## Reviewer 1

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

*The areal extent of thin ice is more sensitive to warming than that of thick ice, because thin ice can more easily melt completely for a given warming. In contrast, thick ice simply becomes a bit thinner for some warming, which then does not lead to a substantial areal change. Hence, the main finding of this paper that pre-industrial sea ice is less sensitive to temperature changes than the much thinner ice of the PlioMIP ensemble is neither surprising nor new.*

We have included a suggestion that the thicker ice in pre-industrial experiments may be a cause of the weaker correlation (section 4.3.3, line 420). We have included a scatter plot of temperatures and sea ice volume alongside the sea ice extent scatter plots, which should take the thicker ice in pre-industrial simulations into account (Figure 14). However, these plots show a similar difference in correlation strength that is seen in the extent plots, and is mentioned in the discussion.

*The thickness of sea ice that is output by climate models is usually the average thickness that the ice would have if it were to cover the entire grid cell while conserving volume. To obtain actual thickness which then could be compared with satellite observations, one simply has to divide this so-called equivalent thickness by sea-ice concentration. This is apparently not done here (at least it is not mentioned), making the comparison to IceSAT simulations somewhat hard to interpret. It also renders some of the other discussion of sea-ice thickness hard to interpret, since this discussion seems to be based on the equivalent thickness but interprets it as if it were actual thickness.*

Calculations of sea ice thickness have been amended to take into account the sea ice concentration.

*The paper suggests a number of times that areal patterns of sea-ice thickness can be tuned for. However, I do not know of a single modeling group that would know a reasonable way of how to achieve this. Tuning of sea-ice models usually only involves a very simple metric, like for example March mean sea-ice thickness or the like, but not a tuning of any patterns. I also find the discussion of the tuning of CICE to most likely not reflect the reality at climate modeling centers. I would expect the developers of NorESM-L to tune CICE according to their needs. CICE itself cannot be tuned meaningfully, because it is a stand-alone sea-ice model that requires a given forcing to produce tunable results. The entire discussion of tuning also fails to appreciate the fact that tuning is necessary for any large-scale model, simply because necessarily the parameterizations cannot fully reflect the physical processes that occur on smaller length scales.*

In section 4.3.1, we have added a caveat reflecting that the strength of the influence of tunings is disputed, and cited relevant literature. The paragraph discussing CICE being tuned for CCSM and not NorESM has been removed. All the tuning references in section 4.4 in the first submission have also been removed.

*Throughout, this paper seems to assume that it is primarily the formulation of the seaice* *model that is responsible for the resulting sea-ice evolution. It fails to acknowledge that in all coupled climate models, it is by far more important to expose the sea-ice model to realistic oceanic and atmospheric forcing to obtain reasonable sea-ice results.*

Throughout section 4.3.1, we make the point that there does not appear to be a strong link between the sea ice rheology or dynamics scheme and the nature of the sea ice produced by different models, so we do not state that the sea ice model is the primary driver of sea ice. In section 4.3.3 and section 5, we stress the importance of understanding the atmospheric and oceanic influences on the sea ice.

*We have pretty reliable observations of sea-ice concentration from 1953 onwards, which should be much closer to pre-industrial sea-ice conditions than those of the past three decades. It would be helpful to compare the simulations against this earlier data set to obtain more robust insights into model quality compared to the recent period with its rapidly changing sea-ice conditions.*

All comparisons of pre-industrial results with observations have been removed, both from the figures and the text. Section 4.1 discusses the CMIP5 results of the PlioMIP ensemble instead, to provide some comparison of models against data.

*For sea-ice thickness, once it is correctly divided by sea-ice concentration, again the comparison of pre-industrial thickness to single-point observations from two months of satellite observations in 2009 is not very meaningful. Over the past decades, summer sea-ice thickness in the Arctic has decreased by roughly 50%, and it will be very hard to gain robust insights into the quality of a pre-industrial simulation based on satellite observations from 2009.*

Comparisons of pre-industrial thickness with recent observations have been removed, and instead a discussion of CMIP5 thickness simulations by PlioMIP models.

*The discussion of albedo vs. warming vs. sea-ice evolution remains unclear. Why should the ice-albedo feedback lead to a stronger relationship between $T$ and extent during the Pliocene? The same ice-albedo feedback acts during the pre-industrial period as during the Pliocene, suggesting that the relationship between a change in $T$ and a change in extent should be similar in both periods if the ice-albedo feedback was indeed the driving mechanism.*

We note that the parameterisations should have similar effects in both simulations (line 481), but suggest the possibility of a temperature threshold above which the feedback strength could become enhanced (line 482). In the conclusions, we suggest that the feedbacks ''may' contribute, to the stronger temperature-sea ice relationships in the mid-Pliocene simulations (line 508), as opposed to saying it is 'likely' that they do.

*I strongly recommend to focus less on statistical relationships, or to at least try to interpret those based on physical grounds. For example, the higher value of CV for Pliocene sea-ice extent is probably simply a reflection of the thinner and smaller mean ice cover, but is geophysically in my opinion not relevant. Geophysically, the actual areal change is much more relevant than the percentage change relative to some mean sea-ice cover.*

CV is a useful metric for measuring variability between datasets with different means, and has been used in other studies, (e.g. Stroeve et al. [2014] described as a 'normalised variability measure'). Whilst there are many other measures of the changes in sea ice cover, which may have greater relevance, we do not believe that CV is then irrelevant.

*Many of the insights found here for the PlioMIP period have been found before by existing studies that deal with CMIP-type ensemble simulations of future sea-ice evolution. These studies should be cited here, and the progress made relative to these studies should be discussed.*

In addition to studies cited in the initial submission, the new version includes discussion of results from Day et al. [2012], Massonnet et al. [2012], Hodson et al. [2013], Stroeve et al. [2014] and Shu et al. [2015].

## Reviewer 2

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

*Why are pre-industrial control-runs compared to observations? Are there no historical simulations or present day simulations done with these model versions? It is quite uncertain how sea ice conditions in the pre-industrial time period were (although very likely that both ice extent and ice volume were larger than at present day/ recent past). It is thus very difficult to judge from comparing pre-industrial simulations to present day observation, if a certain model is simulating realistic ice conditions or not. If historical simulations are available, these should be used. If not, please compare also to ice data sets (e.g. Arctic and Southern Ocean Sea Ice Concentrations; Chapman, W. L. and J. E. Walsh. 1991, updated 1996. Arctic and Southern Ocean Sea Ice Concentrations. [indicate subset used]. Boulder, Colorado USA: National Snow and Ice Data Center. http://dx.doi.org/10.7265/N5057CVT. ), which go further back in time until 1901. HadISST data go even back to 1871. Of course, the authors are right that these data are less certain as data based on satellite observations after 1978 but they probably still provide a better comparison for pre-industrial values.*

We have removed the comparison of pre-industrial simulations with modern observations from the revised version of the paper, and instead refer to comparisons of CMIP5 sea ice and observations in the discussion, in which most of the PlioMIP models have representation (e.g. Shu et al. [2015], Stroeve et al. [2014]).

*It is to my knowledge and to the publications listed in this article relatively uncertain how sea ice conditions looked like in the Pliocene. It is likely that there was less ice but it is unclear how much less. How should we know if models produce realistic Arctic ice conditions in the Pliocene if we do not know how ice conditions in reality were at that time? And what shall we then conclude from such a study for the reliability of models for future climate? I agree some of the ice concentration patterns in a few of the models look strange for the Pliocene and of course it is likely that such models might have difficulties to reliably project sea ice in a future climate but the same conclusions could be drawn from the future simulation of these models. Thus, the added value of performing Pliocene simulation to say something about reliability of models for future sea ice conditions is not getting clear from this study.*

Sections referring to using Pliocene simulations to assess future simulations have been altered so that it is clear that it is not possible with the lack of proxy data available, but may be possible in the future. The title has been altered to ‘Assessment of simulations of Arctic sea ice in the PlioMIP models’ to reflect this change. The paper also stresses the importance of the need for greater proxy data coverage relating to the sea ice conditions in the Pliocene (such as Knies et al. [2014]).

*All three indexes CV, RHO, LAMBDA are not convincing as metrics to measure model performance and model differences. I would suggest focusing this study entirely on possible sea ice conditions in the Pliocene and comparing to the pre-industrial ice conditions and how and why they differ. In order to make this an interesting and publishable study it is not sufficient to study the statistical relationship between two or three variables. Instead, processes in ocean and atmosphere need to be identified, which govern sea ice and sea ice variations in the pre-industrial simulations. Then, one should investigate how and if these processes change/ are different in the Pliocene and if other processes are of importance in the Pliocene for sea ice conditions and variations. Furthermore, more of the existing literature on the topics of sea ice variations and sea ice changes should be used.*

We have removed $\lambda$ and $\rho$ figures from Table 2, and any references to them within the text of the article. We have added to Table 2 the metric of motnthly sea ice extent amplitude (maximum monthly extent - minimum monthly extent), a metric used in Shu et al. [2015], which analyses CMIP5 sea ice output. This has replaced $\lambda$ and $\rho$ as the measure of the cycle of sea ice extent.

We still believe that CV is a useful index to assess variability within the ensemble, given the difference in mean values of the data sets, and so this has remained. CV is used in Stroeve et al. [2014], which analyses the sea ice thicknesses in the CMIP5 ensemble.

Other comments

*Abstract: Page 1265, line 11-15: Tuning discussion*

*The higher correlation between sea ice and T2m in Pliocene might also be due to warmer temperatures and reduced ice thickness, which makes the ice extent more sensitive to small temperature changes compared to a period where ice thickness is 2-3 m almost everywhere in the Arctic Basin. It is not shown at all in this study that the tuning reduces the correlation between temperature and ice. Even though some ice parameters might be tuned in pre-industrial simulations, the dependence of ice on temperature still exists even in a tuned model. I do not understand, why and how a tuned ice model state should in general provide lower correlations of ice to temperature than an untuned model. Especially not, if as I assume, the same tuned model versions has been used to run the Pliocene time slices.*

We have added a caveat in section 4.3.1, citing literature [DeWeaver et al., 2008, Eisenman et al., 2008] that highlights disagreement over the influence of tuning on model sea ice. We have also removed the sentence in the conclusions stating that tuning has affected the correlations between temperatures and ice.

We have added a sentence in the discussion (section 4.3.3, line 420) noting that the thicker pre-industrial sea ice may explain the difference in correlations between the pre-industrial and Pliocene simulations. We have also included scatter plots of sea ice volume against temperatures for both simulation, in addition to extent. These show similar differences in correlation between the pre-industrial and mid-Pliocene simulations as seen for extent, suggesting that there are other causes of the difference in correlations.

*Introduction: P 1266, line 5: Studies using the entire CMIP5 model ensemble should be cited here as well (e.g. Massonnet et al. 2012, Stroeve et al. 2012)*

These papers have been cited in the revised version.

*P1266 Lines 7-15: If it is so uncertain how ice coverage was in the Pliocene, it seems to be very difficult if not impossible to answer the question in the title.*

Title has been changed to 'Assessment of simulations of Arctic sea ice in the PlioMIP models', so the emphasis of the paper has shifted.

*Methods: P1267, lines 8-9: It will only enable a better understanding of the differences in the Pliocene if there is a clear relation between differences in models in pre-industrial and differences in the Pliocene. In the conclusions, the authors state that there is no reliable relationship between pre-industrial performance and Pliocene sea ice conditions.*

This sentence has been changed to 'Pre-industrial results provide an additional climatology against which differences in the models' sea ice outputs can be compared.'

*P1267, lines 13-15: Please make clear what is meant by the 100% sea ice concentration assumption. How are you exactly calculating mean sea ice thickness north of 80N? All models provide both sea ice concentration and ice thickness for each gridpoint and this information should be used. Later on, in the figures also 66-86N is used for ice thickness; for 66-86N, the ice concentration is definitely not "close to 100%".*

Assumptions of areas of 100% sea ice concentration have been removed from the paper, and calculation of sea ice thickness has been adjusted to take into account the sea ice concentration. We have also removed comparisons to observations, and do not use the region 66-86N for any thickness results.

*P1267, lines 16-18: August/ September and February –April are not the "three months" with lowest and largest ice extents. Please correct the "three months"-statement or the period you used for summer.*

This statement has been altered to '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).'

*P 1267, line 20: Please define SD when using the abbreviation the first time (probably standard deviation).*

Done.

*P1267, line 20: CV: I am not entirely convinced by using CV. What is done is calculating* *a type of relative spread instead of absolute spread among ensemble members. I doubt that this is an appropriate measure for sea ice extent. Please clarify, why this is necessary.*

CV allows a comparison of spread in data sets with different mean values, and has been used in other studies (e.g. Stroeve et al. [2014]). As the different months and simulations can have largely different means, then using CV allows a better comparison of the spread than just using standard deviation.

*P 1267, line 24/25: "CV identifies in which months there is greater spread across the ensemble". Greater than what? Maybe rephrase to: The CV identifies the months with large sea ice spread across the ensemble.*

Line has been changed to 'Calculation of the CV identifies the differences in spread in each month in the ensemble'.

*Page 1268, lines 1-3: Where do you provide a correlation between ice metrics and key climatological variables? The only thing I found is figure 15 where a correlation between ice extent and temperature north of 60N is shown. If this is all, you should call it "correlation between ice extent and SAT and SST north of 60N". As it is written now I would expect a detailed investigation of different ice parameters with different important climate variables as SST, SLP or 500hPa geopotential height, northward heat fluxes in ocean and atmosphere and maybe others.*

This has been changed to 'we quantify correlations between the sea ice metrics and sea surface and surface air temperatures.'

*2.2 Page 1269, line 1: The satellite-derived ice concentration can indeed be used as lower bound for the pre-industrial model simulations but they do not tell much about performance of models with much more ice than the observed values. To assume that all models with more ice than in the present day observations are performing well while those with less or similar ice are badly performing, is not a very good criteria for the model performance.*

Comparisons of pre-industrial results with modern observations has been removed. Discussion of model performance compared to observations is based on CMIP5 simulations (e.g. Shu et al. [2015]).

*Page 1269, line 5, equation 1: This assumes that the annual cycle should be generally larger if the model produces more ice. I am not convinced, this is really the case and I do not expect the annual cycle (in absolute values) to grow with larger maximum ice extent. I would suggest as measure for the annual cycle just Emax – Emin. On page 1273 you state yourself that Lambda seems to be dependent on the ice extent. Please explain why you introduce Lambda, and why you think it is better than using absolute values. If someone else already used Lambda, cite the relevant literature.*

$\lambda$ and $\rho$ have been removed. Sea ice extent amplitude ($E_{max}$ - $E_{min}$) is used as the measure of the sea ice extent cycle, as used in Shu et al. [2015].

*Page 129, Equation 2: I am not convinced by equation1, thus, I am of course not either by equation 2 since it is based on Lambda. In my view, table 2 would provide more useful in-*

*formation if only mean, max and min ice extent would be specified instead of these somewhat questionable ice metrics.*

See above response.

*Figure 1: Please extend the area to the south so that it is possible to see how far to the south the ice extends.*

For consistency between all figures we intend to show the same area in sea ice plots. Whilst in some this will mean that the southern extent of the sea ice is not displayed, in most of the figures the sea ice is contained entirely within the area covered by the plot.

*3.1.3/3.1.4 Comparison to observations and overall model performance: In these sections, clear criteria are missing. It appears relatively arbitrary if models are judge as good or weak performer for sea ice. This study introduced several ice metrics' to judge models' performance (it might be discussed if the metrics are well chosen) but if such metrics are defined then there should be a clear procedure how to use the metrics. At least a minimum criterion for each metric needs to be established. Now, e.g. HadCM3 is pointed out as bad model, although both Lambda and Rho are not too far of and the ice extent seems to be quite realistic. MRI instead has an annual ice extent that exceeds the observed extent by more than 50% and almost the entire Nordic Seas are ice-covered but still is judged as being realistic.*

Evaluation of model performance has been removed from the revised version.

*3.2 Pliocene Simulations Page 1277 lines 24ff: maybe it would be better to use the sea ice volume instead sea ice thickness; again it is unclear how sea ice thickness is calculated; is it the ice thickness of ice covered areas?*

Thickness calculations have been changed to take into account the concentration of the sea ice cover. Sea ice volume plot shown in Figure 14.

*Page 1279: Discussion of correlation: The correlations are based on only 8 values, which makes it hard to draw any conclusions. At least, the significance of the correlations should be discussed.*

In section 3.4, we mention that the sample size is low, and give the correlation values required to be significant at 95% and 99% levels.

*Page 1279: CV-discussion: The mean is very small in summer in Pliocene, thus it is not very surprising that CV-values go up as long as some models still show some sea ice. As discussed before, I am not convinced CV is a very good index.*

CV is used to assess the difference in variability between data sets that have different means, and has been used in prior studies (e.g. Stroeve et al. [2014]). Whilst the mean in summer is small, CV takes into account the effect that this might have on the SD (i.e. being smaller), so that the variability can be compared consistently.

*Discussion Page 1287: Since we do not know much about Pliocene ice it is hard to say* *anything about performance of the models in the Pliocene and its relation to performance in the pre-industrial time. However, what this study showed and other CMIP3 and CMIP5 studies showed, is that there is at least some relation between sea ice conditions at present day and sea ice in a warmer climate in the same model. This could indicate that the performance of present day sea ice plays a role but is of course no evidence.*

Section 4.3.2 discusses the areas where the pre-industrial and mid-Pliocene simulations appear to be strongly linked, and cites the study of Massonnet et al. [2012], which looks at the relationship

*Conclusions Page 1289, line 18: All models are tuned. This is not generally negative. Furthermore, tuning of sea ice is not the only factor that affects how a model behaves in a different climate. Changes in ocean and atmosphere are of extreme importance as well.*

Other atmospheric and oceanic influences are discussed, including ocean heat transport and multi-decadal variability sources. Tuning discussion has been altered in section 4.3.1 to reflect that there is some disagreement amongst the literature over the amount of influence model tuning is capable of. The paper no longer states that tuning is likely to be the reason for the differences in correlations between temperatures and sea ice between the pre-industrial and mid-Pliocene simulations.

Changes made to cp-2015-29:

- Title has been changed to 'Assessment of simulations of Arctic sea ice in the PlioMIP models'
- Abstract has removed section relating to the effects of tuning on sea ice model results. Sentences concerning the need for more proxy data and the possibility of different atmospheric and oceanic influences for the pre-industrial and mid-Pliocene simulations.

**Introduction**

- Removed sentence concerning comparison of pre-industrial results to modern observational data
- Other changes are minor, including some extra references and minor linguistical tweaks

**Methods**

- Removal of any passage referring to comparison of pre-industrial results to modern observational data
- Removed description of $\lambda$ and $\rho$ metrics
- Added explanation of calculation of mean sea ice thickness
- Moved paragraph explaining choice of season, and small alteration made to language in this paragraph
- Sentence added to highlight a recent study utilising the CV metric
- Minor language and structural changes

**Results**

- Removal of any reference to $\lambda$ and $\rho$ results
- Replaced by sea ice extent amplitude values
- New calculation of mean sea ice thickness has resulted in changes to many of the thickness results in initial submission
- Comparison to observations section all cut
- Overall model performance section cut
- Minor language and structural changes

**Discussion**

- Removal of passage discussing the differences in the pre-industrial data and observation data time periods
- Addition of paragraphs discussing ideal comparison for pre-industrial data, and unsuitability of modern observations for such purpose

- Sentence mentioning inclusion of discussion of CMIP5 results for PlioMIP models
- Removal of paragraphs discussing comparisons of pre-industrial results and modern observations
- Addition of several paragraphs discussing CMIP5 results
- Addition of paragraphs that discuss disagreement in PlioMIP models on summer sea ice in the mid-Pliocene, and the need for further proxy data to test against
- Sentence added on model tuning, mentioning other studies that question the strength of the influence of tuning
- Paragraph added on control simulation influence, citing results from Massonnet et al. (2012)
- Sentence added concerning whether thicker pre-industrial sea ice is the cause of weaker correlation between sea ice extent and temperatures
- Discussion of possible influence of AMOC on sea ice extent
- Paragraph concerning possible temperature threshold above which albedo feedback has greater influence
- Paragraph on influence of multi-decadal oscillations on sea ice, and the inability to compare for the models due to different run lengths and averaging periods
- Future climate change section cut
- Minor language and structural changes

**Conclusions**

- Removal of any passage concerning comparison of pre-industrial results to data
- Paragraph on need for greater proxy coverage added
- Reduction of passages concerning implication of mid-Pliocene results on quality of future predictions
- Removal of passages suggesting over-influence of tuning on the model results
- Concluding paragraph discussing all possible influences on sea ice cover, and importance to understand these before making judgements about future climate, if more proxy data becomes available

**Tables**

- $\lambda$ and $\rho$ values removed from the table
- Mean annual extent for mid-Pliocene added, along with extent amplitudes for both experiments
- Observations removed

**Figures**

- Removal of observation plot from Figures 1 and 2 (captions updated)
- Observational values removed from Figure 3 (caption updated)
- Figure 7 removed
- Correlation values added to captions for Figures 14 and 15 (now 13 and 14)
- Volume plots added to Figure 15 (now Figure 14)

Manuscript prepared for Clim. Past
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 17 August 2015

# ~~Arctic sea ice in the PlioMIP ensemble: is model performance for modern climates a reliable guide to performance for the past or the future?~~ Assessment of simulations of Arctic sea ice in the PlioMIP models

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, New York, NY, USA
[6]Aix-Marseille Université, 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 (~~mPWP~~ 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 is more than three times larger than the rest of the year. 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.~~It is suggested that the weaker relationship between pre-industrial Arctic sea ice and temperatures is likely due to the tuning of climate models to achieve an optimal pre-industrial sea ice cover, which may also affect future predictions of Arctic sea ice. Model tuning for the pre-industrial does not appear to be best suited for simulating the different climate state of the mid-Pliocene. This highlights~~

~~the importance of evaluating climate models through simulation of past climates, and the urgent need for more proxy evidence of sea ice during the Pliocene.~~

## 1 Introduction

The mid-Pliocene warm period (~~mPWP~~ 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 ~~mPWP~~ mid-Plioceneis 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)). ~~and therefore is a useful interval in which to study the response of sea ice in a warmer world~~ Therefore, this time period is a useful interval in which to study the dynamics and characteristics of sea ice in a warmer world.

September 2012 saw Arctic sea ice fall to a minimum extent of $3.4 \times 10^6$ km$^2$, a reduction of $4.2 \times 10^6$ km$^2$ since the beginning of satellite observations in 1979 (Parkinson and Comiso, 2013; Zhang et al., 2013). The Arctic is widely predicted to become ice free before the end of the 21st century (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 mPWP 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.

The Pliocene Modelling Intercomparison Project (PlioMIP) is a multi-model experiment which compares the output of different models' simulation of the ~~mPWP~~ mid-Pliocene, each following a standard experimental design, set out in Haywood et al. (2011a, b). 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). ~~Two different experiments are defined — Experiment 1 is for atmosphere only simulations, with Experiment 2~~

~~for coupled atmosphere-ocean general circulation models (GCMs).~~ Each modelling group also ran a pre-industrial control simulation.

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 Pliocene outputs. ~~The pre-industrial outputs are compared to observational data in an effort towards determining which models appear to produce more realistic simulations of pre-industrial Arctic sea ice.~~ We quantify the variability of sea ice extent and thickness in both simulations, ~~and determine which factors exert greater amount of influence on the models' sea ice output~~ and identify possible mechanisms that define the result of the sea ice simulations.

## 2 Methods

### 2.1 ~~Analysis of model output~~

The simulation of Arctic sea ice by the individual models in the PlioMIP ensemble (see Table 1 for details)for both their pre-industrial and Pliocene simulations is investigated. ~~Analysis of the outputs of the pre-industrial simulations can demonstrate the relative performance of each sea ice model, which will enable a better understanding of the differences in their simulation of Pliocene Arctic sea ice.~~ 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 reducs the possible causes of disagreement between ensemble members (Haywood et al., 2011a, b).

We focus on the key sea ice metrics of ~~sea ice~~ extent (defined as the area of ocean where sea ice concentration is at least 15%), ~~and sea ice~~ thickness and volume . ~~For our initial comparison between the models~~ We follow the example of Berger et al. (2013), and examine the mean sea ice thickness north of 80°N. 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.~~As the pre-industrial sea ice concentration is close to 100% in this region, then the calculation of the mean sea ice thickness is not distorted by large areas of lower sea ice concentration.~~

~~In our analysis, we define winter as the months February to April, and summer as the months August to September, as these are the three months where the vast majority of models produce the highest and lowest sea ice extents respectively. This is in contrast to the typical seasonal definitions of December to February and June to August.~~

The coefficient of variation (CV), defined as standard deviation (SD) of different simulations divided by the mean, is calculated to assess the variability among the ensemble members for both metrics. ~~CV is considered rather than simply using standard deviation, as it~~ 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 in which months there is greater spread across 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”.

To understand differences in the models’ simulation of sea ice, we quantify correlations between the sea ice metrics and ~~key climatological variables, such as~~ sea surface and surface air temperatures. We also compare the pre-industrial and Pliocene sea ice extents to establish how closely correlated they are. This enables us to determine ~~the degree to which the~~ to which degree the mid-Pliocenesea ice cover is influenced by ~~these factors in the simulations~~ 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)..

### 2.2 ~~Comparisons to observational data~~

~~We compare the output from the pre-industrial simulations with modern day observations of sea ice extent and thickness, to establish which models simulate sea ice extents that better reflect the pre-industrial sea ice cover. Whilst early observations of Arctic sea ice date from as far back as the early 20th century (Walsh and Chapman, 2001; Walsh and Chapman, 2003), it is only since the advent of the satellite era (i.e. 1979 onwards), that spatially and temporally comprehensive coverage of sea ice extent exists (Parkinson et al., 1999)~~.

#### 2.2.1 ~~Sea ice extent~~

~~Sea ice extent observations are obtained from the sea ice index at the National Snow and Ice Data Center (NSIDC). Sea ice extent is calculated from satellite observations of sea ice concentration, obtained from passive-microwave data grids from a scanning multi-channel microwave radiometer (SMMR) (Parkinson et al., 1999; Parkinson et al., 2002). In July 2013, the Sea Ice Index team replaced the original 22-year base period from 1979-2000 with a 30-year version, from 1981-2010.~~

~~As the pre-industrial simulations are designed to model the climate of more than 100 years prior to the satellite era, this temporal disconnection needs to be considered in the discussion and evaluation of our comparison. Sea ice cover has declined rapidly since the beginning of the satellite observational era (Stroeve et al., 2007; Stroeve et al., 2008; Stroeve et al., 2013). In this study we use the observations to identify the models which do not simulate sufficiently extensive sea ice, by utilising the observations as a lower bound on the pre-industrial sea ice extent.~~

~~In addition to the sea ice extent, we define another metric for each simulation, the $\lambda$-value as~~

$$\lambda = \frac{E_{min}}{E_{max}} \times 100 \tag{1}$$

~~where $E_{max}$ and $E_{min}$ are the maximum and minimum simulated monthly sea ice extents respectively, for the particular model. The $\lambda$-value is the minimum sea ice extent as a percentage of the maximum, giving a measure of the magnitude of the annual sea ice extent cycle for each model. We also calculate the $\lambda$-value for observations, as a reference for modelled $\lambda$-values.~~

~~Another sea ice characteristic, $\rho$ is defined as~~

$$\rho = \frac{\lambda}{E_{mean}} \times 10^6 \tag{2}$$

~~where $E_{mean}$ is the annual mean extent of the model or observation. Multiplication by $10^6$ simply provides a scaling of the value to a more convenient value range. We use the $\rho$-values to allow the comparison of $\lambda$-values of models or observations of sea ice extent with differing mean annual extents.~~

#### ~~2.2.2 Sea ice thickness~~

~~Observations of sea ice thickness are less extensive, both spatially and temporally, in comparison to sea ice extent observations. Kwok et al. (2009) produce estimates of Arctic sea ice thickness and volume for ten separate periods (five covering October/November, five covering February/March) from 2003 to 2008, using data from the Ice, Cloud and Land Elevation Satellite (ICESat). These measurements are produced for both first-year and multi-year ice. Uncertainties of up to 0.5 m are associated with these measurements (Kwok and Cunningham, 2008; Kwok and Cunningham, 2009).~~

~~We compare the observations from Kwok et al. (2009) to the monthly mean sea ice thicknesses from 66°N to 86°N in the models, as this is approximately the region covered by the observations. Similarly to extent, the thickness observations represent conditions that are different to the simulations (present day vs. pre-industrial), and are likely to be thinner than pre-industrial sea ice, and consequently they also only act as a guide to a lower bound on the simulated sea ice thickness.~~

~~Due to the discordant time periods and semi-qualitative nature of these comparisons, we will not produce a definitive ranking of each model's ability to simulate pre-industrial sea ice. Rather, we will identify those models which appear to consistently perform well or poorly on each metric, and give a broad grouping of the better and worse models. The relative model performance will be taken into account when interpreting the Pliocene sea ice results.~~

## 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 1 and 2 respectively. Across the eight-member ensemble, the multi-model mean annual sea ice extent is 16.17 $\times 10^6$ km$^2$, with a winter (FMA) multi-model mean of 20.90 $\times 10^6$ km$^2$, and summer (ASO) multi-model mean of 10.98 $\times 10^6$ km$^2$. The individual models' annual means range from 12.27 $\times 10^6$ km$^2$ (IPSLCM5A, hereafter IPSL) to 19.85 $\times 10^6$ km$^2$ (MIROC4m, hereafter MIROC) (see Table 2), and monthly multi-model means range from a minimum of 10.01 $\times 10^6$ km$^2$ (September) to a maximum of 21.24 $\times 10^6$ km$^2$ (March). The lowest individual monthly extent is 7.00 $\times 10^6$ km$^2$ (HadCM3, September), with the highest monthly extent produced by ~~MRI~~ MRI-CGCM (hereafter MRI) (March), measuring 27.01 $\times 10^6$ km$^2$ (Figure 3).

Figure 3 reveals the differences in the annual sea ice extent cycles across the ensemble. ~~The $\lambda$-value is 57% for NorESM-L, and 54% for IPSL (see Table 2), giving relatively small differences between the minimum and maximum extents.~~ The sea ice extent amplitudes of NorESM-L (herafter NorESM) and IPSL are 6.39 and 7.36 $\times 10^6$ km$^2$ respectively (Table 2). These are the only models in the ensemble with seasonal amplitudes below 10 $\times 10^6$ km$^2$. Other models in the ensemble show a much larger seasonal cycle, in particular ~~HadCM3 and GISS which have $\lambda$-values of 36% and 39% respectively. The $\lambda$ for the ensemble mean is 47%.~~ GISS-E2-R (hereafter GISS), MIROC and MRI, which have sea ice extent amplitudes of 14.03, 14.05, and 15.91 $\times 10^6$ km$^2$ respectively (Table 2). The ensemble mean sea ice extent amplitude is 11.18$\times 10^6$ km$^2$.

#### 3.1.2 Sea Ice Thickness

North of 80°N, the multi-model mean annual thickness is ~~2.97 m~~ 3.20 m, with a winter multi-model mean of ~~3.29 m~~ 3.45 mand a summer multi-model mean of ~~2.51 m~~ 2.81 m . Across the ensemble, the annual mean thickness varies from ~~2.27 m (HadCM3) to 3.81 m (CCSM)~~ 2.50 m (NorESM) to 3.98 m (CCSM4, hereafter CCSM) . The winter thicknesses range from ~~2.56 m (NorESM-L) to 4.01 m (CCSM)~~ 2.61 m (NorESM) to 4.08 m (CCSM), with summer between ~~1.27 m (GISS) and 3.60 m (CCSM)~~ 1.66 m (GISS) and 3.84 m (IPSL) .

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 (Figure 4). Also along the Greenwich meridian, between 80°N and 90°N, is a region of thicker sea ice (Figure 4). 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 ~~distribution~~ spatial thickness pattern shows the 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 sea ice thickness that are broadly similar to the overall ensemble mean shown in Figure 4~~., but~~ Yet, there is appreciable variation with respect to the location of maximum ice thickness across the ensemble (Figures 5 and 6). The thickest ice in CCSM is located north of Greenland and the Canadian Arctic Archipelago, and the ice thins consistently with distance from ~~the thickest ice~~ this region. ~~IPSL produces a similar pattern~~ For IPSL a similar pattern is found in the summer, although ~~its winter distribution has a larger region of thicker ice that~~ although for both summer and winter spatial patterns the region of thicker ice extends much further into the Arctic Basin ~~(see Figures 5 and 6).~~ The thickest ice in COSMOS, GISS, MRI and NorESM-L 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 ~~also~~ extends in winter in a band from Greenland towards 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-L 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-L 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 ~~distributions~~ spatial patterns that are noticeably different from the ensemble mean, and the other six models. ~~MIROC displays a pattern which~~ 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 the ensemble mean. The thickest ice is situated in a region north of approximately 70°N, and between 120°W and 150°E and around the North Pole. A smaller area surrounding the rest of the pole also displays a similar thickness. 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 6 illustrates that the PlioMIP ensemble consists of two realisations of pre-industrial summer

(ASO) sea ice, with pronounced sea ice cover in CCSM, IPSL and MRI, and relatively reduced sea ice in the other models.

#### 3.1.3 ~~Comparison to observations~~

~~It would be expected that simulated pre-industrial sea ice extent should generally exceed that of recent observations, but only five of the eight models do so (Figure 3). The HadCM3 sea ice extent exceeds the observations during most of the year, but there is little difference between the HadCM3 and observational sea ice extents from September to December. In modern transient simulations, HadCM3 has produced lower September sea ice extents than modern observations (e.g. Stroeve et al. (2012); Stroeve et al. (2014)), in contrast to the majority of models, which have simulated higher sea ice minima than observations. IPSL produces very similar values to the observations from November to July, whilst NorESM-L simulates a pre-industrial sea ice extent below recent observations from November to June, with very close agreement to observations in July and slightly exceeding the observed sea ice extent in August and October.~~

~~The $\lambda$-value for the mean of the 1981-2010 observations is 42%, compared to the ensemble mean $\lambda$-value of 47% (Table 2). For the most recent five years from the observations, the mean $\lambda$-value is just 32%, and in one year (2012) it is as low as 23%. This decrease in $\lambda$ coincides with the recent sharp decline in the observed minimum sea ice extent, implying that $\lambda$ decreases as the mean sea ice extent also decreases. For each year of the observations, the correlation between the respective $\lambda$-values and mean sea ice extents is 0.71. In contrast, the model-simulated annual mean sea ice extents are negatively correlated with their respective $\lambda$-values, with a coefficient of -0.82. This means that in the observations, the maximum sea ice extent is proportionally greater than the minimum sea ice extent for larger mean annual extents, but the reverse pattern is seen in the pre-industrial sea ice extents simulated by the models.~~

~~As it appears, based on observations, that the $\lambda$-value is somewhat dependent on the overall annual mean sea ice extent, then we also consider the $\rho$-value, to enable a clearer comparison of the different models' annual sea ice cycles (Table 2). The $\rho$-value for the ensemble mean is 2.92, in comparison to 3.49 for the observations, suggesting that the ensemble mean annual sea ice extent cycle is similar to observations. Six of the eight models have $\rho$-values lower than both the observations and the ensemble mean. The COSMOS and CCSM $\rho$-values are closest to the observational values, suggesting that these models are most successful at simulating an appropriate annual cycle of sea ice extent given their respective mean annual extents. The $\rho$-value for NorESM-L is 4.55, more than 50% greater than the observations.~~

~~CCSM and IPSL are the only two models with annual thickness cycles that pass through both the October/November and February/March observation ranges produced by Kwok et al. (2009) (Figure **??**). The sea ice thickness simulated by GISS is at the lower end of the February/March range, but is substantially lower in October/November. Five of the eight models, as well as the overall ensemble~~

~~mean, produce thicknesses for pre-industrial simulations that are lower than observations from the 21st century, suggesting that the majority of models produce sea ice that is too thin. This may have a profound effect on the simulation of Pliocene sea ice.~~

Observations of the sea ice thickness detailed in Kwok et al. (2009) give an indication as to the ~~distribution~~ 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, 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 thickness patterns appear to broadly match the observations from Kwok et al. (2009).

~~Whilst the regions of thickest sea ice are similar in ensemble mean and observations, the simulated pattern for the Arctic basin indicates a reduction in thickness with distance from the pole, rather than from the area of thickest ice, as inferred from the observations. Aside from this small difference, the ensemble mean thickness patterns appear to broadly match the observations from Kwok et al. (2009).~~

The degree to which individual models reproduce the observed thickness patterns is variable. CCSM produces what appears to be the closest pattern to observations, with IPSL ~~matching closely~~ being similar in the summer~~, but~~. Yet, the extension of the large region of thicker ice particularly in ~~its~~ winter ~~distribution~~ prevents ~~it~~ IPSLfrom being as close to the observations as CCSM. ~~As detailed in section 3.1.2, the thickness distributions of~~ The spatial patterns of sea ice thickness simulated by COSMOS, GISS, MRI and NorESM-L show ~~similar~~ some similarity to patterns of ~~to~~ CCSM and therefore also to the observations.. ~~As CCSM has a similar ice thickness pattern to the observations then COSMOS, GISS, MRI and NorESM-L also have similar thickness patterns, with some small differences. as ice thinning from the poles rather than from the areas with thicker ice. Similarly, as~~ As MIROC and HadCM3 show~~ed~~ very different patterns to the other models, their thickness ~~distributions~~ spatial patterns ~~bear no similarity at all~~ are less similar to the observational ~~distributions~~ spatial patterns from Kwok et al. (2009).

#### 3.1.4 ~~Overall model performance~~

~~Our analysis suggested that NorESM-L, IPSL and HadCM3 simulate insufficient sea ice extents in some months in their annual cycle, due to the fact that their results indicate a sea ice cover that does not exceed the observational sea ice extent from 1981-2010. The sea ice extent predicted by HadCM3 is only exceeded by the observations in October, but does show values that are very close to observations from September through to December. The other five models exceed the observations in every month. MRI, MIROC and CCSM display greater mean annual extents than GISS or COSMOS.~~

~~However, without the availability of pre-20th century sea ice extent observations it is difficult to determine the appropriate sea ice extents for the pre-industrial. When considering the overall annual cycle, the $\lambda$ and $\rho$-values for the models and observations demonstrate that the annual cycles of sea ice extent simulated by NorESM-L and IPSL do not appear to be realistic. The models with $\rho$-values closest to the observations are CCSM and COSMOS.~~

~~The majority of the models simulate sea ice that, in comparison with the observations of Kwok et al. (2009), appears to be too thin, particularly during the summer months. Only CCSM and IPSL produce thicknesses that match the observations, when it would be expected that the models should produce sea ice that is thicker than observations from the last decade. HadCM3 simulates substantially thinner ice than the other models. The CCSM thickness distribution is closest in pattern to the observations, followed by IPSL. HadCM3 and MIROC produce patterns that are completely different to observations. The other models reproduce the same broad patterns as the observations, but not as well as CCSM or IPSL.~~

~~CCSM appears amongst the better performers in every metric, suggesting that it has the best all-round performance in terms of simulating pre-industrial sea ice. COSMOS, GISS and MRI perform consistently - the only metric at which they can be said to provide a relatively weak performance is on sea ice thickness, at which the majority of models failed to match or exceed the observations. MIROC performs well at the simulation of sea ice extent, but is less successful in the simulation of sea ice thickness. Like most models, it simulates thinner sea ice than inferred from the observations, but unlike most models it simulates a thickness distribution that does not bear any resemblance to the patterns seen in the observations.~~

~~IPSL's performance is the reverse of MIROC. It performs better than most models with regard to both sea ice thickness and pattern of sea ice thickness, but the simulated sea ice extent is very low, and only NorESM-L simulates a smaller relative difference between minimum and maximum sea ice extent. HadCM3 performs well in the simulation of the overall annual cycle, although the October sea ice extent is exceeded by the observations, albeit by a small amount. NorESM-L appears to provide a comparably weak overall performance in the simulation of sea ice - whilst it reproduces the observed thickness distribution patterns reasonably well, like most models the simulated sea ice is thinner than the observations, and it produces a very low sea ice extent, which is lower than the observational values in four months. In addition to this, the magnitude of the annual cycle of sea ice extent simulated by NorESM-L is low, indicated by its $\lambda$-value of 57%, the highest in the ensemble.~~

### 3.2 Pliocene simulations

#### 3.2.1 Sea Ice Extent

~~Each~~ In agreement with enhanced greenhouse forcing each model in the ensemble simulates a smaller sea ice extent in the mid-Pliocene simulation in comparison to the pre-industrial (Figures

1, 2,7 and 8). The multi-model mean annual extent for the mid-Pliocene simulations is 10.84 $\times10^6$ km$^2$, a reduction of 5.33 $\times10^6$ 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 $\times10^6$ km$^2$ (NorESM-L), to 15.84 $\times10^6$ km$^2$ (MRI) (Table 1).

The lowest multi-model monthly mean extent is 3.15 $\times10^6$ km$^2$ (September), and the highest is 16.59 $\times10^6$ km$^2$ (March). In comparison to the pre-industrial simulation, the lowest multi-model monthly mean extent is reduced by ~~6.91~~6.86$\times10^6$ km$^2$ (69%). The reduction for the highest monthly multi-model mean is 4.65 $\times10^6$ km$^2$ (22%)~~, so the~~ The relative change in the lowest extent is therefore over three times greater than the relative change in the highest extent~~, resulting in an~~ . 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-L) the mid-Pliocene Arctic Ocean is ice-free at some point during the summer (August – September, Figure 9). In contrast to this, CCSM and MRI simulate minimum sea ice extents of 8.90 $\times10^6$ km$^2$ and 8.26 $\times10^6$ km$^2$ respectively, which both exceed the pre-industrial minimum of HadCM3 (7.00 $\times10^6$ km$^2$), with the CCSM minimum also exceeding the NorESM-L pre-industrial minimum (8.34 $\times10^6$ km$^2$). Consequently, there is an overlap in sea ice extents between the mid-Pliocene and pre-industrial simulations.

MRI, CCSM and MIROC simulate the highest maximum Pliocene sea ice extents in the ensemble. Both CCSM and MRI also provide the highest two minimum extents, but MIROC is one of the four models that simulates an ice free Arctic summer. ~~In the pre-industrial simulation, we examined the ratio between the minimum and maximum sea ice extent for each model in order to measure the magnitude of the annual sea ice extent cycle ($\lambda$-value). For the models that are sea ice free in some months, this value will be zero irrespective of the maximum value, so the $\lambda$-value is less useful in this climate scenario.~~ As a result, the sea ice extent amplitude in MIROC 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).

~~For the models that are not ice-free at any time during the year, GISS has a $\lambda$-value of 15.0%, which is much lower than any of the pre-industrial $\lambda$-values. CCSM, IPSL and MRI have respective $\lambda$-values of 46.5%, 42.5% and 37.6%, all of which are smaller than their respective pre-industrial $\lambda$-values, indicating a higher seasonal cycle in the Pliocene.~~

### 3.2.2 Sea Ice Thickness

~~Unlike the pre-industrial simulations, sea ice concentration north of 80°N in the Pliocene is not 100% for the majority of the year. We still show the thickness values for this region, but some calculations, particularly for the summer, are likely to be influenced by a low sea ice concentration.~~

Plots of the mean summer and winter mid-Pliocene Arctic sea ice thicknesses are shown in Figures 10 and 11 respectively. The multi-model mean annual sea ice thickness is ~~1.3~~ 1.48 m, which, compared to the pre-industrial simulations, is a reduction of ~~1.67 m (56.2%)~~ 1.72 m (53.9%). Across the ensemble, the annual mean thicknesses range from ~~0.44 m (NorESM-L) to 2.56 m (MRI)~~ 0.46 m (NorESM) to 2.08 m (MRI). The multi-model winter mean thickness is ~~1.77 m, 1.52 m (46.2%)~~ 1.85 m, 1.60 m (46.4%)less than the pre-industrial, whereas the summer multi-model mean thickness drops by ~~1.78 m (70.9%) to 0.73 m~~ 1.81 m (64.4%) to 1.00 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 thicknesses range from ~~0.79 m (NorESM-L) to 2.79 m (MRI)~~ 0.90 m (NorESM) to 2.80 m (MRI), with the summer thicknesses between ~~0.03 m (NorESM-L) and 2.24 m (MRI)~~ 0.08 m (NorESM) and 2.30 m (MRI).

Many of the models display similar thickness distribution patterns in the Pliocene simulations as they do in the pre-industrial, although the thickness values are reduced, particularly in the summer. The sea ice ~~distributions~~ spatial thickness patternssimulated by CCSM, HadCM3, IPSL and MRI are very similar to their pre-industrial equivalents in both summer and winter. The other four model simulations are ice-free for the majority of the summer in the mid-Pliocene, so no thickness pattern is detectable. In MIROC ~~has~~ the mid-Pliocene simulation showssimilar patterns in the winter to its pre-industrial counterpart~~, as does~~ . Similar findings hold for COSMOS, although in this model the central Arctic sea ice thins by a greater amount ~~in the Pliocene simulation~~ 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 the mid-Pliocene the region of greatest sea ice thickness is in this simulation north of Eastern Siberia. In the simulation with NorESM-L ~~loses~~ all sea ice to the north and east of Greenland is lost, and the thick sea ice, that is in the pre-industrial simulation located to the west of Greenland,thins considerably.

### 3.3 Variability across the ensemble

Figures 7 and 8 ~~appear to show~~ suggest that there is greater variability across the eight PlioMIP models in their mid-Pliocene simulation of summer sea ice compared to winter sea ice. This inference is in the following further studied in Figure 12, which shows the ~~coefficient of variation~~ CV of both the sea ice extent and thickness in the ensemble for each month, for both the pre-industrial and mid-Pliocene simulations.

~~The range of values of the pre-industrial sea ice extent CV is low~~ 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 ~~values are~~ is slightly lower, ~~the~~ 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 to the pre-industrial simulation ~~when the summer months have slightly lower CV values~~ where the summer is characterized by slightly lower CV values than the rest of the year. The large increase in the Pliocene summer CV supports the initial impression~~, given by Figures 7 and 8, that there is much greater variability~~ that across the ensemble there is greater variability of sea ice extent ~~simulation~~ in the Pliocene summer if compared to the remaining months in the mid-Pliocene~~, and~~ or the entirety of the pre-industrial simulation.

~~For each month, the CV of the Pliocene sea ice thickness is greater than in the pre-industrial ensemble (Figure 12).~~ The CV of the mid-Pliocene sea ice thickness is greater than in the pre-industrial ensemble for each month (Figure 12).In both experiments, the highest CV ~~values~~ occursduring 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. ~~The peak CV values for Pliocene sea ice thickness and extent are similar, but there is more variability in simulated sea ice thickness in comparison to sea ice extent variability.~~ For 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.

### 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 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 13 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 13 c,d). Whilst the winter pre-industrial sea ice thickness shows a weak relationship with the Pliocene winter sea ice extent (Figure 13 f), with a correlation coefficient of just 0.30, the relationship between the summer values is stronger, with a coefficient of 0.81 (Figure 13 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.. ~~Scatter plots for these values are shown in Figure 13.~~

The simulated 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) ~~in comparison to the pre-industrial simulations~~ than the pre-industrial sea ice extent and sea ice volume(~~see~~ Figure 14).

The correlation coefficient of the mid-Pliocene mean annual sea ice extent, ~~when compared with the SATs,~~ and the SAT, is -0.76, the ~~equivalent value for the~~ correlation coefficient of thepre-industrial sea ice extent with SAT is -0.18. ~~When compared with mean annual sea surface temperatures, the Pliocene sea ice extents show a correlation of~~ For SST the correlation with mid-Pliocene sea ice extents is -0.73, ~~with a~~ forpre-industrial sea ice extent the correlation coefficient ~~of~~ is -0.26. For the summer, the mid-Pliocene sea ice extents have a correlation coefficient of -0.88 with both SAT~~s~~ and SST~~s~~ (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~~, as Figure 14 suggests,~~ 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 simulations

Before examining the simulations of Arctic sea ice for the mid-Pliocene, ~~we first assess the simulations of pre-industrial sea ice cover by the same models~~ 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. ~~A significant restriction on this analysis is the difference between the climate states represented by models (pre-industrial) and observations (present day). As the observations are from the late 20th and early 21st century, then there is difficulty in using them as a reference to assess model simulations representative of a time period more than 100 years prior to the first observations.~~

Whilst there are ~~earlier~~ observations of sea ice characteristics available ~~that could have been used,~~ dating back ~~as far as~~ tothe early 20th century, that could have been used for the comparison, most, particularly the earliest, are ship-based observations of ice margins. ~~we decided to only use the satellite era observations, due to them being far more comprehensive, both spatially and temporally. Most of the earlier (non-satellite) observations, particularly the earliest, were based on observations of ice margins from ships, and are~~ 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. ~~Therefore, frequency~~ Frequency and location of these observations ~~was~~ are determined by shipping patterns, rather than 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.

~~Using the satellite-era observational data also provides over 30 years of reference data with which we could compare the annual cycles of the models. As the winter extents of earlier observations were achieved via extrapolation, these would not have been suitable for such a comparison.~~

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.

~~Six of the eight simulations show $\rho$-values lower than the observational values. The $\rho$-value can be used to identify models that perform less well, such as NorESM-L or IPSL, but similar to the comparisons of sea ice extent it is harder to distinguish between models that are not hugely different to observational values.~~

~~The difficulties faced in making comparisons between simulations and observations of sea ice extent are also present, to a greater degree, with the sea ice thickness comparisons. As with the observations of sea ice extent, the sea ice thickness observations do not relate to the same period of time as the model simulations represent. However, whereas the sea ice extent observations consist of over 30 years of daily observations covering the whole Arctic, the sea ice thickness measurements from Kwok et al. (2009) were obtained from only ten campaigns spanning a five year period (2003-2008), each campaign providing measurements over one month. Furthermore, sea ice thicknesses are not obtained for the area north of around 86°N.~~

~~The observations from Kwok et al. (2009) are useful in evaluating the patterns of sea ice thickness produced by each model, although the comparisons made have in this case been qualitative rather than quantitative. We do not know if the patterns observed in the last decade are the same as those of the 19th century or earlier.~~

~~Six of the PlioMIP models produce pre-industrial sea ice thicknesses that resemble, at least in some way, the observational patterns. By design, models have been tuned to best reproduce modern observations, although given that most of the models and their sea ice components were designed before the publication of the observations in Kwok et al. (2009), it would seem unlikely that the models have been tuned with these particular patterns in mind.~~

~~Due to these limitations of testing the various metrics to evaluate the skill of the models, it is difficult to justify an absolute ordered ranking for the eight models. For each metric, we are limited to determining only whether a model has or has not performed poorly in comparison to the observational data. The observational data of sea ice extent is, both spatially and temporally, more extensive than the thickness observations, and the associated uncertainties with the measurements are lower. Therefore, it could be argued, comparisons of model output with the observed sea ice extent should be considered more relevant when calculating overall model ranking than a respective~~

~~comparison of the sea ice thickness data. However, there is no clear objective method to determine what weighting each comparison should bear in the overall ranking estimation, so any such decision is likely to be subjective. The overall rankings would, to a certain extent, reflect these choices.~~

~~Whilst we do not produce a ranking for the entire ensemble, the comparisons of model simulations to the observations showed that most of the models reasonably simulated the winter and summer sea ice extents, but exhibited considerable variability in their performances in simulating sea ice thickness. This is particularly true for mean summer sea ice thicknesses, with four of the models simulating very thin sea ice. Overall, CCSM performs well against all the metrics, and HadCM3 and NorESM-L display a weaker performance in some areas.~~

First, we assess the simulated pre-industrial sea ice extent. Shu et al. (2015) provides an analysis of the simulation of Arctic sea ice by the CMIP5 models. Of the 7 PlioMIP models represented in CMIP5, MRI simulates the highest mean annual sea ice extent (15.01 $\times 10^6$ km$^2$), compared to the observational mean of 12.02 $\times 10^6$ km$^2$. MRI simulates the second highest pre-industrial mean annual sea ice extent (just 0.05 $\times 10^6$ km$^2$ less than MIROC), and the highest mid-Pliocene mean annual sea ice extent. The CMIP5 historical extent simulated by MRI is almost 25% greater than the observational mean, which may suggest that Arctic sea ice simulated by MRI is too extensive.

In contrast, MIROC simulates a pre-industrial mean annual sea ice extent that is similar to the MRI simulation, and represents the lowest mean annual sea ice extent of the CMIP5 models that are included in the PlioMIP ensemble (10.66 $\times 10^6$ km$^2$). The NorESM, which simulates both the lowest pre-industrial and mid-Pliocene mean annual sea ice extents, is the CMIP5 model that simulates the sea ice extent that is closest to the observations (12.01 $\times 10^6$ km$^2$) — although, like in the pre-industrial simulations, NorESM simulates the lowest sea ice extent amplitude of the PlioMIP models in CMIP5.

The HadCM3 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 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 and CMIP5, although the CMIP5 simulation is closer to the respective ensemble mean than the other two simulations.

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 simulations

~~Whilst the variability across the ensemble of simulated pre-industrial sea ice extent displays little change throughout the annual cycle, there is a noticeable rise in the variability across the ensemble of the simulated Pliocene sea ice extent during the summer months. The Pliocene Arctic Ocean is ice free at some point during the summer in half of the PlioMIP models (COSMOS, GISS, MIROC and NorESM-L).~~

Four models out of the eight-member PlioMIP ensemble (COSMOS, GISS, MIROC and NorESM) simulate 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 summer sea ice in HadCM3 is confined to the Arctic basin, with concentrations that do not exceed 60%. The summer sea ice margin in MRI, 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 8).

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 simulation of present day sea ice matches observations closely, 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 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.

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 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) implies that the maximum sea ice ~~extent~~ margin during the mid-Pliocene extended southwards beyond these two sites, but the minimum ~~extent~~ 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), although the sea ice concentration at these sites is less than 50% in the CCSM and IPSL 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 the 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 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

### 4.4 ~~Influence of Models' Sea Ice Components~~Influence of the sea ice models

~~GCMs are tuned to best reproduce modern day climate conditions, and parameterisations are based on modern observations (Eisenman et al., 2008; Eisenman et al., 2010). When simulating time periods with different climate states, such as the Pliocene, models tuned to present day may be biased in some regions.~~

The sea ice components of each model ~~can~~ differ in ~~atmospheric and oceanic~~ resolution, ~~the model~~ representation of sea ice dynamics and thermodynamics, and formulations 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 use the same sea ice model, based on CICE4 (Hunke, 2010), ~~CCSM is considered to have been the most successful at simulating pre-industrial sea ice, whereas NorESM-L is identified as performing weakly with respect to several sea ice metrics. The sea ice component has been developed for use with CCSM (Hunke, 2010). If elements of the sea ice model have been tuned, the tuning will be based on the climate state of CCSM, and may well not~~

~~be appropriate for NorESM-L, which~~ although NorESM has a coarser model grid in the atmosphere than CCSM, and ~~uses~~ furthermore employs a completely different ocean component (see Table 1).

The sea ice dynamics of the ensemble members can be categorised into three groups. First, CCSM, NorESM and MIROC use the elastic-viscous-plastic (EVP) rheology of Hunke and Dukowicz (1997). Second, COSMOS, GISS and IPSL have viscous-plastic (VP) rheologies (Marsland et al., 2003; Zhang and Rothrock, 2000; Fichefet and Morales Maqueda, 1999). Third, ~~and~~ HadCM3 and MRI do not consider any type of sea ice rheology, ~~with~~ the seaice 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 produce~~d~~ the two highest annual means for pre-industrial whilst having very different sea ice dynamics. The three models that produced extents lower than some of the observations, i.e. NorESM, IPSL and HadCM3, as well employ different rheology ~~—use~~ 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 the higher-order effects, such as ridging in the ice. However, whilst the spatial pattern of sea ice thickness simulated by HadCM3 ~~is considered to provide the weakest sea ice thickness simulation~~ compares poorly with observations, the spatial patterns simulated by MRI resemble some aspects of the observational patterns, ~~MRI simulates ice thicker than the ensemble mean, and its distribution compares well with the observations,~~ despite the lack of sea ice rheology. The sea ice thickness ~~distribution~~ spatial patterns in MIROC, which uses the more sophisticated EVP rheology, do~~es~~ 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 ~~using~~ 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 all utilise the 'zero-layer' model developed by Semtner (1976). Similarly to the considered sea ice dynamics, there is no clear ~~pattern between the differences in the~~ influence of the thermodynamics schemes used in the models on the simulated pre-industrial sea ice extent~~s and the thermodynamics schemes used in the models.~~

~~GCMs~~ General circulation models are tuned to best reproduce modern day climate conditions, and parameterisations are based on modern observations (Eisenman et al., 2008; Hunke, 2010). When simulating time periods with different climate states, such as the mid-Pliocene, models tuned towards present day 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).

#### 4.4.1 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 13 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 PlioMIP models' simulation of sea ice in CMIP5 is not the same as in the pre-industrial or mid-Pliocene simulations in the PlioMIP ensemble.

~~With the exception of HadCM3~~ 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 ~~previously been demonstrated to be~~ a stronger influence~~, in comparison to sea ice extent,~~ on the climate state of the Northern Hemisphere polar region in simulations of future climates, than sea ice extent ~~(Holland and Bitz, 2003)~~. 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 Pliocene sea ice extents and thicknesses, respectively. Mean summer pre-industrial sea ice extents show on the other hand 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 ~~the Pliocene summer~~ PlioMIP by COSMOS, GISS, MIROC and NorESM therefore appears to be an important factor in ~~each~~ the ability of those models ~~simulating~~ to simulate an ice-free mid-Pliocene summer~~, although~~. An exception is HadCM3, that simulates perennial sea ice in the Pliocene, despite simulating ~~the thinnest~~ relatively thin (within the PlioMIP ensemble) pre-industrial sea ice ~~of the ensemble~~.

#### 4.4.2 Influence of atmosphere and ocean on the sea ice simulation

~~From a physical point of view, it would seem likely that surface temperatures in the Arctic would have a strong influence on the state of the sea ice cover.~~ In the mid-Pliocene simulations, the correlation between Arctic surface temperatures and simulated sea ice extent is much stronger than the corresponding correlation ~~for~~ inthe pre-industrial simulations (Figure 14 a,b). ~~This is particularly noticeable in the summer months.~~ 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 14, 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 covered fully with sea ice, where the SSTs will be ~~now lower than~~ -1.8°C. The uniformity of the SSTs in these 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. ~~which are shown to have a much stronger correlation with the simulated sea ice extents (Figure 14).~~ This is potentially the reason for a much stronger correlation with the simulated mid-Pliocene sea ice extents (Figure 14). This explanation does not apply, however, to the SATs, ~~where~~ for which a similar difference in correlation strengths with sea ice extent between the pre-industrial and mid-Pliocene ~~in correlation strengths with sea ice extent is seen~~ is present.

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. (2013) 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), only two (COSMOS and GISS) 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 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 the same four that simulate an ice-free Pliocene summer (COSMOS, GISS, MIROC, and NorESM). The NorESM shows the largest warming due to clear sky albedo, ~~with~~ CCSM on the other hand shows ~~showing~~ the smallest clear sky albedo effect. Both ~~these models~~ NorESM and CCSM use the same sea ice component, based on CICE4 (Hunke and Lipscomb, 2008).~~, which uses~~ 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 contributions to high latitude warming by clear sky albedo in NorESM and CCSM4 is reflected in the large difference in their simulations of summer mid-Pliocene sea ice. ~~Due to the nature of the sea-ice albedo feedback mechanism~~ 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 result of reduced sea ice extent. Models with parameterisations that produce lower sea ice albedos have thereforea greater potential to amplify the warming ~~from other factors~~ that originates from other sources in simulations of the mid-Pliocene, such as greenhouse gas emissivity~~, that is seen in simulations of the Pliocene~~. The low sea ice albedo ~~generated by~~ assumed in NorESM is a likely explanation for the low sea ice extents it simulates (Figures 3 and 9), both in mid-Pliocene and pre-industrial simulations.

~~After~~ Second to NorESM, for MIROC clear sky albedo has the highest contribution to high latitude ~~warming from clear sky albedo~~. In MIROC ~~has~~ there is a fixed albedo of 0.5 for bare ice, with higher albedo for snow-covered sea ice, ~~which vary~~ 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 a lower albedo minimum than 0.5~~, but it~~ Yet, this model allows the albedo to vary ~~from~~ between 0.44 ~~to~~ and 0.84 (Schmidt et al., 2006). All other models allow the sea ice albedo to vary, and ~~so~~ consequently MIROC has a lower overall albedo. This may help to explain ~~how~~ the ability of MIROC to simulate~~s~~ 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.

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 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.

### 4.5 ~~Implications For Future Climate Predictions~~

~~Evidence that models have been tuned in order to simulate a desired pre-industrial sea ice extent, and the weak correlation between sea ice extent and surface temperatures in these simulations, could suggest that the models may not be ideally suited to simulating sea ice in a climate state different to modern. The Pliocene Arctic sea ice simulated by HadCM3 has the most favourable comparison with the proxy data evidence of the entire PlioMIP ensemble, despite its pre-industrial Arctic sea ice simulation comparing relatively poorly with the observational data. The relationship between the simulated pre-industrial and Pliocene sea ice extents (Figure 13) also suggests that a model's simulation of pre-industrial sea ice is not necessarily a good predictor of its simulation of Pliocene sea ice.~~

~~If the performance of a model at simulating pre-industrial sea ice is not a reliable indicator of how well it simulates Pliocene sea ice, then it may also not be a reliable indicator of how well a model simulates future changes in Arctic sea ice. Sea ice output from CMIP5 simulations shows greater consistency with satellite observations of Arctic sea ice in comparison to previous modelling studies (Stroeve et al., 2012), but if this improvement is due to greater tuning, and not a better overall representation of sea ice processes, then it may have a detrimental effect on more accurately predicting Arctic sea ice in the future. This uncertainty highlights the importance of palaeoclimate studies, to test the performance of models' simulations of climate states different to modern, and thus the pressing need for greater proxy data evidence of sea ice coverage in the past.~~

## 5 Conclusions

We have presented a detailed analysis of the simulation of Arctic sea ice in the PlioMIP model ensemble, for both the pre-industrial control and Pliocene simulations. The sea ice in the Pliocene simulations is overall less extensive and thinner than the pre-industrial sea ice, with a 33% drop in mean annual sea ice extent for the ensemble mean, and a ~~56~~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, for both sea ice extent and sea ice thickness.

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

The simulated mid-Pliocene sea ice extents are strongly negatively correlated with the Arctic temperatures. In contrast, ~~to the~~ there is only a weak correlation between ~~the~~ pre-industrial sea ice extents and temperature~~s~~. Hill et al. (2014) identified clear sky albedo as the dominant driver of high latitude warming in the mid-Pliocene simulations of PlioMIP, ~~in particular in those that become ice-free in the summer~~ particularly in those models that simulate an ice-free mid-Pliocene summer. Sea-ice albedo feedbacks ~~are therefore likely to have~~ may contribute~~d~~ to the stronger relationship between surface temperatures and sea ice in the mid-Pliocene simulations, as the feedback mechanism enhances the warming ~~due to~~ that originates from increased greenhouse gas concentrations. ~~In contrast, tuning of sea ice components in some of the PlioMIP models is likely to have contributed to the weak relationship between pre-industrial sea ice and Arctic temperatures.~~ 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.

~~Three models (HadCM3, IPSL and NorESM-L) produce sea ice extents that in some months are exceeded by the observational mean extents. Given the decline in sea ice during and prior to the period of observations (1981-2010), this appears to indicate that the simulated pre-industrial sea ice in these models is insufficiently extensive. NorESM-L and IPSL are also identified as producing a weaker simulation of the annual cycle of sea ice, based on their high minimum to maximum sea ice extent ratio, and the low overall extent.~~

~~Most models simulated pre-industrial thickness thinner than observational measurements from the 21st century. In particular the summer sea ice is demonstrably far too thin in over half of the models.~~ 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 ~~a~~ difficulty in simulating observed sea ice thickness. ~~Despite this, it does not simulate an ice-free Pliocene Arctic during the summer months, unlike half of the models in the ensemble, and so is~~ 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 Pliocene by Darby (2008).

~~Despite simulating pre-industrial sea ice that compares poorly with observations,~~ The HadCM3 is the only model that simulates both perennial mid-Pliocene Arctic sea ice and a minimum sea ice extent that ~~has completely retreated beyond~~ 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, ~~at which~~ where IP$_{25}$ proxy data indicates the presence of a sea ice margin in the ~~mPWP~~ mid-Pliocene. This appears to suggest that HadCM3 produces the ~~best simulation of~~ mid-Pliocene simulation that is in best agreement with both inferences from the proxy record, 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.~~sea ice in the ensemble, but the data is from just two sites in the same region, and understanding of Pliocene sea ice is still too low to have confidence in this simulation.~~

~~Simulations of past climates such as the Pliocene can be used to test the sensitivity of models to conditions which are very different to those experienced in the instrumental period. Due to model tuning, we can not assume that the high performance in pre-industrial simulations of models such as CCSM and COSMOS is still valid for the simulation of a different climate state. This is reinforced by the differences between the sea ice simulated by PlioMIP models and the indications of Pliocene sea ice margins and perennial sea ice in Knies et al. (2014) and Darby (2008).~~

~~If a model's performance simulating pre-industrial sea ice is not a reliable indicator of its ability to simulate the sea ice of past climates, then it may also not be a reliable indicator of the model's skill in simulating future sea ice. This underlines both the importance and value of palaeoclimate modelling studies in assessing model skill and improving predictions of future climate, and the urgent requirement for more sea ice proxy data that allow a grading of the models against an independent benchmark.~~

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 4.4.2, 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. 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.C. 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 model. Z.Zhang would like to thank Mats Bentsen, Jerry Tjiputra, Ingo Bethke from Bjerknes Center for Climate Research for the contribution to the development of NorESM-L.

## References


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.

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.

Comiso, J. C., Parkinson, C. L., Gersten, R., and Stock, L.: Accelerated decline in the Arctic sea ice cover, Geophysical Research Letters, 35, L01 703, doi:10.1029/2007gl031972, 2008.

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.

Fetterer, F., Knowles, K., Meier, W., and Savoie, M.: Sea Ice Index, National Snow and Ice Data Center, Boulder, CO, digital media. [Available online at http://nsidc.org/data/g02135.html.] Last access 17/9/2014, 2002.

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.

Haywood, A. M., Dowsett, H. J., Otto-Bliesner, B. L., Chandler, M. A., Dolan, A. M., Hill, D. J., Lunt, D. J., Robinson, M. M., Rosenbloom, N., Salzmann, U., and Stoll, D. K.: Pliocene Model Intercomparison Project (PlioMIP): experimental design and boundary conditions (Experiment 1), Geosci. Model Dev., 3, 227–242, 2011a.

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, 2011b.

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., 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.

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. and Cunningham, G. F.: ICESat over Arctic sea ice: Estimation of snow depth and ice thickness, Journal of Geophysical Research: Oceans, 113, C08 010, doi:10.1029/2008jc004753, 2008.

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.: Sensitivity of sea ice to physical parameterizations in the GISS global climate model, Journal of Geophysical Research, 108, 3053, 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.

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. 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., Cavalieri, D. J., Gloersen, P., Zwally, H. J., and Comiso, J. C.: Arctic sea ice extents, areas, and trends, 1978-1996, Journal of Geophysical Research-Oceans, 104, 20 837–20 856, 1999.

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.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, Journal of Geophysical Research-Atmospheres, 108, 4407, doi:10.1029/2002jd002670, d14, 2003.

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, 2013.

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 ~~details~~ and references for each of the eight PlioMIP Experiment 2 simulations.

| ~~GCM~~ 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 | $0.9 \times 1.25$ | $1 \times 1$ | 1300/100 | 550/100 | EVP rheology, melt ponds Hunke and Dukowicz (1997); Hunke (2010); Holland et al. (2011) | Rosenbloom et al. (2013) |
| COSMOS | $3.75 \times 3.75$ | $3 \times 1.8$ | 3000/30 | 1000/30 | VP rheology, leads Marsland et al. (2003) | Stepanek and Lohmann (2012) |
| GISS-E2-R | $2 \times 2.5$ | $1 \times 1.25$ | 950/30 | 950/30 | VP rheology, leads Zhang and Rothrock (2000); Liu et al. | Chandler et al. (2013) |
| HadCM3 | $2.5 \times 3.75$ | $1.25 \times 1.25$ | 200/50 | 500/50 | Free drift, leads Cattle and Crossley (1995) | Bragg et al. (2012) |
| IPSLCM5A | $3.75 \times 1.9$ | $0.5 - 2 \times 2$ | 2800/100 | 730/30 | VP rheology, leads Fichefet and Morales Maqueda (1999) | Contoux et al. (2012) |
| MIROC4m | $2.8 \times 2.8$ | $0.5 - 1.4 \times 1.4$ | 3800/100 | 1400/100 | EVP rheology, leads K-1 Model Developers (2004) | Chan et al. (2011) |
| MRI-CGCM | $2.8 \times 2.8$ | $0.5 - 2 \times 2.5$ | 1000/50 | 500/50 | Free drift, leads Mellor and Kantha (1989) | Kamae and Ueda (2012) |
| NorESM-L | $3.75 \times 3.75$ | $3 \times 3$ | 1500/200 | 1500/200 | Same as CCSM4 | Zhang et al. (2012) |

**Table 2.** ~~Mean annual sea ice extents for the pre-industrial simulations for each participant model in PlioMIP Experiment 2, the ensemble mean and for sea ice extent observations from 1981-2010 (Fetterer et al., 2002). The $\lambda$ and $\rho$-values for each model, as well as for the ensemble mean and for observations are also shown.~~ 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, for each participant model in PlioMIP Experiment 2 and for the ensemble mean.

| Model | PImean annual extent ($\times10^6$ km$^2$) | ~~$\lambda$~~ PI extent amplitude ($\times10^6$ km$^2$) | ~~$\rho$~~ mid-Pliocene mean annual extent ($\times10^6$ km$^2$) | mid-Pliocene extent amplitude ($\times10^6$ km$^2$) |
|---|---|---|---|---|
| CCSM4 | 18.35 | ~~53.42~~10.94 | ~~2.95~~14.99 | 10.26 |
| COSMOS | 15.52 | ~~45.74~~11.66 | ~~2.91~~7.72 | 12.75 |
| GISS-E2-R | 17.30 | ~~38.87~~14.03 | ~~2.25~~9.63 | 15.43 |
| HadCM3 | 13.76 | ~~36.05~~12.42 | ~~2.62~~ 10.38 | 14.17 |
| IPSLCM5A | 12.27 | ~~54.15~~7.36 | ~~4.42~~ 9.06 | 7.05 |
| MIROC4m | 19.85 | ~~46.88~~14.05 | ~~2.36~~ 11.48 | 21.98 |
| MRI-CGCM | 19.80 | ~~41.11~~15.91 | ~~2.08~~ 15.84 | 13.69 |
| NorESM-L | 12.52 | ~~56.95~~6.39 | ~~4.55~~7.60 | 12.86 |
| Ensemble mean | 16.17 | ~~47.35~~11.18 | ~~2.92~~ 10.84 | 13.44 |
| ~~Observations~~ | ~~12.01~~ | ~~41.96~~ | ~~3.49~~ | |

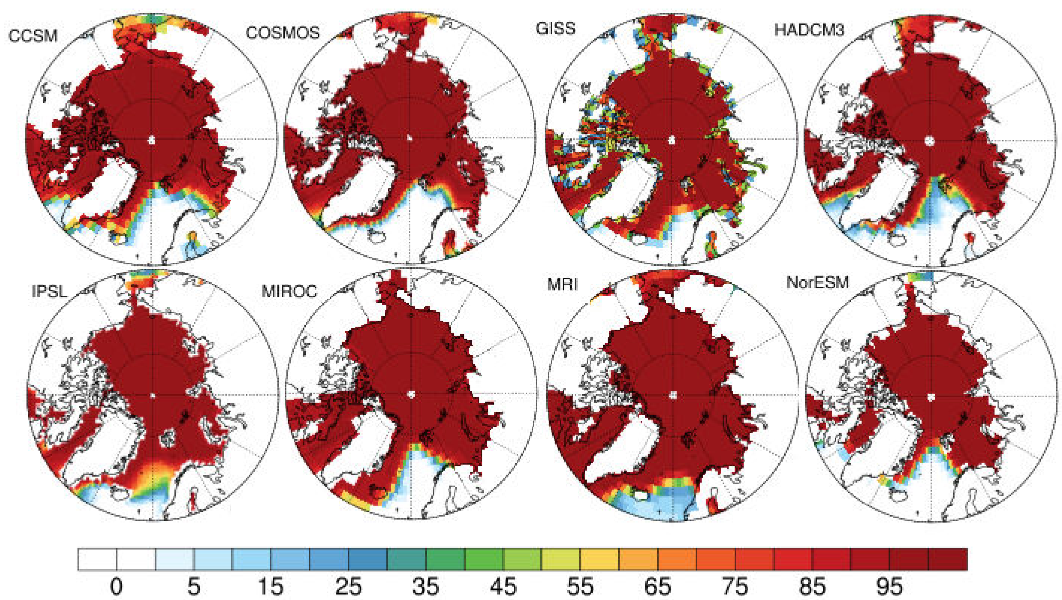


**Figure 1.** Mean winter (FMA) sea ice concentrations (%) in the pre-industrial control simulations for each PlioMIP Experiment 2 model, ~~and observations from 1981-2010 (Fetterer et al., 2002).~~ Missing data at the poles in each plot is a plotting artefact (seen also in Figures 2, 4, 5, 6, 7, 8, 10 and 11).

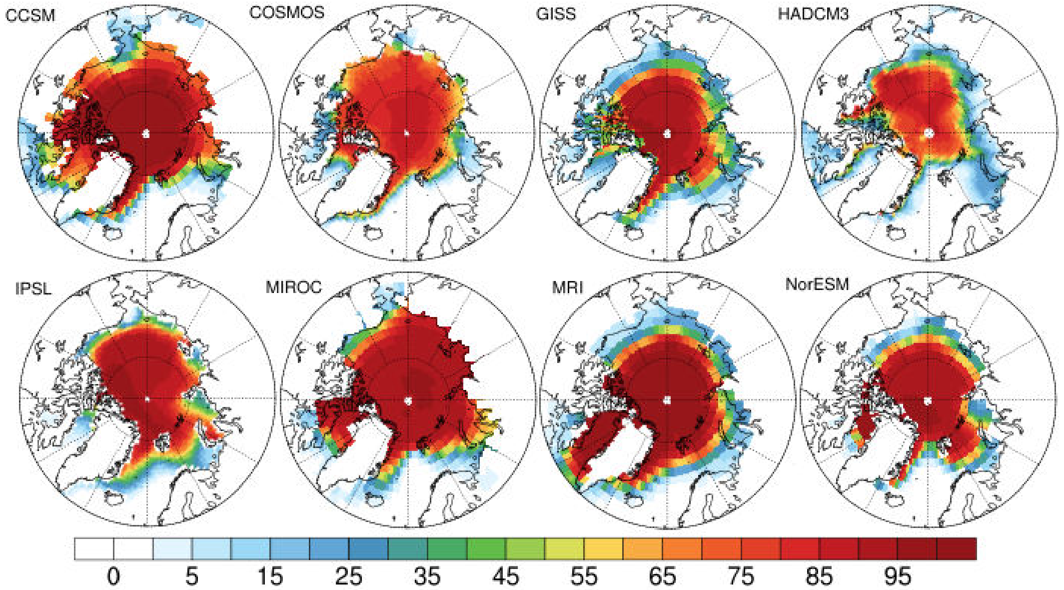


**Figure 2.** ~~Mean summer (ASO) sea ice concentrations (%) in the pre-industrial control simulations for each PlioMIP Experiment 2 model, and observations from 1981-2010 (Fetterer et al., 2002).~~As Figure 1, but for mean summer (ASO) sea ice concentrations (%).

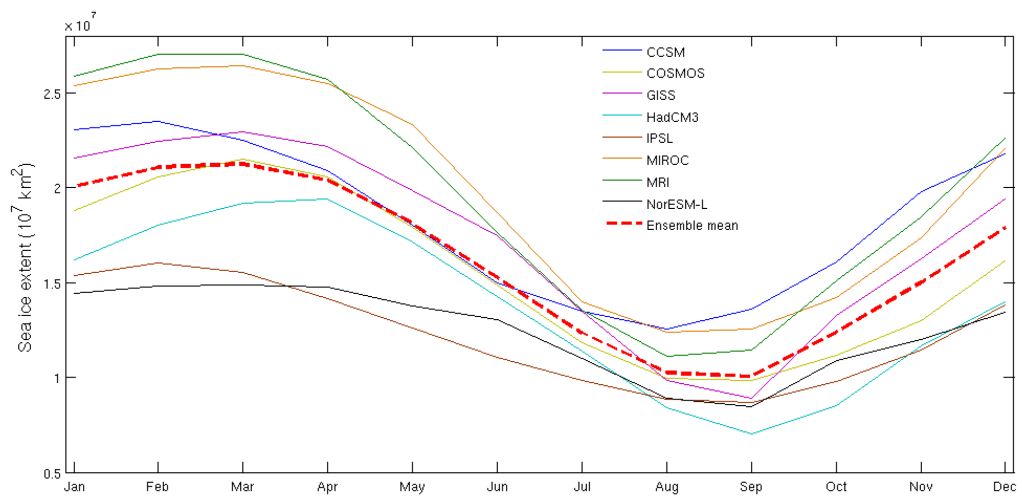


**Figure 3.** Annual cycle of total Arcticsea ice extent in the pre-industrial simulations for each participating model in PlioMIP Experiment 2, ~~together with~~ andthe ensemble mean ~~extent cycle. Also shown is the annual cycle of sea ice extent for the mean of observations from 1981-2010 (Fetterer et al., 2002).~~

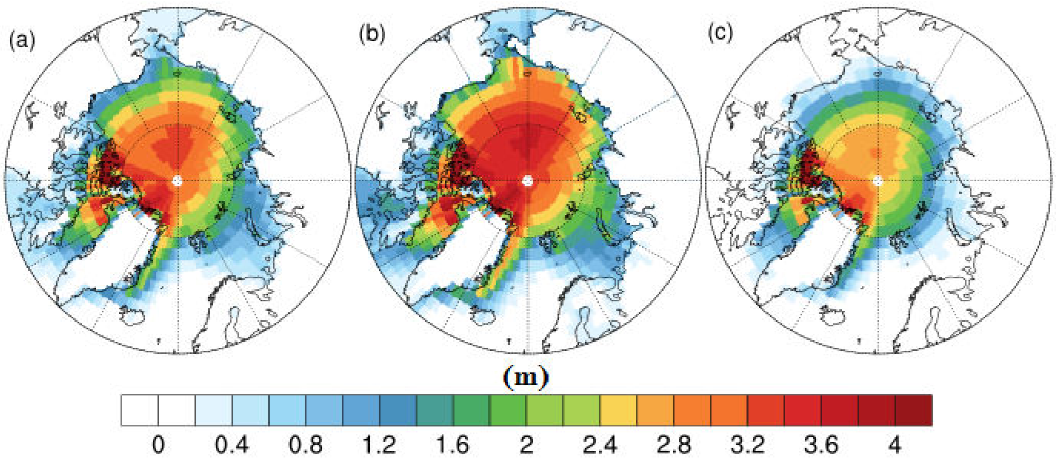


**Figure 4.** Mean sea ice thickness (m) in the pre-industrial simulations for the entire PlioMIP Experiment 2 ensemble, for (a) annual, (b) winter (FMA), and (c) summer (ASO).

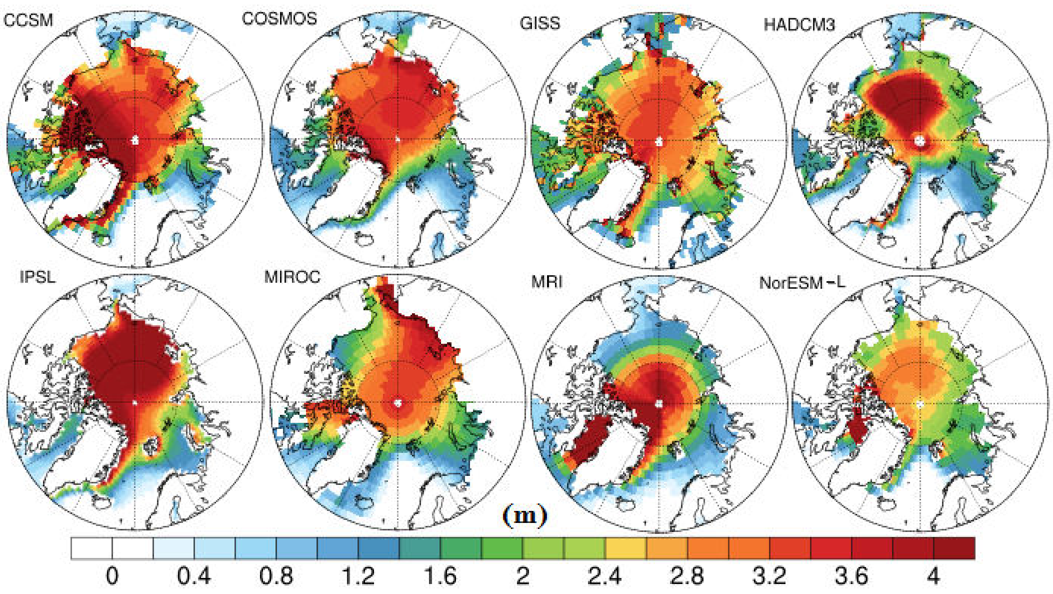


**Figure 5.** Mean winter (FMA) sea ice thicknesses (m) in the pre-industrial control simulations for each PlioMIP Experiment 2 model.

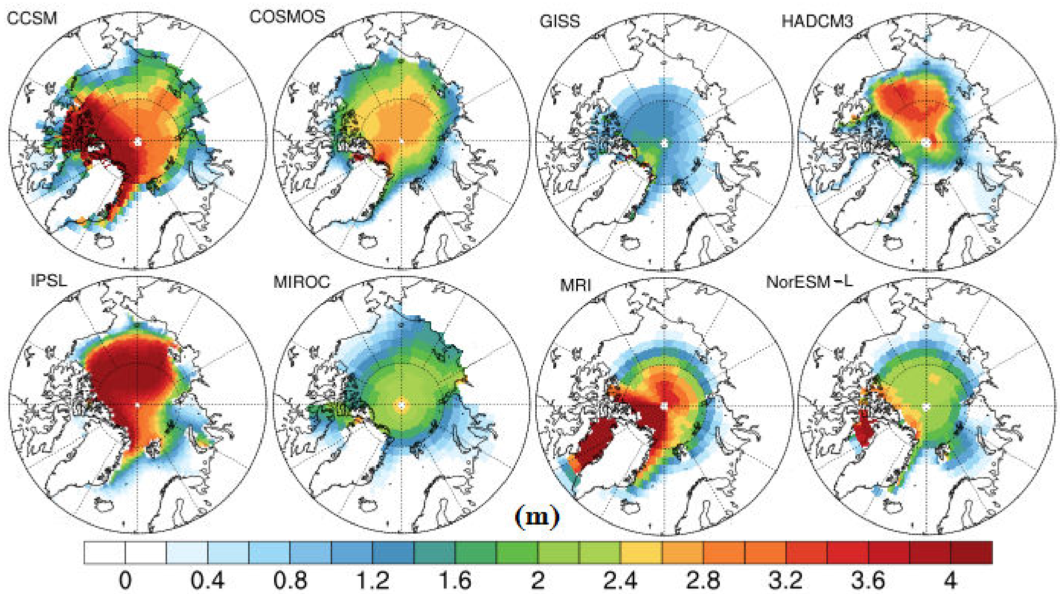


**Figure 6.** ~~Mean summer (ASO) sea ice thicknesses (m) in the pre-industrial control simulations for each PlioMIP Experiment 2 model.~~ As Figure 5, but for mean summer (ASO) sea ice thicknesses (m).

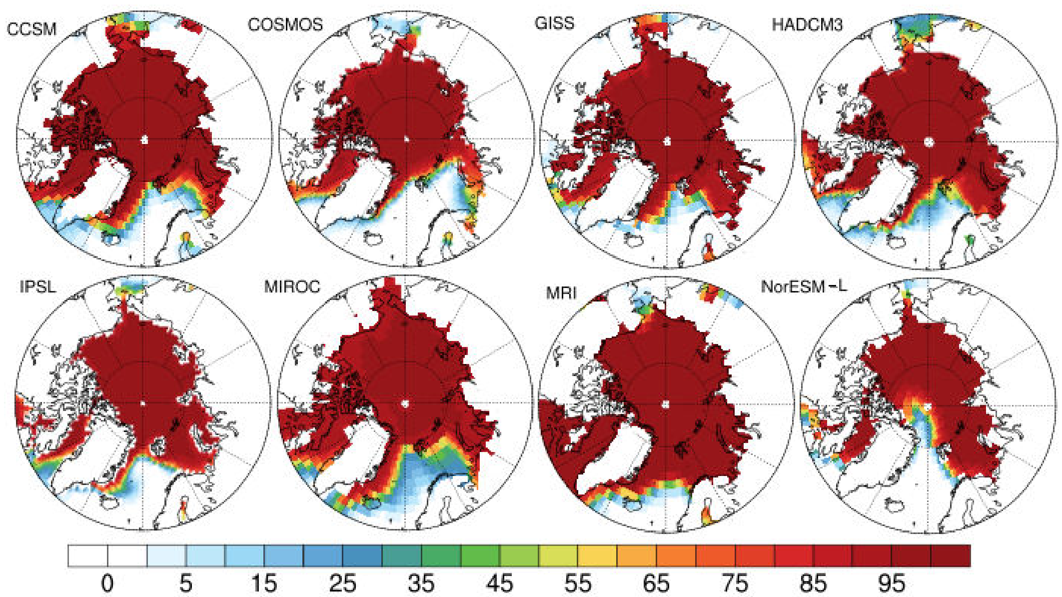


**Figure 7.** Mean winter (FMA) sea ice concentrations (%) in the Pliocene simulations for each PlioMIP Experiment 2 model.

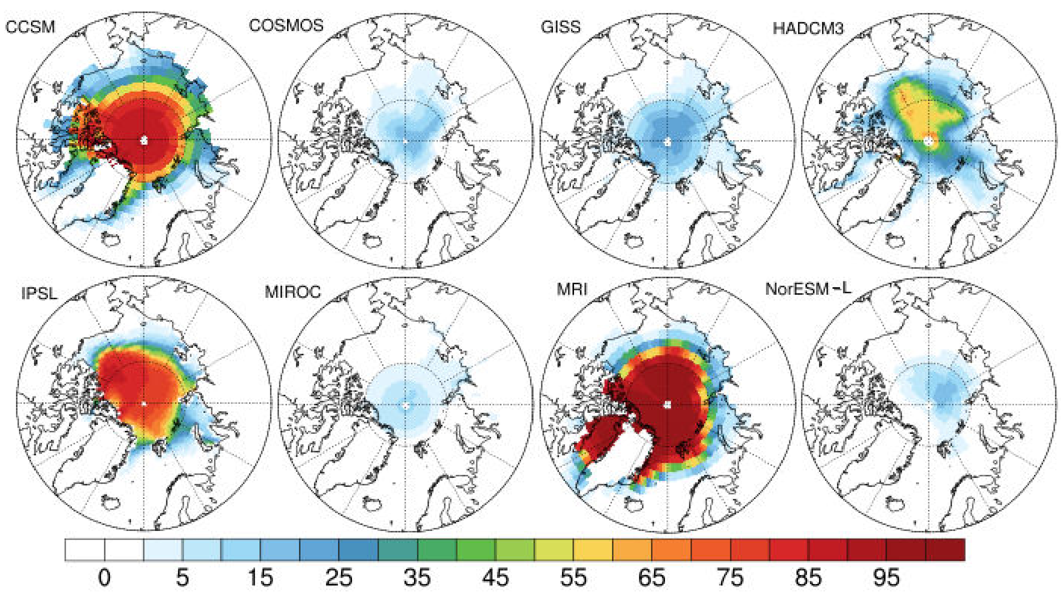


**Figure 8.** ~~Mean summer (ASO) sea ice concentrations (%) in the Pliocene simulations for each PlioMIP Experiment 2 model.~~ As Figure 7, but for mean summer (ASO) sea ice concentrations (%).

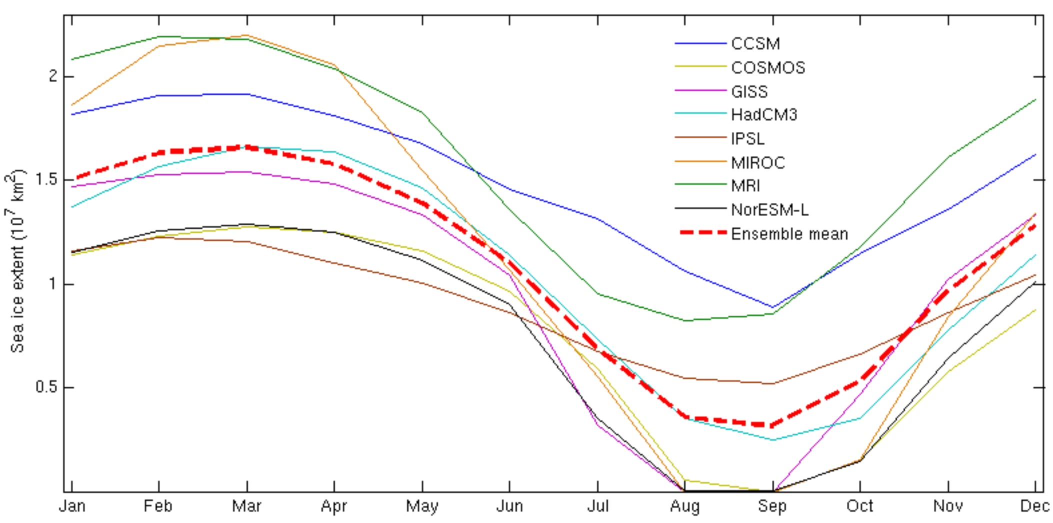


**Figure 9.** Annual cycle of sea ice extent in the mid-Pliocene simulations for each participating model in PlioMIP Experiment 2, and for the ensemble mean.

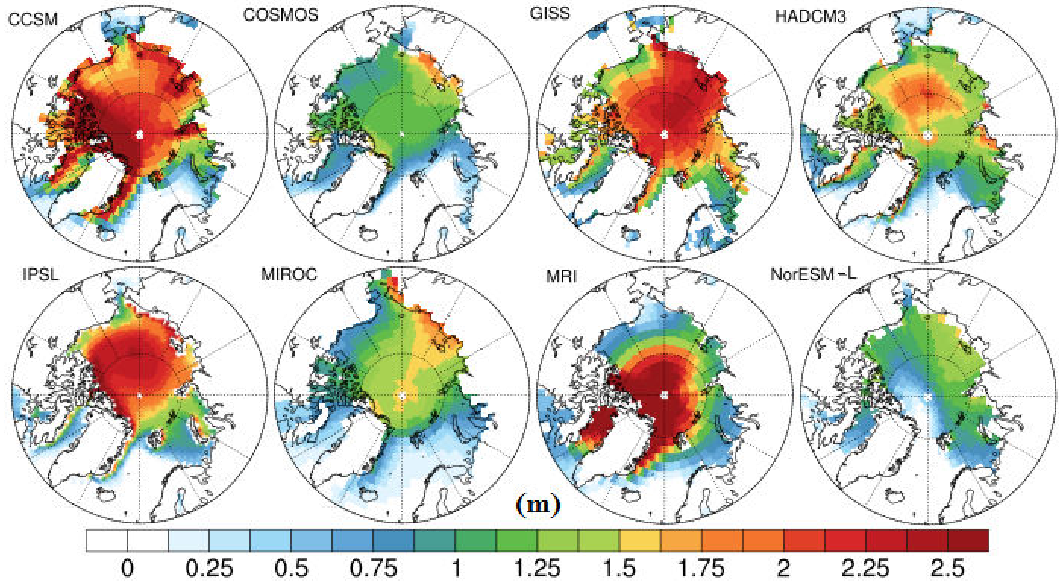


**Figure 10.** Mean winter (FMA) sea ice thicknesses (m) in the mid-Pliocene simulations for each PlioMIP Experiment 2 model.

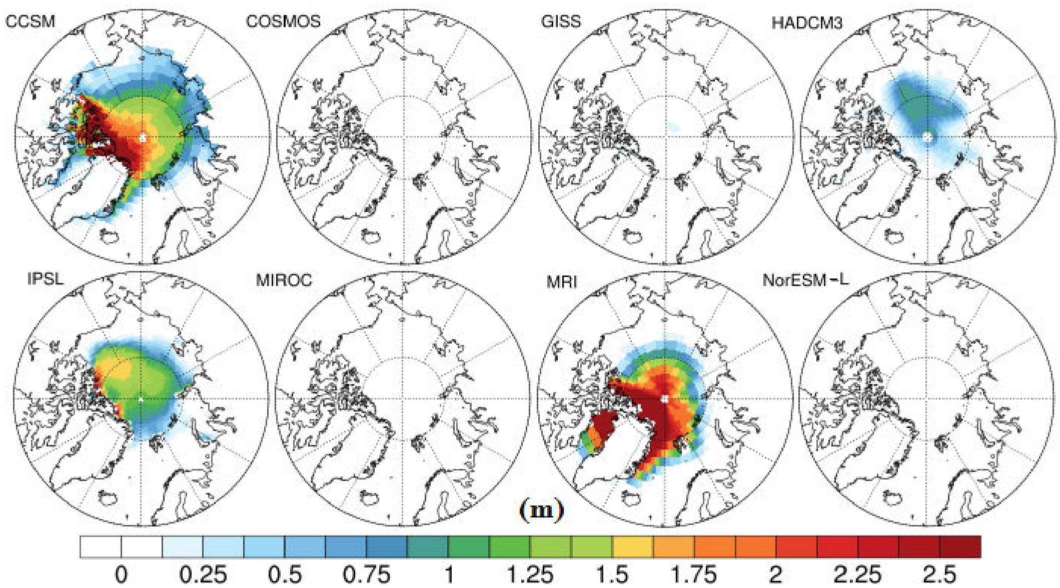


**Figure 11.** ~~Mean summer (ASO) sea ice thicknesses (m) in the Pliocene simulations for each PlioMIP Experiment 2 model. Low sea ice concentrations in COSMOS, GISS, MIROC and NorESM result in mean thicknesses very close to zero in each model grid cell.~~ As Figure 10, but for mean summer (ASO) sea ice thicknesses (m). Low sea ice concentrations in COSMOS, GISS, MIROC and NorESM result in mean thicknesses very close to zero in each model grid cell.

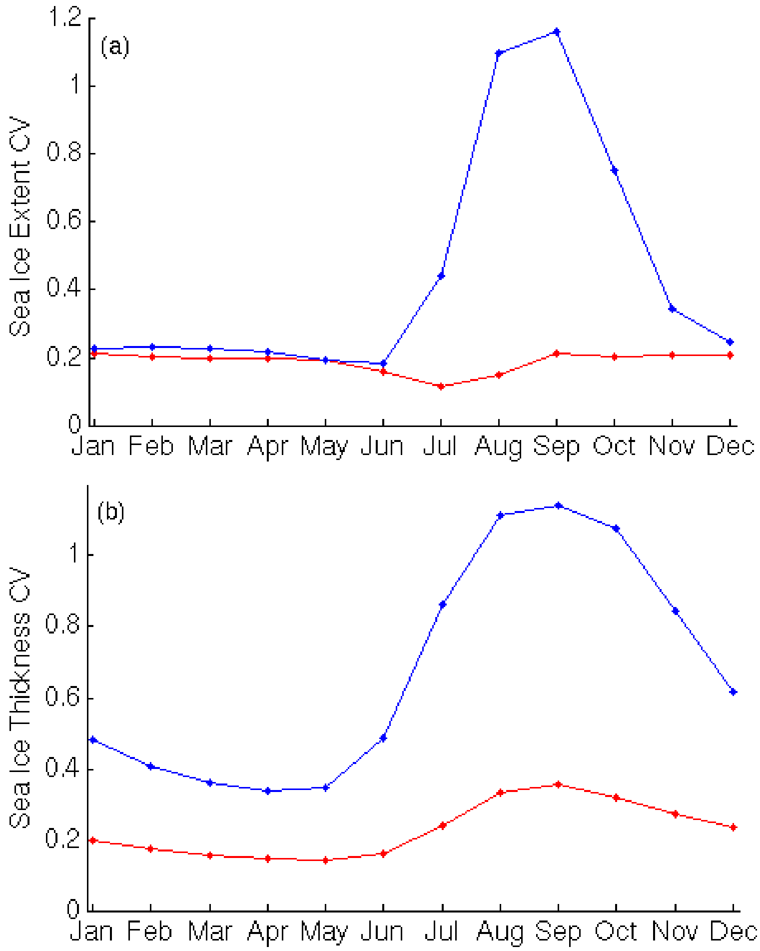


**Figure 12.** Annual cycles of the coefficient of variation (CV) of (a) sea ice extent and (b) sea ice thickness for the PlioMIP Experiment 2 ensemble. Red lines represent the pre-industrial annual cycle, blue lines represent the mid-Pliocene annual cycle.

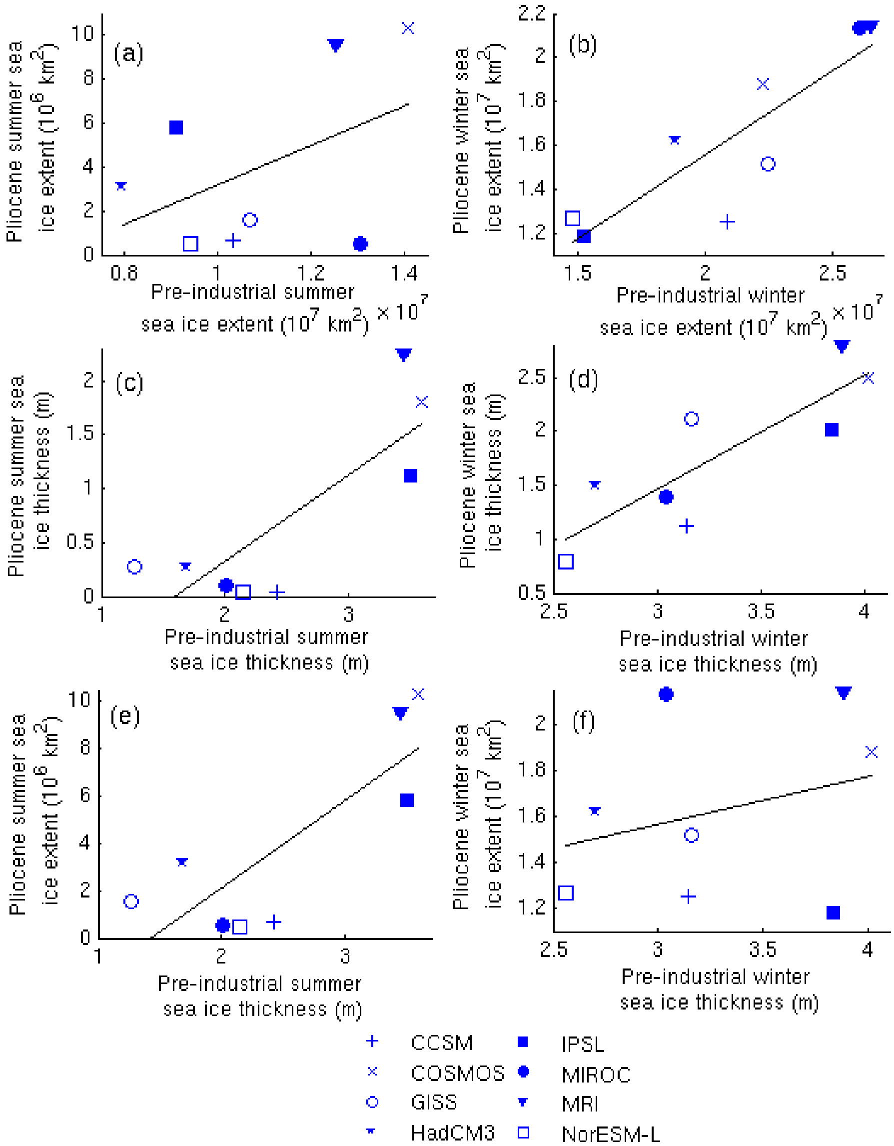


**Figure 13.** ~~Pre-industrial values vs. Pliocene values showing (a) and (b) sea ice extent vs. sea ice extent, (c) and (d) sea ice thickness vs. sea ice thickness, (e) and (f) sea ice thickness vs. sea ice extent. (a), (c), and (e) show summer values, (b), (d), and (f) show winter values.~~ Relationship 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

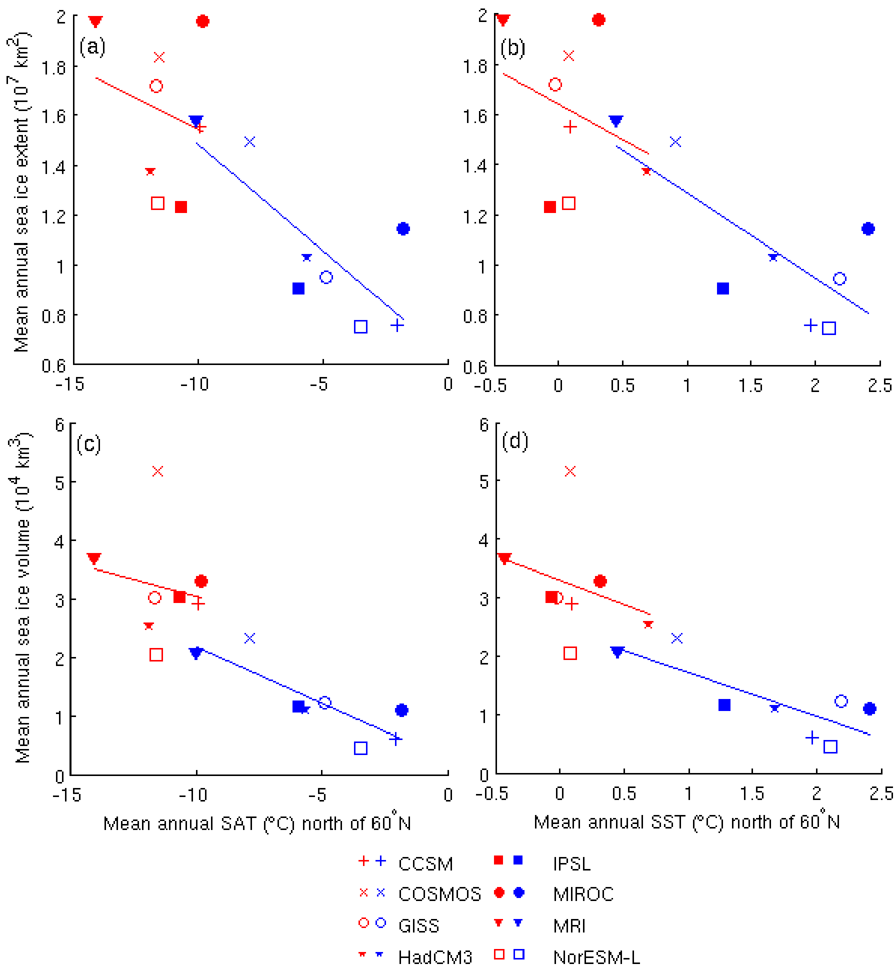


**Figure 14.** ~~Mean annual surface temperatures north of 60°N vs. mean annual sea ice extent, in both pre-industrial and Pliocene simulations, for (a) SAT and (b) SST. Pre-industrial experiments are marked in red, and Pliocene experiments are marked in blue.~~ Mean annual surface temperatures north 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