Review of *"A high-resolution pan-Arctic meltwater discharge dataset from 1950 to 2021"*
by Adam Igneczi et al.

The authors present long-term timeseries of daily pan-Arctic land ice and tundra runoff (1950-2021). Daily outputs from the regional climate model MAR at 6 km are first statistically downscaled to 250 m spatial resolution in sub-regions, i.e., Greenland, Canadian Arctic, Svalbard, Iceland and Russian Arctic, and further spatially integrated at the basin scale. The authors find that spatially integrated land ice and tundra runoff are larger than previous estimates at lower spatial resolution (pan-Arctic or Greenland only), while sharing overall similar variability and trends. The authors suggest that the larger runoff at 250 m stems from enhanced spatial resolution relative to previous products, the result of small glaciers and rugged tundra regions being better resolved.

While this data set will be of interest to the cryosphere community, I have major concerns regarding its evaluation. The authors claim that higher spatial resolution relative to previous estimates results in improved runoff representation. However, without a thorough data evaluation as e.g., in Mankoff et al. (2020) for the Greenland ice sheet, it is impossible to verify this statement. In addition, the actual impact of statistical downscaling on MAR integrated runoff is not assessed. Runoff increase relative to previous lower-resolution products could potentially originate from an overall runoff overestimate in the downscaled MAR product. This must be examined in more detail to support the authors claim that the presented data set is an improvement upon previous products. Based on the above and the following general and point comments, I deem that major revisions are required before considering this study for publication in ESSD.

General comments

1. Data evaluation is crucial to assess whether land ice and tundra runoff timeseries are robust, and an actual improvement upon previous products. To do so, the authors could use discharge measurements (e.g., for Greenland rivers in Mankoff et al., 2020) and modelled runoff estimates from e.g., (statistically downscaled) regional climate models that have been thoroughly evaluated in previous publications (i.e., using in-situ and remote sensing data). Such data sets exist for most ice masses in the Arctic, but a thorough evaluation for the well-studied Greenland ice sheet would be highly beneficial.
2. The authors do not assess the impact of their statistical downscaling technique, i.e., how does downscaling MAR at 6 km to 250 m affect integrated runoff in different regions? This is particularly important for smaller Arctic ice masses that may not be well resolved in low-resolution MAR.
3. The authors claim that higher spatial resolution improves runoff representation based on a previous study (Noël et al., 2016). However the latter work uses a different regional climate model combined with a different statistical downscaling technique, which does not imply that similar improvements will hold for MAR. For instance, Tedesco et al. (2023) found better agreement with in-situ surface mass balance measurements in Greenland after statistically downscaling MAR at 6 km to 100 m using a mass conservative approach (i.e., no runoff increase between resolutions). Model evaluation is therefore essential to ensure that enhanced runoff at 250 m found in this study does not reflect an overall overestimate in downscaled MAR.
4. The MAR version used in this study is never mentioned. This is important information for cross-study comparisons e.g., is it the same MAR version as in Mankoff et al. (2020) and/or Tedesco et al. (2023)?
5. The authors should elaborate on the difference between their statistical downscaling method and that of e.g., Franco et al. (2012), Noël et al. (2016), and Tedesco et al. (2023).
6. The paper and its Figures are mostly focused on Greenland. It would be beneficial to show downscaled outputs from smaller glaciers, to provide insights on the technique performance in different regions.

Point comments

L1: The authors could consider "basin scale" instead of "high-resolution" in the title.
L16: I am concerned about using the term "improve" as no detailed model evaluation is performed. I strongly recommend evaluating the presented data set with existing (high-resolution) runoff products

(modelled and observed) to show that the larger runoff found in this study compared to previous ones e.g., Bamber et al. (2018) and Mankoff et al. (2020), is an actual improvement.
L16: I suggest “basin scale meltwater discharge data product”.
L19: “Modèle Atmosphérique Régional (MAR)”.
L67-68: The authors state that the higher spatial resolution in this study (250 m) improves runoff estimates (L16, L374-377, L392-395). However, the data set in Mankoff et al. (2020) is available at 100 m. Please clarify.
L71: Do the authors mean “1950-2021”, the year 2022 is not shown or discussed elsewhere.
L79: As downscaling uses elevation gradients, this implies that a relationship between MAR ice albedo and surface elevation exists. Is this the case? How well does MAR at 6 km represent ice albedo and the location of the snowline? This is important as the location of the snowline, and hence the bare ice extent and associated albedo, have a strong impact on the modelled runoff amount and spatial distribution. For instance, Ryan et al. (2019) showed that MAR overestimates bare ice extent in Greenland (and potentially the runoff production?).
L124: After interpolating the GIMP/RGI ice masks onto the 250 m grid, how did the authors cope with fractional ice cover? Did you discard e.g., grid-cells with <50% of ice coverage? Please elaborate. Do the resulting ice/land mask area align with observations, especially for smaller glaciers?
L140: Please specify which MAR version is used here.
L141: Do the authors mean “six hourly”?
L145-146: I count six MAR domains in Figure 2. The Svalbard region is not mentioned. Same comment in L155 “4 domains”.
L146-149: What about the tundra area? Are they well captured by MAR? How robust are runoff estimates over the tundra regions?
L172-174: Why not using the original grid instead of interpolating MAR on a regular 6 km grid? This may lead to additional uncertainties. Please clarify.
L197-202: In fact, data mix has already been applied to create Greenland masks, i.e., combining Copernicus GLO-90 DGEDEM for surface elevation and land mask, GIMP DEM v1 for ice sheet mask, and RGI v6 for peripheral GIC mask. For instance, the GIMP DEM could have been used for surface elevation, land and (peripheral) ice masks. Please clarify or reformulate.
L237-238: How are these data gaps filled? Please clarify. Do the authors mean “Section 4.3”?
L238-240: How does discarding smaller basins affect the total land ice/tundra integrated areas?
L244: The paper refers to python tools and options that readers may not be familiar with. The authors could briefly explain what these consist of.
L255-258: The authors should elaborate on differences with previously published downscaling techniques, notably that of Tedesco et al. (2023) using MAR at 6 km as input. I am not sure to understand how downscaled ice albedo is used in the calculation of runoff.
L262: Are tundra and land ice runoff gradients computed separately? How do the authors downscale runoff at the interface between tundra and land ice?
L263-265: Discarding vertical gradients for elevation difference < 50 m may affect runoff production nearby the equilibrium line, i.e., towards the flatter glacier interior.
L269-271: This is valid for RACMO, but does it hold for MAR?
L281-285: What are the differences with previous downscaling techniques?
L319: I think Fig. 5 shows the opposite: “COP-250 DEM minus MAR DEM” with outlet glaciers (ice divides) elevation being overestimated (underestimated) in low-resolution MAR.
L337-343: The upper threshold for bare ice albedo is commonly set to 0.55 and is used to discriminate snow from ice. Could you please clarify why a snow value of 0.70 is used instead?
L362-364: I strongly recommend that the authors compare both the original and downscaled MAR runoff, i.e., to assess the impact of the statistical downscaling technique.
Section 4: Differences in Greenland ice sheet and peripheral GIC runoff can reach up to ~50 Gt between this study and the two previous products (Fig. 7a), i.e., 10 to 15% of the total. This is significant. For Greenland tundra and non-Greenland land ice, these differences are even larger. This calls for a proper data evaluation using discharge measurements (Mankoff et al., 2020) or previously published (high-resolution) land ice runoff products.
L374-377: This is somewhat speculative. The statement can only be verified by statistically downscaling MAR to different spatial resolutions and comparing the outputs with MAR at 6 km. In addition, the other data sets are based on a different regional climate model combined with different downscaling techniques, which may also explain the discrepancies.

L380-383: This statement suggests that the current data set outperforms that of Bamber et al. (2018). However, without proper evaluation, it is impossible to assess.
L383-384: Different reanalysis forcing will indeed impact the results. The input regional climate model will also strongly affect the results as they may not produce identical runoff amount and distribution.
L396-399: Trends and variability are similar for Greenland land ice, tundra and non-Greenland land ice, i.e., although with a positive runoff offset in the current study. However, it is not the case for non-Greenland tundra. Could you elaborate on this?
L400-403: I am not sure to understand how statistical downscaling in tundra regions is important for Greenland (L392-395) but not for other Arctic regions? Is the tundra region rougher in Greenland? Please explain.
L407: I am confused, I understood that MAR at 6 km was used as input. Was it formerly downscaled to higher resolution before applying your downscaling technique? Please clarify.
L411-413: I strongly recommend performing a thorough data evaluation, otherwise it is impossible to assess whether this new data set is robust, or an improvement upon other products.
L415-420: The fact that improvement was found in Noël et al. (2016) using RACMO, with a different downscaling technique, does not imply similar results when using MAR as input. For instance, Tedesco et al. (2023) suggest that mass conservative statistical downscaling (i.e., no runoff increase) is required to better align with in-situ measurements in Greenland.
L420-422: The variability and trends are mostly similar between products, but the runoff offset in this study remains important, calling for a proper model evaluation.

Figures

Fig. 2: Please add a scale bar for surface elevation.

References

Mankoff et al. (2020): https://essd.copernicus.org/articles/12/2811/2020/
Noël et al. (2016): https://tc.copernicus.org/articles/10/2361/2016/tc-10-2361-2016.html
Tedesco et al. (2023): https://tc.copernicus.org/articles/17/5061/2023/
Franco et al. (2012): https://tc.copernicus.org/articles/6/695/2012/
Bamber et al. (2018): https://iopscience.iop.org/article/10.1088/1748-9326/aac2f0/meta
Ryan et al. (2019): https://www.science.org/doi/10.1126/sciadv.aav3738