The Cryosphere Discuss., 8, C821–C830, 2014
www.the-cryosphere-discuss.net/8/C821/2014/

The Cryosphere
Discussions
Open Access

Interactive comment on "Using records from submarine, aircraft and satellite to evaluate climate model simulations of Arctic sea ice thickness" *by* J. Stroeve et al.

F. Massonnet (Referee)

francois.massonnet@uclouvain.be

Received and published: 28 May 2014

1 General comments

The paper proposes a detailed evaluation of CMIP5 modeled Arctic sea ice thickness over the past four decades, using a hierarchy of observational sea ice thickness data. To my knowledge, no earlier study has ever engaged in such a comprehensive evaluation of models using sea ice thickness data, and this paper is in this respect very welcome. The authors find that CMIP5 models simulate the average sea ice thickness reasonably well, but that only few models simulate spatial patterns of Arctic sea ice

thickness correctly. Finally, the authors discuss trends in modeled sea ice volume as compared to the PIOMAS sea ice reanalysis. The multi-model mean trend is found to be underestimating the PIOMAS trend, but to lie within the range of uncertainty of the PIOMAS statistic.

The paper has a several positive points. First, it is novel in the use of so many observational thickness data. Second, it proposes to use sea ice thickness to evaluate coupled models, which I agree with the authors is a more physical metric than sea ice extent, and certainly important for constraining projections. Third, The authors do not hesitate to test some of their hypotheses extensively, as e.g. the use of multiple atmospheric reanalyses data set to examine the skill of sea level pressure in CMIP5. Finally, the paper is well structured and has a clear scope.

Using sea ice thickness to evaluate models has certainly a better physical basis than evaluations based on sea ice extent alone; however, the price to pay is that thickness products are subject to larger uncertainties: sampling in time and space is not uniform, instrumental and methodological errors are large (Zygmuntowska et al., 2014, doi:10.5194/tc-8-705-2014; the authors should cite this very informative study); PIOMAS is certainly useful but is a model, for which long-term trends can be sensitive to the atmospheric forcing used. On top of that, natural variability is pronounced and makes the evaluation a delicate task, especially for short periods of time. The authors are aware of these individual sources of uncertainty as discussed nicely in the text. However, understanding the interplay between all these sources of uncertainty, and how large is the resulting total uncertainty, is key to making a clean model evaluation. In the diagnostics, the uncertainties in observational data are probably underestimated because not treated as a whole: for instance, in Fig. 3, the authors co-locate the model and observations thickness in space (very good choice) but do not co-locate model and observations in time (models statistics span 1981-2010, one of the products spans 2011-2013, the other 2004-2005, ...). In addition, it is not clear if instrumental uncertainty and methodological uncertainties (related e.g. to assumptions on snow load, snow and ice densities when ice thickness is retrieved) are taken into account. This introduces additional uncertainty in the comparison, which is not displayed in the error bars. If the authors think it is not the case, they should then argue why. I have listed below (in the Specific Comments) several places where I think uncertainties could be larger than displayed. Thus, in my opinion, statements such as "The climate models as whole also tend to underestimate the rate of ice volume loss from 1979 to 2013" (Abstract) must be tempered by the recognition that uncertainties are much larger than for the less-physical, but more reliable ice extent metric.

I agree 100% with the authors that model evaluation based on sea ice thickness and its distribution in space and time has a clear physical meaning and would be a good choice to constrain projections. Yet, the conclusion that the CMIP5 models have low ability to replicate sea ice thickness is, to me, too strong given the large cumulative uncertainties in observational data or reanalyses of sea ice thickness and volume.

I list below several comments related to my main point. I also list several other points that deserve more detailed information in the text (I pointed several inconsistencies that need to be looked at in more detail). In particular, I would not be able to replicate several figures myself just based on the information given in the text, so that some clarifications are needed. I hope that my review of this paper will help the authors. If my comments/questions are addressed and my remarks taken into account, I strongly recommend the paper for publication.

2 Specific comments

1. p. 2180, line 19: Please cite the source for the trends reported, and include uncertainties.
2. p. 2181, line 1 : The September 2013 sea ice extent anomaly is thought to be "partly a result of anomaly cool summer conditions". Are there studies that have

been investigating the causes for this unusually high minimum compared to the trend line? If so, could you refer to those studies?

3. p. 2181, line 15: In order to stick to the CMIP3 assessment made in line 13 (67% of the models...), I would not use "most" here, but rather a quantitative estimate as well.
4. p. 2183, line 20: The CMIP5 database is complete since more than one year now; why are only 27 climate models analyzed (out of 39 available)? Did the authors apply a first filtering on the models before the analysis was conducted? Could the conclusions be sensitive to the inclusion of the models not taken into account?
5. p. 2184, lines 21-26 and Fig. 1: The results are extremely interesting, and probably worth investigating (perhaps not in this paper!). It appears from first-order inspection of Fig. 1 that the three models with the most intrinsic variability in sea ice thickness comprise an ice-thickness distribution (ITD) framework, and the three others don't. That is, it looks like models that resolve the statistical sub-grid scale distribution of sea ice thickness (EC-Earth, CCSM4, HadCM3) produce grid-cell thicknesses that are more likely to be influenced by natural variability than models without ITD. Could there be a physical reason for that? Anticipating that most models of the next generation will include sea ice models with an ITD, the evaluation of mean thickness will perhaps be even more difficult in CMIP6 than it is today with CMIP5.
6. p. 2184, line 27-29: The spatial correlations of thickness between individual ensembles are found to be very high ($>$0.9). The authors infer that evaluation based on thickness patterns is not too much affected by natural variability. This statement relies on the hypothesis that the models simulate the correct natural variability; was this hypothesis tested, and how? In line with my previous comment, models comprising more realistic sea ice physics simulate more spatial

variability. Does that mean that the other models may underestimate the natural variability in sea ice thickness? Given the short period of time of the ICESat campaigns (a few years) that are used for the evaluation of spatial patterns (Fig. 5), is this evaluation really robust and free of impacts from natural variability?

7. p. 2188, lines 7-9: The satellite thickness fields were regridded using a drop-in-the-bucket approach. Please specify how you treated instrumental/methodological uncertainties (related, e.g., to assumptions on snow and ice densities when thickness is retrieved), how you propagated uncertainties from the 25km level to the 100 km during this interpolation, and whether you accounted for these uncertainties in the evaluation. These uncertainties are maybe much lower than the interannual variability, in which case they can be ignored as a first approximation, but then please show that this is the case.

8. p. 2190, lines 9-11. I would temper this statement. I can accept that PIOMAS estimates for the mean sea ice thickness compare well with observational estimates (as seen in Fig. 2, and discussed in Laxon et al., 2013 or Schweiger et al., 2011). That the trends in PIOMAS volume may be used with confidence to evaluate CMIP5 trends should be tempered by the recognition (i) that the PIOMAS trends are sensitive to the atmospheric forcing used (Lindsay et al., doi:10.1175/JCLI-D-13-00014.1), but also that the evaluation is strongly impacted by natural variability. If these two sources of uncertainty are independent, the error bars displayed in Fig. 8 are probably larger than depicted.

9. p. 2190, line 19: "uncertainty of decadal PIOMAS trends of 1×10^3km^3 ": the units are confusing for characterizing trends. Write "uncertainty in PIOMAS trends of 1×10^3 km^3/dec"?

10. p. 2190, lines 19-21: "Given the large observed volume trend ..., PIOMAS is a suitable tool for assessing long-term trends in CMIP5 models". I don't understand the logical articulation of this sentence. The suitability of a reanalysis to assess

models is not related to the magnitude of the trend, rather to the confidence we have in this trend.

11. p. 2190, line 20: Remove "observed". PIOMAS is a model.

12. p. 2191, lines 16-17: What is meant by "spread"? The 10-90% interval, the range, ...?

13. p. 2192, line 17: "PIOMAS facilitates more robust comparisons". Again, I would temper this sentence (see my comment [p.2190, lines 9-11]): using PIOMAS brings the advantage of long and homogeneous records, at the expense of using a model instead of observations.

14. p. 2193, lines 6-13: This diagnostic is extremely interesting. If I follow the authors and inspect Fig. 5, models resemble more each other than they resemble observations. Is this an indication that models share the same biases (rheology, thermodynamics, winds)?

15. p. 2194, line 24: The authors mention the range of 14470 km3 to 87000 km3 for simulated ice volume in March and refer to Fig. 7 - dashed lines. The dashed lines in Fig. 7 are at the $\sim$19000 km3 and $\sim$43000 km3 levels and are supposed to represent the minimum and maximum volumes in the model ensemble. Did I miss something?

16. p. 2194, line 24: The value of 87000 km3 for GISS-E2-R is clearly unrealistic. It turns out that the GISS-E2-R model output has sea ice thickness of $\sim$1 m and sometimes more over a large fraction of Northern Hemisphere continents. Did the authors correctly mask the continents when calculating sea ice volume? What is the impact on the multi-model mean volume/trends?

17. p. 2195, line 23-25: "The majority of ensemble member trends ... can therefore be considered compatible with PIOMAS". If the null hypothesis is "H0: CMIP5

trends are consistent with PIOMAS" (as stated p. 2195, line 12), then the fact that the majority of CMIP5 2 sigma ranges overlap with the PIOMAS does not allow to reject H0. Thus, I would turn the sentence in "The majority of trends cannot be considered incompatible with PIOMAS".

18. p. 2196, line 1: The individual ensemble members are averaged together to produce the multi-model ensemble mean trend in March ice volume. If I understand well, more weight is thus given to model with more ensembles. Is there a particular reason for that? Why was the evaluation of mean thickness carried out by giving equal weight to each model by pre-averaging members (p. 2184, line 4) ?

19. p. 2196, lines 14-15: Units are 10^3km^3/dec, not 10^3 km^3.

20. p. 2198, lines 1-2. I cannot follow the sequence of arguments here. It is said that only two models have the correct spatial thickness patterns but have very different trends in sea ice volume, so that constraining models based on sea ice thickness patterns is not promising. I think it is, as the distribution of ice thickness has been shown to be a source of spread in projections (Holland et al., 2010, doi:10.1007/s00382-008-0493-4). It is, probably, not sufficient to filter projections based on thickness patterns only. Is that what the authors meant?

21. p. 2202, Table 2: In the caption: "Mar"$-->$"March".

22. p. 2202, Table 2: In the caption: "Trends are listed as km^3" should be replaced by "Trends are listed as 10^3 km^3/decade" or "Trends are listed as 10^2 km^3 per year" (according to the table header).

23. p. 2202, Table 2: "NorEMS1-M" $-->$ "NorESM1-M"

24. p. 2202, Table 2: I suggest to include a brief description of the sea ice model used in each CMIP5 model. Since the paper evaluates sea ice thickness, it seems important to me to specify what thermodynamic scheme is used, whether the

model includes the sub-grid scale ice thickness distribution or not, and the type of rheology that is used. To increase the impact of this paper and help subsequent groups identifying how biases in sea ice thickness relate to the physical sea ice model used, this step seems instructive to me.

3 Comments on the figures

1. Fig. 1: Over which time period are the "stddev" and "average" statistics computed? For what month are the diagnostics shown (March, September, annual average)? For a given model and given grid cell, how is computed "stddev": by first averaging thickness in time for each member, then taking the standard deviation over members, or by first taking the standard deviation of thickness over members for each year and then averaging over years? The order has an importance.

2. Fig. 3: In the "IceSat" panel (third from the top), at least 10% of the data was sampled in open-water since the 10% percentile line (green) is superimposed on the zero-line. Returning to the paper of Kwok et al. (2009, doi:10.1029/2009JC005312). I can read that IceSAT samples with ice draft less than 10 cm are considered to be open water. Is that the explanation, or the $>$10% of data with ice thickness equal to 0 m are really open water? In the former case, did you also mask the model output below 10 cm to ensure consistency in the comparison?

3. Fig. 5: It would be good, at least for the correlations, to specify which ones are significantly greater than 0. Given that a large number of grid points is used to compute the correlations (the grid resolution is 100 km by 100 km, the area covered is approximately 10x10^6 km^2, so I would expect about 1000 grid points),

the correlations are probably significant even for low values. Providing the significance would also allow to point out which models have a totally unrealistic sea ice thickness.

4. Fig. 5: There are only 25 models evaluated in this figure but in the text 27 models are presented. That is, correlations and RMSE scores are not shown for CanCM4 and GFDL-ESM2M in Fig. 5. Why leaving these models aside?
5. Fig. 7: The title ("March") is cropped.
6. Fig. 8: In the legend, please change "Observed" by "Reanalyzed" or "PIOMAS". PIOMAS is a model.
7. Fig. 8: How is the confidence interval for the multi-model mean constructed? Is its width equal to the average width of all confidence intervals, or is its width calculated directly from the time series of multi-model mean sea ice volume? Referring to my comment [p.2196, line 1], is this confidence interval biased towards models with more members?

4 Technical corrections, wording, typos, etc.

1. p. 2181, line 5: I think "but" is not necessary
2. p. 2182, lines 25-26: What do you mean by "mean distribution of sea ice thickness"? As I understand from criterion (1), it is rather the "(statistical) distribution of mean thickness". In the abstract, the wording "mean thickness distribution" is used; is the meaning equivalent?
3. p. 2187, line 2: "similar same" $-->$ "similar", or "same"
4. p. 2191, line 17: "fall" $-->$ "falls" ("the spread... falls")

5. p. 2191, line 18: "Fig. 2" $-->$ "Fig. 3".
6. p. 2193, lines 9-10: "Fig. 5, top" and "Fig. 5, bottom" should be replaced by "Fig. 5, left" and Fig. 5, right", respectively.
7. p. 2193, lines 23: "annual mean annual" $-->$ "annual mean"
8. p. 2194, line 1: "FGOALS" $-->$ "FGOALS-g2"
9. p. 2194, line 13: "the decline" is not necessary in the sentence "sea ice volume is declining faster than the decline in ice extent"
10. p. 2197, line 16: "maybe become" $-->$ "may become" ?

Interactive comment on The Cryosphere Discuss., 8, 2179, 2014.