Review of: "Reconstruction of the Greenland Ice Sheet surface mass balance and spatiotemporal distribution of freshwater runoff from Greenland to surrounding seas", by *S. H. Mernild et al.*, submitted to *The Cryosphere*.

General comments

Using the updated version of SnowModel/HydroFlow, the authors simulate the surface mass balance (SMB) and components, i.e. runoff, melt and retention, of the Greenland ice sheet (GrIS) at 5 km resolution for the period 1979-2014. Precipitation is downscaled from ERA-Interim re-analysis for the same period. The model includes a snow module accounting for meltwater retention in snow, an energy balance scheme and a meltwater runoff routing module. This routing module allows to quantify the runoff contribution of 6 GrIS sectors, further refined to over 3,000 individual catchments. The authors first discuss the modeled contemporary (1979-2014) GrIS SMB and recent trends (2005-2014) in these 6 sectors. The analysis is further extended to the numerous individual catchments to show that about 80% of these have experienced increasing meltwater runoff since 1979. Then, the authors correlate this recent runoff increase with the natural variability of the Atlantic Multidecadal Oscillation (AMO) and North Atlantic Oscillation (NAO).

The manuscript is overall well written, but the results presented seem inaccurate and raise questions and concerns on the model ability to reproduce the contemporary GrIS SMB, potentially altering the conclusions drawn in this paper. These concerns are summarized in the Substantive Comments. In brief, I suspect that the current version of SnowModel/HydroFlow has several issues resulting in inaccurate SMB estimates over the GrIS. Compared to other studies, runoff is significantly overestimated likely due to inaccurate representation of meltwater retention in firn. This study suggests that only 25% of melt is retained in the firn pack while most recent efforts demonstrated that it is closer to ~45%. Such a retention underestimation has severe implications on runoff calculation and SMB. For instance, the authors present negative modeled GrIS-integrated SMB for the period 2005-2014 as opposed to recently published GrIS mass balance and GRACE studies. This makes these results potentially unreliable and the conclusions drawn inconsistent.

For these reasons, I judge that the manuscript **cannot be published** in the current form and needs **major revisions**, unless the authors prove that their estimates of precipitation, runoff, melt, refreezing and SMB are reasonably accurate. To achieve this, the authors **must** perform a thorough model evaluation against in situ SMB measurements and compare their results to remote sensing mass change records, and mass balance estimates compiled in previously published studies. This would highlight potential issues in their snow model and provide some insight on how to solve them. If these evaluations/comparisons and validation of the model results can be successfully achieved, and if the suggested corrections listed hereunder are applied, I would be happy to reassess this manuscript.

Substantive Comments

1 Throughout the manuscript, the authors discuss changes in SMB components and recent trends without providing a thorough evaluation of their modeled SMB estimates. Nowadays, comprehensive in situ ablation (Machguth et al. [2016]) and accumulation (Bales et al. [2001, 2009]) data sets are available over Greenland to evaluate modeled SMB in time and space. Such evaluation must be performed and discussed in the manuscript to provide some insight on how well SMB components, i.e. notably precipitation and runoff, are represented in the SnowModel/HydroFlow model. Such evaluation is now systematically performed in Greenland mass balance publications, i.e. Fettweis et al. (2017), Noël et al. (2016), Niwano et al. (2017) or Langen et al. (2017).

2 GrIS-integrated SMB components presented in this study, i.e. notably runoff, meltwater retention and SMB (see Table 1), do not generally agree with recent GrIS SMB studies, e.g. Fettweis et al. [2017], van den Broeke et al. [2016], Noël et al. [2017], Mottram et al. [2017] or Vizcaìno et al. (2013), suggesting potential issues in SnowModel/HydroFlow. For instance, Table 1 shows that only 25% of meltwater is retained in snow while other recent studies suggest 45%, e.g. van Angelen et al. (2013), Noël et al. (2017), Steger et al. (2017a). Steger et al. (2017b) performed a similar basin analysis of SMB components (8 sectors) of the GrIS using another state-of-the-art firn model (SNOWPACK); the authors must compare their results to that study and discuss the differences.

3 Figure 2f strikes me as the ablation zones in north Greenland, i.e. notably in northwest and northeast Greenland are by far larger than in other studies (Fettweis et al., 2017; Noël et al., 2016; Mottram et al., 2017; Vizcaìno et al., 2013; Cullather et al., 2014). In addition, Table 1 suggests that these northern basins contribute equally or more runoff than southern basins, e.g. NW and NE contribute 70 Gt/yr and 63 Gt/yr on average (1979-2014). The study also suggests that these northern basins showed a negative SMB on average (1979-2014), meaning that these have been losing mass for more than 30 years. This is not supported by other studies, e.g. Mouginot et al. (2015) and Steger et al. (2017b).

4 The authors state that they use an updated version of the SnowModel/HydroFlow model but they never discuss the relevant changes implemented, or their impact on the modeled data. The authors should at least list and discuss relevant model updates, compare results from their previous and current model version, and explain where and why changes occur. This would highlight the novelty of the presented data set. To me, it is not clear whether this new data set is a general improvement on previous versions.

5 Tables and figures are sometimes very difficult to read and interpret, especially Figure 5 and Table 1. These may potentially confuse the reader with too much information. Suggestions to improve the text, figures and tables are provided below.

Point Comments

L59-61: In Wilton et al. (2016), Fig. 2 only shows SMB as low as ~100 Gt/yr in years 2010 and 2012, the same applies for runoff. This should be clarified in the manuscript.
L65-66: I would reformulate as follows: "(Chen et al. 2017), and up to 43% for the GrIS and peripheral glaciers and ice caps in 2010-2012 (Noël et al. 2017)".
L70: I think the authors mean "melt season duration" instead of "surface ablation duration", see Tedesco et al. (2016).
L100-102: I think this sentence is somewhat misleading as Enderlin et al. (2014) compiled estimates of solid ice discharge from ~178 marine-terminating glaciers around Greenland. Could the authors reformulate to clarify this?
L121: In the abstract and at L134, the authors state that they use ERA-I reanalysis to force their model, while automatic weather stations are mentioned here. Could the authors clarify and further elaborate on how their model was forced? It would also be useful to learn more about how the snowpack was initialized at the beginning of the simulation.
Section 2.1: Here, the authors state that they use an updated version of the SnowModel/HydroFlow model but do not discuss the changes implemented in this new version, nor the impact on the modeled results. See also our Substantive Comment 4.
L169: Could the authors clarify what they mean by "water-equivalent evolutions"? I assume this is related to meltwater retention/refreezing in the firn pack and runoff production. Please, elaborate.
L171: Could the authors explain what they mean by "hypothetical gridded topography and ocean-mask datasets"?
L184: Could the authors briefly explain how the 6 or 12-hourly forcing fields were downscaled to 3-hourly data in MicroMet? Replace "and" by "on" before "a 5-km". Could the precipitation underestimation suggested at L338-343 be the result of this downscaling?
L188: Could the authors mention the original resolution of the DEM presented in Levinsen et al. (2015)?
L213-215: Why do the authors obtain more catchments? Is this a result of the DEM and/or model updates? Please, clarify.
L216-223: Here, the authors describe MicroMet and then resume their discussion on HydroFlow. I would therefore suggest moving these sentences to L201 on page 9.
L221-222: Could the authors provide a reference that corroborates this assumption on blowing snow?
L244: Could the authors provide a reference for the 10 hydrometric monitoring stations?
L252: For model evaluation, the authors refer to a paper that is not published yet. As the authors analyze average SMB components and recent trends, I feel that a proper evaluation, as suggested in the Substantial Comment 1, of the modeled SMB data set must be added here. This would provide some insight on the model performance across the GrIS ablation and accumulation zone, i.e. how well the model simulates runoff and precipitation, respectively. Without model evaluation, the discussion on SMB components and recent trends in Section 4 is somewhat insubstantial.
L308-317: These lines are somewhat descriptive, additional insight on the model performance could be gained by performing the SMB evaluation against in situ measurements in the accumulation zone using Bales et al. (2001, 2009).
L338-351: Here the authors discuss uncertainties and underestimation of precipitation in the ERA-I forcing field based on, e.g. Fettweis et al. (2017). Again, a proper evaluation of SMB in the accumulation zone would be more convincing to evaluate precipitation and quantify a potential bias that could be further corrected. See also Subtantial Comment 1.
L366-371: I don't think that discussing "ablation" adds relevant knowledge to the paper. The authors should better discuss and evaluate melt, runoff and refreezing in more detail. I would also replace ablation by runoff fields in Figs. 2, 4 and Table 1. The authors could consider including these information about "ablation" in a Supplementary Material.
L375: The authors probably mean "573.7 +- 119.8 Gt yr-1", which is relatively high compared to other estimates. See also the next comment and Substantial Comment 2.
L389-392: The ablation zone in north Greenland, especially in northeast and northwest, are relatively large compared to e.g. RCMs estimate from Fettweis et al. (2017, Fig. 6a), Mottram et al. (2017, Fig. 5) and Noël et al. (2016, Fig. 1), or GCMs estimate from Cullather et al. (2014, Fig. 9) or Vizcaìno et al. (2013, Fig. 7), suggesting ablation overestimate in SnowModel/HydroFlow. Could the authors

elaborate on this? Again, it would be very useful to perform a proper SMB evaluation in the ablation zone of the GrIS. This would allow for estimating SMB (runoff) uncertainty and make the following regional SMB (runoff) analysis more robust. See also Substantive Comment 1.
L404-406: I'm not sure to understand these lines, could the authors reformulate?
L407-419: The authors obtain a refreezing-retention fraction of 25% for the period 1979-2014, which is by far lower than other estimates of ~45%, e.g. Steger et al. (2017a) or Noël et al. (2017). This could likely explain why runoff is so high compared to other studies, e.g. Van den Broeke et al. (2016). The authors should stress this as these inaccuracies may strongly impact the discussion of contemporary runoff production and recent trends discussed in the paper. In addition, the recent study of Steger et al. (2017a) also integrated refreezing (Gt/yr) and fraction (%) from two state-of-the-art firn models (IMAU-FDM and SNOWPACK) over the same 6 GrIS sections discussed here. For all these sections, the refreezing fraction is lower by almost a factor of 2 compared to Steger et al. (2017a). See also Substantive Comment 2.
L415: The authors should refer to Steger et al. (2017a) rather than Ettema et al. (2009).
L438-439: Here the authors state that their runoff and SMB product is improved, but no comparison with a previous version or with observations has been conducted. Could the authors elaborate on how they draw these conclusions.
L442-443: Given the potential underestimated precipitation, and overestimated runoff, the obtained SMB product is unrealistic. Van den Broeke et al. (2016, Fig. 9) show a reconstruction of GrIS mass balance, solid ice discharge and SMB, clearly refuting an average SMB of ~120 Gt/yr for 1979-2014. This is also supported by other studies: e.g. Fettweis et al. (2017, Fig. 8) and Mottram et al. (2017, Fig. 3). In addition, they obtain a negative SMB after 2005, which is again not supported by other studies.
L459: It is misleading to say that Wilton et al. (2016) obtained a GrIS SMB of ~100 Gt/yr in the late-2000s, as this is only true for years 2010 and 2012.
L480 and 482: The authors probably refer to Figures 5a and 5b.
L498-500: I agree that many catchments show this out-of-phase pattern but there are still quite some that don't. Could the authors quantify this, e.g. as a percentage of the number of catchments in southeast Greenland.
L509-514: While I agree that some correlation exists between EOF1 and AMO, it is not so clear for NAO (see Fig. 9b, r =0.4 or r^2 =0.16). I would suggest some more caution when drawing firm conclusions on these teleconnections, as at e.g. L44-45, L519 or L544-548.
L538-540: This sentence is misleading, to my knowledge no other studies show average SMB of ~120 Gt/yr for 1979-2014, nor a negative SMB in the period 2005-2014. I think the authors should reformulate to stress this.
L545: I would replace "indicates" by "suggests" as the correlation obtained for AMO and notably NAO are relatively low. I suggest: "This suggests that runoff variations are related to large-scale natural variability of AMO and NAO in Greenland."
L549: My main concern on using the data set presented in this study is that the modeled runoff and SMB are by far overestimated and underestimated, respectively, when compared to other studies. It is therefore questionable whether this data set accurately reproduces the contemporary SMB of the GrIS, if it can be used to force ocean models or to quantify mass changes over Greenland.

Stylistic comments

L33: I would suggest 'resolution' instead of 'increments'. This holds for the whole manuscript.
L34: I suggest: 'Compared to previous studies, simulated SMB is low whereas the GrIS surface conditions remain similar.' In addition, the authors should use the present tense here and at L34-40. Using the past tense is confusing as it suggests that the authors discuss previously published model results.
L49: Present tense should also be used in the introduction and following sections when referring to the data discussed in this study.
L71-72: I would suggest: "[…] because meltwater may be retained or refrozen in the porous […]".
L89: "particularly common".
L92: "[…] understanding is used to explain […]".
L104: Remove the "of" before "catchments".
L109: Remove "the" before "link".

L110: I would suggest: "This has further implications […]" as the "unaddressed knowledge gap" is already mentioned in the previous sentence.
L127: Maybe replace "land area" by "tundra region".
L140: Remove "conditions".
L141: Remove "(the last decade […])".
L150: I would suggest: "spatiotemporal patterns of runoff".
L155: I would replace "verification" by "evaluation".
L161: I would suggest: "interpolation scheme. Interpolation fields were adjusted […]"
L172: Remove "from" before "catchment outlets".
L173: Replace "tested" by "evaluated".
L189: I would suggest "resolution" instead of "increment".
L201: Replace "of" by "with" before "glacier ice".
L230: Remove ", which include a part of the GrIS".
L231: I would suggest: "[…], feedbacks from a thinning ice […] will not influence the catchment […]".
L235-236: Remove ", not by the glacial drainage system." as this is already mentioned in the previous sentence. L236: Maybe "obtained" instead of "gained".
L240: "Evaluation" instead of "Verification". I would also suggest this throughout the whole section (L249, 251).
L253: "SMB" instead of "surface mass balance".
L285: I would suggest: "The latter analysis enables to link changes in, for example, NAO or AMO with GrIS outlet catchments mass loss and runoff."
L299 and 301: Remove "balance" before "loss".
L319: Refer to Fig. 1b after "six sections".
L323-324: I would suggest: "[…] towards the steep slopes of the southern coast of Greenland, generating orographic enhancement […]".
L341: "between 642.0-747.0 Gt yr-1".
L349: Maybe: "This highlights the importance of accurately representing precipitation for estimating the energy […]".
L352-354: I would suggest: "Besides precipitation, melt (including extent, intensity and duration) and ablation are other […] and understanding GrIS SMB. Surface melt can influence albedo, as wet snow absorbs […]".
L357: "[…] affect total runoff, but also ice dynamics […]".
L366: The authors certainly mean "northern and southwestern sections".
L374: "in southwest Greenland" and for clarity add "over the GrIS" after "period".
L394: Maybe "Therefore, in that region the snowpack persists longer compared to […]".
L402: Maybe "within the range of our previous study (Mernild et al., 2008)".
L448: This sentence could be removed.
L455: "24.7 Gt decade-1".
L484: Maybe "variance" instead of "variation".
L489-490: Replace "goes down/up" by "decreases and increases".
L515: This sentence could be removed.

Figures and Tables

Figure 2c and e: The authors should display regions showing surface melt = 0 in white. This would highlight the dry snow zone of Greenland. The authors should also show runoff instead of ablation using a color scale similar to the one used for melt, i.e. runoff = 0 in white.
Figure 3: The scale of SMB components is too small, and numbers are difficult to read. I would also suggest showing values ≤ 0 in white.
Figure 4: The authors should better show time series of runoff instead of ablation.
Figure 5: This figure is rather overwhelming and confusing. It is very difficult to interpret the data or identify any spatial pattern. In addition, the representation of individual catchment in color is somewhat redundant as it is already shown on Fig. 1c. Therefore, I would suggest to display runoff, variance and trends for each catchment using a color scale instead of circles. For trends, a blue-to-red scale, centered on 0, could be used to distinguish negative from positive values.

Figure 6: EOF2 and EOF3 are not significant and the associated figures are not discussed in the paper. Therefore, these could be moved to a Supplementary Material. A new Figure 6 could consist of 3 subpanels combining Figs. 6a, 7 and 8a, all referring to EOF1.
Figure 8b and c: These figures could be shown in a Supplementary Material as they are not discussed in the main manuscript.
Figure 9: The x-axis of the lower Fig. 9a and b should read "years".
Table 1: This Table is rather overwhelming and shows too much information. I think that ablation and "E + Su" could be removed as they are not discussed in detail. In addition, the description of refreezing and retention could be included in the figure caption instead of within the Table itself.

Additional references and DOI

- Steger et al. (2017a): https://doi.org/10.3389/feart.2017.00003
- Steger et al. (2017b): https://doi.org/10.5194/tc-11-2507-2017
- Langen et al. (2017): https://doi.org/10.3389/feart.2016.00110
- Mottram et al. (2017): 10.14943/lowtemsci.75.105
- Noël et al. (2016): https://doi.org/10.5194/tc-10-2361-2016
- Niwano et al. (2017): https://doi.org/10.5194/tc-2017-115
- Van den Broeke et al. (2016): https://doi.org/10.5194/tc-10-1933-2016
- Vizcaìno et al. (2013): 10.1175/JCLI-D-12-00615.1
- Cullather et al. (2014): 10.1175/JCLI-D-13-00635.1
- Machguth et al. (2016): 10.1017/jog.2016.75
- Bales et al. (2001): 10.1029/2001JD900153
- Bales et al. (2009): 10.1029/2008JD011208
- Mouginot et al. (2015): 10.1126/science.aac7111