Clim. Past Discuss., 11, C944–C955, 2015
www.clim-past-discuss.net/11/C944/2015/

Climate of the Past Discussions Open Access

Interactive comment on "Arctic sea ice in the PlioMIP ensemble: is model performance for modern climates a reliable guide to performance for the past or the future?" *by* F. W. Howell et al.

F. W. Howell et al.

eefwh@leeds.ac.uk

Received and published: 3 July 2015

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:

(Reviewer 2) *The article "Arctic sea ice in the PlioMIP ensemble: is model performance for modern climates a reliable guide to performance for the past or the future?" by F.W. Howell et al. uses pre-industrial and mid-Pliocene model simulations from 8 different models to assess the reliability of sea ice simulations in the different models at present day and the spread of sea ice simulations in the Pliocene. The authors try to find*

connection between performance at present day/ pre-industrial and the results for the Pliocene. As metrics they use ice extent and thickness and create three additional ice metrics based on the amplitude of the annual cycle of sea ice and the ice extent. Results indicate an increased cross-model spread in the Pliocene and limited dependence of the Pliocene ice conditions on the pre-industrial ice conditions. The topic of this article is interesting and the question posed in the title of high relevance. Unfortunately, the investigations performed in this study are not sufficient to contribute answering the question. The analysis of causes for the different behavior between models is not convincing and quite superficial. There are many articles around, which have discussed possible differences in Arctic climate and sea ice in present day and future simulations in much more detail than this article. In order to become publishable, this article needs a number of substantial major revisions and needs to be rewritten in larger parts.

Main comments:

1. 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.

In section 2.2 (pg 1268) we set out why we chose the post-1979 observation data for comparison. Whilst we felt that this offered the most useful comparison when taking into account the advantages and disadvantages of other data sets, we accept that the discrepancy in the time represented by the models and data makes the comparison unsatisfying. We intend to remove this direct comparison 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., The Cryosphere, 2015)

2. 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.

We accept that there is not sufficient proxy data evidence relating to the sea ice coverage during the mid-Pliocene to assess the model performance at simulating mid-Pliocene sea ice, although it is improving. Parts of the paper and the title will be altered so that the emphasis is not on using the Pliocene sea ice results to assess model performance. Instead, the paper will stress the importance of the need for greater proxy data coverage relating to the sea ice conditions in the Pliocene (such as Knies et al. (Nature Communications, 2014)), so that in the future results of PlioMIP or

subsequent model intercomparisons may be used to assess future model performance.

3. 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 intend to remove the λ and ρ figures from Table 2, and any references to them within the text of the article. We will add to Table 2 the metric of motnthly sea ice extent amplitude (maximum monthly extent - minimum monthly extent), a metric used in Shu et al. (The Cryosphere, 2015), which analyses CMIP5 sea ice output. This will replace λ and ρ 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. CV is used in Stroeve et al. (The Cryosphere, 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 accept that the influence of tuning may have been overstated in the paper, and we will change the emphasis in the discussion, so that we do not explicitly state that the tuning may have reduced the correlation between temperatures and extent. We still feel it is important to mention the potential impact of tuning on sea ice. We will cite both Eisenman et al. (GRL, 2007), which discusses the influence of tuning of sea ice albedo on model sea ice thickness, and DeWeaver et al. (Comment on GRL, 2008), which questions the strength of the impact of the tuning. Mauritsen et al. (Journal of Advances in Modeling Earth Systems, 2012), which mentions more explicitly the aims when tuning sea ice in a model, will also be cited.

The discussion will be amended to reflect the fact that the pre-industrial sea ice is thicker, and thus may be affecting the correlation of temperature with extent. Relationships between sea ice volume and temperature will also be included, which should support this assertion.

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 and others will be 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.

As mentioned in the response to main comment #2, the paper and title will be altered, so that there is not an emphasis on assessing model future performance based on Pliocene results.

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 will be rephrased (e.g. from ‘will enable’ to ‘might enable’)

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%”.

Calculation of sea ice thickness will be adjusted to take into account the sea ice concentration. As we will remove the comparison to observations, then we will no longer use the 66-86N region for the figures.

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.

Whilst ASO and FMA are not the minimum and maximum three months in all models, they are in at least half the models for both pre-industrial and Pliocene simulations. The statement will be amended so it is clear this is not the case in all models.

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

This will be 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. (The Cryosphere, 2014))

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.

The line will be changed.

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.

Wording will be changed to be more specific than it is now.

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 will be removed. Discussion of model performance compared to observations will be based on CMIP5 simulations (e.g. Shu et al. (The Cryosphere, 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.

λ and ρ will be removed. Sea ice extent amplitude (E_{max} - E_{min}) will be used as the measure of the sea ice extent cycle, as used in Shu et al. (The Cryosphere, 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 information 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.

The revised version of the paper will omit an explicit evaluation of model performance, as we are unable to formulate clear and justifiable criteria. Discussion of model performance will be based on CMIP5 simulation comparisons to modern observations, as published in Shu et al. (The Cryosphere, 2015).

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 figures will also be shown.

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.

As there are only 8 models then nothing can be done to increase the values in the correlation, but discussion will be amended so that this is taken into account, and significance of the correlations will be looked at.

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. (The Cryosphere, 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.

As mentioned previously, the revised version of the paper will not look at assessing model performance based on the Pliocene sea ice simulations.

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.

As mentioned previously, other atmospheric and oceanic influences will be explored,

and the discussion of tuning will be altered and relevant literature cited.