We are grateful to the Editor and the reviewers for their helpful comments, which have led to an improved version of our original manuscript. In addition to these edits, we decided to add the table on the ice model physics given in our previous response to reviewer's comments in the main body of the manuscript. Since this information is difficult to find, others will appreciate having a table such as this more easily accessible. Below are our responses to the remaining comments.

Response to Editor's comments:
l. 135ff: How is the coefficient of variability shown in Figure 1 calculated? On the previous page you refer to monthly means calculated over the whole year but looking at coefficients of variation that are less than 5% for some of the models this looks like it is just for the month of March? Please clarify.
The coefficients of variability were computed for each month. Figure 1 is just for March.

l. 453: "Day et al. (2012)"
Done

l. 486: Missing period at end of sentence.
Done

l. 617ff: Please add brief reference to the time period (1981-2010) and mention of the monthly mean that this data is based on to the figure caption. Also what does "is attached" refer to? Please clarify.
Figure caption has been modified accordingly.

l. 622: "each campaign's period of record"
Done

Figure 2: Please change the axis labels of the scatter plots to indicate that these are ice thickness data, rather than simply stating, e.g., "PIOMAS (m)"; you could simply do this by writing zi[PIOMAS] (m) or something like that. Also, the axis labels for the x-axes are closer to the plot below than the one they belong to, please correct.
Done

Response to Reviewer's comments:
I thank the authors for considering my points. Most of them have been taken into account. The sample of 33 models is OK to me, and the table with model information is welcome.

I can see from the authors response that they are aware that the treatment of uncertainties during model evaluation is the main unknown. I wish this could be more reflected in the text. For example, the abstract is very definitive, in particular the last sentence. Adding "although large uncertainties in observational products complicate model evaluation" could temper the conclusions. I would also repeat that these uncertainties are not well quantified and make the evaluation challenging in the conclusions. I'm asking this because the abstract and the conclusions will be the two parts of the paper that will be read the most, and I cannot accept that a reader is not warned about how large these uncertainties are. I noted that Reviewer#2 (who submitted his review after mine) makes the same point, so I think the abstract and conclusions must include a recognition that it is not as easy to

We have added the modification to the abstract as suggested by the reviewer, and reiterated it in the conclusions.

Two more comments

In the new manuscript:
p.2, l. 51-52: "Simulating the sea ice distribution has emerged as a key issue". I think you mean simulating the *spatial* distribution of sea ice thickness, not the sub-grid scale distribution of ice thickness, right? If so, please use "spatial" (the evaluation of the CMIP5 subgrid scale distribution of sea ice thickness has to my knowledge not be conducted yet).

-

Changed as suggested

Regarding author response
- p. 6. The authors write that "Simply put, PIOMAS does not necessarily reflect the true observed thickness distributions. Compared to observations, biases in PIOMAS are similar to those for the models in the CMIP5 archive". This is an interesting information that should appear in the new manuscript version. That the product used to make the evaluation is closer to the models than the truth is an issue (although I recognize PIOMAS is the best that we have so far) and should be acknowledged somewhere.

We do mention in the main body of the manuscript that the spatial correlations between CMIP5 and PIOMAS are higher than between CMIP5 and ICESat, and that this is because both PIOMAS and many of the CMIP5 models show a tongue of fairly thick ice extending across the Arctic Ocean towards the Chukchi and East Siberian seas (lines 375-379). Thus, we feel that we have already addressed this concern in our revised version.

p. 12, remark 2.: The authors write that there is no mention in Kwok's paper of setting IceSAT thicknesses less than 0.1 m to open water. There is (Kwok et al., doi:10.1029/2009JC005312, p. 5: "All samples with ice draft less than 10 cm are considered to be open water"). I'm just wondering whether this is a big issue in model evaluation. Apparently it is not, but please mention at least that IceSAT masks data with less than 10 cm.

In Ron's paper we found it unclear if this was done only for the lead detection or if it was also used in deriving the ice thickness fields. We asked Ron to confirm, and found that the 10 cm was used to derive lead width statistics to compensate for the footprint size of ICESat. It is not used in thickness calculations.