Ashik, I., and de Cuevas, B.: Evaluation of Arctic sea ice thickness simulated by Arctic Ocean Model Intercomparison Project models, *Journal of Geophysical Research: Oceans*, 117, C00D13, doi:10.1029/2011JC007257, 2012.
- 790 K-1 Model Developers: K1 Coupled Model (MIROC) Description: K1 Technical Report 1, edited by: Hasumi, H. and Emori, S., 34 pp., Center for Climate System Research, University of Tokyo, 2004.
- 795 Kamae, Y. and Ueda, H.: Mid-Pliocene global climate simulation with MRI-CGCM2.3: set-up and initial results of PlioMIP Experiments 1 and 2, *Geoscientific Model Development*, 5, 793–808, doi:10.5194/gmd-5-793-2012, 2012.
- Knies, J., Cabedo-Sanz, P., Belt, S. T., Baranwal, S., Fietz, S., and Rosell-Melé, A.: The emergence of modern sea ice cover in the Arctic Ocean, *Nat. Commun.*, 5:5608, doi:10.1038/ncomms6608, 2014.
- 800 Kwok, R.: Recent changes in Arctic Ocean sea ice motion associated with the North Atlantic Oscillation, *Geophysical Research Letters*, 27, 775–778, doi:10.1029/1999GL002382, 2000.
- Kwok, R., Cunningham, G. F., Wensnahan, M., Rigor, I., Zwally, H. J., and Yi, D.: Thinning and volume loss of the Arctic Ocean sea ice cover: 2003-2008, *Journal of Geophysical Research*, 114, C07005, doi:10.1029/2009jc005312, 2009.
- 805 Liu, J., Schmidt, G. A., Martinson, D., Rind, D. H., Russell, G. L., and Yuan, X.: Sensitivity of sea ice to physical parameterizations in the GISS global climate model, *Journal of Geophysical Research*, 108, 3053, doi:10.1029/2001JC001167, 2003.
- Mahajan, S., Zhang, R., and Delworth, T.: Impact of the Atlantic Meridional Overturning Circulation (AMOC) on Arctic Surface Air Temperature and Sea Ice Variability, *Journal of Climate*, 24, 6573–6581, 2011.
- 810 Marsland, S. J., Haak, H., Jungclaus, J. H., Latif, M., and Röske, F.: The Max-Planck-Institute global ocean/sea ice model with orthogonal curvilinear coordinates, *Ocean Modelling*, 5, 91–127, 2003.
- Massonnet, F., Fichefet, T., Goosse, H., Bitz, C. M., Philippon-Berthier, G., Holland, M. M., and Barriat, P.-Y.: Constraining projections of summer Arctic sea ice, *The Cryosphere*, 6, 1383–1394, doi:10.5194/tc-6-1383-2012, 2012.
- 815 Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Haak, H., Jungclaus, J., Klocke, D., Matei, D., Mikolajewicz, U., Notz, D., Pincus, R., Schmidt, H., and Tomassini, L.: Tuning the climate of

- a global model, *Journal of Advances in Modeling Earth Systems*, 4, M00A01, doi:10.1029/2012MS000154, 2012.
- 820 Mellor, G. L. and Kantha, L.: An ice-ocean coupled model, *Journal of Geophysical Research-Oceans*, 94, 10 937–10954, doi:10.1029/JC094iC08p10937, 1989.
- Miles, M. W., Divine, D. V., Furevik, T., Jansen, E., Moros, M., and Ogilvie, A. E. J.: A signal of persistent Atlantic multidecadal variability in Arctic sea ice, *Geophysical Research Letters*, 41, 463–469, doi:10.1002/2013GL058084, 2013GL058084, 2014.
- 825 Moran, K., Backman, J., Brinkhuis, H., Clemens, S. C., Cronin, T., Dickens, G. R., Eynaud, F., Gattacceca, J., Jakobsson, M., Jordan, R. W., Kaminski, M., King, J., Koc, N., Krylov, A., Martinez, N., Matthiessen, J., McInroy, D., Moore, T. C., Onodera, J., O'Regan, M., Pälike, H., Rea, B., Rio, D., Sakamoto, T., Smith, D. C., Stein, R., St John, K., Suto, I., Suzuki, N., Takahashi, K., Watanabe, M., Yamamoto, M., Farrel, J., Frank, M., Kubik, P., Jokat, W., and Kristoffersen, Y.: The Cenozoic palaeoenvironment of the Arctic Ocean, *Nature*, 441, 601–605, doi:10.1038/nature04800, 2006.
- 830 Pagani, M., Liu, Z., LaRiviere, J., and Ravelo, A. C.: High Earth-system climate sensitivity determined from Pliocene carbon dioxide concentrations, *Nature Geoscience*, 3, 27–30, doi:10.1038/geo724, 2010.
- Parkinson, C. L.: Recent trend reversals in arctic sea ice extents: possible connections to the north Atlantic oscillation, *Polar Geography*, 31, 3–14, doi:10.1080/10889370802175945, 2008.
- Parkinson, C. L. and Comiso, J. C.: On the 2012 record low Arctic sea ice cover: Combined impact of pre-conditioning and an August storm, *Geophysical Research Letters*, 40, 1356–1361, doi:10.1002/grl.50349, 835 2013.
- Parkinson, C. L., Vinnikov, K. Y., and Cavalieri, D. J.: Evaluation of the simulation of the annual cycle of Arctic and Antarctic sea ice coverages by 11 major global climate models, *Journal of Geophysical Research: Oceans*, 111, C07 012, doi:10.1029/2005JC003408, 2006.
- 840 Polyak, L., Alley, R. B., Andrews, J. T., Brigham-Grette, J., Cronin, T. M., Darby, D. A., Dyke, A. S., Fitzpatrick, J. J., Funder, S., Holland, M. M., Jennings, A. E., Miller, G. H., O'Regan, M., Savelle, J., Serreze, M., St John, K., White, J. W. C., and Wolff, E.: History of sea ice in the Arctic, *Quaternary Science Reviews*, 29, 1757–1778, doi:10.1016/j.quascirev.2010.02.010, 2010.
- Rosenbloom, N. A., Otto-Bliesner, B. L., Brady, E. C., and Lawrence, P. J.: Simulating the mid-Pliocene Warm 845 Period with the CCSM4 model, *Geoscientific Model Development*, 6, 549–561, doi:10.5194/gmd-6-549-2013, 2013.
- Schmidt, G. A., Reto, R., Hansen, J. E., Aleinov, I., Bell, N., Bauer, M., Bauer, S., Cairns, B., Canuto, V., Cheng, Y., Del Genio, A., Faluvegi, G., Friend, A. D., Hall, T. M., Hu, Y., Kelley, M., Kiang, N. Y., Koch, D., Lacis, A. A., Lerner, J., Lo, K. K., Miller, R. L., Nazarenko, L., Oinas, V., Perlwitz, J. P., Perlwitz, J., Rind, D., 850 Romanou, A., Russell, G. L., Sato, M., Shindell, D. T., Stone, P. H., Sun, S., Tausnev, N., Thresher, D., and Yao, M.-S.: Present-Day Atmospheric Simulations Using GISS ModelE: Comparison to In Situ, Satellite, and Reanalysis Data, *Journal of Climate*, 19, 153–192, doi:10.1175/jcli3612.1, 2006.
- Seki, O., Foster, G. L., Schmidt, D. N., Mackensen, A., Kawamura, K., and Pancost, R. D.: Alkenone and boron-based Pliocene pCO₂ records, *Earth and Planetary Science Letters*, 292, 201–211, 855 doi:10.1016/j.epsl.2010.01.037, 2010.

- Semtner, A. J.: A model for the thermodynamic growth of sea ice in numerical investigations of climate, *Journal of Physical Oceanography*, 6, 379–389, doi:10.1175/1520-0485, 1976.
- Shu, Q., Song, Z., and Qiao, F.: Assessment of sea ice simulations in the CMIP5 models, *The Cryosphere*, 9, 399–409, doi:10.5194/tc-9-399-2015, 2015.
- 860 Stepanek, C. and Lohmann, G.: Modelling mid-Pliocene climate with COSMOS, *Geoscientific Model Development*, 5, 1221–1243, doi:10.5194/gmd-5-1221-2012, 2012.
- Stroeve, J., Holland, M. M., Meier, W., Scambos, T., and Serreze, M.: Arctic sea ice decline: Faster than forecast, *Geophysical Research Letters*, 34, L09 501, doi:10.1029/2007gl029703, 2007.
- Stroeve, J., Barrett, A., Serreze, M., and Schweiger, A.: Using records from submarine, aircraft and satel-
865 lites to evaluate climate model simulations of Arctic sea ice thickness, *The Cryosphere*, 8, 1839–1854, doi:10.5194/tc-8-1839-2014, 2014.
- Stroeve, J. C., Kattsov, V., Barrett, A., Serreze, M., Pavlova, T., Holland, M. M., and Meier, W. N.: Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations, *Geophysical Research Letters*, 39, L16 502, doi:10.1029/2012gl052676, 2012.
- 870 Thomsen, H.: The Annual Reports on the Arctic Sea Ice issued by the Danish Meteorological Institute, *Journal of Glaciology*, 1, 140–141, 1947.
- Walsh, J. E. and Chapman, W. L.: 20th-century sea-ice variations from observational data, *Annals of Glaciology*, 33, 444–448, doi:10.3189/172756401781818671, 2001.
- Wang, M. and Overland, J. E.: A sea ice free summer Arctic within 30 years: An update from CMIP5 models,
875 *Geophysical Research Letters*, 39, L18 501, doi:10.1029/2012gl052868, 2012.
- Zhang, J. and Rothrock, D.: Modeling Arctic sea ice with an efficient plastic solution, *Journal of Geophysical Research*, 105, 3325–3338, 2000.
- Zhang, J., Lindsay, R., Schweiger, A., and Steele, M.: The impact of an intense summer cyclone on 2012 Arctic sea ice retreat, *Geophysical Research Letters*, 40, 720–726, doi:10.1002/grl.50190, 2013a.
- 880 Zhang, Z., Nisancioglu, K., Chandler, M., Haywood, A., Otto-Bliesner, B., Ramstein, G., Stepanek, C., Abe-Ouchi, A., Chan, W., Bragg, F., Contoux, C., Dolan, A., Hill, D., Jost, A., Kamae, Y., Lohmann, G., Lunt, D., Rosenbloom, N., Sohl, L., and Ueda, H.: Mid-pliocene Atlantic Meridional Overturning Circulation not unlike modern, *Climate of the Past*, 9, 1495–1504, 2013b.
- Zhang, Z. S., Nisancioglu, K., Bentsen, M., Tjiputra, J., Bethke, I., Yan, Q., Risebrobakken, B., Andersson, C.,
885 and Jansen, E.: Pre-industrial and mid-Pliocene simulations with NorESM-L, *Geoscientific Model Development*, 5, 523–533, doi:10.5194/gmd-5-523-2012, 2012.

Table 1. Technical details of the PlioMIP model ensemble: atmosphere and ocean resolutions, details of the sea ice component, and references for each of the eight PlioMIP Experiment 2 simulations.

model	Atmosphere	Ocean	Length of run/ averaging period (years)		Sea Ice components and references	Reference
	resolution (° lat × ° long)	resolution (° lat × ° long)	Pre-industrial	mid-Pliocene		
CCSM4	0.9 × 1.25	1 × 1	1300/100	550/100	EVP rheology, melt ponds Hunke and Dukowicz (1997); Hunke (2010); Holland et al. (2011)	Rosenbloom et al. (2013)
COSMOS	3.75 × 3.75	3 × 1.8	3000/30	1000/30	VP rheology, leads Marsland et al. (2003)	Stepanek and Lohmann (2012)
GISS-E2-R	2 × 2.5	1 × 1.25	950/30	950/30	VP rheology, leads Zhang and Rothrock (2000); Liu et al. (2003)	Chandler et al. (2013)
HadCM3	2.5 × 3.75	1.25 × 1.25	200/50	500/50	Free drift, leads Cattle and Crossley (1995)	Bragg et al. (2012)
IPSLCM5A	3.75 × 1.9	0.5 – 2 × 2	2800/100	730/30	VP rheology, leads Fichefet and Morales Maqueda (1999)	Contoux et al. (2012)
MIROC4m	2.8 × 2.8	0.5 – 1.4 × 1.4	3800/100	1400/100	EVP rheology, leads K-1 Model Developers (2004)	Chan et al. (2011)
MRI-CGCM	2.8 × 2.8	0.5 – 2 × 2.5	1000/50	500/50	Free drift, leads Mellor and Kantha (1989)	Kamae and Ueda (2012)
NorESM-L	3.75 × 3.75	3 × 3	1500/200	1500/200	Same as CCSM4	Zhang et al. (2012)

Table 2. Mean annual sea ice extents and amplitude of sea ice extent (maximum annual sea ice extent minus minimum annual sea ice extent) for the pre-industrial (PI) and mid-Pliocene simulations from PlioMIP, and historical (1979-2005) simulations from CMIP5, for each participant model in PlioMIP Experiment 2 and for the ensemble mean. All values are in 10^6 km². Starred CMIP5 values are from a different version of the model than used in PlioMIP (see section 2.1 for details).

Model	PI mean annual extent	PI extent amplitude	mid-Pliocene mean annual extent	mid-Pliocene extent amplitude	CMIP5 mean annual extent	CMIP5 extent amplitude
CCSM4	18.35	10.94	14.99	10.26	12.33	8.56
COSMOS	15.52	11.66	7.72	12.75	-11.10*	-7.95*
GISS-E2-R	17.30	14.03	9.63	15.43	13.65	15.17
HadCM3	13.76	12.42	10.38	14.17	13.94	13.59
IPSLCM5A	12.27	7.36	9.06	7.05	12.72	10.07
MIROC4m	19.85	14.05	11.48	21.98	10.66*	9.65*
MRI-CGCM	19.80	15.91	15.84	13.69	15.01*	15.27*
NorESM-L	12.52	6.39	7.60	12.86	12.01*	5.96*
Ensemble mean	16.17	11.18	10.84	13.44	12.90 -12.68	11.18 10.78

Table 3. Correlation between AMO and NAO indices, and mean annual and decadal Arctic sea ice extent for three PlioMIP models. Starred values are significant at the 90% level.

Model	Pre-industrial (annual)	Pre-industrial (decadal)	Mid-Pliocene (annual)	Mid-Pliocene (decadal)
	$r(\text{AMO}, \text{SIE})$	$r(\text{AMO}, \text{SIE})$	$r(\text{AMO}, \text{SIE})$	$r(\text{AMO}, \text{SIE})$
CCSM4	-0.036	-0.16	-0.23*	-0.27
HadCM3	-0.069	-0.17	-0.022	-0.22
NorESM-L	-0.10	-0.076	-0.035	0.12
	$r(\text{NAO}, \text{SIE})$	$r(\text{NAO}, \text{SIE})$	$r(\text{NAO}, \text{SIE})$	$r(\text{NAO}, \text{SIE})$
CCSM4	-0.18*	-0.099	-0.033	0.18
HadCM3	-0.24*	-0.33	-0.0063	-0.093
NorESM-L	-0.14*	-0.28	0.07	0.24

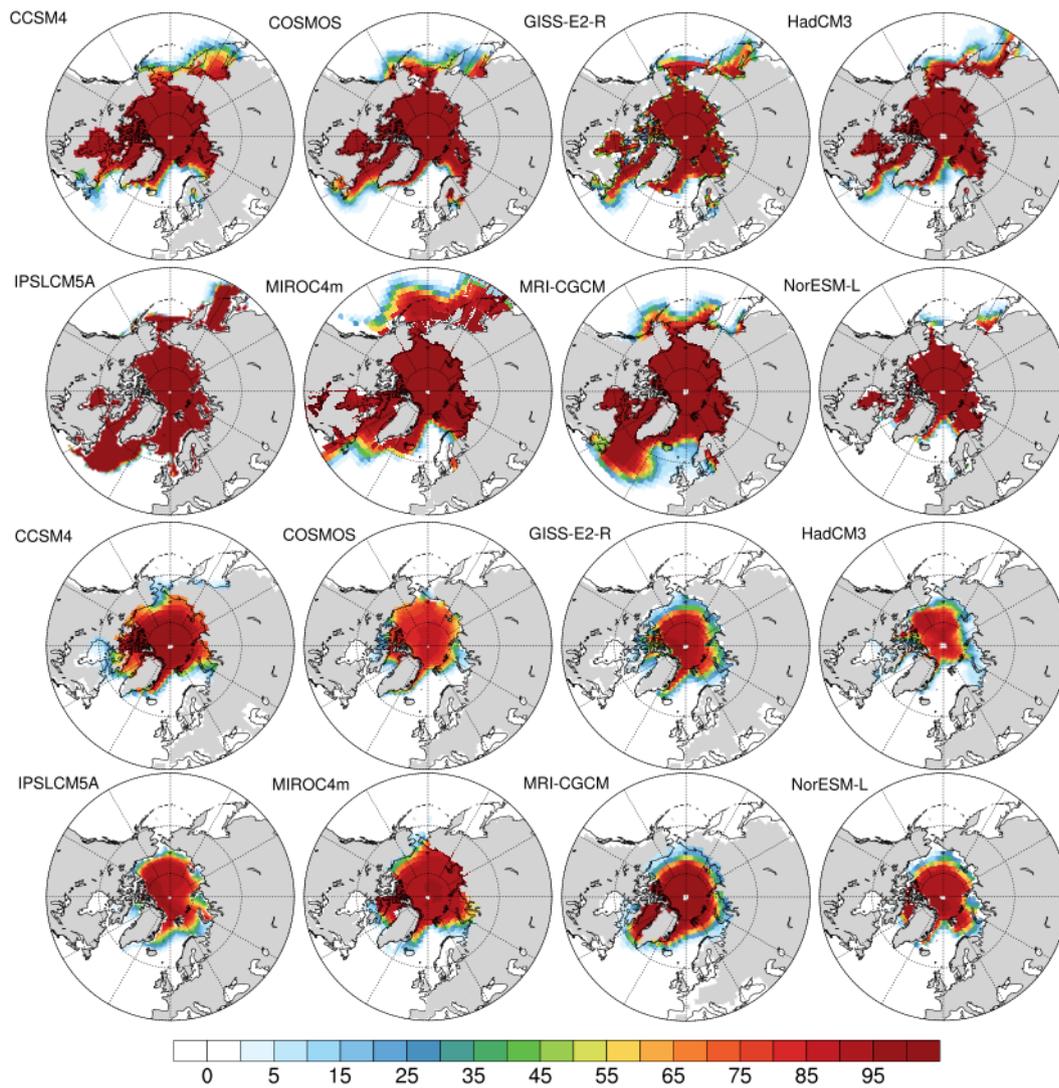


Figure 1. Mean sea ice concentrations (%) for winter (FMA, upper half) and summer (ASO, lower half) in the pre-industrial control simulations for each PlioMIP Experiment 2 model. Missing data at the poles is a plotting artefact (seen also in Figures 1, 3, 5, and 7).

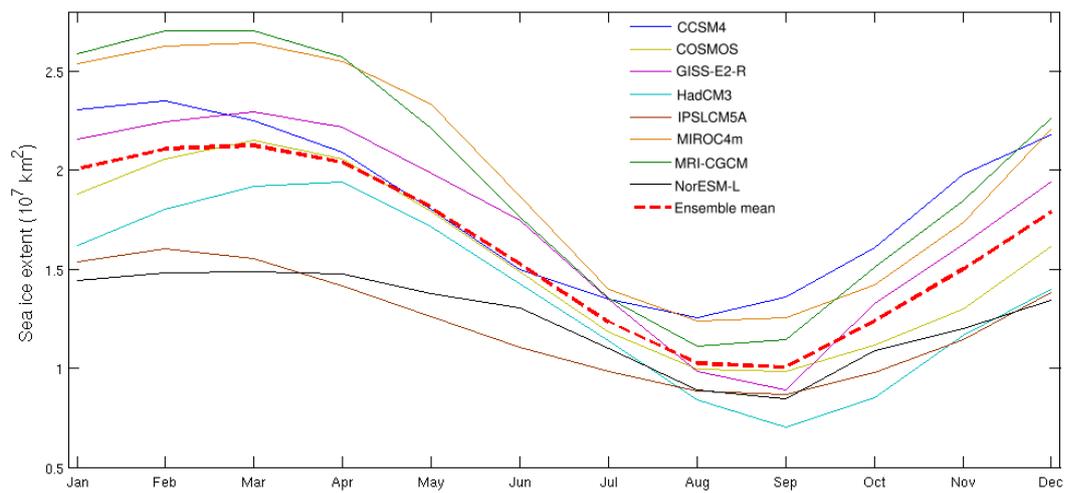


Figure 2. Annual cycle of total Arctic sea ice extent in the pre-industrial simulations for each participating model in PlioMIP Experiment 2, and the ensemble mean.

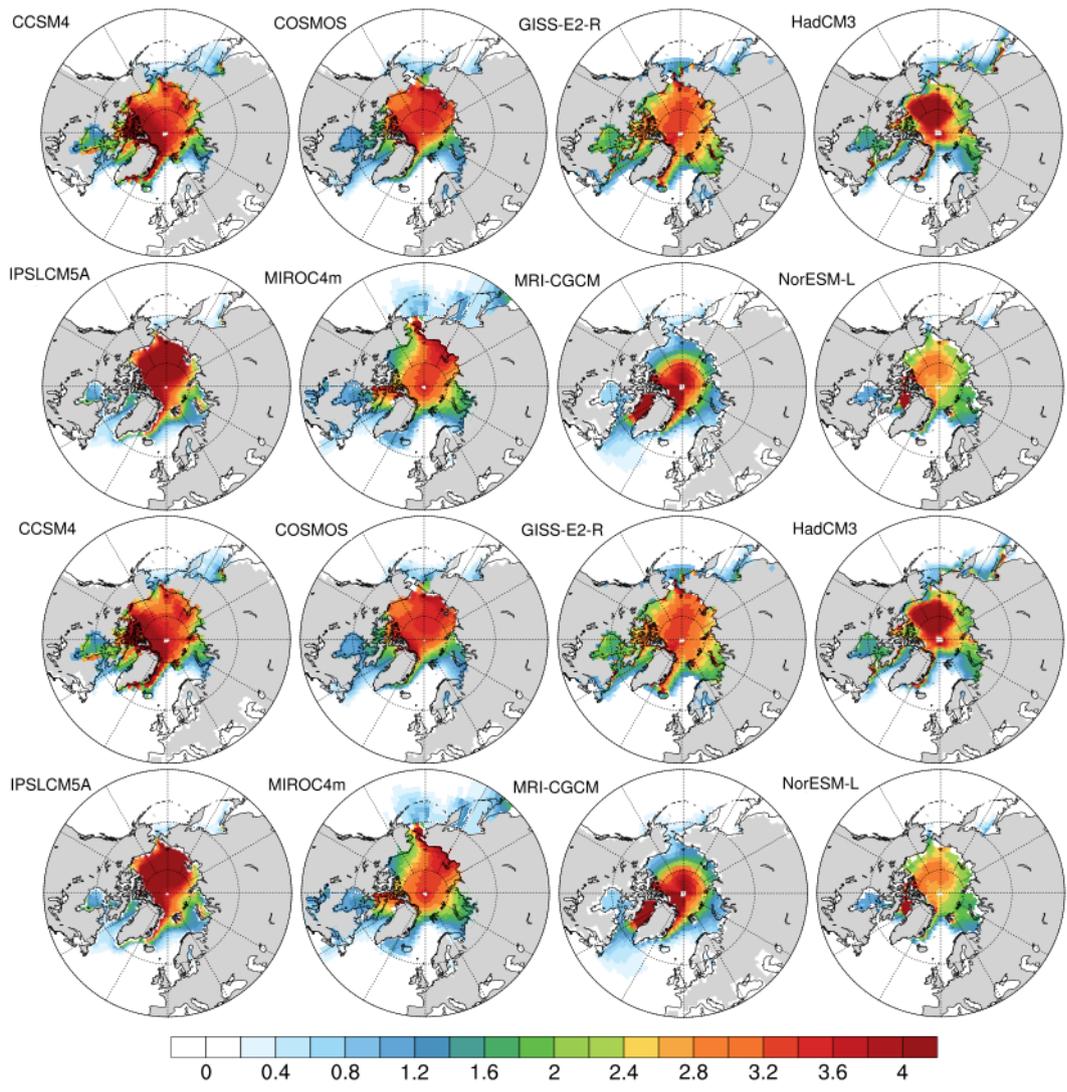


Figure 3. Mean sea ice thicknesses (m) for winter (FMA, upper half) and summer (ASO, lower half) in the pre-industrial control simulations for each PlioMIP Experiment 2 model.

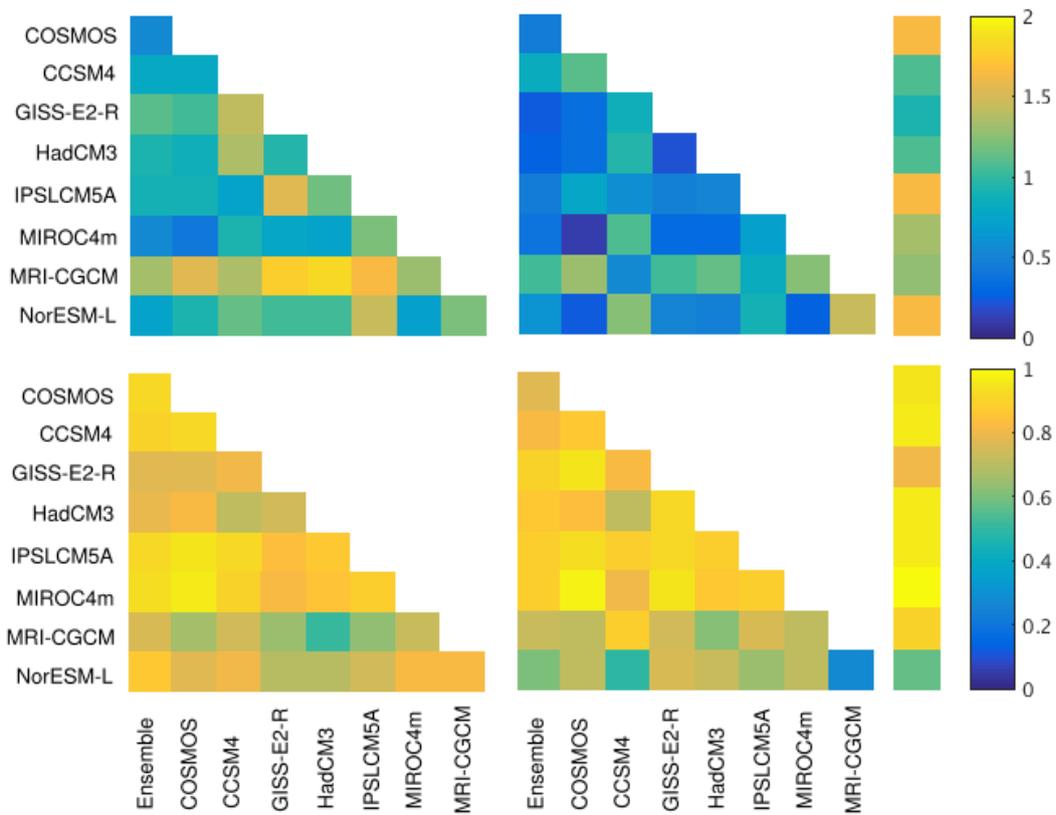


Figure 4. Root-mean-square error deviation (RMSE/MSD, m) (top) and spatial pattern correlations (SPCs) (bottom) of mean annual Arctic sea ice thickness (for ice-covered areas north of 60°N) in the pre-industrial (left) and mid-Pliocene (right) simulations by the PlioMIP models and ensemble mean. The single columns to the right show the RMSE/MSDs and spatial pattern correlations for SPCs between each model's pre-industrial and mid-Pliocene mean annual Arctic sea ice thickness.

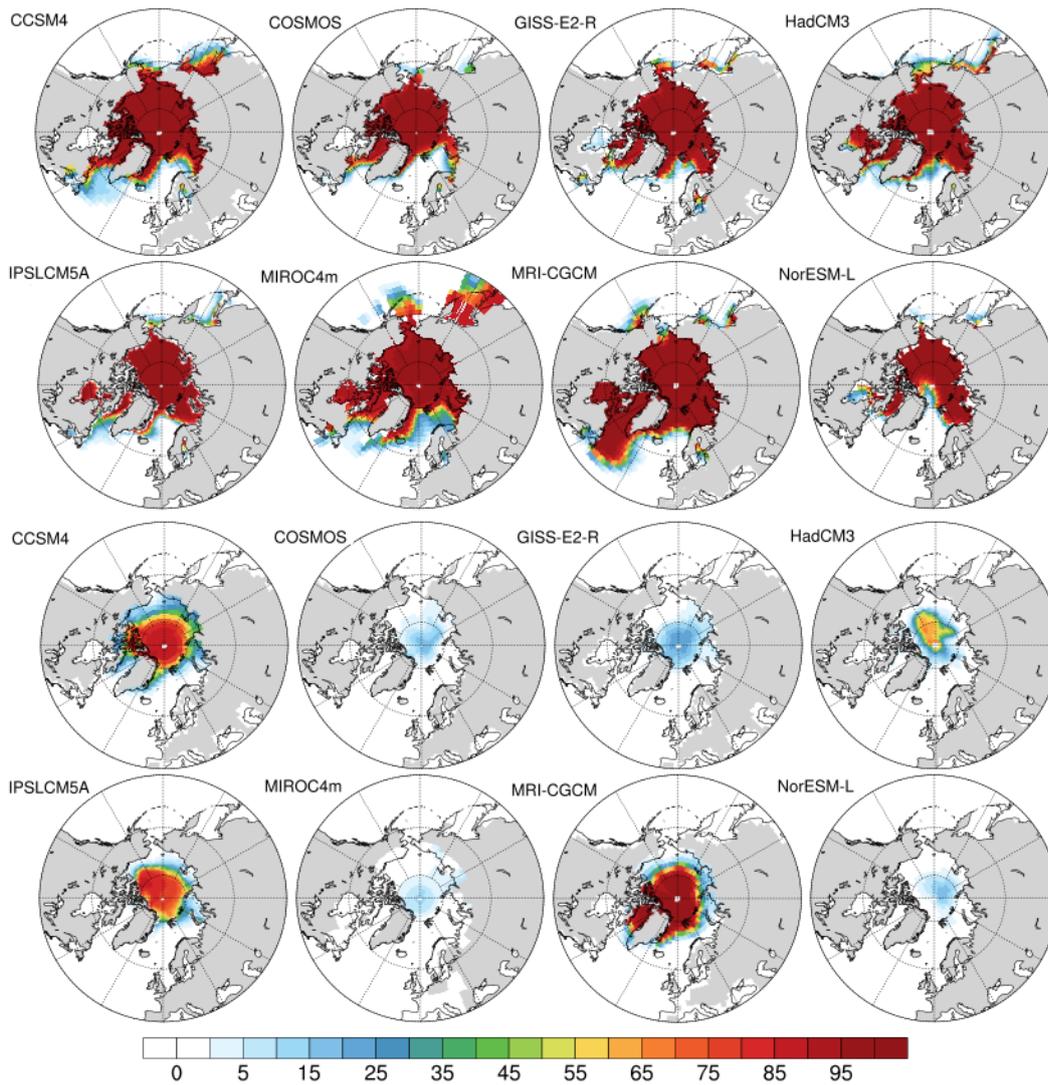


Figure 5. Mean sea ice concentrations (%) for winter (FMA, upper half) and summer (ASO, lower half) in the mid-Pliocene simulations for each PlioMIP Experiment 2 model.

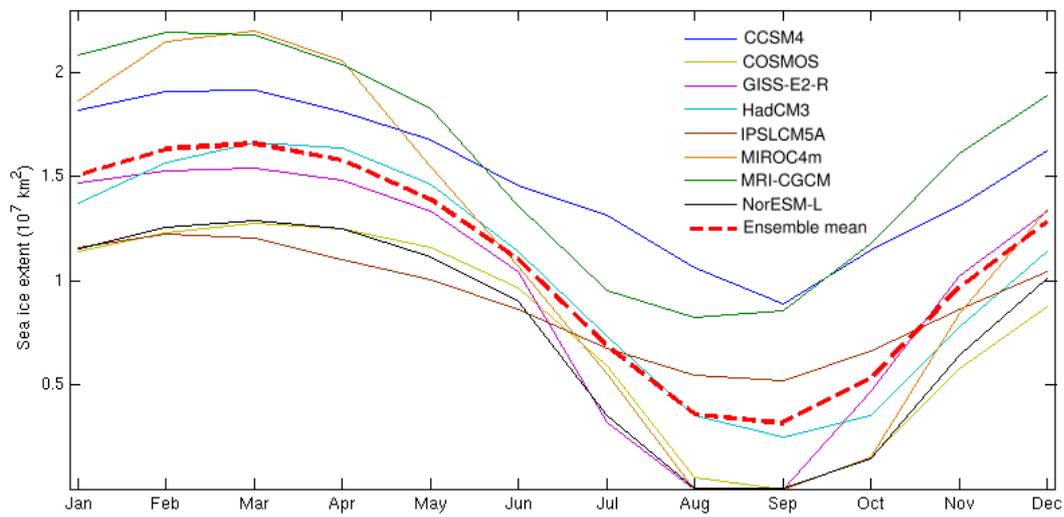


Figure 6. Annual cycle of sea ice extent in the mid-Pliocene simulations for each participating model in PlioMIP Experiment 2 and for the ensemble mean.

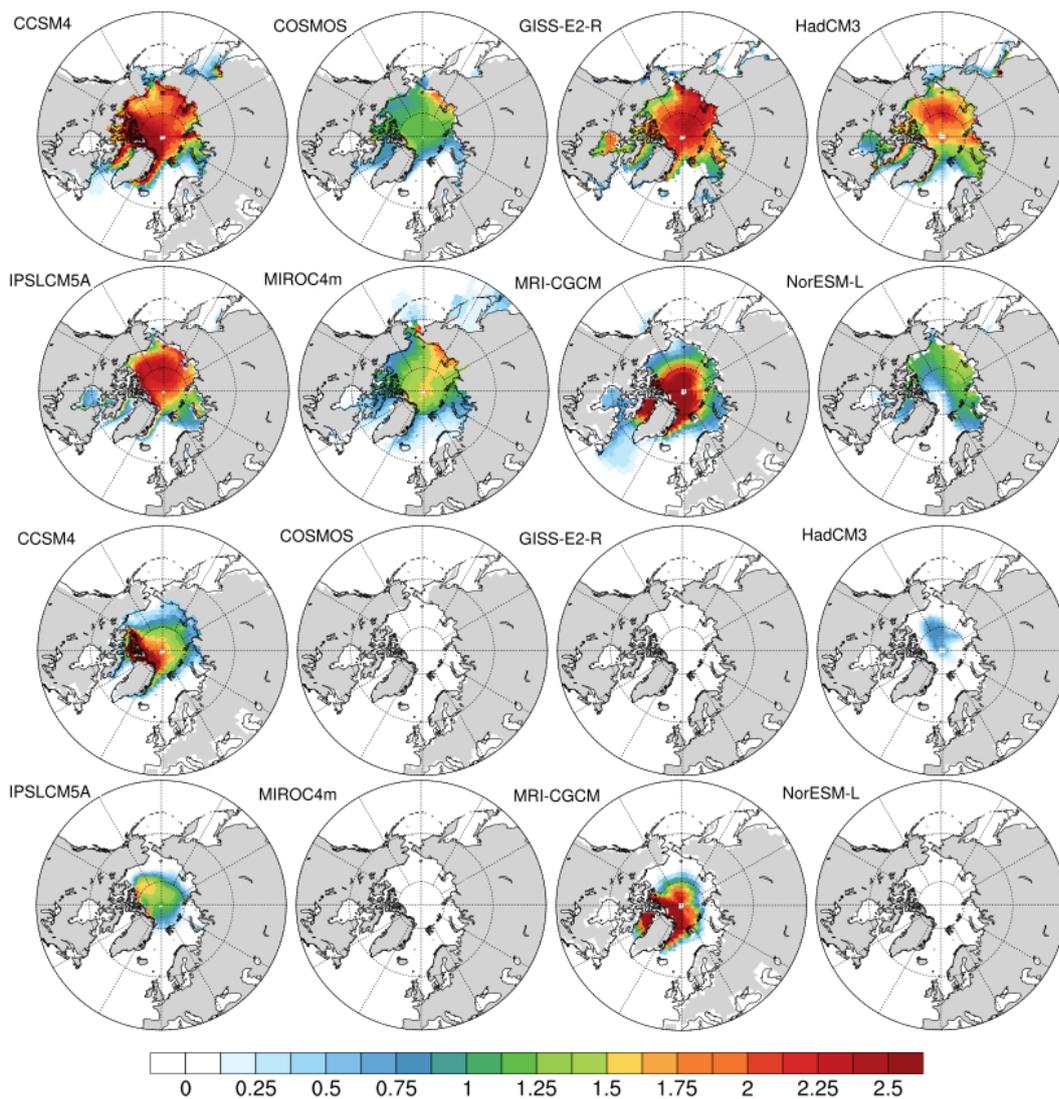


Figure 7. Mean sea ice thicknesses (m) for winter (FMA, upper half) and summer (ASO, lower half) in the mid-Pliocene simulations for each PliomIP Experiment 2 model. **Low sea ice concentrations in the summer plots for COSMOS, GISS-E2-R, MIROC4m and NorESM-L result in mean thicknesses very close to zero in each model grid cell.**

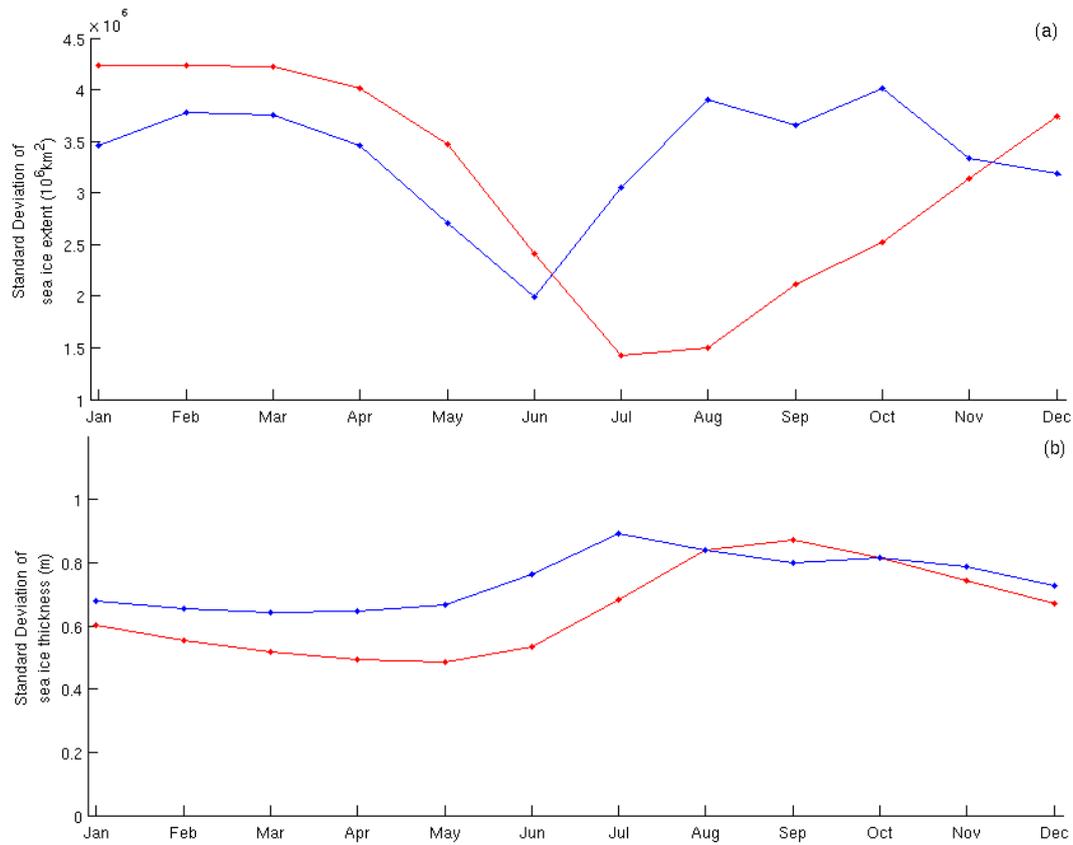


Figure 8. Annual cycle of the standard deviation of (a) sea ice extent and (b) sea ice thickness for the PlioMIP Experiment 2 ensemble. Red lines represent the pre-industrial annual cycle, blue lines represent the mid-Pliocene annual cycle.

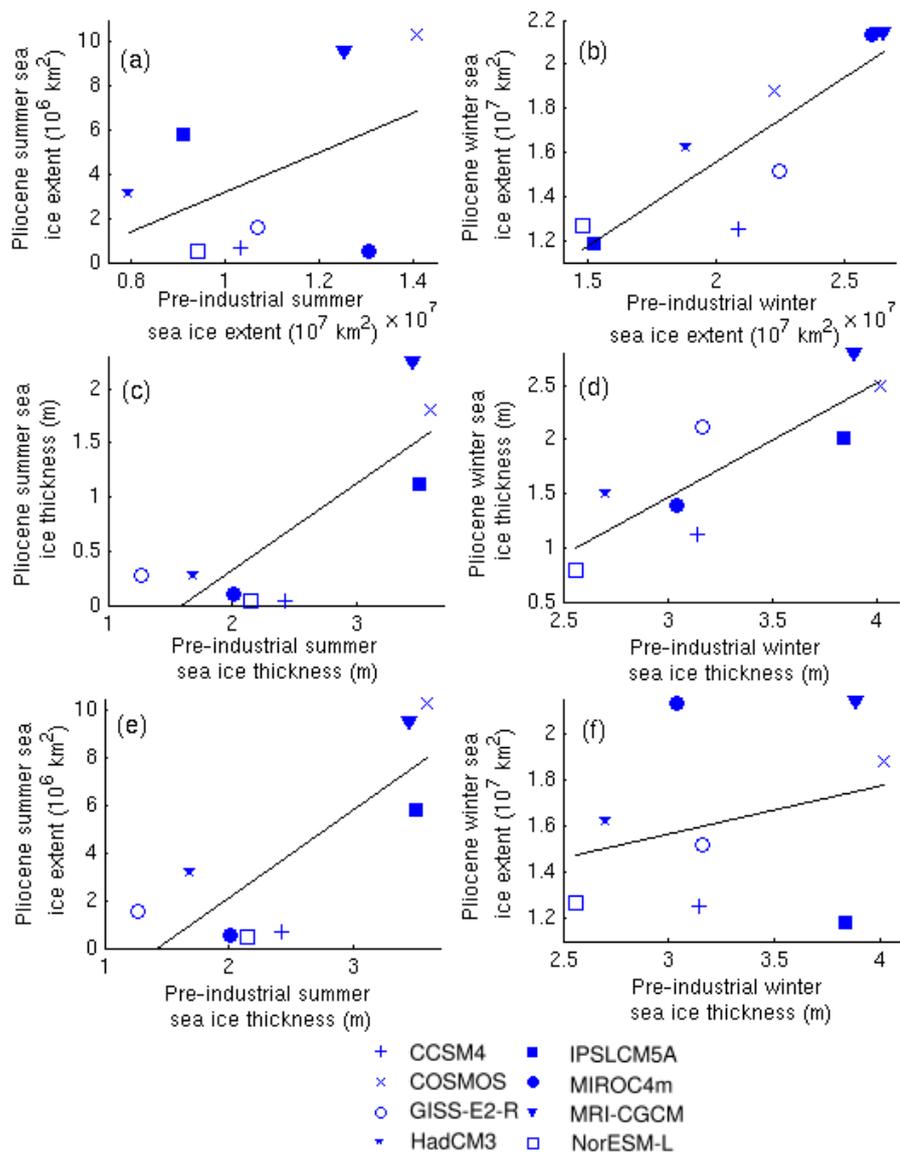


Figure 9. Relationship between various sea ice characteristics. Shown are pre-industrial values vs. mid-Pliocene values for (a) and (b) sea ice extent vs. sea ice extent, (c) and (d) sea ice thickness vs. sea ice thickness, (e) and (f) sea ice thickness vs. sea ice extent. (a), (c), and (e) illustrate summer conditions, (b), (d), and (f) illustrate winter conditions. Correlation coefficients for each plot are (a) 0.47, (b) 0.87, (c) 0.82, (d) 0.85, (e) 0.81, (f) 0.30

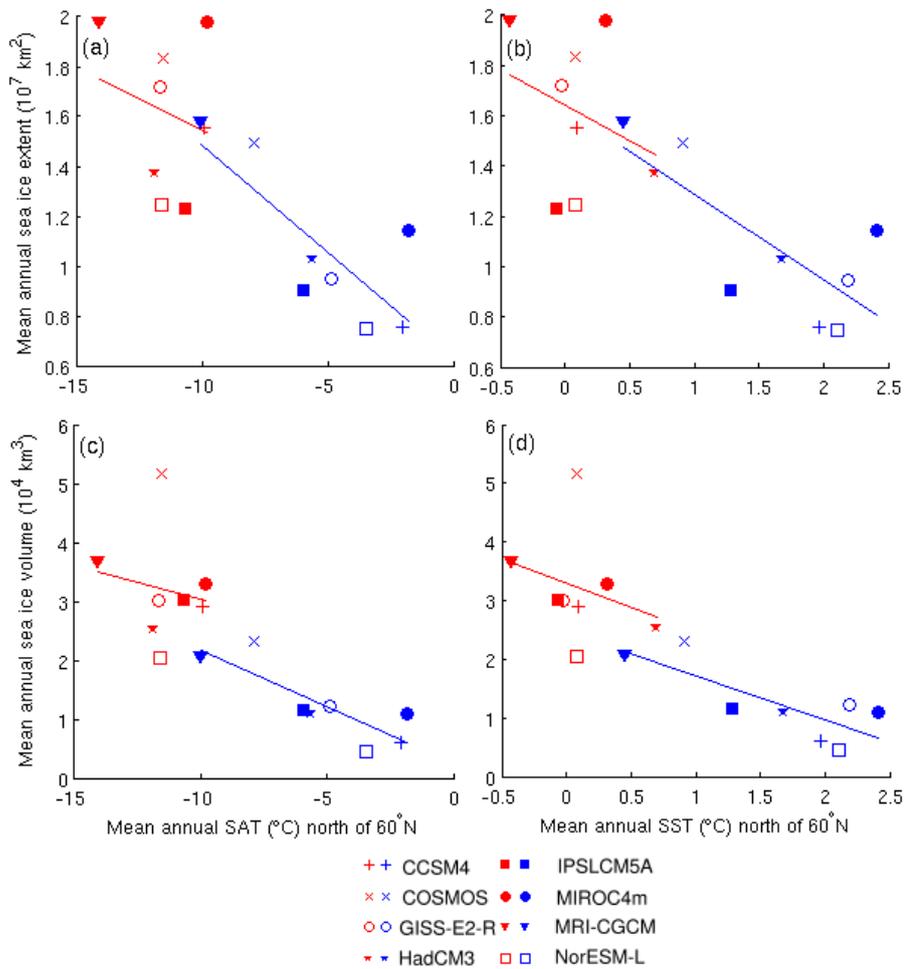


Figure 10. Mean annual surface temperatures north of 60°N vs. mean annual total Arctic sea ice extent (a,b), and mean annual surface temperatures north of 60°N vs. mean annual total Arctic sea ice volume (c,d) in both pre-industrial and mid-Pliocene simulations, for (a,c) SAT and (b,d) SST. Pre-industrial experiments are marked red, mid-Pliocene experiments are marked blue. Correlation coefficients for the pre-industrial simulations in each plot are (a) -0.18, (b) -0.26, (c) -0.12, (d) -0.29. Correlation coefficients for the mid-Pliocene simulations in each plot are (a) -0.76, (b) -0.73, (c) -0.83, (d) -0.82

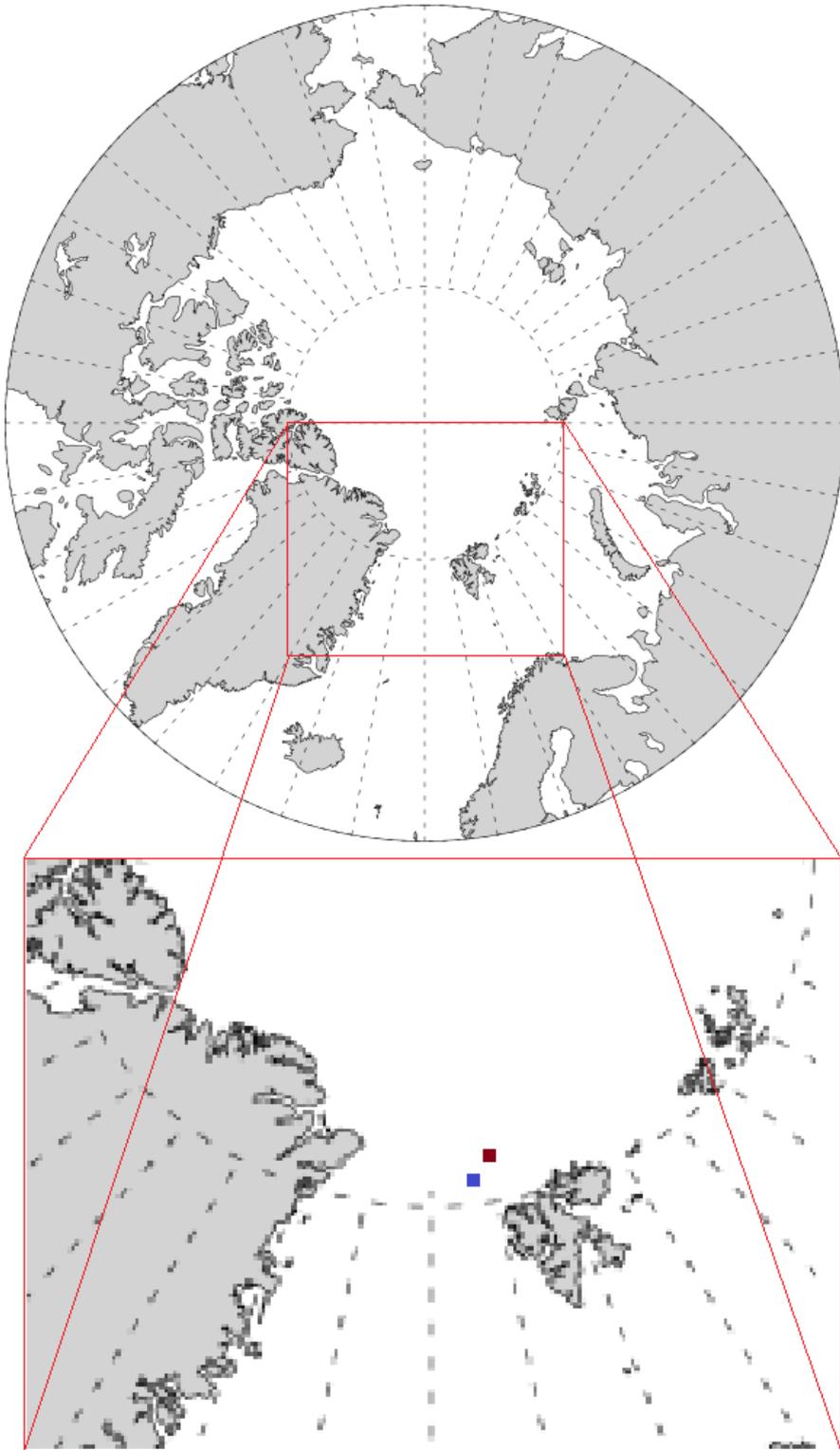


Figure 11. Location of Ocean Drilling Program (ODP) sites 911A (brown), and 910C (blue), used by Knies et al. (2014) for IP_{25} analysis.

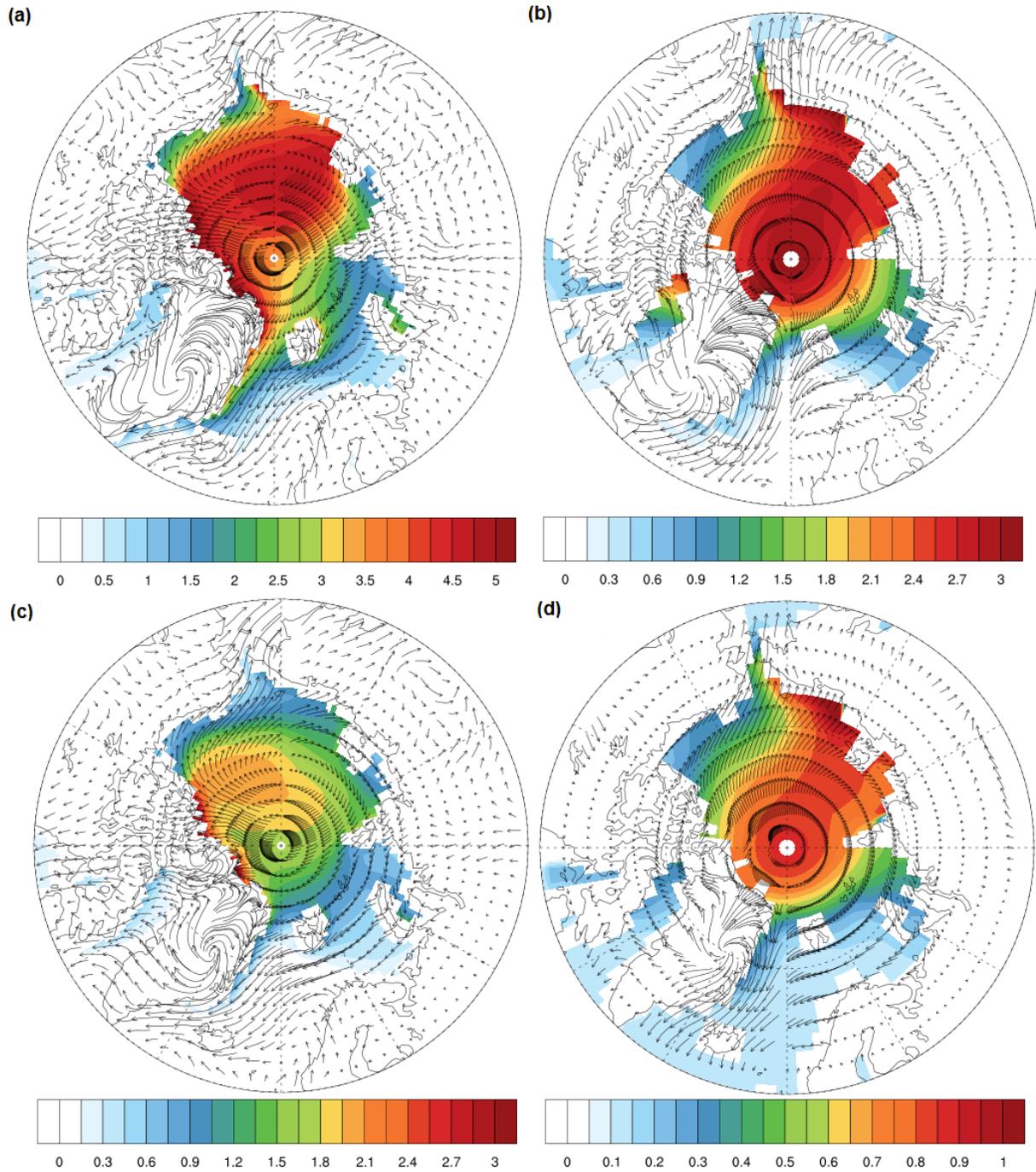


Figure 12. Mean annual 10 m winds and sea ice thicknesses (m) for (a) **COSMOS mid-Pliocene/IPSLCM5A pre-industrial**, (b) **MIROC4m mid-Pliocene/pre-industrial**, (c) **IPSLCM5A pre-industrial and mid-Pliocene**, (d) **NorESM-L pre-industrial/MIROC4m mid-Pliocene**. Vector length is proportional to wind speed.

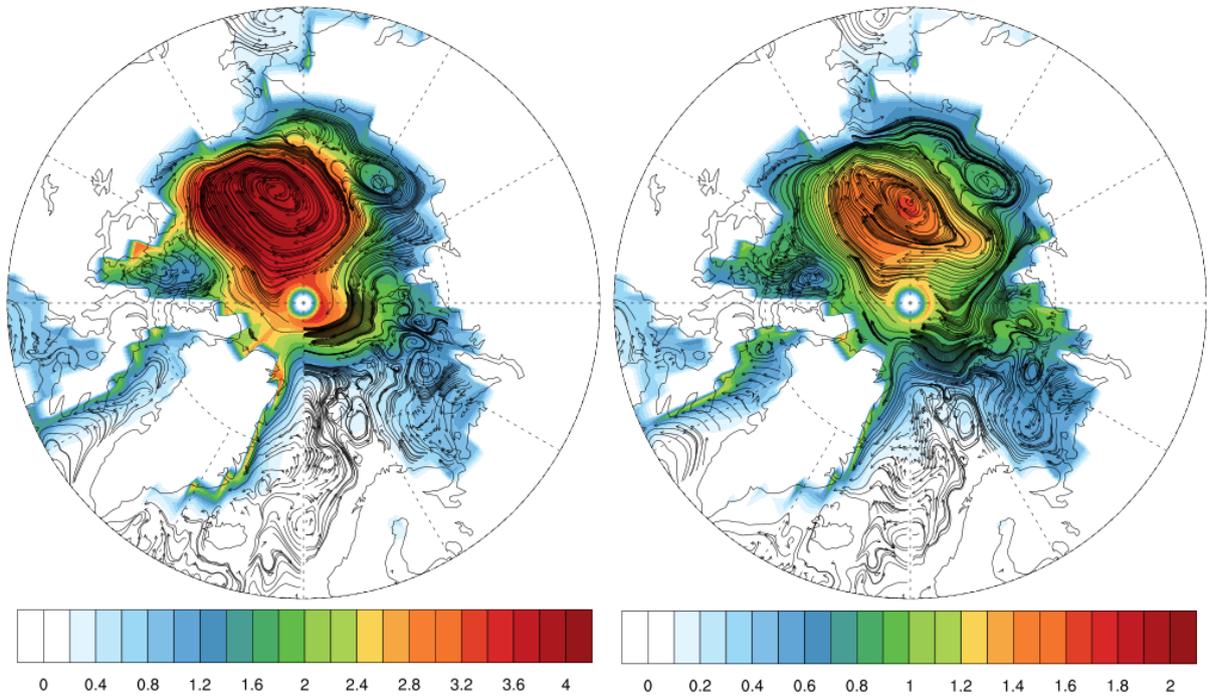


Figure 13. Mean annual ocean surface currents and sea ice thicknesses (m) for HadCM3 pre-industrial (left) and mid-Pliocene (right) simulations. Vector length is proportional to ocean current speed.