Response to Dr Collier

Thank you for the comments and great suggestions. Below in blue are our responses and actions.

Thank you to the authors for their responses and for the revisions, which have improved the manuscript. I support the publication of the revised manuscript subject to a few small clarifications and changes:

1. Lines 43-45: “Therefore, the proper simulation of the non-homogenous, non-stationary evolution of a glacier requires atmospheric processes at much finer resolution than typical global or regional climate models can provide (Collier et al., 2013, Collier et al., 2015, Aas et al., 2016).”
I appreciate that the authors have cited the additional studies. However, only Aas et al. (2016) looked at the resolution requirements of atmospheric forcing fields for adequate glacier simulations, as cited. The important, but still missing, point to contextualize the authors’ work is that these studies provided the first efforts to integrate a physically based glacier mass balance model into WRF for improved simulations.

Thank you for pointing this out. We have now added this sentence (underline indicate changes)

“Glacier mass balance parameterizations have been implemented in atmospheric models such as the regional climate model (REMO, Kotlarski et al. 2010b) <u>and a climate mass balance model with feedback to the atmosphere was implemented into WRF by Collier et al. (2013).”</u>

We have also removed the Collier et al citations in the sentence the reviewer pointed to.

2. Line 109: “Furthermore, this exposed glacier ice cannot melt as the glacier is only a land surface category”
Please rephrase to be consistent with the revised lines 74-75.
We added to this sentence: “Furthermore, this exposed glacier ice cannot melt as the glacier is only a land surface category <u>(though the glacier is represented in the soil layer with a two-meter layer of water/ice but does not provide runoff to WRF-Hydro).”</u>

3. Lines 134 to 138: “Importantly, the Crocus model interacts with the atmosphere by providing fluxes between the surface of the glacier and the atmosphere. These fluxes are total absorbed solar radiation, total reflected solar radiation, total net longwave radiation, total sensible heat, evaporation heat flux (and rate) from snow, and ground heat flux. Some diagnostic outputs, such as the 2m temperature and 2m vapor mixing ratio are calculated by the original Noah-MP snow model, but with the snow information (snow surface temperature and albedo) provided by Crocus.”

Since the forcing data come from an offline WRF simulation, there is no interaction between the Crocus solution and the atmosphere. This sentence can mislead readers into thinking interactive simulations have been performed between all three [atmospheric-cryospheric-hydrological] components. Please remove or rephrase to indicate that atmospheric interactions are planned for future work.

We agree with this comment, and this paragraph has been removed.

4. Line 442 “This is likely due to lack of using the baseflow/groundwater module in these specific WRF-Hydro/Glacier simulations.”
Why was this module not used and what is the potential impact?
The potential impact is that some groundwater/baseflow that could potential contribute to the streamflow is not included in the model. I did not use this option because I was not provided with the necessary input data for running this option. Any new simulation should include this option.

We have added to this sentence: “This is likely due to lack of using the baseflow/groundwater module in these specific WRF-Hydro/Glacier simulations, <u>which could add some water to the surface streamflow.”</u>

5. Line 477: “And while we have calibrated some parameters in Crocus, we used all default values in Noah-MP”
Can the authors provide the calibrated parameter values or include the information about finding the code on GitHub in the manuscript, so that the results could be reproduced?

Thank you for this question. In our first model results we had to adjust the roughness length to a rather low value for both snow and ice in order to compare better with observations. These are the results used when writing the initial drafts of the manuscript. However, a bug was found in the code (from the coupling, not in the Crocus code itself). When we reran the simulations, we went back to the original roughness values (but the text about calibration remained in the manuscript). We have therefore removed this paragraph.

6. Figure 13: I suggest merging the relevant panels with Figure 11 and 12, to aid with visual comparison.

This is a good suggestion. However, while merging the figure together, I found the figure to be “overwhelming”. I also believe that the two merged figures have the potential to being smaller in the final manuscript, than two individual figures. So I will keep them separate. However, I rearranged the rows and columns, so that the two figures now represent the year and model in the same row/column format.

Response to reviewer #2

Thank you for the comments and great suggestions. Below in blue are our responses and actions.

The authors have carefully revised the manuscript based on the comments raised during the first round of review. Therefore, the manuscript had been made clearer in many respects. I would like to thank the authors for this work. I have listed some comments about their answers followed by more specific and technical comments

Comments about the answers

- WRF-Hydro Glacier at 100-m.

I understand the argumentation about the potential limited impact of meteorological downscaling for glacier mass balance simulation in the context of this study. I still think that it would be good in the conclusion to have a paragraph about the benefits and limitations of the configuration at 100-m resolution compared to a configuration at 1-km. From my understanding, it allows a high-resolution glacier initialization and routing. On the other hand, it does not benefit from any topographic-based meteorological downscaling and does not represent physical processes affecting snowpack evolution at 100-m grid spacing.

We added this sentence in the conclusion

"Finally, the forcing at 1 km does not account for any topographic variations in the 100 m domain, thus snowpack evolution at 100 m scale is not included."

- Evaluation of winter precipitation

The comparison between WRF and the DFAR data from Haukeliseter reveals similar precip. biases to those obtained in Finse and Midstova. However, the DFAR is supposed to be less impacted by wind-undercatch of precipitation than the precip. gauge in Finse and Midstova. The paragraph (P 10 L 294-305) added by the authors suggests that the comparison with the DFAR data has little value and is not really reliable. Does it mean that the authors question the quality of the DFAR data? I think the argumentation could be improved in this paragraph.

This is a very good and interesting question. I took a closer at the DFAR data and the bias, and the bias is lower than initial stated in the manuscript.

Regarding the DFAR, the bias in windy conditions is about 5-10 %. This comes from comparisons with bush-sheltered Tretyakov gauges (e.g. Rasmussen et al. 2012): " The DFIR configuration was extensively compared to a bush-sheltered Tretyakov gauge, considered to be a true representation of snowfall, at the hydrological research station near Valdai, Russia, from 1970 to 1990. Although the large octagonal double fence was shown to catch less snowfall than the bush gauge, the differences were relatively small (<10%) "

This underestimation does not explain the about 20% higher overestimation in WRF. After conversation with other modelers that focus on Norway, it is often seen that in westerly conditions, models at times underestimate precipitation at the coast and therefore produce more precipitation further inland.

We have now rephrased the text so the quality of the DFAR is not questioned:

"We compared the WRF model results with the DFAR data (Smith et al. 2019), and WRF is predicting more precipitation compared to these observations, with a bias typically at ~30% (not show). About 10% of this bias could potentially be attributed to underestimation with the DFAR (Rasmussen et al. 2012). The bias in WRF is opposite and higher compared to what is found at locations with little impact of snow. In regards to transfer functions (correcting for under catch in observations), Smith et al (2020) stated this is their study: "Although the application of transfer functions is necessary to mitigate wind bias in solid precipitation measurements, especially at windy sites and for unshielded gauges, the inconsistency in the performance metrics among sites suggests that the functions be applied with caution." We are therefore not adjusting the observed observations on Finse for our evaluation, and rather stress the well comparison between model and observations at Fet and summer season precipitation on at Finse. "

- Variability of snow accumulation

The new figure 13 added to the manuscript is quite interesting and illustrates well the impact of blowing snow sublimation on the simulated snow depth at the glacier scale. However, it also raises a question which is not answered by the authors. Figure 11 shows a clear difference between Noah-MP and Crocus in terms of pixel-to pixel variability of simulated snow depth in 2017 and 2018. Figure 13 (left) shows that the simulated snow depth also presents this pixel-to pixel variability (alternance of orange and yellow/green colors) when the blowing snow sublimation is not activated. This variability is at sub-kilometer scale. In these experiments, Crocus at 100 m is driven by bilinearly interpolated smooth atmospheric forcing obtained from WRF at 1-km grid spacing. In addition, Crocus does not account for lateral-snow redistribution that could create small-scale variability of snow accumulation. Therefore, the author should better explain which processes are generating this small-scale variability in the model. Does it result from the implementation of Crocus in WRH-Hydro? It seems the blowing snow sublimation is not the explanation. Or maybe, it is just a visual effect of the plotting library.

I agree with this comment. This is an issue that is under investigation and my suspicion is that some error constraints in the model had to be relaxed in the current version when implemented into WRF-Hydro. Though this is something that should be looked into, the overall conclusion in the manuscript does not change.

- Station selection

The authors considered in their new figure 4 the differences between the station elevation in the model and the actual station elevation when selecting the stations used for evaluation.

However, the criteria of 100-m mentioned in the response to the reviewers is not mentioned in the revised manuscript. This should be added.

We have added this sentence to the text: “Furthermore, stations that are located in the model over 100 m above the actual elevation are not included.”

Specific comments

P 16 L 477: the author mentions here that some Crocus parameters have been calibrated. Could they list the parameters that has been calibrated and which calibration strategy was used? Maybe it could be briefly described in Sect. 2.1.

Thank you for this question. In our first model results we had to adjust the roughness length to a rather low value for both snow and ice in order to compare better with observations. These are the results used when writing the initial drafts of the manuscript. However, a bug was found in the code (from the coupling, not in the Crocus code itself). When we reran the simulations, we went back to the original roughness values (but the text about calibration remained in the manuscript). We have therefore removed this paragraph.

P 17 L 502-507: I think it is important to mention here that the observed density has been measured at one single point over the glacier. This is a limitation for the comparison between observed and simulated snow density.

We added this sentence: “Also note that the snow density is only measured at one location and assumed to be the same over the entire glacier.”

P 25 Table 2: the signification of the values appearing in bold is still not really clear to me. It seems that for some year and some location, the highest correlation value is not written in bold. See for example the edge site in 2015 and the value of 0.90 for Noah-MP with respect to MODIS Terra.
We have removed the bold font as we do not mention this in the text

Technical comments

P11 L 323: space between “-0.13” and “m ..”. use “m s^{-1}” instead of “m/s”
Done

P13 L 400: “turning off”
Done

P 24 Table 1: “Domain 2”
Done

P 24 Table 1: “grid points”
Done