

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.

Anonymous Referee #2

Received and published: 12 May 2015

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 arti-

C360

cle 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.

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

C361

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.

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.

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

C362

than an untuned model. Especially not, if as I assume, the same tuned model versions has been used to run the Pliocene time slices.

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).

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.

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.

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 grid-point 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%".

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.

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

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.

P 1267, line 24/25: "CV identifies in which months there is greater spread across the

C363

ensemble". Greater than what? Maybe rephrase to: The CV identifies the months with large sea ice spread across 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.

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.

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 $E_{max} - E_{min}$. On page 1273 you state yourself that Λ seems to be dependent on the ice extent. Please explain why you introduce Λ , and why you think it is better than using absolute values. If someone else already used Λ , cite the relevant literature.

Page 129, Equation 2: I am not convinced by equation 1, thus, I am of course not either by equation 2 since it is based on Λ . 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.

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

C364

south the ice extends.

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 Λ and ρ 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.

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?

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.

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.

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.

C365

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.

Interactive comment on Clim. Past Discuss., 11, 1263, 2015.

C366