Review of "Meltwater runoff and glacier mass balance in the high Arctic: 1991-2022 simulations for Svalbard" by Louise Steffensen Schmidt et al.

This study presents a new dataset of modelled surface energy and mass balance fluxes over Svalbard using CryoGrid, and the new, high resolution reanalysis product CARRA as atmospheric forcing. The model is then extensively evaluated against a large database of weather station and mass balance observations, and finally used to quantify the changes in near-surface meteorology, glacier mass balance, and surface runoff over Svalbard during 1991-2022.

The methods used in this study are very suitable, and are also innovative in the sense that it is the first time that such a high-resolution atmospheric reanalysis product (CARRA 2.5km) is used to estimate the energy and mass balance using a surface/subsurface energy balance model for a period of 30 years. The evaluation of the new dataset is very detailed, and supports the overall capability of an energy balance model to simulate region-wide mass balance components. The significance of this work is large, since it describes a new modelling framework to estimate region-wide surface fluxes, where until now rather coarse and uncertain climate models are used for instead. Something that is not explicitly mentioned by the authors, yet increases the significance even more, is that such a database may serve as a unique benchmark dataset for climate models, since it contains many more variables than those available in an atmospheric reanalysis product like CARRA or ERA5. The paper is well written and well structured. The data is presented with clear figures and tables, but the manuscript is quite long (26 pages without counting the reference and the supplementary material).

In its present form the manuscript does not explicitly raise a research question, which I believe makes it harder for the reader to find the scientific significance of this work. Instead, this manuscript is a rather detailed description of a new, high quality dataset, and would therefore be very suited for a dataset journal such as ESSD.
Given the significance of this work for the community, I would nevertheless recommend publication in The Cryosphere, yet I would also recommend that both the abstract and the introduction are reformulated such that it becomes more clear what are the motivation and the scientific questions of this work.
In the conclusion it appears that an answer is given to these questions :

- How accurately can we model the energy and mass balance components using CARRA ?
- What is the mass balance of Svalbard glaciers and how did it change in the period 1991-2022 ?
- Can we realistically forecast in near-real time the surface fluxes in Svalbard using weather forecast model AROME-ARCTIC ?

Explicitly mentioning the research questions with a clear 'motivation → obstacle → solution' structure in the introduction would be very helpful. A clear obstacle I can think of, and that strongly motivates this work, is that climate models are possibly too coarse to accurately model the mass balance of Svalbard glaciers, and that statistical downscaling methods do not explicitly resolve the physical processes. In the next part some further suggestions are given with the aim to further improve this manuscript. Overall I would like to congratulate the authors on generating this dataset and on performing such a detailed analysis.

Major comments

- The introduction does a good job in describing the scientific literature, yet as mentionned above it could be more clear in describing the scientific challenge, and also more clear in explaining why using CryoGrid forced by CARRA offers a solution to the scientific challenge.
- It is not clear why AROME-ARCTIC is also used as a forcing for CryoGrid. It seems that CARRA and AROME-ARCTIC differ quite significantly, which is interesting yet not very surprising given the fact that CARRA assimilates some observations and that AROME is a weather forecast model (as mentioned by the authors in the conclusion as well). Regional difference between the two products may average out, but this still reduces the applicability/usefulness of a near-real time forecast of mass balance. Perhaps the authors could more clearly motivate the choice of using AROME-ARCTIC, or just briefly mention this possibly in the discussion and then remove section 6.2.
- How did the authors compute the temperature and relative humidity from automatic weather stations at 2 m, and wind speed at 10 m ? These variables are often measured at some height above the surface that changes between each maintenance, and also due to snowfall. Because of the large vertical gradients on a glacier, the deviation of T,WS and q in using a wrong sensor height can be quite large, which could make the evaluation with CARRA or ARMOE-ARCTIC inaccurate.

Minor comments

- Figure 1: Perhaps it would be useful for the non-expert reader to add the locations of each glacier/region mentioned in the text (Etonbreen, Austre Brøggerbreen, etc ...). It would also be very useful to add the coordinates and elevation of each AWS, or to refer to studies where this information is available.
- L69-71, It is not clear why the Russian Arctic is mentioned here since it does not re-appear in the methods or in the results. This interesting idea should be better mentioned in the discussion.
- Lines L83-84 can be made more consistent with L27-28.
- L186: Please clarify what modules of CryoGrid are part of Westermann et al (2022), and what modules have been specifically added for this study. Also, please explain how the 'water percolation and runoff modules' (4.2.1) differ from the original CryoGrid setup.
- L192: Why was a constant bare ice albedo of 0.4 chosen ? This seems very simplistic yet it appears that the errors in albedo are very small nonetheless (Table S2.).
- L205 I believe that this is commonly called a "bucket scheme". Then the variable θ_{fc} in Eq (1) is the irreducible water content (or " maximum liquid water-holding capacity " as defined in CROCUS by Vionnet et al 2012.), which is not necessarily the same as the field capacity (typically defined for soils, as far as I know).

- Also, if "Water is not allowed to flow into an impermeable layer, here defined as layers with a density higher than 830 kgm3" (L219), it is interesting and perhaps counter-intuitive to see in the results that "the average annual internal accumulation is 0.11 mw.e., and thus accounts for almost half of the total refreezing (Fig. 9c)". Perhaps the authors could comment on this in the discussion ?

- L209: Consider replacing "water saturation" by " volumetric water content in the snow" to be consistent with other snow models.

- L225 How is the "water in excess of the field capacity" defined ?

- L256-259. I acknowledge that the manuscript will be a bit long if the AWS evaluation is further detailed here. Yet I find the statement "both models generally fit well with observations" a bit simplistic, not very specific and possibly also inaccurate. It would be very useful for future studies to better understand what causes the limitations of a SEB model forced by CARRA. For instance it seems that summer ablation is overestimated at Nordenskiöldbreen (figure 4, red lines), which could partly be explained by an underestimated albedo by 0.06 (Table S1). Also, it appears that there is a systematic underestimation in both incoming and outgoing longwave radiation components at all glacier AWS.

Technical comments

- L10 "an"

References

Vionnet V, Brun E, Morin S, et al (2012) The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2. Geosci Model Dev 5:773–791. https://doi.org/10.5194/gmd-5-773-2012