A review of "GrSMBMIP: Intercomparison of the modelled 1980-2012 surface mass balance over the Greenland Ice sheet" By X. Fettweis, et al.

This manuscript presents a surface mass balance (SMB) model inter-comparison for Greenland over the period 1980-2012. Varying types of SMB models are compared, and assessed against a number of observational datasets. The manuscript also reports a mean trend, mean, and standard deviation of SMB over the assessed period. Additionally, regional comparisons are presented, and suggest that the largest discrepancies between models are found in proximity to the ice sheet margin. Regional climate models (RCMs) compare the best against observations, especially with respect to their skill in capturing SMB components independently, but are computationally intensive. Results suggest that the less-expensive, simpler models may be appropriate for simulating SMB for longer simulations, as they have comparable biases to the RCMs. The manuscript also notes that the ensemble mean produces the best estimate of SMB compared to observations, and no systematic biases are exposed by this exercise. This inter-comparison is an important step in evaluation of the scientific community's understanding of the physical processes driving mass trends in Greenland and in the ability to determine what types of models are appropriate for answering key scientific questions, particularly with respect to future projections of spatial ice sheet change and sea level contribution.

The current revision of the manuscript is much improved over the first version. The introduction is now extremely helpful to the reader in providing details about this important inter-comparison effort and the consequences of its outcomes. The authors have appropriately responded to concerns from both reviewers, and the revisions to the tables and figures make the results and discussion more readable. For these reasons, I recommend publication in The Cryosphere with minor revisions, specifically an expansion of the discussion section.

General comments:

In general, I find that the Discussion section lacks *discussion* of scientific implications of the many interesting results presented here. Most importantly, a discussion about what the results imply about the ability of different types of models to project into the future, with consideration to the assumptions/tuning that needs to be made in order to match present-day conditions, would be appropriate. For instance, the last sentence of the conclusions implies that PDD/EBM may be better for future simulations in terms of general skill and cheaper computations, but this point is more complex than stated here. For instance, do these models have just as much skill during strong melt years (i.e. 2012), as other neutral years? Though they seem to have skill historically, are there any indications in this analysis that suggest they will do as well in an extreme future scenario? Since 2012 is the last year assessed here, a preliminary hypothesis about future skill could be made within this manuscript. If not, it would be important to state that no such conclusion can be made. Another example is the comparison against Bougamont et al., 2007, concerning PDD sensitivity (see note in comments below). Pointing out to the reader through discussion that historical results may not translate directly to skill in to future projections, as runoff becomes more important in the future, would result in the richer discussion that I would like to see presented. Currently, the discussion section, is only two paragraphs, so I suggest dedication of an additional paragraph to reflect on such questions and support the final conclusion statement of the manuscript.

Specific comments and suggestions are noted below:

Line 73, To me, the end of this statement, "but have never been evaluated until now" implies that GS is being evaluated. Earlier, the text clearly states that GS is not simulated by the models. Because GS is clearly not simulated, I do not know what "until now" refers to. Either this is a miscommunication and should be rephrased, or please be more explicit about what it meant by this sentence.

Line 90, Missing "s": model"s"

Line 92, awkward phrasing: perhaps, "the" then

Line 146, It is unclear how this exercise specifically reduces uncertainty. Please clarify why this is so or rephrase. Maybe "quantifies" could be used instead of "reduces"?

Line 196 and Line 244, Nowicki

Line 266, missing word: "the" sum

Line 290, "no" significant differences, instead of "not"

Line 382, Please include a reference to the accumulation being constant over the last decades

Lines 408-409, Does this sentence refer to the fact that in the King paper, RACMO is used to evaluate discharge estimates? The way it is written, it sounds like this current manuscript uses RACMO to compare total mass balance with GRACE. Perhaps just a simple rephrasing to clarify that you refer to the methods of the King et al., 2018 here would help alleviate this confusion.

Line 432-434, Could this be because Bougamont et al., 2007 is comparing the results of these model in an extreme warming scenario, while here, you are considering only the historical period? This point seems like something worth elaborating on with regards to SMB models and their ability to make future projections. It would strengthen the discussion to add some sentences dedicated to this subject. (See point on Discussion above).

Line 466, awkward phrasing: maybe, disallowing "the representation of" the spatial variability

Line 511, awkward: perhaps, matches well "with" (or "against") the

Line 550, It would be appropriate here, to note that these results are consistent with past SMB assessments using GRACE, e.g. "is consistent with past assessments of SMB seasonal variability using GRACE", with reference to e.g. Velicogna et al., 2014; Alexander et al., 2016; Schlegel et al., 2016

Velicogna, I., Sutterley, T. C., and van den Broeke, M. R.: Regional acceleration in ice mass loss from Greenland and Antarctica using GRACE time-variable gravity data, Geophys. Res. Lett., 41, 8130–8137, doi:10.1002/2014GL061052, 2014.

Alexander, P. M., Tedesco, M., Schlegel, N.-J., Luthcke, S. B., Fettweis, X., and Larour, E.: Greenland Ice Sheet seasonal and spatial mass variability from model simulations and GRACE (2003–2012), The Cryosphere, 10, 1259–1277, https://doi.org/10.5194/tc-10-1259-2016, 2016.

Schlegel, N.-J., Wiese, D. N., Larour, E. Y., Watkins, M. M., Box, J. E., Fettweis, X., and van den Broeke, M. R.: Application of GRACE to the assessment of model-based estimates of monthly Greenland Ice Sheet mass balance (2003–2012), The Cryosphere, 10, 1965–1989, https://doi.org/10.5194/tc-10-1965-2016, 2016.

Line 578, missing space: "quarters of"

Line 582-583, This is a very important statement, and the amount of tuning that each type of model does may be an important aspect of its ability to make projections in the future. I suggest touching on this idea earlier in the manuscript, with a couple of sentences of additional discussion (i.e. general Discussion comments above) to help the reader reflect on how assumptions and tuning to present day may affect future results.

Figure 4, caption – Fig. 4 refers to being the same as Fig. 23. Likely this should be Fig. 2 referred to here?