Reply to Xavier Fettweis (Reviewer#3)

We sincerely appreciate Xavier Fettweis for taking the time to review our paper and providing the MAR model output data as reference information. Below we describe our responses (in blue text) point-by-point to each comment (in black text). In addition, we indicate revisions in the updated manuscript together with new line numbers. Please also refer to the revised marked-up manuscript uploaded in the discussion board.

This paper presents a new RCM based simulation over the Greenland ice sheet. While the scientific interest of this paper is generally poor, this "model validation" paper deserves to be published in TC and opens the door to future applications over the GrIS using a new RCM in addition to the wide commonly used RCMs family (MAR, RACMO, HIRHAM). In addition to the justified remarks from both other reviewers, I have additional remarks that should be resolved before publication if it is not a too big job for the authors.

Thanks to insightful comments and suggestions provided by all the reviewers, we believe the manuscript has been improved and scientific quality of the revised paper has been increased.

pg2, line 67: site rather Fettweis et al. (2017) here

Revised as suggested. (P. 2, L. 67)

pg5, section 2.2.1: What is the sensibility of the model results to the fresh snowfall density? With MAR, the sensibility is very small and MAR uses a minimum snowfall density of 200kg/m3. 300kg/m3 is a bit high for me.

Thank you for the comment. In fact, NHM-SMAP's sensitivity to the fresh snowfall density has not been investigated yet. The reason why we used the parameterization by Lenaerts et al. (2012a) is simple: this is based on in-situ measurements in polar region. If smaller fresh snowfall density is set in NHM-SMAP, underestimation of snow surface height discussed in Sect. 4.5 can be solved; however, I think we don't have enough measurement-based information for fresh snowfall density to change the model scheme now.

MAJOR: pg 7, line 231: As the JRA-55 surface conditions are bad (Section 4.1, line 325), is an atmospheric spin-up of 6h enough to be independent of the initial near surface atmospheric conditions? How are the results sensitive to this spin-up time? For me, performing 48h long simulations by keeping only the last 24h will be more robust.

Please note that insufficient conditions of JRA-55 surface analysis was unraveled through the present study. In addition, it should be noted that an appropriate spin-up period has not been established yet. An appropriate spin-up period can be found by performing a large number of simulations. The reason why we employed 6h spin-up time in the present study is that it is a typical model configuration in Japan. However, we agree with the point that further consideration of an atmospheric spin-up time can be effective to improve the model performance. The 6h spin-up period might not be suitable in the GrIS, whereas the setting seems to be effective empirically in Japan. In Sect. 4.1, we have added the following discussion:

"This result in turn suggests that making every day atmospheric spin-up period (6h; Sect. 2.3.2) longer than 6h can improve the performance of NHM-SMAP. Finding an appropriate spin-up period in the GrIS is a future issue to be coped with." (P. 10, L. 353-355)

pg9, section 4.1: As SMAP seems to underestimates the ablation (see Fig 8), the statistics over summer (JJA) should be provided at least in supplementary material? Is the model too warm or too cold in summer?

Thank you for the constructive comment. In Sect. 4.7 entitled as "Surface mass balance", we have added the following discussion:

"As presented in Sect. 4.1, the on-line version of NHM-SMAP successfully reproduced 2m air temperature at SIGMA-A during summer. Because surface mass loss during the summer is affected by near-surface (2m) temperature, model performance in terms of simulating JJA 2m air temperature at each AWS on the GrIS were re-examined (Table S8). As indicated in the table, significant or systematic error are not found, and obtained ME and RMSE are well (around –0.2 and 2.1 °C, respectively). Therefore, ---" (P. 16, L. 565-570)

MAJOR: pg 10, line 341: If a RCM is totally free, it should be normally independent of the surface biases in the forcing fields. A too short spin-up time of 6h starting from too warm JRA-55 based surface conditions explains likely these biases because MARv3.5.2 forced by JRA-55 is colder in winter than MARv3.5.2 forced by ERA-Interim. Therefore, extending the spinup time should better resolve this bias than changing of forcing reanalysis. Finally, SMAP seems to underestimate LWD in winter but overestimates temperature? This is very strange?? This issue should be discussed in the paper.

We think that a RCM cannot be totally free, because RCM-simulated atmospheric field is generally

constrained by a parent reanalysis data in lateral and upper boundaries of the RCM model domain. We imagine that simulated atmospheric field can be "almost" independent of the forcing data if we employ the "climate simulation mode", where the atmosphere is initialized only at the beginning of the simulation period, as employed by MAR. It seems to us that the present NHM-SMAP model configuration called "weather forecast mode" that initializes the atmospheric profile every day by referring to the forcing data is affected strongly by the parent data compared to the climate simulation mode. Based on this consideration, we agree with the reviewer's point that extending the spin-up time can resolve the reported bias. We have added the following discussion:
"At the same time, extending the atmospheric spin-up period discussed above can also resolve the issue, because simulation results are expected to less susceptible to a parent reanalysis data." (P. 10, L. 371 – P. 11, L. 373)

In the summary and conclusions section, it is mentioned again as follows:
"At the same time, extending the atmospheric spin-up period (6h) can also resolve the issue, because simulation results are expected to less susceptible to a parent reanalysis data." (P. 17, L. 636-637)

pg 10, section 4.2 : I do not see the interest of showing here the ability of SMAP only to simulate a single wind event. Outputs from JRA-55 should be added in the comparison to show the interest of SMAP in respect to JRA-55. MARv3.5.2 (at a resolution of 20km) forced by JRA55 underestimates also this event by a factor of 10-15m/s. The interest of using a non-hydrostatic model at 5 km should be highlighted here.

Thank you for the comment. We have included 10m wind speed data from JRA-55 in Fig. 4 and added the following discussion:
"In the figure, 10m wind speed from the parent JRA-55 reanalysis with a horizontal resolution of TL319 (~55 km) is depicted together. Clearly, JRA-55 could not reproduce the strong wind event and an advantage of a high-resolution non-hydrostatic atmospheric model is successfully demonstrated." (P. 11, L. 401-403)

In connection with this point, we thought horizontal resolution of JRA-55 should be mentioned in Sect. 2.3.2: "Dynamical downscaling of atmospheric field from reanalysis data with JMA-NHM". Therefore, it has been described in Sect.2.3.2 as follows:
"Horizontal resolution of JRA-55 is TL319 (~55 km)." (P. 7, L. 240)

pg 12, lines 409-423: the fact that SMAP overestimates surface temperature but underestimates both LWD/SWD fluxes suggests that SMAP is likely too dependent of the forcing data. What about the latent and sensible heat fluxes? The authors suggests that near-surface snow density is likely too high. I am very sceptic about this explanation. The sensibility of the results to the near-surface snow density can be tested offline. For me, the problem comes from the JRA-55 fields which are too warm and which are used every day to reinitialise the SMAP atmospheric fields.

Thank you for the insightful comment. First of all, regarding the underestimation of downward longwave radiant flux, we think that observation data also has error that affects model evaluation significantly. At the end of Sect. 4.3, we have added the following discussion:
"On the other hand, observation data for downward longwave radiant flux can also have error especially during the winter period due to riming, which may act to increase measured values. In SIGMA-A, measured 2m air temperature often decreased to about –40 °C during the 2013-2014 winter (Fig. 3a). Although such reductions in 2m air temperature during March and April 2014 were followed by significant reductions in downward longwave radiant flux (Fig. 3e), they did not synchronize in December 2013 and January 2014. These results suggest that observed downward longwave radiant flux especially during December 2013 and January 2014 were affected by riming and forced to increase. A reliable quality control technique for automatic downward longwave radiant flux measurements in the polar region should be developed in the future to perform not only model validation but also climate monitoring accurately." (P. 12, L. 443 – P. 13, L. 452)

In the summary and conclusions section, an additional summary regarding this issue has been added as follows:
"On the other hand, observation data for downward longwave radiant flux can also have error especially during the winter period due to riming, which might affect the evaluation." (P. 18, L. 651 - 653)

During the revision, we performed additional data quality control for downward longwave radiant flux. What we performed is that rejecting such data as downward and upward longwave radiant fluxes agree exactly. This situation is caused when extreme riming occurs and these two properties are diagnosed only from sensor temperature. However, our discussion was not affected by the reassessment of measurement data.

Based on these, we now agree with the reviewer's point that the problem comes from the JRA-55 fields which are too warm and which are used every day to reinitialize the SMAP atmospheric fields. At the same time, overestimation of relatively low surface wind speeds (Sect. 4.2) might affect this problem, because it acts to increase sensible heat flux. As a result, we have revised the sentence as follows:

“One possible cause of the model’s overestimation of surface temperature is overestimation of the near-surface snow density profile, which would increase the conductive heat flux to the surface (see Sect. 4.5).”

----->

“One possible cause of the model’s overestimation of surface temperature is overestimation of the surface wind speeds when they are relatively low (see Sect. 4.2), which acts to heat the surface through increases in sensible heat flux. Of course, overestimation of 2m temperature by the model (see Sect. 4.1) especially during winter (November to March) also may contribute to the error.” (P. 13, L. 465-468)

Related to this revision, the following description in the summary and conclusions section has been revised as follows:

“A possible cause for this overestimate is overestimation of the near-surface density profile, as suggested by validation of snow surface height changes.”

----->

“A possible cause for this overestimate is overestimation of the surface wind speeds when they are relatively low, which acts to heat the surface through increases in sensible heat flux. In addition, overestimation of 2m temperature by the model especially during winter (November to March) also may contribute to the error.” (P. 18, L. 657-660)

pg 12, lines 424-439: it is true that MAR overestimates albedo but as it also overestimates SWD. Due to error compensations (as explained in Fettweis et al., 2017), the MAR surface fields are OK. Here, it is strange that SMAP overestimates temperatures but overestimates albedo and underestimates SWD and LWD.

In the original manuscript, we mentioned two possible reasons for the overestimation of albedo by NHM-SMAP as follows:

“Therefore, future models should consider this process as well as the possibility that NHM-SMAP overestimates snowfall during the summer period.” (P. 13, L. 483-484 in the revised manuscript)

In the revision process, we conducted additional model sensitivity tests where ice albedo is set to 0.2 following the suggestion by the reviewer, which is detailed below. The results from the sensitivity tests indicate that simulated SMB did not change significantly compared to the control RE setting (Fig. 8). Based on the result, we reached a conclusion that overestimation of surface albedo by NHM-SMAP can be attributed mainly to overestimates snowfall. These results are mentioned in Sect. 4.7 entitled as “Surface mass balance”, and they can also be found in this answer file (our answer to “pg 14, line 513:”).

Pg 13, section 4.6 : the comparison with the melt extent is excellent! Adding here a 2D comparison (nbr of melt days in 2012 for example) should be interesting to evaluate if this agreement is also OK locally. The simulated total melt extent could be good for bad reasons and local overestimation/underestimation of melt can be compensated.

Thank you very much for the encouraging comment. In Fig. S1 of the supplementary file, we have added the 2D comparison figure. At the end of Sect. 4.6, we have added the following explanation regarding the figure:
“Figure S1, which shows observed and simulated total numbers of surface melt days in 2012, supports this argument.” (P. 14, L. 520 - 521)

pg 13, line 479: A 2D comparison with other RCM based estimations (RACMO, MAR, ...) is needed here for me. The raw 20km MARv3.5.2 daily outputs forced by JRA55 are available here:
ftp://ftp.climato.be/fettweis/MARv3.5.2/Greenland/JRA-55_20km/
and could be used in this paper just by citing Fettweis et al. (2017).

Thank you for the suggestion and providing the data. We considered whether we should use other RCM based estimations or not, and decided to include simulation results by MAR v3.5.2 forced by the same reanalysis data JRA-55 as used in the present study. At the beginning of Sect. 4.7, we have indicated it as follows:

“In addition, simulated SMB data from MAR v3.5.2 forced by JRA-55 (Fettweis et al., 2017) were employed as reference information.” (P. 15, L. 525-527)

At present, there are many different points in model formulations and configurations of MAR and NHM-SMAP, namely, resolution, ice sheet mask, dynamic core of atmospheric model, albedo model, water percolation scheme for snow/firn, etc. Therefore, detailed model inter-comparison should be beyond the scope of this paper; however, we do hope to perform such a comparison in the near future.

MAJOR: pg 14, line 507: MAR at 20km is generally able to resolve the ablation zone. The 5 km resolution used here is not an issue here to explain the systematic SMB overestimation in the ablation zone by SMAP. RACMO at 11km works also already very well. Significant biases in energy balance fluxes could explain the underestimation of ablation.

Thank you for the comment. We think the reason why MAR at 20km successfully resolves the ablation area is the introduction of a sub-grid mask, which is not considered by the present version of NHM-SMAP. Based on this consideration, we added the following discussion:
"On the other hand, MAR v3.5.2 with a horizontal resolution of 20km is generally able to resolve the ablation zone well (Fettweis et al., 2017). A possible cause for this success can be attributed to the introduction of sub-grid mask, which is not employed by NHM-SMAP. It appears that statistical downscaling or further dynamical downscaling or introduction of sub-grid mask is inevitable to obtain more realistic SMB estimates." (P. 16, L. 582-586)

Also, in the final section, we have mentioned it again as follows:
"Moreover, statistical downscaling or further dynamical downscaling to a higher spatial resolution than used here, e.g. 1 km (Noel et al. 2016, Wilton et al. 2017) or introduction of sub-grid mask (Fettweis et al., 2017) may inevitably be required to improve the SMB estimates." (P. 18, L. 673 - 675)

pg 14, line 513: to test the problem of the overestimation of albedo in SMAP, an offline simulation using a bare ice albedo of 0.2 could be carried out here and results should be shown in Fig 8.

It is a very nice suggestion. We have performed the suggested model sensitivity tests and discussed the results as follows:
"According to the PROMICE data in the ablation area, ice albedo often decreases to around 0.2 during summer. Therefore, additional model sensitivity tests, where ice albedo is set to 0.2, were performed. Obtained results indicate that simulated SMB did not change significantly compared to the control Richards equation setting (Fig. 8), suggesting that overestimation of surface albedo by NHM-SMAP can be attributed mainly to overestimates snowfall as pointed out in Sect. 4.4." (P. 16, L. 571 - 576)

In accordance with this, Fig. 8 has been updated. In the original manuscript, we did not refer Fig. 8 explicitly, therefore, it has been referred at the beginning of Sect. 4.7 as follows:
"During the study period, 55 measurements were available, and comparison results are presented in Fig. 8." (P. 14, L. 524 – P. 15, L. 525)

Accordingly, the following sentence in the original manuscript (P. 14, L. 512-513) has been removed:
"Moreover, it is imperative that we develop a realistic albedo model for high-density firn and ice that

incorporates the effects of cryoconite."

Also, the following sentence in the original manuscript (P. 16, L. 592-594) has been removed as well:
"This finding underscores the need to develop a realistic albedo model for high-density firn and ice that allows us to consider the effects of darkening of the GrIS by cryoconite and so on.", and the following sentence has been added in the revised manuscript instead:
"It was attributed to overestimation of snowfall." (P. 18, L. 662)

pg 14, line 522, explicit comparison with MAR or RACMO is needed here for me. RACMO or MAR time series could be added in Fig 9.

As mentioned above, we have included simulation results by MAR v3.5.2 forced by JRA-55. In the revised manuscript, we have compared the data with the NHM-SMAP-simulated GrIS SMB in Fig. 10a. The related description are as follows:
"According to simulation results by MAR v3.5.2 forced by JRA-55 (Fettweis et al., 2017) that uses the bucket schemes with irreducible water contents of 8 %, the GrIS SMB during the 2011-2012 mass balance year was relatively low (147 Gt $year^{-1}$), then increased greatly in 2012-2013 (473 Gt $year^{-1}$) and decreased slightly in 2013-2014 (403 Gt $year^{-1}$). Our model, which tends to simulate lower SMB compared to MAR v3.5.2, produced a similar sequence in those years, with accumulated SMBs at the end of each mass balance year of –23, 420, and 312 Gt $year^{-1}$, respectively (Fig. 10a)." (P. 16, L. 591 - 596)

pg 15, lines 532-540: such sensitivity to the irreducible water content is also simulated by MAR which uses a value of 8%.

Thank you for the information. The provided information has been included in the revised manuscript as mentioned in the previous answer.