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

The article "Assessment of simulations of Arctic sea ice in the PioMIP models" by F.W. Howell et al. analyzes sea ice representation in pre-industrial and mid-Pliocene climates in an ensemble of 8 different global coupled models. The inter-model spread is discussed and possible relations between pre-industrial and Pliocene simulations are investigated.

Since data are very sparse for the Pliocene and basically no robust information about sea ice distribution in the Pliocene is available, we do not gain much information about model performance by simultaneously comparing pre-industrial and mid-Pliocene model simulations. There are several studies existing, which either evaluate a larger number of global models for present day climate (better suited for model evaluation than pre-industrial) or relate behavior of ice properties in present day and future studies in a much more comprehensive way than this study does. Thus, I appreciate that this new version of the article focus more on sea ice conditions under pre-industrial and mid-Pliocene conditions and less on a potential ranking of model's ability to simulate sea ice.

Unfortunately, large parts of the manuscript are very descriptive and mentioning trivialities; it is not necessary to describe in detail the ice thickness pattern of every single participating model. Instead more analysis should focus on the causes for the model spread and the causes for different pre-industrial – Pliocene differences. Why is the seasonal amplitude of ice extent different? How does the atmospheric circulation look like in the different models both in pre-industrial and pliocene? How does this affect the sea ice thickness distribution in the different models? How does the atmospheric and oceanic heat transport into the Arctic look like and how might it affect sea ice representation?

The argumentation that the representation of NAO or AMO could not be investigated at all due to too short time series is not entirely convincing. I agree 30 years are short but NAO and AMO are strongly affecting sea ice and if we can not say anything about NAO and AMO, then it might also be difficult to make robust conclusions on the sea ice conditions themselves. Here, the question arises why only 30 years from the 500-1000 year simulations are used for the analyses? To make results more robust, a longer period should be used.

We have removed the paragraphs describing in detail the thickness patterns simulated. We have performed root mean square error and spatial pattern correlations between all models, these are displayed graphically (Figure 4), and discussed.

We have included plots of surface winds and ocean currents (Figures 12 and 13, plus supplementary information) to demonstrate potential causes of differences between the models. For some there is a very clear indication of the causes of the sea ice patterns (e.g. HadCM3 ocean pattern). With other models the causes of certain features can be identified, but not every aspect. A more detailed analysis may be able to demonstrate with more certainty the exact causes for model differences due to atmospheric and oceanic circulation, but the level of analysis required for this would be far beyond the scope of this paper.

Sufficiently long time series were not available to perform analysis of NAO and AMO with all the PlioMIP models. We have done so for three where enough data was available (see section

4.3.3, and Table 3), and found that the influence of variability modes disappears when looking at longer-term averages (at least 10 years), whilst still having some influence on a year-to-year basis. Again, this is an area for which there is scope for further work, but that level of analysis would be worthy of a separate paper.

We also looked at the effect of using different averaging periods (30, 50, 100 and 200 years) for the models with longer timeseries output, and found that the differences were extremely small, and made no difference to the analysis, so this can be ruled out as a cause of model-model differences.

I am still not convinced by using CV as measure for the variability even it has been used by Stroeve et al. 2014 – your CV is also not exactly what Stroeve et al. (2014) used and they used it in a somewhat different context. CV assumes that sea ice concentration and thickness variability should linearly increase with ice concentration and thickness, respectively. To my knowledge, no one ever showed this. In any case, CV can not be a reliable measure for ice variations if ice thickness/ concentration are near 0 as it is in four of your Pliocene-summer simulations. Thus, the conclusions that especially summer ice variability increases in Pliocene compared to the pre-industrial simulations should not be taken based only on CV-values. I would even call this statement wrong. CV could be used if a) the "normal" standard deviation/ variance of the models is shown as comparison and b) evidence is shown that we can assume a linear growth of ice concentration and ice thickness variability with increasing ice concentration and thickness, respectively (e.g. cite an article that shows this).

We have removed discussion of CV from the paper. Standard deviation is instead shown (Figure 12), and discussed in its place.

Specific comments

1. Line 6: I would say that the statement that the model spread is 3 times larger in summer than in the rest of the year is just wrong: Figure 13 shows that ice varies between 0.7 and 2.7m thickness in winter (quite evenly distributed) and between 0 and 2 m in summer (5 out of 8 models between 0 and 0.3 m, the other three about 1, 1.7 and 2.2 m). See discussion above.

This line has been removed from the abstract.

2. Line 40 and following, line 306-310: I would suggest making a single section discussing model-setup and the experiments a bit more in detail. What is the difference to the CMIP5-model versions? How have sea ice and SST been prescribed in Pliocene-Experiment 1 (AGCM-simulations) if observations are so uncertain?

Section 2.1 now describes the PlioMIP experimental designs. The basis for the prescribed SSTs and sea ice is described, with the appropriate references. Reference to models which simulated CMIP5 simulations has been moved to this section.

3. Line 54: additional to what?

Due to the restructuring of the method section, this line has been removed.

4. Line 54: "reduces"

See above.

5. Lines 59-61: Are you sure all models provide ice thickness as grid-box mean and not as mean over the ice-covered part only (which is the case in most CMIP5 models)?

This change from the previous version was implemented due to a comment in review 1 in the initial review stage stating that sea ice thickness output was grid-box mean. However it appears that this is not correct, and that thickness is in fact ice-covered thickness. Values have been amended accordingly.

6. Line 83: Figure 1: Extend the plotted area to the south – if this figure does not show the position of the ice edges, it is useless. It does not matter that the other figures show a different area, you could indicate this in the figure caption.

This has been done.

7. Lines 92-98: You should also indicate the observed amplitude for comparison.

Given that all comparison with observations were dropped from the previous version, we have not included this, as we do not believe it adds anything.

8. Line 106: I do not see any ice thickness anomaly at 0E, 80-90N, to be clear it is maybe better to say 180E, 80-90N instead of "Greenwich meridian".

9. Line 110: mention that ice thickness is likely overestimated at the Siberian coast in MIROC, COSMOS.

10. Lines 111-113: This is not needed. However, if you want to mention that sea ice is thinner in Nordic Seas compared to Siberian coast you should also shortly mention why this is the case.

11. Lines 120-143: This discussion is too detailed and difficult to read. The section should be shortened and the most important points mentioned. The reader can find details in the figure.

Passages referred to above have all been removed.

12. Line 164: In contrast to this, in the discussion it seems to be stated that CCSM shows a larger ice extent in the mid-Pliocene – please check.

The line to which is being referred (line 300) should only have said 'CCSM CMIP5', and not 'CCSM CMIP5 pre-industrial'. This has been amended to avoid any confusion.

13. Line 187: delete one "amplitude"

Done.

14. Line 205: "Many" sounds a bit strange with a total of 8 models whereas 4 do not show any ice in summer.

The sentence in which this appears has been removed.

15. Figure 7: As figure 1: please show a larger area in the plot.

Done.

16. Line 218/219: difficult to judge from the Figures 7 and 8, there are very large differences across models in winter as well and Figure 9 shows about the same max-min difference in winter compared to summer. As mentioned earlier I do not think CV should be used and comparing Figures 9 and 12 clearly indicates why. Just from looking at Figure 9: do you really want to suggest that model spread is 3 times larger in summer than in winter?

This statement has been removed.

17. Line 284-296: Please make clearer if this is a summary from Shu et al. 2015. Please specify which time period Shu et al. (2015) analyzed. I would suggest adding these numbers, at least annual mean ice extent to table 2.

CMIP5 mean annual extent and extent amplitude have been added to the table. Additional Shu et al. citations have been added to make clear it is a summary of results from that paper.

18. Lines 299-302: I do not understand this sentence: Please clarify.

Sentence has been reworded in an attempt to make the meaning clearer (line 259).

19. Lines 303-305: This sentence is not clear.

This part of the paragraph has been removed.

20. Lines 326-327: Of course they are not the same: maybe better: "...vary strongly: The summer sea ice ..."

Edited as suggested.

21. Line 324-325: maybe better: "almost ice free" or "ice free in late summer"

Changed to 'almost' ice-free.

22. Line 326-327: sounds like HadCM3 simulates summer sea ice in the entire Arctic Basin. Ok, not really wrong but actually ice concentration is very low along all ice edges.

Mention of low concentration along ice edges has been added.

23. Line 405: Please specify what you mean with "CMIP5". Here and elsewhere CMIP5 is compared to pre-industrial simulations: Pre-industrial simulations are also part of CMIP5 as *historical and future and many more simulations are. It seems you mean a certain time-period with CMIP5 (historical, satellite period ...)?*

Have clarified that 'CMIP5' refers to the historical (1979-2005) simulations.

24. Line 435-440: SST and SAT are not necessarily drivers of sea ice variations but could also be driven by ice variations: One reason for better correlation in Pliocene could be that larger parts of the ocean north of 60N are ice-free for longer periods in the year and could thus warm up much stronger than in the pre-industrial period. The longer, the ice-free period, the more the ocean can warm. In the pre-industrial period instead, summer SST and SAT in the Arctic will almost be very near melting temperature of ice, it does not matter, if sea ice concentration is 100% or 50% in a certain gridbox or smaller region: as long as some ice is left, the ocean can hardly warm up.

This has been added to the discussion (line 439).

25. Lines 453ff: The albedo discussion would fit better into section 4.3.1 "Influence of sea ice models"

Section moved as advised.

26. Lines 521-526: According to the introduction, it is debated if sea ice was seasonal or perennial? Yes, HadCM3 agrees with the findings of perennial Arctic sea ice but the other models agree with findings of Cronin et al., Moran et al and Poyak et al.. Is there any particular reason to believe more in the perennial assumption? Furthermore, HadCM3 shows a very unrealistic sea ice concentration distribution in both pre-industrial and mid-Pliocene summer, thus even if HadCM3, probably by chance, keeps the points of Knees (2014) at 80N ice-free year around, we can be quite sure that Arctic sea ice distribution will not look like HadCM3. All the years with very low observed ice concentrations (e.g. 2007, 2012) still showed the thickest ice with highest concentration north of Greenland and the Canadian Archipelago.

Lines have been added emphasising the models which agree with the other proxy interpretation, and the sea ice pattern simulated by HadCM3 has also been referenced.

Reviewer 2

l.5 and section 3.3: I am still not convinced that much can be learned from using CV in the current context. What is the geophysical relevance of CV that makes this measure preferable over simply using ensemble spread? If in a warmer climate all simulations are ice free, but one simulation still has a tiny ice floe of 2 m lying around somewhere, then CV will be more than 10. But this high value would be totally irrelevant, as is expressed by the geophysically more relevant ensemble spread given by ensemble standard deviation. I disagree in particular with the statement tat standard deviation does not allow one to compare data sets with different mean values (l.64). Why not? If the authors decide to keep the analysis of CV, it'd be helpful to give geophysical reasons for its relevance - rather than simply stating that others have used this metric before. Please also note that "ensemble spread" is very different from "variability", but currently these terms are used as if they were to describe the same thing.

Analysis of CV has been removed, and replaced with standard deviation.

l.11: "suggesting that the dominant atmospheric and oceanic influences may be different in the [two] simulations": This is one example of the speculative language. All data is there to test this suggestion, so why not do it? In particular since I doubt that this is true.

This line has been removed.

l.24: The Arctic is only "widely predicted to become seasonally ice free before the end of the 21st century" for a specific evolution of CO2

'Under RCP 4.5' added to make this clear.

Introduction in general: This should include some short discussion of what we do know from previous studies on sea-ice ensemble spread, correlations between individual sea-ice metrics and drivers, temporal correlation of sea-ice evolution, generally evolution of sea ice in a warmer climate, etc., which is necessary to allow the reader to identify the open questions that are addressed by the present study.

Paragraph added to introduction.

l.56 leading to Figure 14: I was wondering if some of the results of this study are simply related to the fact that sea-ice extent is used to describe the areal coverage of sea ice, rather than sea-ice area. If in a cold climate sea-ice concentration reduces because of some warming from, say, 90% to 45%, sea-ice extent would remain the same, even though the area decreases by 50%. This then renders the correlation of extent and temperature very weak. Sea-ice extent is only a useful metric when comparing data to observations, since it allows one to account for some observational uncertainty. In the present context, where most of the analysis is only carried out in the model realm, sea-ice area would give much more robust results, in particular given the very low sea-ice concentration that is obtained in the warm climate runs.

Figure 10 shows that the same difference in temperature correlation with extent for pre-industrial and Pliocene applies to mean sea ice volume as well, which takes into account total sea ice to a greater degree than sea ice area would. Correlations with sea ice area are similar to extent, and by including the volume correlations we have demonstrated the difference cannot be explained due to the way extent is measured, and so we have retained use of sea ice extent, which has been used in modelling studies with no observational comparison previously (e.g. **?**).

Section 3: I found this section unnecessarily long. The reader can simply look at the figures, and doesn't need a detailed description of every single panel. In particular since much of the language remains very vague, repetitive and sometimes contradictive, such as "Most of the models display patterns that are broadly similar to ensemble mean - but there is appreciable variation with respect to the location of maximum ice thickness". Either the patterns are broadly similar (which includes their key characteristics), or they are not (as given by the location of maximum thickness). Or: "The thickest ice in COSMOS [...] is located in approximately the same region as in the ensemble mean." followed by "In COSMOS, the thickest ice is concentrated into a smaller area." I found this entire section very cumbersome to read.

This section has been removed, and in its place results of RMSE and spatial pattern correlation calculations are presented (shown graphically in Figure 4 also).

l.143: What is "relatively" reduced ice?

l.144ff: Why should multi-year pre-industrial ice-thickness patterns match two months of observational record from 2009?

l.152: Another example for very vague language: "The ensemble mean thickness patterns appear to broadly match the observations."

The section which the above comments refer to has been removed.

l.187: I did not understand the logic (and meaning) behind: "The finding that sea-ice extent amplitude in the mid-pliocene is 64% greater than the pre-industrial simulation amplitude holds for the ensemble mean at a lower amplitude extent amplitude."

This line has been removed.

l.210: Another example for vague and somewhat contradictive language: "A similar finding to the fact that MIROC has similar patterns in winter in both simulations holds for COSMOS, where the central Arctic sea ice thins by a greater amount in comparison to sea ice in other regions."

This has been removed.

section 4.1: There is no assessment of pre-industrial simulations in this section, hence the title is misleading. Instead, this section primarily summarizes results from other studies on the historcial simulations from CMIP5.

Section title changed to 'Pre-industrial simulations'.

l.289: Another example for very vague language: "The fact that historical extent simulated by MRI is almost 25% greater than observations may suggest that its Arctic sea-ice cover is too extensive."

Wording changed to 'MRI consistently simulates Arctic sea ice extent larger than the ensemble mean'.

l.292: Why is it a contradiction that a model has a sea-ice extent closest to observations "although" it has the lowest sea-ice extent amplitude?

Wording of this sentence altered.

section 4.2: Again, this section does not really give an assessment of mid-Pliocene simulations, but instead comes to the conclusion that such assessment is not possible.

Heading title changed to 'mid-Pliocene simulations'.

l.324: Unnecessary repetition, I find.

Second part of sentence removed to avoid repetition.

l.335: This is not very clearly spelled out: Why may a reasonable performance of a model relative to mid-Pliocene sea ice improve confidence into this model, while a the same time a match to present-day observations does not necessarily mean that the model is good?

This paragraph has been amended to make it clear that a model which simulates one climate state 'well', but another less well, may be due to chance, but a good simulation of two different climate states should improve confidence in the model. The emphasis is on the need for a good mid-Pliocene and present day performance.

l.359: Why does HadCM3 only appear to be in closest agreement with proxy-data indications? Either it is, or it isn't.

Line now says HadCM3 has nest agreement, not appears to have.

l.365: Why is it a contradiction that CCSM and NorESM use the same sea-ice component "although" NorESM has a coarser atmosphere and a different ocean?

'Although' does not imply a contradiction. It merely serves to emphasise that while the two models have the same sea ice component, there are differences in the atmosphere and ocean components.

l.406: In section 4.1, there is no analysis of pre-industrial or mid-Pliocene performance, which would require some comparison against data to actually assess performance.

The word performance has been removed.

section 4.3.3: Much of this section remains unnecessarily vague. All data to support or reject the suggestions is in the data that the authors have available, so I find that the analysis should move beyond quoting existing studies by Hill et al., Zhang et al., etc.

We have added sections analysing the influence of variability modes on the model simulations (for models for which sufficiently long time series were available), and a discussion of the influence of winds and ocean current on the sea ice thickness patterns.

l.424ff: I did not fully understand what is meant by "stronger correlation": A higher slope of the linear fit, or less spread around the fit?

Changed to 'higher correlation coefficient'.

l.473: I found this confusing: Models with lower sea-ice albedo have less ice-albedo feedback. Why would they have greater potential to amplify warming from greenhouse gas emissivity?

This has been changed to emphasise that it is models with a lower sea ice albedo minimum - i.e. under warmer conditions, the sea ice albedo will take values from the lower range more. Those models with a lower minimum will therefore absorb more SW radiation, and thus have an enhanced feedback.

l.505: What is a "relatively consistent level of variabililty"?

This section has been removed.

l.518: Again, the data is there to examine this, rather than having to say that "If models see an enhanced ice-albedo feedback, than this is likely to affect those models predictions of future Arctic sea-ice change".

This paper is focused on the PlioMIP simulations, and the output from these. Output from future simulations are not within the scope of this paper.

l.521: Why does the fact that HadCM3 produces the thinnest pre-industrial sea ice imply that this model generally has difficulty in simulating observed sea-ice thickness?

This paragraph is removed.

l.530: see l.359

Altered to say 'HadCM3 therefore produces the mid-Pliocene simulation that is in best agreement with the proxy inferences'.

Bibliography

Manuscript prepared for Clim. Past
with version 2015/11/06 7.99 Copernicus papers of the LaTeX class copernicus.cls.
Date: 9 December 2015

~~Assessment of simulations of Arctic sea ice in the PlioMIP models~~ Arctic sea ice simulation in the PlioMIP ensemble

F.W. Howell[1], A.M. Haywood[1], B.L. Otto-Bliesner[2], F. Bragg[3], W.-L. Chan[4], M.A. Chandler[5], C. Contoux[6], Y. Kamae[7], A. Abe-Ouchi[4,8], N.A. Rosenbloom[2], C. Stepanek[9], and Z. Zhang[10]

[1]School of Earth and Environment, University of Leeds, Leeds, UK
[2]National Center for Atmospheric Research, Boulder, Colorado, USA
[3]School of Geographical Sciences, University of Bristol, Bristol, UK
[4]Atmosphere and Ocean Research Institute, University of Tokyo, Kashiwa, Japan
[5]Columbia University - NASA/GISS-E2-R, New York, NY, USA
[6]Aix-Marseille Universit, CNRS, IRD, CEREGE UM34, Aix en Provence, France
[7]Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan
[8]Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan
[9]Alfred Wegener Institute - Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany
[10]Bjerknes 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 displays greater variability within the ensemble compared to the pre-industrial. This variability is highest in the summer months, when the model spread in the mid-Pliocene~~ , and the model spread of extent is more than ~~three times larger than during the rest of the year~~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 simulate ice-free conditions in the mid-Pliocene summer, in contrast to proxy data evidence that suggests the possibility of perennial sea ice. Correlations between mid-Pliocene Arctic temperatures and sea ice extents are almost twice as strong as the equivalent correlations for the pre-industrial simulations~~, suggesting that the dominant atmospheric and oceanic influences on the sea ice may be different in the pre-industrial and mid-Pliocene simulations~~. The need for more comprehensive sea ice proxy data is highlighted, in order to better compare model performances.

1 Introduction

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 CO_2 concentrations resembling those seen in the 21st century, with concentrations estimated to be 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 3.4×10^6 km^2, a reduction of 4.2×10^6 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 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 (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 perennial. 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). 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 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 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. (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 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) ~~. Two different experiments are defined — Experiment 1 is for atmosphere only simulations with prescribed sea ice, with Experiment 2 for coupled atmosphere-ocean general circulation models (GCMs)where the sea ice is explicitly simulated. All simulations use for the mid-Pliocene a modern orbital configuration, 405 ppm atmospheric CO_2, and PRISM3D boundary conditions (Dowsett et al., 2010) . Each modelling group also ran a pre-industrial control simulation.~~

(further details in section 2.1). In this study we analyse the simulation of Arctic sea ice in each of the 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 ~~identify possible mechanisms that define the result of the sea icesimulations~~present an overview of some of the important mechanisms influencing the simulation of sea ice.

2 Methods

~~The simulation of Arctic sea ice by the individual models in the PlioMIP ensemble (see Table 1 for details)for both their~~

2.1 PlioMIP experimental design

Two experimental designs for the PlioMIP simulations are described, Experiment 1 in Haywood et al. (2010) and Experiment 2 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~~is investigated. Pre-industrial results provide an additional climatology against which differences in the models' sea ice outputs can be compared. The consistent experimental design followed by each model reduces the possible causes of disagreement between ensemblemembers (Haywood et al., 2010, 2011) .~~. The PRISM3D reconstruction provides the boundary conditions for the mid-Pliocene simulations, which 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 the SST reconstruction, and evidence from diatoms and sedimentological data (Dowsett et al., 2010) . In both Experiment 1 and Experiment 2, atmospheric CO_2 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. 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.

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, and volume. We follow the example of Berger et al. (2013) and examine the mean sea ice thickness north of 80°~~. Mean sea ice thickness is calculated by dividing the modelled sea ice thickness in each grid cell by the corresponding sea ice concentration. Mean sea ice volume is computed by multiplying the modelled sea ice thickness in each grid cell by the area that the grid cell covers.~~

~~The coefficient of variation (CV), defined as the standard deviation (SD) of different simulations divided by their mean, is calculated to assess the variability among the ensemble members for both metrics. Unlike the standard deviation, the CV allows comparisons of data sets with different mean values, which is a necessity due to offsets in the mean sea ice characteristics between members of the PlioMIP model ensemble. Calculation of the CV identifies the differences in spread between models in each month in the ensemble. The CV has been used in other studies of sea ice simulations, such as Stroeve et al. (2014) , who use the CV to evaluate variability in March sea ice thickness in the ensemble, describing it as a "normalized measure of variability so that variability can be compared spatially and between models.~~

N. To understand differences in the models' simulation of sea ice, we quantify correlations between 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 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

3.1.1 Sea ice extent

Plots of the mean summer and winter pre-industrial Arctic sea ice concentrations are shown in ~~Figures ?? and ?? respectively~~Figure 1. Across the eight-member ensemble, the multi-model mean annual sea ice extent is 16.17 $\times 10^6$ km^2 (Table 2), with a winter (FMA) multi-model mean of

20.90 $\times10^6$ km^2, and a summer (ASO) multi-model mean of 10.98 $\times10^6$ km^2. The individual models' annual means range from 12.27 $\times10^6$ km^2 (IPSLCM5A~~, hereafter IPSL~~) to 19.85 $\times10^6$ km^2 (MIROC4m~~, hereafter MIROC~~) (Table 2), and monthly multi-model means range from a minimum of 10.01 $\times10^6$ km^2 (September) to a maximum of 21.24 $\times10^6$ km^2 (March, Figure ~~3~~2). The lowest individual monthly extent is 7.00 $\times10^6$ km^2 (HadCM3, September), with the highest monthly extent produced by MRI-CGCM (~~hereafter MRI) (~~March), measuring 27.01 $\times10^6$ km^2 (Figure 2).

Figure 2 reveals the differences in the annual sea ice extent cycles across the ensemble. The sea ice extent amplitudes of NorESM-L ~~(herafter NorESM) and IPSL~~ and IPSLCM5A are 6.39 and 7.36 $\times10^6$ km^2 respectively (Table 2). These are the only models in the ensemble with seasonal amplitudes below 10 $\times10^6$ km^2. Other models in the ensemble show a much larger seasonal cycle, in particular GISS-E2-R~~(hereafter GISS), MIROC and MRI~~, MIROC4m and MRI-CGCM, which have sea ice extent amplitudes of 14.03, 14.05, and 15.91 $\times10^6$ km^2 respectively (Table 2). The ensemble mean sea ice extent amplitude is 11.18$\times10^6$ km^2.

3.1.2 Sea ice thickness

North of 80°N, the multi-model mean annual thickness is ~~3.20~~ 2.97 m, with a winter multi-model mean of ~~3.45~~ 3.29 m and a summer multi-model mean of ~~2.81~~ 2.52 m. Across the ensemble, the annual mean thickness varies from ~~2.50 m (NorESM) to 3.98~~ 2.27 m (HadCM3) to 3.81 m (CCSM4~~, hereafter CCSM~~). The winter thicknesses range from ~~2.61 m (NorESM) to 4.08 m (CCSM~~2.56 m (NorESM-L) to 4.01 m (CCSM4), with summer between ~~1.66 m (GISS) and 3.84 m (IPSL).~~ 1.27 m (GISS-E2-R) and 3.60 m (CCSM4). Plots of mean winter and summer pre-industrial Arctic sea ice thicknesses are shown in Figure 3.

~~In the ensemble mean, the regions of thickest sea ice are located polewards from the northern coast of Greenland, and surrounding the more northerly isles of the Canadian Arctic Archipelago. Also along the Greenwich meridian, between 80°N and 90°N, is a region of thicker sea ice (Figure ??). The annual thickness in these regions differs little from the winter sea ice thickness, with only slightly thinner summer sea ice, suggesting a very consistent year round sea ice coverage in these regions.~~

~~The winter spatial thickness pattern shows that sea ice in the Beaufort, Chukchi and East Siberian seas is particularly thick, with thicknesses of 2-4 m, which is thicker in comparison to other regions of comparable latitude — such as the Kara and Barents seas, and in particular the Norwegian sea, where the ice is often less than 1 m thick, if present at all. The annual and summer thicknesses also broadly show this qualitative pattern.~~

~~Most of the models display patterns of~~ Root mean square errors (RMSE) and spatial pattern correlations for mean annual Arctic sea ice thickness ~~that are broadly similar to the overall ensemble mean shown in Figure ??. Yet, there is appreciable variation with respect to the location of maximum ice thickness across the ensemble (Figures ?? and ??). The thickest ice in CCSM is located north of~~

~~Greenland and the Canadian Arctic Archipelago, and the ice thins consistently with distance from this region. For IPSL a similar pattern is found in the summer, although for both summer and winter spatial patterns the region of thicker ice extends much further into the Arctic Basin. The~~ are shown by Figure 4. MIROC4m has the highest spatial pattern correlation with the ensemble mean (0.93), despite the thickest ice in ~~COSMOS, GISS, MRI and NorESM is located in approximately the same region as the thickest ice in the ensemble mean. In COSMOS, the thickest ice is concentrated into a smaller area, and with the exception of this region, the ice thickness reduces with distance from the pole, in contrast to CCSM. For GISS, the region of thickest ice extends in winter in a band from Greenland towards~~ its simulation being located north of Eastern Siberia, ~~passing over the pole. The thinner ice is seen in the Barents Sea and the region north of Alaska and the Canadian mainland. Like in COSMOS, the sea ice in MRI generally thins outwards from the pole, with the areas of greatest thickness also extending further south into the region between western Greenland and Baffin Island. This is also seen in the NorESM simulations, where the winter sea ice is thicker in the region to the west of Greenland than in the band to the north. The sea ice in NorESM generally also thins with distance from the pole, a clear deviation from this trend being the region of maximum sea ice thickness between the North Pole and the Chukchi Sea.~~

~~The MIROC and HadCM3 models simulate thickness spatial patterns that are noticeably different from~~ opposite the region of thickest ice in many of the models (see Figure 3). It also has the lowest RMSE (0.55), marginally lower than COSMOS (0.56). MRI-CGCM displays the lowest spatial pattern correlation with the ensemble mean (0.76) and the highest RMSE (1.33). The lowest spatial pattern correlation between two models is 0.51 (HadCM3 and MRI-CGCM), which have a RMSE of 1.83, the ~~other six models. The pattern displayed by MIROC is almost a 180°-rotation of the ensemble mean sea ice distribution with respect to the location of sea ice extremes. The thickest ice is present north of Eastern Siberia in winter, and thins gradually outwards from a wedge bounded by the 170°E and 130°W lines of longitude. There is also a small patch of thicker ice in the region between Greenland and Baffin Island. The HadCM3 sea ice pattern is not at all similar to~~ highest of the ~~ensemble mean. The thickest ice is situated in a region north of approximately~~ 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 ~~between~~ latitude line, and 120°W and 150°E ~~and around the North Pole. In winter, the ice thickness reduces dramatically outside of this region, dropping by around 2 m, with further thinning southwards. In the summer the contrast is not quite as large, but the general pattern is replicated. Figure **??** illustrates that the PlioMIP ensemble consists of two realisations of pre-industrial summer sea ice, with pronounced sea ice cover in CCSM, IPSL and MRI, and relatively reduced sea ice in the other models.~~

~~Observations of the sea ice thickness detailed in Kwok et al. (2009) give an indication as to the spatial pattern of sea ice thickness within the Arctic and enable an evaluation of modelled pre-industrial sea ice in the PlioMIP ensemble. Figure 6 in Kwok et al. (2009) shows that the thickest sea ice is~~

~~situated in a narrow band north of Greenland and the most northerly islands of the Canadian Arctic Archipelago, resembling the pattern simulated by CCSM. In general, the observed ice becomes thinner with greater distance from the region of highest thickness. Whilst the regions of thickest sea ice are similar in ensemble mean and observations,~~ (see Figure 3). However, it has a greater spatial pattern correlation with the ~~simulated pattern for the Arctic basin indicates rather a reduction in thickness with distance from the pole. Aside from this difference, the~~ ensemble mean than GISS-E2-R or MRI-CGCM, and the RMSE between the ensemble mean thickness ~~patterns appear to broadly match the observations from Kwok et al. (2009) .~~

~~The degree to which individual models match the observed thickness patterns is variable. CCSM produces what appears to be the closest pattern to observations, with IPSL being similar in the summer. Yet, the extension of the large region of thicker ice particularly in winter prevents IPSL from being as close to the observations as CCSM. The spatial patterns of sea ice thickness simulated by COSMOS, GISS, MRI and NorESM show some similarity to patterns of CCSM, and therefore also to the observations. As MIROC and~~ and HadCM3 ~~show very different patterns to the other models, their thickness spatial patterns are less similar to the observational spatial patterns from Kwok et al. (2009)~~ is lower than GISS-E2-R or MRI-CGCM when compared to the ensemble mean (Figure 4).

3.2 Pliocene simulations

3.2.1 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 ~~??, ??, ??, and ??~~1 and 5). The multi-model mean annual extent for the mid-Pliocene simulations is 10.84 $\times 10^6$ km^2, a reduction of 5.33 $\times 10^6$ km^2 (33.0%) in comparison to the respective multi-model mean of the pre-industrial simulations. Annual means in the ensemble range from 7.60 $\times 10^6$ km^2 (~~NorESM~~NorESM-L), to 15.84 $\times 10^6$ km^2 (~~MRI~~MRI-CGCM) (Table 1).

The lowest multi-model monthly mean extent is 3.15 $\times 10^6$ km^2 (September), and the highest is 16.59 $\times 10^6$ km^2 (March). In comparison to the pre-industrial simulation, the lowest multi-model monthly mean extent is reduced by 6.86 $\times 10^6$ km^2 (69%). The reduction for the highest monthly multi-model mean is 4.65 $\times 10^6$ 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, the mid-Pliocene is characterized by an enhanced seasonal cycle of sea ice extent, with severely reduced sea ice during boreal summer.

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

~~MRI, CCSM and MIROC~~ MRI-CGCM, CCSM4 and MIROC4m simulate the highest maximum mid-Pliocene sea ice extents in the ensemble. Both ~~CCSM and MRI~~ CCSM4 and MRI-CGCM also provide the highest two minimum extents, but ~~MIROC~~ MIROC4m is one of the four models that simulates an ice-free Arctic summer. As a result, the sea ice extent amplitude in ~~MIROC~~ MIROC4m in the mid-Pliocene simulations is $\approx 64\%$ greater than the pre-industrial simulation extent amplitude (Table 2). ~~This finding also holds for the ensemble mean, although at a lower amplitude extent amplitude.~~ 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 in the mid-Pliocene simulations. Not all of the models, however, show this trend. Only five models (the four with ice-free summers and HadCM3) simulate a higher mid-Pliocene sea ice extent amplitude, the remaining three models simulate a (slightly) lower annual cycle in the mid-Pliocene simulations (Table 2).

3.2.2 Sea ice thickness

Plots of the mean summer and winter mid-Pliocene Arctic sea ice thicknesses are shown in ~~Figures ?? and ?? respectively~~Figure 7. The multi-model mean annual sea ice thickness is ~~1.48~~ 1.30 m, which, compared to the pre-industrial simulations, is a reduction of ~~1.72 m (53.9~~1.7 m (56%). Across the ensemble, the annual mean thicknesses range from ~~0.46 m (NorESM) to 2.08 m (MRI~~0.44 m (NorESM-L) to 2.56 m (MRI-CGCM). The multi-model winter mean thickness is ~~1.85 m, 1.60 m (46.4~~1.77 m, 1.5 m (46%) less than the pre-industrial, whereas the summer multi-model mean thickness drops by ~~1.81 m (64.4~~1.8 m (71%) to ~~1.00~~ 0.74 m. Similarly to the sea ice extent, the summer sea ice thickness 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 ~~0.90 m (NorESM) to 2.80 m (MRI~~0.79 m (NorESM-L) to 2.78 m (MRI-CGCM), with the summer sea ice thicknesses between ~~0.08 m (NorESM) and 2.30 m (MRI~~0.3 m (NorESM-L) and 2.24 m (MRI-CGCM).

~~Many of the models display similar spatial patterns of sea ice thickness in the~~ Spatial pattern correlations and RMSEs between the pre-industrial and mid-Pliocene simulations ~~as they do in the~~ are shown in Figure 4. All but five of the mid-Pliocene RMSEs are lower than the equivalent RMSE for the pre-industrial ~~, although the thickness values are reduced, particularly in the summer. The sea ice thickness spatial patterns simulated by CCSM, HadCM3, IPSL and MRI are very similar to their~~ simulations. This trend is not seen in the spatial pattern correlations, where just over half (19 out of 36) of the mid-Pliocene correlations are higher than the corresponding pre-industrial

~~equivalents in both summer and winter. The other four model simulations are ice-free for the majority of the summer~~ correlation. These results show that the differences in thicknesses between the models are lower in the mid-Pliocene ~~, so no thickness pattern is detectable. In MIROC the mid-Pliocene simulation shows similar patterns in the winter to its pre-industrial counterpart. Similar findings hold for COSMOS, although in this model the central Arctic sea ice thins by a greater amount in comparison to the ice in other regions. In GISS, the ice north of Greenland and the Canadian Arctic Archipelago thins more than in other regions, so during~~ simulations, but the thickness patterns are overall no more or less similar. Lower overall RMSEs 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 with the ~~region of greatest sea ice thickness is in this simulation north of Eastern Siberia. In the simulation with NorESM all sea ice to the north and east of Greenland is lost,~~ 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 compared to the ensemble of all the simulations (1.05), and the RMSE of 1.46 between MRI-CGCM and NorESM-L is the ~~thick sea ice, that is in the~~ highest between any two models. The highest spatial pattern correlation between two models is 0.97, between COSMOS and MIROC4m, which also have the lowest RMSE, at 0.11.

Figure 4 also shows RMSEs and spatial pattern correlations between each model's pre-industrial ~~simulation located to the west of Greenland, thins considerably.~~ and mid-Pliocene runs. All but two models had spatial pattern correlations 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 between the ensemble means is 0.79.

3.3 Variability across the ensemble

~~Figures ?? and ?? suggest that there is greater variability across the eight PlioMIP modelsin their mid-Pliocene simulation of summer sea ice compared to winter sea ice. This inference is in the following further studied in Figure ??, which shows the CV of both the sea ice extent and thickness in the ensemble for each month, for~~ The standard deviation (SD) of the monthly ensemble sea ice extents and thicknesses for both the pre-industrial and mid-Pliocene simulations ~~.~~

~~The~~ is shown in Figure 8. In each month from December to June, the mid-Pliocene extent SD is lower than the pre-industrial ~~sea ice extent CV is low and relatively stable throughout the year, with nine of the months having values between 0.19 and 0.22. The June to August CV is slightly lower, a minimum of 0.116 occurring in July. The mid-Pliocene simulation shows a much greater contrast between the monthly extremes, with a minimum of 0.181 in June , and a maximum of 1.16 in September. There is a sharp increase in CV during the summer months, which contrasts with the pre-industrial simulation where the summer is characterized by slightly lower CV values than the~~

~~rest of the year. The large increase in~~ extent SD. During these months, the maximum extent SD in both simulations occurs in February, and SD decreases each month from Feburary 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 ~~summer CV supports the initial impression that across the ensemble there is greater variability of sea ice extent~~ 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 ~~summer if compared to the remaining months in the mid-Pliocene or the entirety of the~~ extent. The annual cycle of of pre-industrial ~~simulation.~~

~~The CV of the mid-Pliocene~~ sea ice thickness ~~is greater than in the pre-industrial ensemble for each month (Figure ??). In both experiments, the highest CV occurs during the summer months, which is also when the difference between the mid-Pliocene and pre-industrial CV is greatest. The pre-industrial thicknesses show greater overall variation in comparison to the pre-industrial extent. For~~ SD has a minimum in May, and maximum in September. The mid-Pliocene sea ice thickness ~~and extent the peak CV values are similar , but over the year there is more variability in simulated sea ice thickness than in sea ice extent~~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.

3.4 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 ~~Assessment of pre-industrial~~ 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. ~~Hence, the historical data sets are ignored here, and the analysis is performed with satellite-based recent sea ice data.~~

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.

~~First, we assess the simulated pre-industrial sea ice extent.~~ Shu et al. (2015) provides an ~~analysis of the~~ assessment of the historical simulation of Arctic sea ice by the CMIP5 models ~~. Of the~~ for the period 1979-2005. Their results show that for the historical simulations by the 7 PlioMIP models represented in CMIP5, ~~MRI~~ MRI-CGCM simulates the highest mean annual sea ice extent (15.01×10^6 km^2), compared to the satellite observational mean of 12.02×10^6 km^2 ~~. MRI~~ for the comparable period (1979-2005). MRI-CGCM simulates the second highest PlioMIP pre-industrial mean annual sea ice extent (just 0.05×10^6 km^2 less than ~~MIROC~~MIROC4m), and the highest mid-Pliocene mean annual sea ice extent. The CMIP5 historical extent simulated by ~~MRI~~ MRI-CGCM is almost

25% greater than the observational mean, ~~which may suggest that~~ showing MRI-CGCM consistently simulates Arctic sea ice ~~simulated by MRI is too extensive~~extent larger than the ensemble mean.

In contrast, ~~MIROC simulates a~~ MIROC4m simulates a PlioMIP pre-industrial mean annual sea ice extent that is similar to the ~~MRI~~ 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 $\times 10^6$ km^2, Shu et al. (2015)). The ~~NorESM~~NorESM-L, which simulates both the lowest PlioMIP pre-industrial and mid-Pliocene mean annual sea ice extents, is the CMIP5 model ~~that simulates the~~ which simulates the closest historial mean annual sea ice extent ~~that is closest~~ to the observations (12.01 $\times 10^6$ km^2~~) — although, like in the~~, just 0.01×10^6 km^2 lower than the observations). As with the PlioMIP pre-industrial simulations, ~~NorESM~~ NorESM-L simulates the lowest sea ice extent amplitude of the PlioMIP models in CMIP5 (Shu et al., 2015) .

~~The HadCM3~~ In addition to the mean annual sea ice extent simulated by each model in the CMIP5 ~~pre-industrial simulation has a greater mean annual extent than the observations and exceeds the mean CMIP5 extent of the PlioMIP models. This contrasts to its pre-industrial and mid-Pliocene simulations in PlioMIP that are lower than~~ historical and PlioMIP simulations, Table 2 shows the ensemble mean ~~. Similarly, the CCSM CMIP5 pre-industrial mean annual sea ice extent is less than the PlioMIP mean sea ice extent , whereas CCSM simulates an extent that is above the mean in both pre-industrial and mid-Pliocene simulations. The GISS CMIP5 pre-industrial extent is greater than the PlioMIP mean , but its mid-Pliocene simulation is below the PlioMIP ensemble mean. For IPSL it is found that the simulation is below the mean in pre-industrial, mid-Pliocene~~ 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 ~~, although the CMIP5 simulation is closer to the respective ensemble mean than~~ historical simulations, the ~~other two simulations.~~ ensemble mean annual extent is greater than the CCSM4 mean annual extent, and lower than the mean annual extent simulated by HadCM3.

~~It is important to note that the versions of MIROC, MRI and NorESM used for CMIP5 are slightly different to the versions used for PlioMIP. The version MIROC4h (Sakamoto et al., 2012) is used for CMIP5. It represents a higher resolution version of MIROC4m, which was used for PlioMIP. The version of NorESM used for CMIP, NorESM-M (Bentsen et al., 2013) , is similarly a higher resolution version of NorESM in PlioMIP. The MRI-CGCM3 (Yukimoto et al., 2012) of CMIP5 is an updated version of the MRI-CGCM2.3 model used in PlioMIP.~~

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.

4.2 ~~Assessment of mid-Pliocene~~ Mid-Pliocene simulations

Four models out of the eight-member PlioMIP ensemble (COSMOS, ~~GISS, MIROC and NorESM) simulate~~ GISS-E2-R, MIROC4m and NorESM-L) simulate almost ice-free conditions in the mid-Pliocene summer~~, whereas the remaining four models simulate year-round sea ice coverage~~. For those models that simulate summer sea ice in the mid-Pliocene the summer sea ice conditions vary ~~. The~~ 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~~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 km^2, 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 km^2, 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 1 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~~. If on the other hand a model~~ , 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). ~~A~~ However, a 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.

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 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°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 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 (~~CCSM, IPSL and MRI~~CCSM4, IPSLCM5A and MRI-CGCM) — although the sea ice concentration at these sites is less than 50% in the ~~CCSM and IPSL~~ 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 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 ~~appears to be in~~ has the closest agreement with the proxy data indications from Darby (2008) and Knies et al. (2014), but greater data coverage may provoke a different conclusion.

4.3 Causes of PlioMIP ensemble variability

4.3.1 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 details of each model's sea ice component are summarised in Table 1. The models ~~CCSM and NorESM~~ CCSM4 and NorESM-L use the same sea ice component, based on CICE4 (Hunke, 2010), although ~~NorESM~~ NorESM-L has a coarser model grid in the atmosphere than ~~CCSM~~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, ~~CCSM, NorESM, and MIROC~~CCSM4, NorESM-L, and MIROC4m, that all use the elastic-viscous-plastic (EVP) rheology of Hunke and Dukowicz (1997). Second, COSMOS, ~~GISS, and IPSL~~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~~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 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 and MIROC~~ 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~~lower than some of the observations~~, i.e. ~~NorESM, IPSL~~NorESM-L, IPSLCM5A, and HadCM3, ~~as well employ different rheology~~ employ different rheologies — EVP, VP and no rheology respectively.

~~The dynamics also do not appear to be a strong influencing factor on the simulated sea ice thickness. We might expect the models with the most basic sea ice dynamics to simulate thickness most poorly, as the model would not account for higher-order effects, such as ridging in the ice. However, whilst the spatial pattern of sea ice thickness simulated by HadCM3 compares poorly with observations, the spatial patterns simulated by MRI resemble some aspects of the observational patterns, despite the lack of sea ice rheology. The sea ice thickness spatial patterns in MIROC, which uses the more sophisticated EVP rheology, do not compare favourably to the sea ice observations.~~

Most of the models use a leads parameterisation in their sea ice thermodynamics component, with only ~~CCSM and NorESM~~ 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, ~~MIROC and MRI~~ 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 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 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, 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 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 latitudes 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.

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 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.

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 influence of the sea ice parameterisation on pre-industrial and mid-Pliocene simulations.

General circulation models are tuned to best reproduce modern day climate conditions, and parameterisations are based on modern observations (Hunke, 2010; Mauritsen et al., 2012). When simulating 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 observational 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).

4.3.2 Influence of the control simulation

Massonnet et al. (2012) describe the characteristics of Arctic sea ice simulated by the CMIP5 ensemble for the time period from 1979-2010 as being related in a 'complicated manner' to the simulated 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 4.1 it was highlighted that ~~in CMIP5 the relative performance of the~~ the differences in the PlioMIP models' simulation of sea ice for 1979-2005 in CMIP5 ~~is not the same as in the~~ are not consistent with the differences in pre-industrial or mid-Pliocene simulations in the PlioMIP ensemble.

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. Holland 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 climates than sea ice extent. Massonnet et al. (2012) find that those CMIP5 models that predict an 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 and 0.51. The relatively thin pre-industrial summer sea ice simulated in PlioMIP by COSMOS, ~~GISS, MIROC and NorESM~~ 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.

4.3.3 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 ~~stronger~~ 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. 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 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). ~~This explanation does not apply, however, to the SATs, for which a similar difference in correlation strengths with~~ In the models, the presence of ice in a grid box, even at low 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 ~~between the pre-industrial and~~ and surface temperatures in the mid-Pliocene ~~is present~~simulations.

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. (2013a)~~ Zhang et al. (2013b) analyse the simulation of the AMOC in both pre-industrial and mid-Pliocene simulations of 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, IPSL and MRI~~GISS-E2-R, IPSLCM5A and MRI-CGCM), only two (COSMOS and ~~GISS~~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.

~~The simulation of~~ An analysis of multi-decadal variability influence on Arctic sea ice ~~by means of GCMs has been demonstrated to be very sensitive to the parameterisation of sea icealbedo. 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 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, MIROC, and NorESM). The NorESM shows the largest warming due to clear sky albedo, CCSM on the other hand shows the smallest clear sky albedo effect. Both NorESM and CCSM use the same sea ice component, based on CICE4 (Hunke and Lipscomb, 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 (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 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 and CCSM4 is reflected in the large difference in their simulations of summer mid-Pliocene sea ice. One cause is certainly the nature of~~ 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 ~~sea-ice albedo feedback mechanism (Curry et al., 1995) . Reduced albedo at high latitudes can be both a cause of and a result of a reduced sea ice extent. Models with parameterisations that produce lower sea ice albedo 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 is a likely explanation for the low sea ice extents it simulates (Figures 2 and 6), both in mid-Pliocene and pre-industrial simulations~~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.

~~Second to NorESM, for MIROC clear sky albedo has the highest contribution to high latitude warming. In MIROC 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 and CCSM), only GISS has an albedo minimum 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 MIROC has a lower overall albedo. This may help to explain the ability of MIROC to simulate an ice-free~~ 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 ~~summer, despite simulating one of the highest winter sea ice extents for both~~ Arctic sea ice extents and NAO indices are significant at the 90% level. The correlations between pre-industrial ~~and~~ 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.

~~As the parameterisation of~~ 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 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 ~~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~~ 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) . ~~However, if there is a temperature threshold above which~~

Figure 12 shows the mean annual 10 m surface winds for the ~~ice-albedo feedback becomes more dominant in some of~~ COSMOS and MIROC4m mid-Pliocene simulations, where 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 IPSLCM5A pre-industrial simulation (Figure 12), the ~~models, then this could explain the different influence~~ 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. Mean annual 10 m winds and sea ice thicknesses for all simulations (excluding CCSM4, for which 10m 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, 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 ~~parameterisation on pre-industrial and mid-Pliocene simulations.~~

~~Finally, atmospheric and oceanic variability, such as the North Atlantic Oscillation (Hurrell et al., 2001) and Atlantic multi-decadal oscillation (Schlesinger and Ramankutty, 1994) , have been demonstrated to influence Arctic sea ice extent (Kwok, 2000; Day et al., 2012) . Further study of their effect on~~ margins simulated by the model. Mean annual surface ocean currents and sea ice ~~simulation in PlioMIP is not possible since run lengths and averaging periods of the PlioMIP simulations are not equal (Table 1). This makes determining the effect, that any multi-decadal variability has on the simulations, difficult to determine~~thicknesses for all simulations are included in the supplementary information.

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.

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 54% reduction in the ensemble 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.

~~For the pre-industrial simulations there is a relatively consistent level of inter-model variability in the simulation of sea ice extent over the year, with only a slight decrease in the summer. In contrast, the inter-model variability in the simulated mid-Pliocene sea ice extent is much enhanced in the summer months. Thickness variability is highest during summer in both climate states, and is higher for the mid-Pliocene simulations throughout the year.~~

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 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 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.

~~Most models show similar patterns in the distribution of relative ice thickness, with HadCM3 and MIROC being obvious exceptions. HadCM3 also produces the thinnest pre-industrial sea ice, suggesting that the model generally has difficulty in simulating observed sea ice thickness. It is particularly noteworthy that this general difficulty does not prevent the model from simulating perennial sea ice in the mid-Pliocene Arctic Ocean, which is in contrast to half of the models in the ensemble. HadCM3 is therefore consistent with the findings of perennial Arctic sea ice in the mid-Pliocene by Darby (2008) .~~

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 IP_{25} proxy data indicates the presence of a sea ice margin in the mid-Pliocene. ~~This appears to suggest that~~ HadCM3 therefore produces the mid-Pliocene simulation that is in best agreement with ~~both inferences from the proxy record~~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 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 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 is 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 judgements 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 4.3.3, 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 Laboratory 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 Promotion of Science and computing resources at the Earth Simulator Center, JAMSTEC. Y. Kamae acknowledges S. Yukimoto, O. Arakawa, and A. Kitoh in Meteorological Research Institute in Japan for providing source code of the ~~MRI~~ MRI-CGCM model. Z. Zhang would like to thank M. Bentsen, J. Tjiputra, and I. Bethke from Bjerknes Center for Climate Research for the contribution to the development of ~~NorESM~~NorESM-L.

References

Arzel, O., Fichefet, T., and Goosse, H.: Sea ice evolution over the 20th and 21st centuries as simulated by 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 IP$_{25}$: a review of current understanding, recommendations for future research and applications in palaeo sea ice reconstructions, Quaternary Science Reviews, 79, 9–25, 2013.

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, Ø., Drange, H., Roelandt, C., Seierstad, I. A., Hoose, C., and Kristjánsson, J. E.: The Norwegian Earth System Model, NorESM1-M – Part 1: Description and basic evaluation of the physical climate, Geoscientific Model Development, 6, 687–720, doi:10.5194/gmd-6-687-2013, 2013.

Berger, M., Brandefelt, J., and Nilsson, J.: The sensitivity of the Arctic sea ice to orbitally induced insolation 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, 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 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 Warm Period using two versions of the NASA/GISS ModelE2-R Coupled Model, Geoscientific Model Development, 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.

Cronin, T. M., Whatley, R., Wood, A., Tsukagoshi, A., Ikeya, N., Brouwers, E. M., and Briggs, W. M.: Microfaunal 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.

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 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.

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, l04502, 2008.

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.

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.

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.

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.

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.

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, l18501, 2011.

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.

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.

Hurrell, J. W., Kushnir, Y., and Visbeck, M.: The North Atlantic Oscillation, Science, 291, 603–605, doi:10.1126/science.1058761, 2001.

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.

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.

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.

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, C07 005, doi:10.1029/2009jc005312, 2009.

Liu, J., Schmidt, G. A., Martinson, D., Rind, D. H., Russell, G. L., and Yuan, X.: Journal of Geophysical Research, doi:10.1029/2001JC001167.

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.

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 Discussions, 6, 2931–2959, doi:10.5194/tcd-6-2931-2012, 2012.

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.

Mellor, G. L. and Kantha, L.: An ice-ocean coupled model, Journal of Geophysical Research-Oceans, 94, 10 937–10 954, 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.

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.

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/ngeo724, 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 preconditioning and an August storm, Geophysical Research Letters, 40, 1356–1361, doi:10.1002/grl.50349, 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.

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 Period with the CCSM4 model, Geoscientific Model Development, 6, 549–561, doi:10.5194/gmd-6-549-2013, 2013.

Sakamoto, T., Komuro, Y., Nishimura, T., Ishii, M., Tatebe, H., Shigoama, H., Hasegawa, A., Toyoda, T., Mori, M., Suzuki, T., Imada, Y., Nozawa, T., Takata, K., Mochizuki, T., Ogochi, K., Emori, S., Hasumi, H., and Kimoto, M.: MIROC4h - A New High-Resolution Atmosphere-Ocean Coupled General Circulation Model, Journal of the Meteorological Society of Japan. Ser. II, 90, 325–359, doi:10.2151/jmsj.2012-301, 2012.

Schlesinger, M. and Ramankutty, N.: An oscillation in the global climate system of period 65-70 years, Nature, 367, 723–726, 1994.

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., 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 pCO_2 records, Earth and Planetary Science Letters, 292, 201–211, 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.

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 satellites 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.

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, Geophysical Research Letters, 39, L18 501, doi:10.1029/2012gl052868, 2012.

Yukimoto, S., Adachi, Y., Hosaka, M., Sakami, T., Yoshimura, H., Hirabara, M., Tanaka, T. Y., Shindo, E., Tsujino, H., Deushi, M., et al.: A new global climate model of the Meteorological Research Institute: MRI-CGCM3—model description and basic performance—, Journal of the Meteorological Society of Japan, 90, 23–64, 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.

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., Risebrobrakken, B., Andersson, C., 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 resolution (° lat × ° long)	Ocean resolution (° lat × ° long)	Length of run/ averaging period (years) Pre-industrial	mid-Pliocene	Sea Ice components and references	Reference
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.	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	$0.5 - 2 \times 2$	2800/100	730/30	VP rheology, leads Fichefet and Morales Maqueda (1999)	Contoux et al. (2012)
MIROC4m	2.8×2.8	$0.5 - 1.4 \times 1.4$	3800/100	1400/100	EVP rheology, leads K-1 Model Developers (2004)	Chan et al. (2011)
MRI-CGCM	2.8×2.8	$0.5 - 2 \times 2.5$	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 km^2.

Model	PI mean ~~annual extent~~ ~~($\times 10^6$ km^2)~~ annual extent	PI extent ~~amplitude~~ ~~($\times 10^6$ km^2)~~ amplitude	mid-Pliocene mean~~annual~~ ~~extent ($\times 10^6$ km^2)~~ annual extent	mid-Pliocene extent amplitude	CM ann
CCSM4	18.35	10.94	14.99	10.26	
COSMOS	15.52	11.66	7.72	12.75	
GISS-E2-R	17.30	14.03	9.63	15.43	
HadCM3	13.76	12.42	10.38	14.17	
IPSLCM5A	12.27	7.36	9.06	7.05	
MIROC4m	19.85	14.05	11.48	21.98	
MRI-CGCM	19.80	15.91	15.84	13.69	
NorESM-L	12.52	6.39	7.60	12.86	
Ensemble mean	16.17	11.18	10.84	13.44	

~~Mean winter (FMA) sea ice concentrations () 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 ??, ??, ??, ??, ??, ??, ?? and ??).~~

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(AMO,SIE)	*r*(AMO,SIE)	*r*(AMO,SIE)	*r*(AMO,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(NAO,SIE)	*r*(NAO,SIE)	*r*(NAO,SIE)	*r*(NAO,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

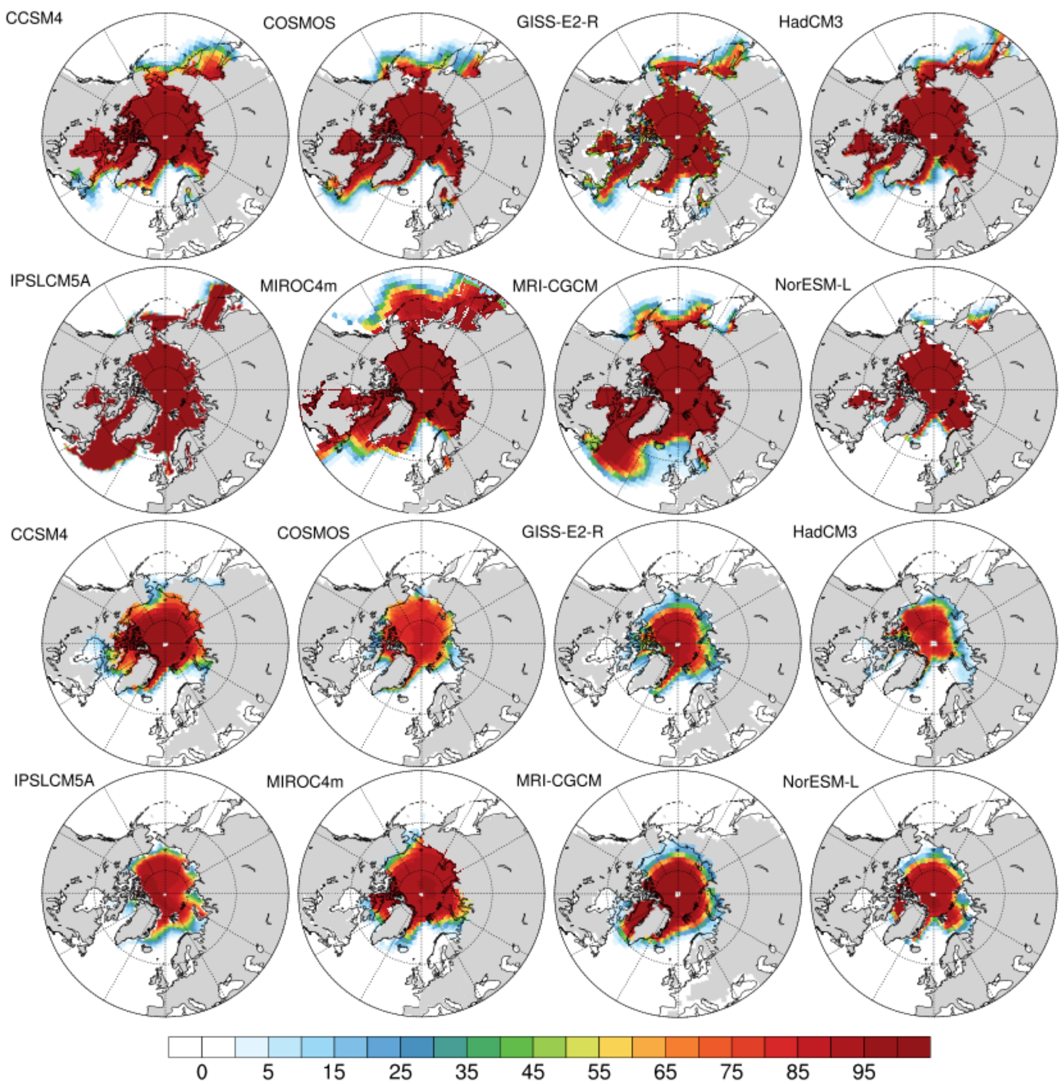


Figure 1. ~~As Figure ??, but~~ Mean sea ice concentrations (%) for ~~mean~~ winter (FMA, upper half) and summer (ASO, lower half) ~~sea ice concentrations~~ in the pre-industrial control simulations for each PlioMIP Experiment 2 model. Missing data at the poles is a plotting artefact (seen also in Figures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

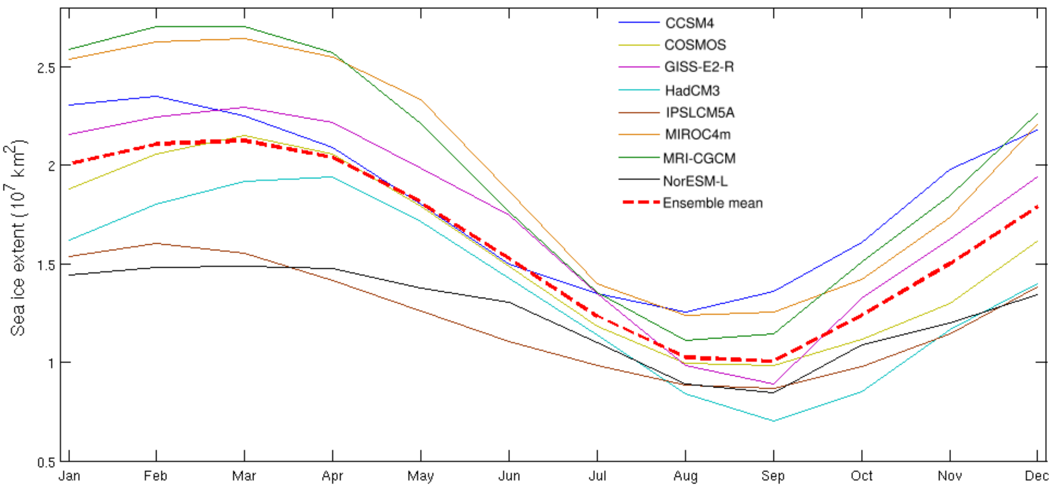
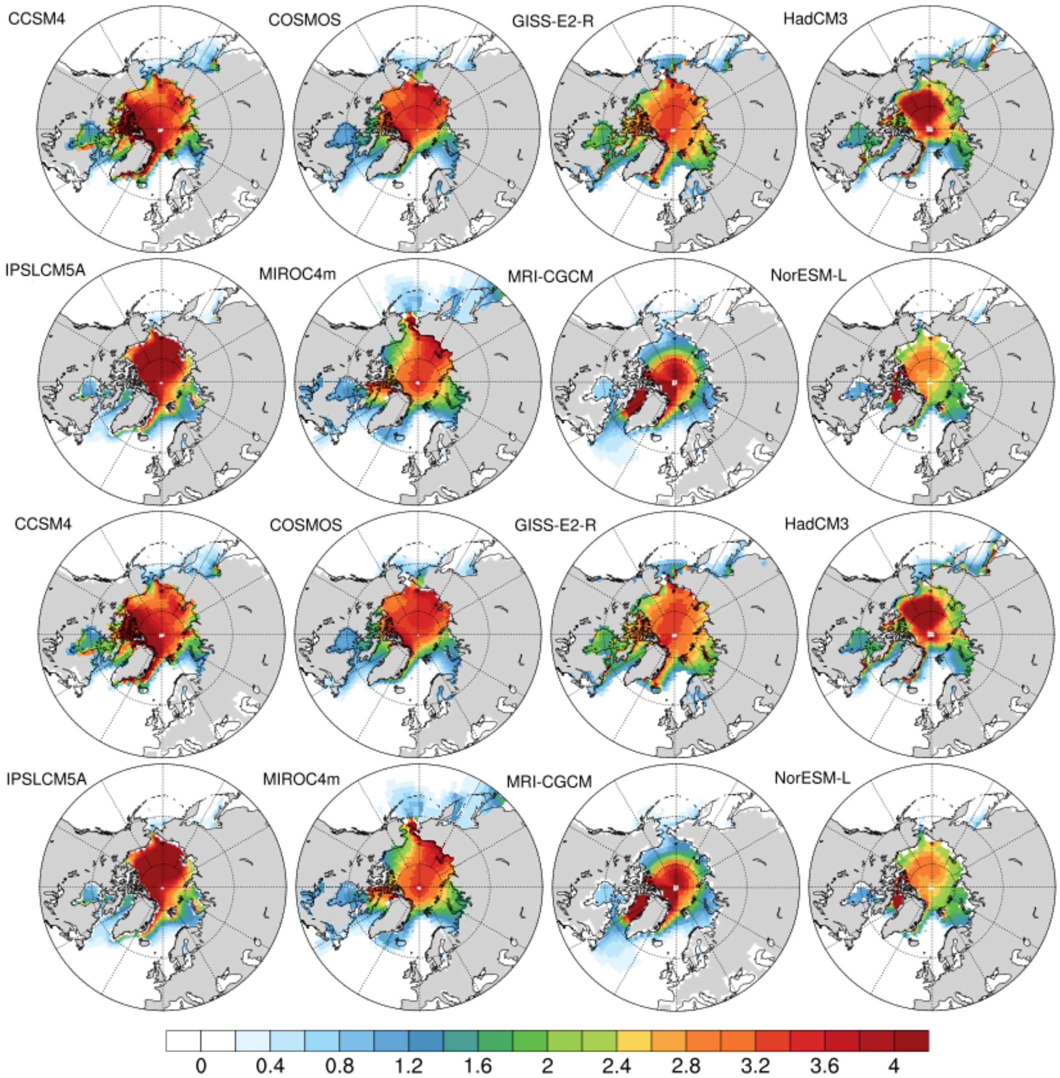


Figure 2. Annual cycle of total Arctic sea ice extent in the pre-industrial simulations for each participating model in PlioMIP Experiment 2, and the ensemble mean.



~~Mean winter (FMA) sea ice thicknesses (m)~~ in the pre-industrial control simulations for each PlioMIP Experiment 2 model.

~~Mean winter (FMA) sea ice thicknesses (m)~~ in the pre-industrial control simulations for each PlioMIP Experiment 2 model.

Figure 3. Mean sea ice ~~thickness~~ thicknesses (m) ~~in the pre-industrial simulations~~ for ~~the entire PlioMIP Experiment 2 ensemble, for (a) annual, (b)~~ winter (FMA~~)~~, ~~and (c~~upper half) and summer (ASO, lower half)
~~.~~

~~Mean winter (FMA) sea ice thicknesses (m)~~ in the pre-industrial control simulations for each PlioMIP Experiment 2 model.

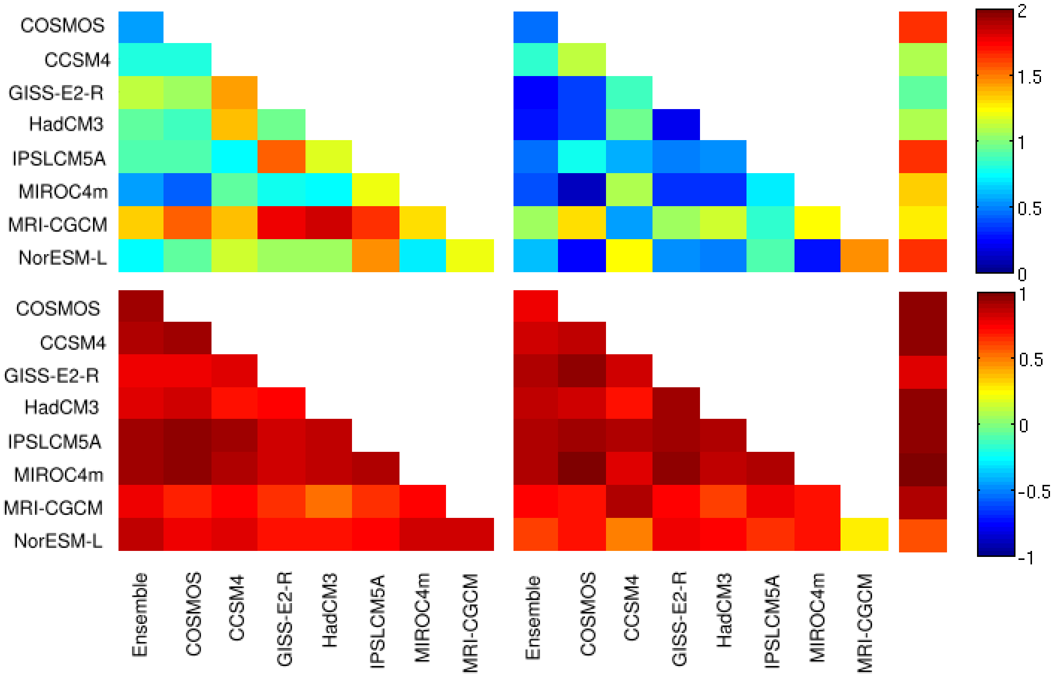


Figure 4. ~~As Figure ??, but for mean summer~~ Root-mean-square error (~~ASO~~RMSE) ~~sea~~ (top) and spatial pattern correlations (bottom) of Arctic ice ~~thicknesses~~ thickness in the pre-industrial (~~m~~left) and mid-Pliocene (right) simulations by the PlioMIP models and ensemble mean. The single columns to the right show the RMSE and spatial pattern correlations for between each model's pre-industrial and mid-Pliocene mean annual Arctic sea ice thickness.

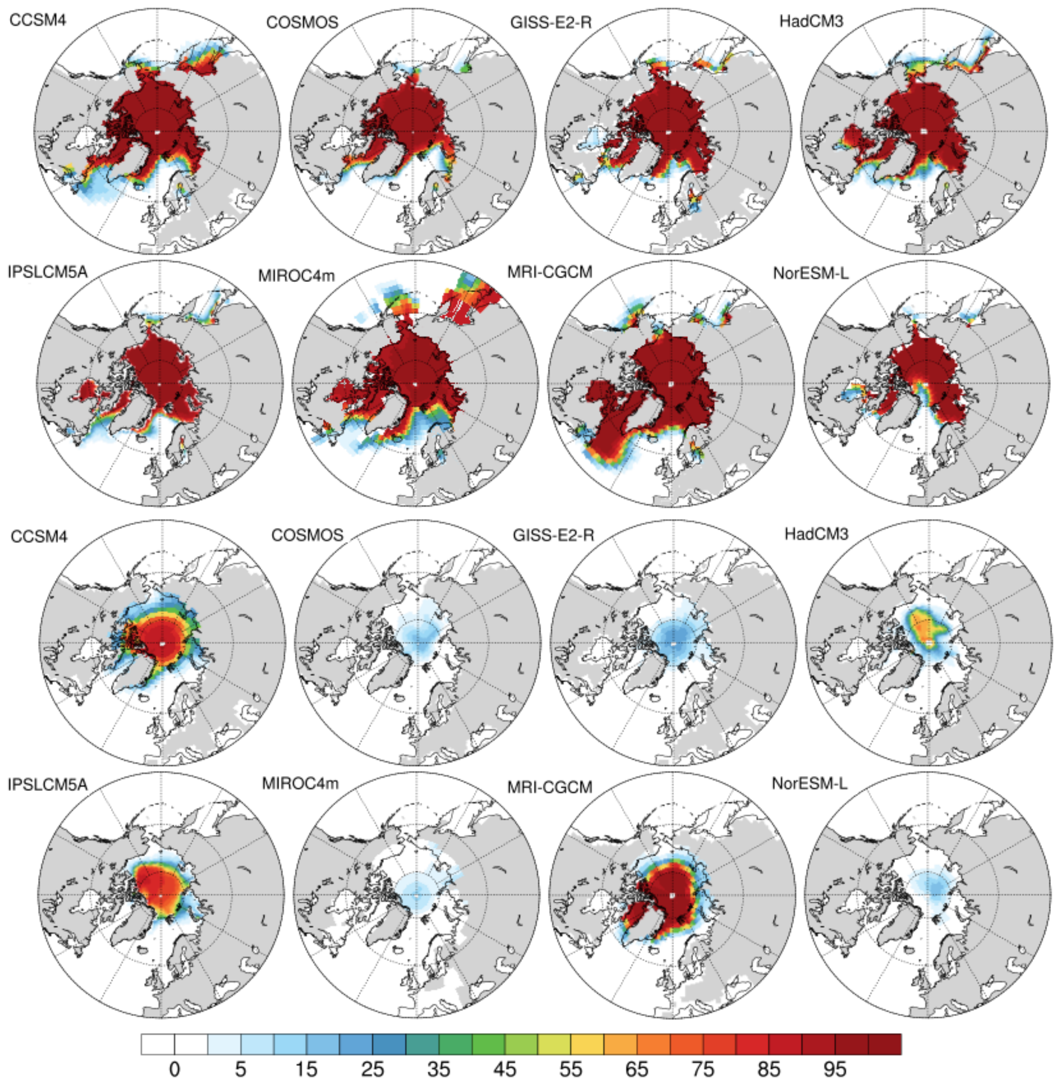


Figure 5. Mean ~~winter (FMA)~~ sea ice concentrations (%) for winter (FMA, upper half) and summer (ASO, lower half) in the mid-Pliocene simulations for each PlioMIP Experiment 2 model.

~~As Figure ??, but for mean summer (ASO) sea ice concentrations ().~~

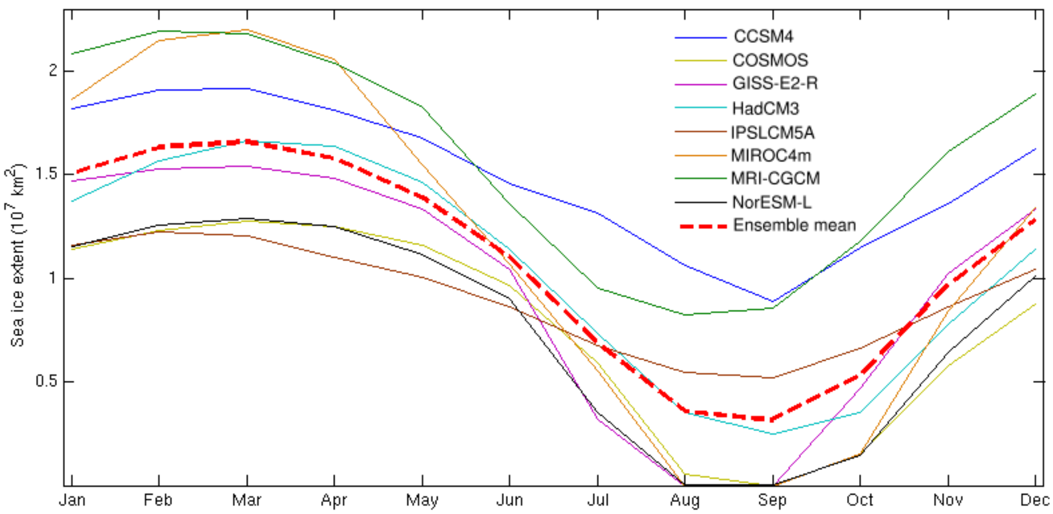
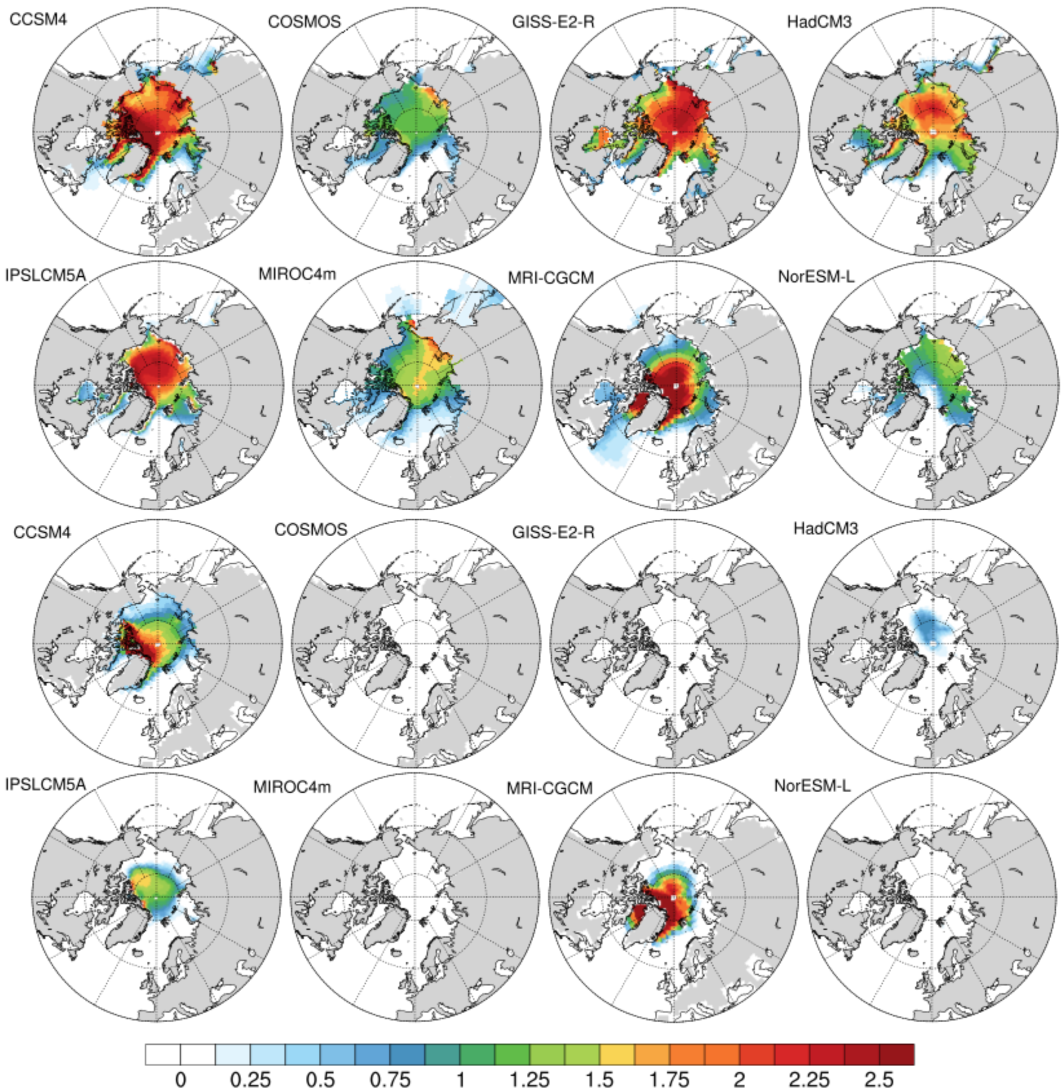


Figure 6. Annual cycle of sea ice extent in the mid-Pliocene simulations for each participating model in PlioMIP Experiment 2 and for the ensemble mean.



~~As Figure ??, but for mean summer (ASO) sea ice thicknesses (m).~~ Low sea ice concentrations in the summer plots for COSMOS, ~~GISS~~GISS-E2-R, ~~MIROC~~ MIROC4m and ~~NorESM~~ NorESM-L result in mean thicknesses very close to zero in each model grid cell.

~~As Figure ??, but for mean summer (ASO) sea ice thicknesses (m).~~ Low sea ice concentrations in the summer plots for COSMOS, ~~GISS~~GISS-E2-R, ~~MIROC~~ MIROC4m and ~~NorESM~~ NorESM-L result in mean thicknesses very close to zero in each model grid cell.

Figure 7. Mean ~~winter (FMA)~~ 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.

~~As Figure ??, but for mean summer (ASO) sea ice thicknesses (m).~~ Low sea ice concentrations in the summer plots for COSMOS, ~~GISS~~GISS-E2-R, ~~MIROC~~ MIROC4m and ~~NorESM~~ NorESM-L result in mean thicknesses very close to zero in each model grid cell.

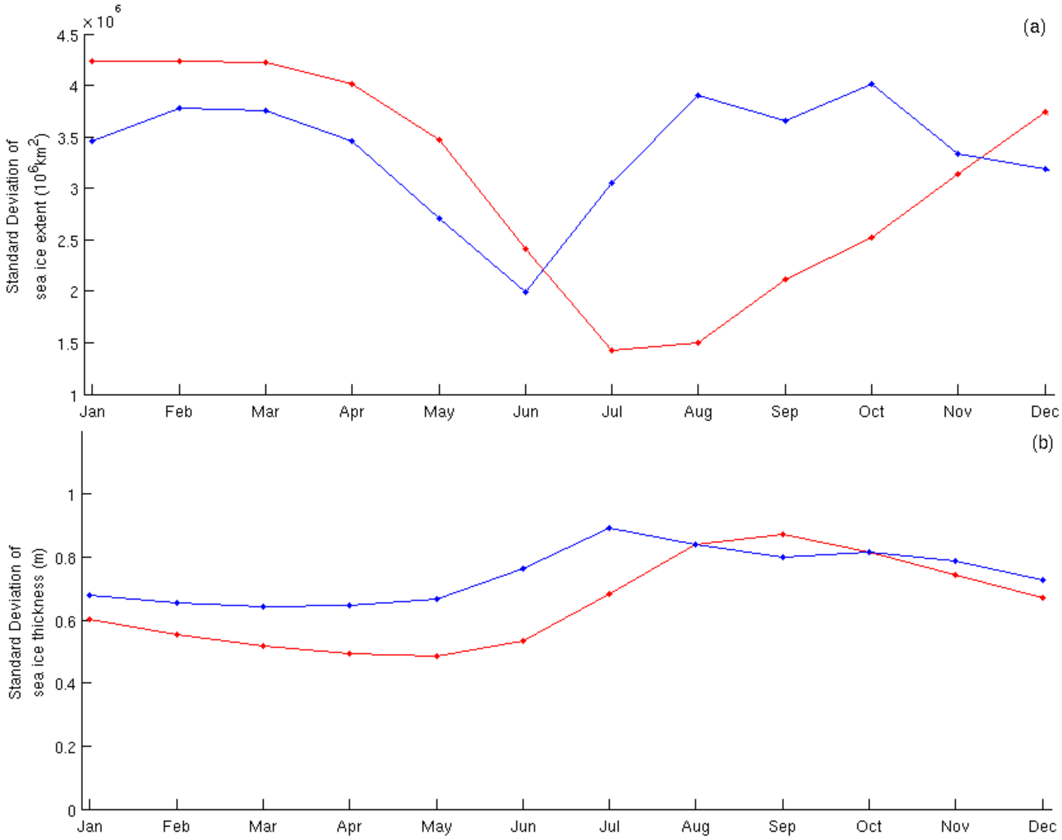


Figure 8. Annual cycle of the ~~coefficient~~ standard deviation of ~~variation (CV) of (~~a) sea ice extent and (b) sea ice thickness for the PlioMIP Experiment 2 ensemble. Red lines represent the pre-industrial annual cycle, blue lines represent the mid-Pliocene annual cycle.

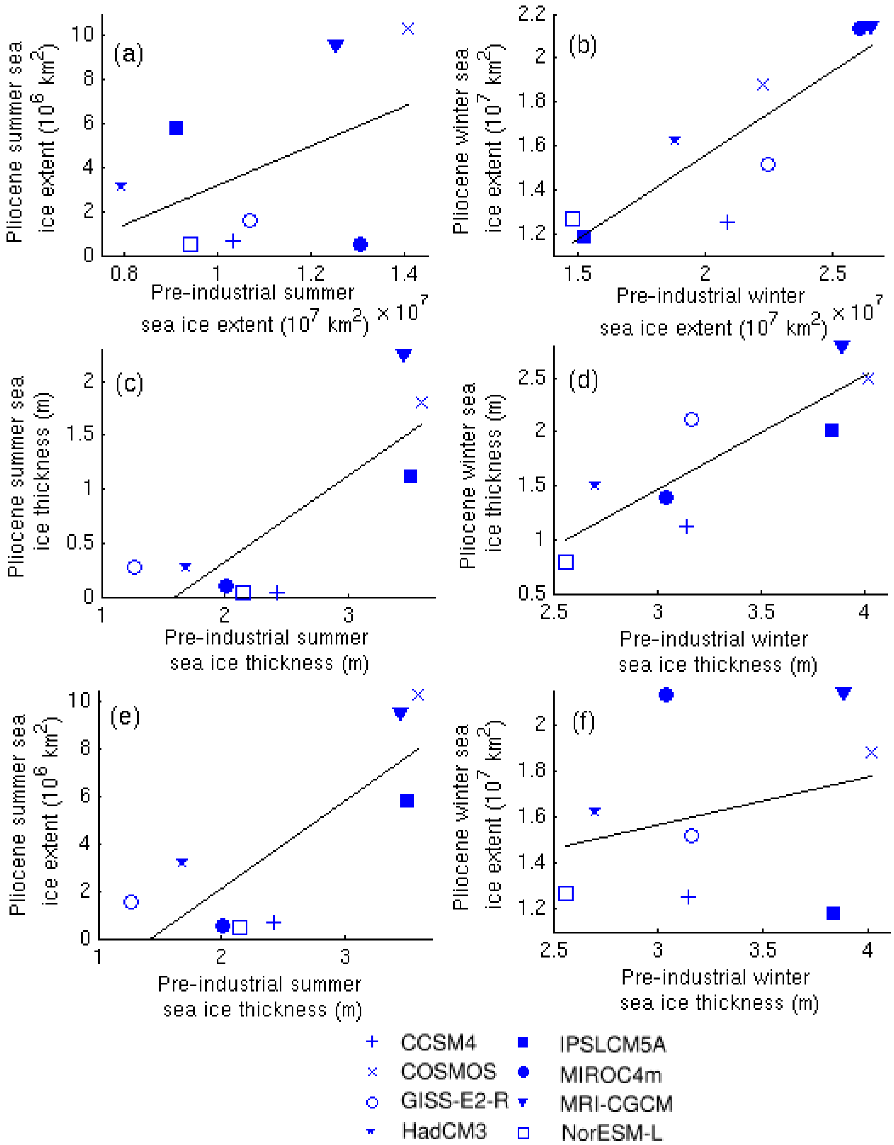


Figure 9. Relationship between various sea ice characteristics. Shown are pre-industrial values vs. mid-Pliocene values for (a) and (b) sea ice extent vs. sea ice extent, (c) and (d) sea ice thickness vs. sea ice thickness, (e) and (f) sea ice thickness vs. sea ice extent. (a), (c), and (e) illustrate 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

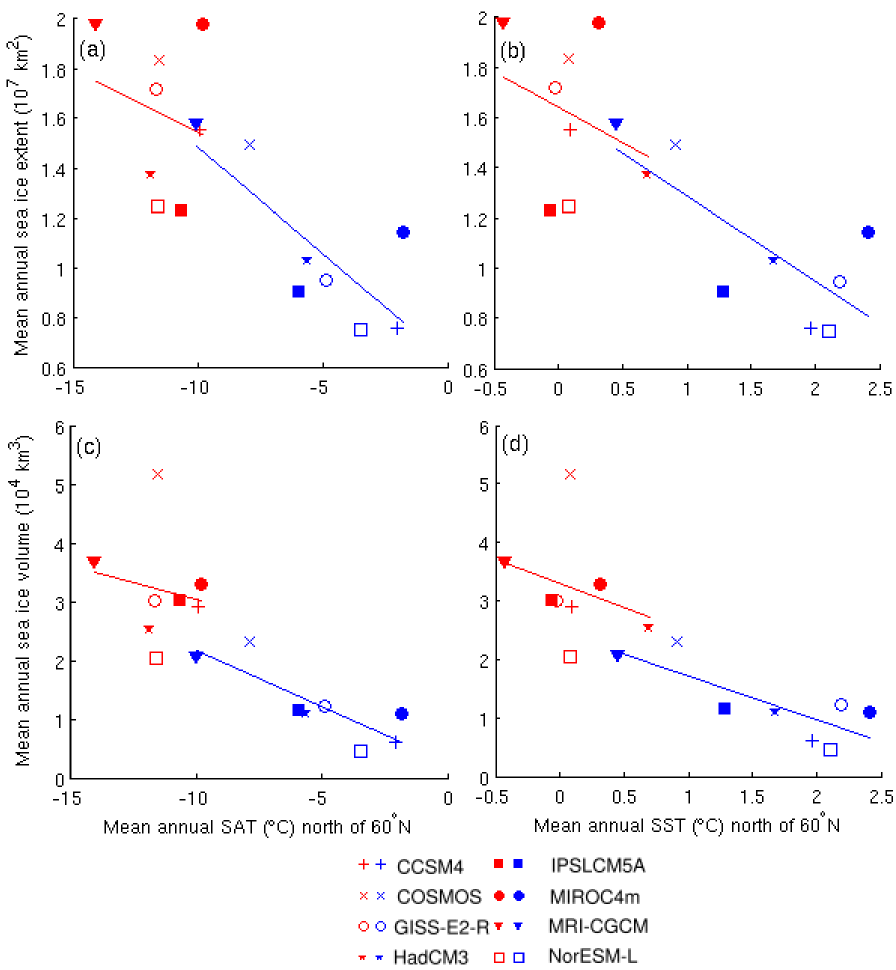


Figure 10. Mean annual surface temperatures north of 60°N vs. mean annual total Arctic sea ice extent(a,b), and mean annual surface temperatures north of 60°N vs. mean annual total Arctic sea ice volume(c,d) in both pre-industrial and mid-Pliocene simulations, for (a,c) SAT and (b,d) SST. Pre-industrial 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

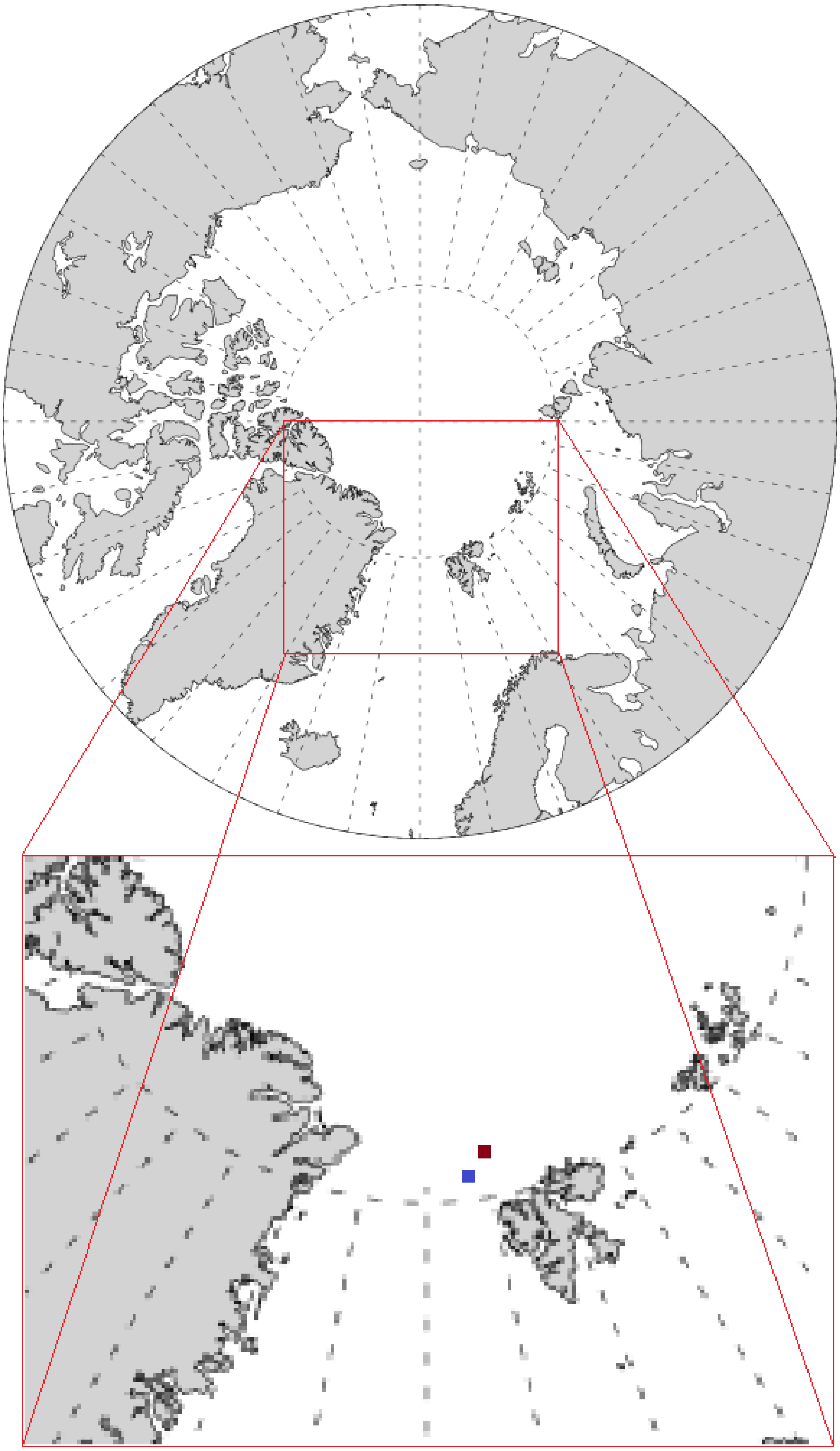


Figure 11. Location of Ocean Drilling Program (ODP) sites 911A (brown), and 910C (blue), used by Knies et al. (2014) for IP_{25} analysis.

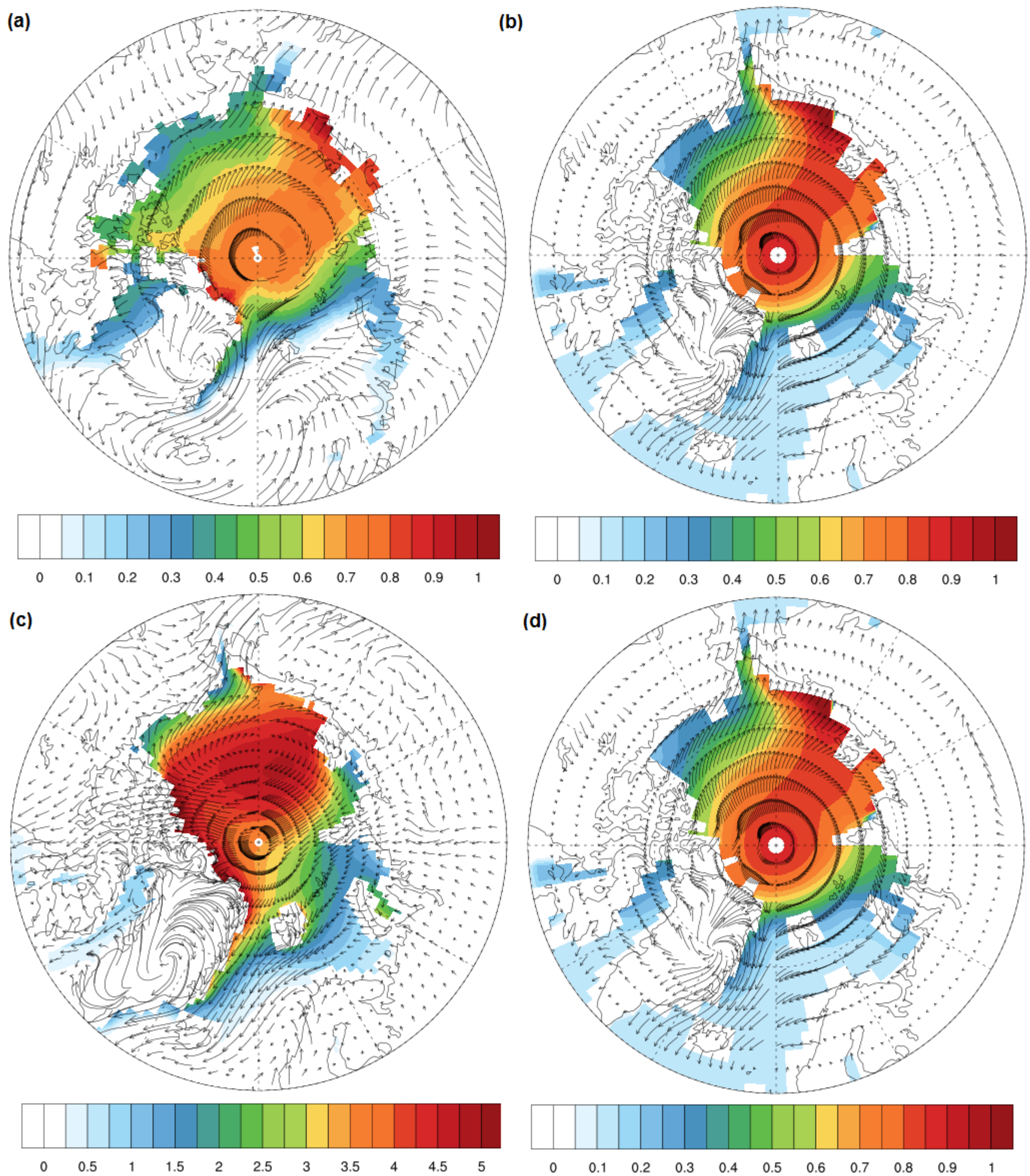


Figure 12. Mean annual 10 m winds and sea ice thicknesses (m) for (a) COSMOS mid-Pliocene, (b) MIROC4m mid-Pliocene, (c) IPSLCM5A pre-industrial and (d) NorESM-L pre-industrial. Vector length is proportional to wind speed.

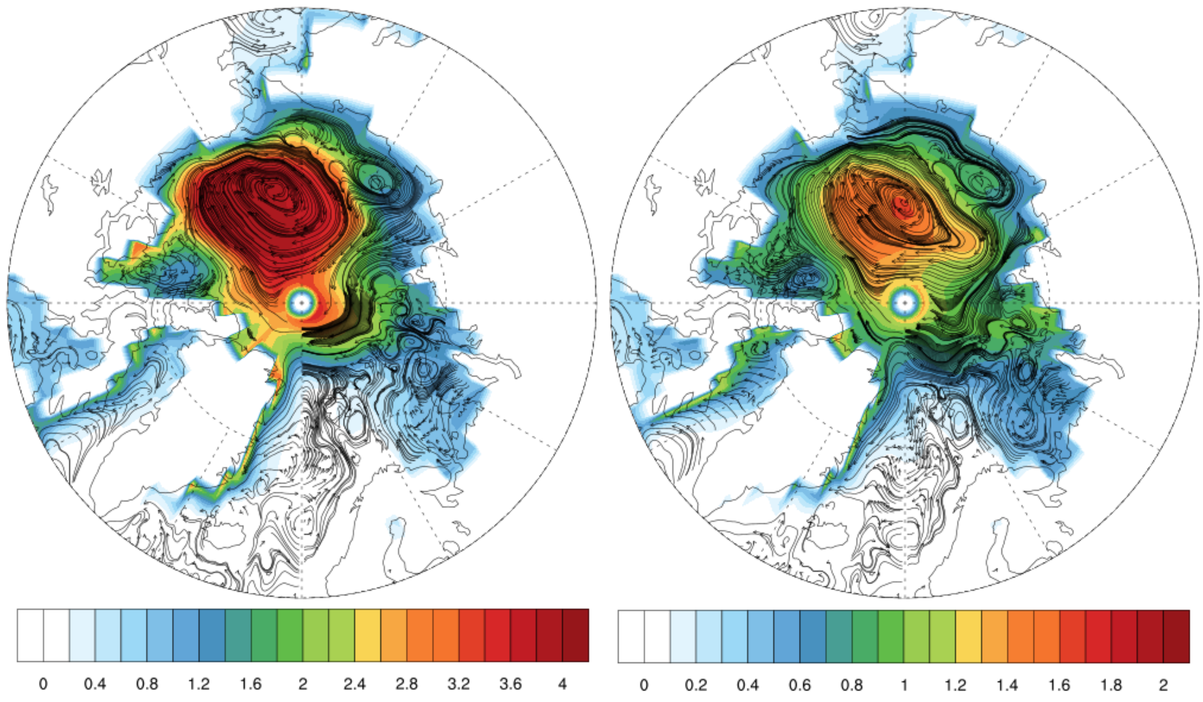


Figure 13. Mean annual ocean surface currents and sea ice thicknesses (m) for HadCM3 pre-industrial (left) and mid-Pliocene (right) simulations. Vector length is proportional to ocean current speed.