Review of '*High-resolution modeling of glacier mass balance and surface runoff in western Norway driven by bias-corrected climate forcing*' by Gong and Rogozhina
Under review for '*Hydrology and Earth System Sciences'*

In this manuscript, Gong and Solomina present a study in which they model the glacier mass balance and related surface runoff in western Norway. In contrast to most other studies, the modelling is performed at a very high resolution and by relying on a bias-corrected climate forcing. The authors perform simulations over past time periods for two types of climate forcing: the bias-corrected NORA10 hindcast (2000-2014, considered as the reference / benchmark case) and the uncorrected CORDEX outputs (2000-2020). They find that simulations forced with the NORA10 product can match various types of observations well, which is not the case for the CORDEX-forced simulations. The latter do not correctly reproduce the runoff in some catchments. The authors suggest that this mismatch is likely to be transferred to future simulations, thereby questioning the capability of uncorrected products for future simulations.

This work is generally well presented, and I found the text and figures easy to follow. The research question is also interesting, and worth investigating. However, I do have some concerns about the analyses performed and the conclusions that are drawn from this. I hope these concerns can be addressed by the reviewers, which is likely to require a substantial reworking of the manuscript and the results presented.

General comments

- Using high-resolution products to model glacier mass balance and runoff is interesting, but a central question here is: **how are these products downscaled to the glacier scale?** This question is particularly relevant for the CORDEX data, which comes at a lower resolution. A technique (/trick) that has been used in other studies that mainly focus on glacier mass balance is to ensure a consistency between observed and modelled glacier mass balance through a calibration of various mass balance parameters (e.g. Huss and Hock, 2015). This calibration procedure thus acts as a kind of downscaling. In the manuscript you present here, it is not entirely clear how (elaborately) the calibration of the model is performed. Without such a thorough calibration, it is not very surprising to have a lower model performance when the climate data used is somewhat rougher (e.g. in terms of resolution) / less specific for the region (e.g. by not being bias-corrected to detailed measurements over region of interest). **However, this does not imply that future projections with a rougher/less specific product will produce less good results for future simulations: most of this will depend on how the data is downscaled** (or bias-corrected to local observations if you will). An important study in this regard is the one by Compagno et al. (2021), which specifically analyzed the effect of using various climate forcing products to model the future evolution of glaciers (including over the region covered in your study). Although this study cannot directly be compared to yours, e.g. given that focus in that study is specifically on glaciers (vs. more broad scope, including runoff, in your study here), their main findings are in contrast with yours. They suggest that as long as a detailed downscaling occurs (i.e. a step in which the climate data is in fact bias-corrected to local observations on glacier mass balance), the impact of the climate product used on the modelled future glacier evolution is very limited.
- Besides this fundamental question related to the transfer of climate data to the high resolution of the modelling framework, also other elements were not entirely clear to

me (which may of course be related to my misunderstanding, but it would then still be nice to have this further clarified in an updated manuscript):

  - Why are different time periods considered for NORA10 and CORDEX? I understand that NORA10 is not available until 2020, but to have a more direct comparison between both approaches, it would make sense to compare them over the same time period. See also next comment.
  - It is not clear why different RCPs are considered for the CORDEX data and whether this is a correct approach for the goal you want to reach. RCPs are projections, and over short time periods they will mainly produce a random signal, that will (of course) not closely match the real climatic data over this period. This is especially the case when comparing CORDEX to NORA10, where the latter was specifically downscaled to match observations for the period that you considered here. From this perspective, it does not seem to make a lot of sense to compare RCP4.5 to RCP8.5 results. Again, it would seem more logical to only perform your comparison between NORA10 and CORDEX data over an observational time period / period for which e.g. reanalyses data exists, which can be used to constrain these simulations / bias-correct them to match these (in the case of NORA10). Without being constrained, it does not really make sense to compare the CORDEX simulations with NORA-10. The former will be able to reproduce general trends, but of course not the year-to-year variability..
  - Why is only one CORDEX GCM-RCM couple considered? The EUROCORDEX framework contains many dozens of simulations, which would allow you to go for an ensemble approach and really explore in greater detail how well CORDEX data can be used to reproduce local observations. All data is available, so it seems difficult to justify why only one specific GCM-RCM couple is used.
- Throughout the text: can you 'validate' your model setup? Or is this rather an 'evaluation'? Is strictly speaking not the same, and what you are doing here (and people in general who use models in Earth Sciences) is probably rather an evaluation than a validation.

Specific and technical comments

- l. 16: 'the variables mentioned above': not entirely clear which variables this refers to.
- l. 27: 'casting doubts on the applicability of bias-corrected…': not entirely sure: see first general comment above. You will probably only be able to know if it reasonable by performing future simulations. And as said, if the climatic data is downscaled/calibrated to local data (on e.g. mass balance), you may end up finding that the climatic data used does not have such a huge influence on your modelled future simulations in the end..
- l.31-32: suggest adding a reference to the landmark paper by Immerzeel et al. (2020)
- l.36-37: 'extensive snow and glacial ice melting': is this still the case now? Or are we now already past the peak? See also 'peak water' concept, suggesting that regions like Scandinavia may have pass their peak in runoff already.
- l.42: global trend in glacier retreat. Here makes sense to refer to recent study by Hugonnet et al. (2021).

- l.45-46: advance of glaciers and link to climatic conditions: possibly refer to study by Trachsel and Nesje (2015) here.
- l. 46-47: mass loss in Norway inevitable. Suggest adding reference to Compagno et al. (2021) that focuses on this region (vs. Cogley et al., 2011, on Himalaya and Karakoram glaciers..) and to the most recent GlacierMIP effort as well (Marzeion et al., 2020), which combines future glacier simulations (including Norwegian glaciers) from various groups around the world.
- l.49: 'Regi<u>o</u>nal Climate Models'
- l.50: 'constraints from observations availability' → yes, but is not really a constraint anymore, definitely over the time periods you consider in your study, given the availability of glacier-specific observations for every glacier on Earth (Hugonnet et al., 2021)
- l.56: '…and glacier dynamics': true. But you do not include this in your study, do you?
- l.62: 'regional and national scales': bit confusing. Typically, regional scales are referred to as being over entire regions (i.e. > national scale mostly). Maybe change this 'local' scales?
- l.75: '2000-2014 and 2000-2020': see earlier comment. Would make sense to have this over same time period + problem related to use of rcps…
- l. 103-104: 'full surface energy balance': is this justified over such a large domain? Probably uncertainties over some of the input variables must be very large, no? (even for the NORA10 product). Not sure I entirely understand, as a bit later (l. 122) you mention that 'leaves the surface temperature as the only unknown'. But in this case you are not really solving a full energy balance model, are you?
- l. 107: five submodules → four submodules? (or maybe I am missing one..)
- l. 125: SnowPack. What about ice / how is this treated? I understand that glaciers are considered to be static in your approach (see below), but what about the part of the glacier ice that is exposed at the surface (in ablation area during spring and summer)
- l. 130: sublimation. Must be very limited here? If not, could this be quantified? (e.g. vs. melt)
- l. 152-153: glaciers do not change over time. But in reality glaciers over this region have changed substantially over the past two decades (again, refer to Hugonnet et al., 2021). Can this not be accounted for? Probably worth mentioning that other approaches that focus on glacier mass balance and runoff over recent and future time periods have explicitly accounted for glacier changes over time (e.g. Laurent et al., 2020; Muelchi et al., 2021)
- l. 168: "they are corrected against": not entirely clear who did this. Did you do this or was this readily available? Good if you could be clear here to avoid confusion.
- Table 1: why this particular CORDEX simulation chosen and not one of the many others available? Ideal would be to have a large ensemble for this…
- l.176-183: great to have this info! Is often missing, and really good to explain. Will be useful for others attempting a similar modelling effort.
- l.187: to validate our SMB results: so no calibration performed for this? SMBs obtained 'out of the box' and compared to measurements? Results will strongly depend on how the data is downscaled to the very high resolution, no?
- l.213: "which are validated against": do I understand it correctly that this was done by others in their study? If so, maybe "which were validated"?
- l.239: why not compare over the same time period?
- Figure 3b and c, right panels: confusing to use the same color scheme, but representing a different extent (-4 to 10 mm/day and -2 to 7 mm/day). Make this consistent

throughout all panels? Or use different color schemes for every figure if this does not represent the same?

- l. 247: under different RCPs. Does not seem to make sense. Or is this meant to reproduce a random variation around exiting climate? But then again, would be more advisable to work with an ensemble based on various CORDEX simulations.
- l. 298: results align well with observed glacier retreat. Strange. Would expect a bias if you keep the glacier geometry constant over entire time period. In reality, the glacier has retreated, thereby increasing its mass balance (losing lower parts where mass balance is very negative). As you do not account for glacier change, I would therefore expect a negative bias in your results (your model "sees" the glaciers as being too big, with lower parts with a very negative mass balance, which in reality do not exist anymore)
- l. 303: some basins that have seen an increase in SMB towards 2014: what do you mean with 'towards 2014'? Has it been growing over the period 2000-2014? And is this confirmed in the observations by Hugonnet et al. (2021)?
- When describing these variations, does NAO play a role? (Marzeion and Nesje, 2012; Trachsel and Nesje, 2015)
- Sections 3.4 and 3.5 are very long. In the end we are mostly interested in comparing the outcome of both approaches, rather than going into the specific findings for the various regions. Would suggest making this more compact, potentially by having some of the figures and explanation in suppl. mat?
- l.379: "is by far the best…": well of course, as this product was made specifically for this region. But question is how it performs if both approaches are downscaled to local observations. Is there then an added value in using the detailed product vs. rougher CORDEX simulations?
- l. 383: good agreement. Not a negative bias because keep glaciers constant in time? See comment above (l. 298).
- l.410: "confirm the relationship between increased glacier cover and delayed peak runoff": increased glacier cover compared to what? Or is this relative in space (vs. other, less glaciated regions)? Not sure to entirely understand.
- l.426: "we are reluctant to carry out future projections": but then, with the material presented, I am afraid that it is difficult to make some sound conclusions about this. As mentioned before: would be interesting to compare large ensemble of COREDEX simulations, downscale each of them to a higher resolution as part of calibration procedure to reproduce observed glacier changes, and then see if the choice of the used climate product has a large effect on the modelled future results.
- l.440: simulations with RCPs do not reproduce specific runoff. Well, not that surprising. Would make more sense to have over observational time period / without relying on random RCPs.
- l.454: 'hampered by keeping the glacier geometry fixed in time': may indeed be the case. But could potentially circumvent this by modelling the evolution of glaciers, or using observations on glacier changes and impose these over the 2000-2015 time period.
- l.459-460: future glacier projections: could lose even more mass than mentioned here. Refer to reference work on future glacier projections and analyze the results over Scandinavia (Marzeion et al., 2020)
- l.479-480: link with NAO?

- l.480-495: quite specific results given here. For the conclusion, as message of general interest, would suggest focusing more on your main message: role of NORA-10 vs. CORDEX.

- General comment for figures:
    - Missing labelling of panels (Figures 1, 2, 8) or only partly labelled (Figures 3, 4, 5, 6, 7, 9, 10). By adding labels, avoid having descriptions in the text like: "according to the left column of…" (l. 295), "…left column of figure 6" (l. 311),..etc.
    - Often the figure cannot be read as standalone and need to refer to caption to know the content (e.g. Fig 2 right panel). Suggest adding this information directly in the figure, which will also allow using this figure direcly in a presentation for instance.

References

Compagno, L. et al., 2021. Limited impact of climate forcing products on future glacier evolution in Scandinavia and Iceland. Journal of Glaciology 67, 727-743, https://doi.org/10.1017/ jog.2021.24

Hugonnet, R., et al. 2021. Accelerated global glacier mass loss in the early twenty-first century. Nature 592, 726–731. https://doi.org/10.1038/s41586-021-03436-z

Huss, M. and Hock, R., 2015. A new model for global glacier change and sea-level rise. Frontiers in Earth Science 3, 1–22. https://doi.org/10.3389/feart.2015.00054

Immerzeel, W.W. et al., 2020. Importance and vulnerability of the world's water towers. Nature 577, 364–369. https://doi.org/10.1038/s41586-019-1822-y

Laurent, L., et al., 2020. The impact of climate change and glacier mass loss on the hydrology in the Mont-Blanc massif. Scientific Reports 10, 10420. https://doi.org/10.1038/s41598-020-67379-7

Marzeion, B., et al. 2020. Partitioning the Uncertainty of Ensemble Projections of Global Glacier Mass Change. Earth's future 8, e2019EF001470. https://doi.org/10.1029/2019EF001470

Marzeion, B. and Nesje, A., 2012. Spatial patterns of North Atlantic Oscillation influence on mass balance variability of European glaciers. The Cryosphere 6, 661–673. https://doi.org/10.5194/tc-6-661-2012

Muelchi, R., et al., 2021, River runoff in Switzerland in a changing climate – runoff regime changes and their time of emergence. Hydrology and Earth System Sciences 25, 3071–3086. https://doi.org/10.5194/hess-25-3071-2021

Trachsel, M. and Nesje, A., 2015. Modelling annual mass balances of eight Scandinavian glaciers using statistical models. The Cryosphere 9, 1401–1414. https://doi.org/10.5194/tc-9-1401-2015