Dear editor and referee #2,

We thank referee #2 for his/her constructive comments and suggestions of interesting directions in which this manuscript can and will develop to become a stronger scientific study. Whether or not these comments were meant to set a stage for manuscript modifications, we conclude that some issues raised by referee #2 deserve a thorough investigation to make a definitive case on the general applicability limits of downscaled, bias-corrected climate products, both during instrumental periods and future decades, for glacier surface mass balance (SMB) impact studies in Norway. While working on our responses, we have realized that apparent disparities and contradictions in the three reviews may be related to different scientific "cultures" in our native field - glaciology - as compared to other fields, such as hydrology and climatology. For example, more complex bias-correction methods, such as analyzed in our study, are traditionally used in hydrological impact studies, but not in glacier SMB studies, where climate forcing is commonly corrected using either simple Delta Change (DC) approach (e.g., Huss and Hock, 2018).

In attempt to accommodate "cultural" differences across scientific disciplines, including very different approaches and methodologies used within parallel fields of potential readers, we have decided to substantially restructure the manuscript to focus on an in-depth present-day evaluation of different bias-corrected climate forcings dynamically downscaled by the Regional Climate Model (RCM) for glacier change impact studies. On the one hand, limitations of this study imposed by the lack of feedbacks from changing glacier geometries do not allow us to quantify differences in the future SMB projections driven by bias-corrected versus raw climate products. On the other hand, our focus on the model validation over the instrumental record provides an opportunity to carry out a thorough analysis of the issues and open questions raised by the referee. In particular, we have introduced several new directions in the study by zooming in on the differences in SMB model reconstructions arising from different bias correction methods and RCP scenarios, an identification of their intrinsic drivers and an interpretation of what they mean for the performance of bias-correction methods and downscaled model products, with general conclusions being applicable to both the instrumental period and future projections.

In response to the referee's interesting observations, this new version of the manuscript is set to accommodate a range of goals and objectives that will complement the objectives of our original study:

- To quantify the uncertainties brought by climate forcing corrected for large-scale biases by different methods in our specific applications to Norwegian glacier SMB studies, with an outlook into future glacier impacts studies.
- To estimate to which extent the lack of seasonal and interannual signals in the CORDEX climate forcing could contaminate glacier SMB calculations in the future projections.

As a result, the new version of the manuscript does not simply present the results from the benchmark and CORDEX-driven simulations but zooms in on the intricacies of different members of the climate model chain, including divergent RCP scenarios and bias-corrected outputs, and their direct validation against a combination of glacier SMB simulations and in situ observations. Below is the list of major changes (mainly in Sect. 3 and 4) we will implement in the current manuscript:

- A more detailed evaluation of the climate model outputs (NORA10 vs. observations, NORA10 vs. CORDEX, raw CORDEX vs. bias-corrected CORDEX as well as bias-corrected CORDEX with RCP4.5 vs. RCP 8.5 scenarios) and of the modeling results in terms of statistical quantities (see the response to the 1st major comment),
- Analysis of the surface runoff on the glacier-covered regions and its components instead of the surface runoff of the entire catchment (drainage basin), as it allows for a more robust model validation against observations, without the need to compare with the observed discharge regimes at the gauge stations. The latter is complicated by the fact that this study

does not use a runoff routine model to route the surface runoff to the stream flow, which has been developed and included in the follow-up article,

- Quantification and discussion of the uncertainties that are likely to be inherited by the future simulations.

To clearly reflect our goals and objectives listed above, we have also decided to change the title of the manuscript to ‘Synopsis of the uncertainties introduced by bias-corrected climate forcings in regional glacier surface mass balance evolution studies - A case study using a CORDEX chain envelope in western Norway’.

The specific response to the major comments (written in light blue and *italic* font) of referee #2 are presented below.

1. Comparison of climate scenarios and observations: It is not clear how you compare the climate forcing with observations in 3.1, but it sounds like you’re calculating the RMSE on the time series. However, this cannot be done as there is no correspondence between the dates: the climate projections are disconnected from the actual weather system evolution, and they cannot be compared in terms of time series even on the control period. Only overall statistics can be compared between the two. It is also the case for Fig. 8, where time series are plotted for CORDEX (climate projections) and NORA10 (reanalysis-driven).

We agree with the referee #2 those raw outputs of climate model projections are not expected to represent the actual weather system evolution. However, bias-correction of RCM outputs is, at least, trying to provide accurate model forcing not only of time-average conditions, but also of the day-to-day (and even sub-daily) variability to certain impact studies, e.g. hydrological modeling (Portoghese et al., 2011). Following this comment of the reviewer, we have decided to analyze actual impacts of bias corrections on the climate signal reproduction in climate model products and to which extent failures to reproduce such signals have adverse impacts on the glacier SMB and runoff simulations. While many other surface processes may not be as sensitive to climate variability, glacier ice and snow melt rates are directly related to the correctness of climate forcing that activates such processes and introduces a range of feedback mechanisms through e.g., energy balance, albedo, glacier geometry, etc.

We would like to point out that we have calculated the annual and seasonal (winter and summer) RMSE and mean differences between the basin-wide (averaged over the entire catchment area) temperature and precipitation assimilated by SnowModel and the gridded seNorge dataset (observations) in Sect. 3.1 (see Fig. 3). We did not compare the daily time series to the observations, but rather the annual and seasonal means. Now we think that these details have not been sufficiently described and will introduced a more detailed description of the evaluation process in the revised manuscript.

We also agree that we should add a comparison of more overall statistics between the assimilated and observed temperature and precipitation data as well as the modeled outputs between the benchmark and CORDEX driven simulations. We will therefore include more statistical quantities relevant to bias-corrections and SMB evolution, such as rainfall and snowfall frequency, annual cycle of daily temperature, seasonal mean temperature, etc., as well as, for instance, multi-annual mean and seasonal distribution of SMB, maximum snow and glacier melt and the month of peak glacier surface runoff.

2. Comparison of RCP scenarios: It does not make much sense to compare RCP scenarios for the present, as the scenarios did not diverge for the past/present. They only diverge in the future. Thus, all analyses of the role of the RCP scenario (l.247-249, l.264-265, l.271-273, l.350, l.375-376, …) do not make sense.

The climate forcings under different scenarios do diverge from 2005 (e.g. Schwalm et al., 2020). In addition, our modeled SMB and surface runoff also show divergence (Fig. 8 and 9); we therefore believe it is useful to look at how much different scenarios cause the climate to diverge from the ‘real’ present day climate for a control period (Yang et al., 2010) and what kind of uncertainty this will bring to the future projections of glacier change. We will clarify and discuss the choice of different bias-corrected datasets and scenarios more in detail in the revised manuscript, including an analysis of how different bias corrections work with departures in present-day climate simulations under dissimilar RCPs.

3. You used only one GCM-RCM chain. However, it is nowadays recognized as a best practice to not use a single model chain but to account for the uncertainty of the climate models by using several climate forcing chains. You state in 4.2 that "the first and foremost concern lies in the choice of future climate forcing from GCMs and RCMs". Well, you shouldn't pick only one in the first place… Using different bias correction methods is a good idea to account for the uncertainties related to the downscaling/correction, but it does not replace the consideration of the uncertainties from the climate modeling chain.

The purpose of the analysis presented is not to outline uncertainties in all existing climate model projections, but rather to zoom in on the performance of bias-corrected climate products that are expected to be regionally calibrated for the use across Norway. It is also important to keep in mind that such evaluation is only viewed through the prism of glacier SMB impact studies as opposed to a general climatological context. Hence, our study has a purely glaciological orientation, where we analyze pros and cons of bias corrections for studies of the Norwegian cryosphere. We realize that our motivation for the choice of the climate forcing datasets can benefit from further expansion of the text in Sect. 2.4, where we explain the rationale behind our experimental design.

Firstly, we use a single GMC/RCM chain to reduce the complexity of the evaluation, because our goal is to assess the uncertainties brought by different bias-corrections to glaciological studies over the time interval that can be validated against observations – i.e., the instrumental period. Therefore, future projections are not carried out at this stage.

Secondly, we combine analyses of EC-Earth/RCA (the only GCM-RCM chain) and NORA10 outputs to drive our simulations, because of the proximity between these two products: NORA10 is produced by HIRLAM driven by ECMWF IFS outputs, whereas EC-Earth uses ECMWF IFS for the atmosphere-land component, and RCA is based on a parallel coding of HIRLAM with some modifications in the model formulation. This choice is made deliberately to decrease the complexity of our analysis, to exclude model runs that have dissimilar origin and to focus solely on the evaluation of bias corrections. We will not only work on emphasizing our rationale but also on clarifying our main objectives early in the text.

4. You show that the bias-corrected CORDEX outputs still have a high bias. Thus, it seems that the bias correction was not optimal. It is not clear if you did the bias correction yourself or if you used an already bias-corrected product. The analysis of the climate forcing is then enough to identify that the data cannot be used directly in a climate impact study. There is no need to go all the way through the snow/glacier/hydro models. Maybe these errors you identified in the bias-corrected CORDEX can be related to how you computed the comparison, such as by directly comparing the time series (see above)

Even though the bias-corrected CORDEX outputs still have a high bias compared to the reanalysis product NORA10, it does not mean that we cannot use them for impact studies. Indeed, it is unclear how significant such bias may be for the glaciological studies, as long as we have not put any numbers on their direct impacts. Existing articles on bias-corrections of RCM outputs have pointed out that all bias-correction methods have their limitations (e.g. Maraun, 2016; Holthuijzen et al., 2021). What is important at this stage is to address these limitations, quantify the uncertainties they might bring to impact studies and motivate development of new bias-correction methods that are more suitable to this particular application – glacier SMB impact studies. This has been our goal in this study, but we agree that the presentation of our results may have interfered with the clarity of the narrative. As we have pointed out in responses to earlier comments, in the revised version we will move away from the simple presentation of the results from the benchmark and CORDEX driven simulations to a detailed statistical evaluation of the climate products and their significance for the SMB model experiments.

5. You analyze spatial and temporal patterns from the model outputs, while this can (and should) be first retrieved from the data. Also, even when you used the benchmark model (reanalysis-based), you never compare the runoff outputs to observations. We have no way to assess what your model outputs are worth. The model results of the runoff are the basis of several analyses. However, there is no metric regarding the calibration/validation with reference to observed discharge (e.g., NSE, bias, …). Thus, we cannot know if these analyses rely on plausible results.

The current version of the manuscript presents the catchment wide surface runoff results that could not be compared with the measurements from gauge stations, because this study did not utilize a runoff routine model, which had been later developed for the follow-up manuscript. However, we did compare our modeled SMB with direct measurements, which is a common-place procedure in glacier SMB model studies. Thus, following this common practice, we have decided to only limit our analysis to the surface runoff in the glacier covered region, which mainly consists of the glacier ice and snow melt water as well as rainfall, instead of the surface runoff of the entire catchment as it was presented in the initial manuscript.

6. I'm not so keen on the correlation analysis (4.1) based on the model outputs, with no comparison with observations. We retrieve the model behavior more than the natural system behavior. You infer several conclusions based on the model outputs, while we have no clue what they are worth.

We did compare our modeled glacier SMB with measurements (see the response to the point #5 above).

The response to the other elements of review #2 follow. **Due to the substantial re-framing of the manuscript some of these comments might not apply to the new manuscript**:

1. You state that "the ability of high-resolution modeling to accurately project glacial hydrological changes into the future is hampered by keeping the glacier geometries fixed in time" (l. 453). You also suggest in your last sentence to consider "the evolution of the glacier geometries and extents". There is a whole bunch of literature on that topic. Approaches to account for a change in the glacier geometry exist.

We agree. We emphasis it because the geometry change in this specific study is not counted. Thus, we need to mention it as a limitation of current study and a reason why we did not do future projection at this stage. But we will make this clear in the revised manuscript.

2. The unit used for the runoff is Gt/km2, which is quite uncommon in hydrology. A unit of mm/yr is much more common.

Gt is used in glacier mass balance studies such as van Pelt et al. (2012). We use Gt/km^2 to allow the comparison between different catchments with different area. We do agree that we should use mm/yr/km^2 if we are talking to the hydrological community.

3. The fact that glacierized catchments have a hydrological cycle with a peak discharge later in the year is well established in the literature. References should be added, and Fig. 11 can be removed.

But what we are trying to say is that there is a difference in peak discharge between catchments with different glacier coverage, but not that glacierized catchments have a hydrological cycle with a peak discharge later in the year. Anyhow, it will be removed as we have decided to change our research focus.

4. 176-183: the original issue and your computation are not clearly explained. Please better explain the problem.

Thanks for pointing out. We will better explain the problem here in the revised manuscript.

5. In Fig 5b the axes are reversed, but you analyze the results as if it was not the case… 336-339 (Fig. 7): it is not clear how significant these trends are. Analyses of the peak discharge (4.1) are a bit out of scope here and somehow reinventing the wheel…

The y-axe is reversed because we want to emphasis the magnitude of melt, which has a negative sign. But we will significantly re-write Sect. 3 and 4 and only present the runoff on glaciers.

The response to the comment of referee #2 on the figures are followed. **Due to the substantial re-framing of the manuscript some of these comments might not apply to the new manuscript**:

1: Scales are difficult to read. Try to make them larger and white. The insert with the whole country is very small and of poor quality, we do not see much. You also mention in the caption the "Conrad's continentality index" with no explanation nor reference.

We will improve Fig.1 and explain Conrad's continentality in the text.

2: difficult to see as quite small.

We will improve Fig.2.

3/4: it does not make much sense to compare the different RCPs for the past.

We will present the comparison in a different way in the revised manuscript.

5: the maps are too small, we cannot see the patterns. For panel (b), the y-axes are

reversed!

We will present the benchmark results in a different way. The y-axe is reversed because we want to emphasis the magnitude of melt, which has a negative sign.

6: the maps are too small

We will exclude the runoff results in the revised manuscript.

8/9: analyzing the time series of climate model outputs does not make sense. Also, plotting outputs for 2 catchments in the same figure makes it impossible to read.

Fig 8 and 9 will not be included in the manuscript.

10: mm/d is more frequent than m/day.

We use m/day to match the unit with the SMB, but we will change it to mm/d when comparing precipitation and snowfall.

11: should be removed.

It will be removed as we have decided to reframe our research focus.

Reference

Holthuijzen, M. F., Beckage, B., Clemins, P. J., Higdon, D., and Winter, J. M.: Constructing High-Resolution, Bias-Corrected Climate Products: A Comparison of Methods, Journal of Applied Meteorology and Climatology, 60, 455–475, https://doi.org/10.1175/JAMC-D-20-0252.1, 2021.

Huss, M. and Hock, R.: Global-scale hydrological response to future glacier mass loss, Nature Climate Change, 8, 135–140, https://doi.org/10.1038/s41558-017-0049-x, 2018.

Maraun, D.: Bias Correcting Climate Change Simulations - a Critical Review, Current Climate Change Reports, 2, 211–220, https://doi.org/10.1007/s40641-016-0050-x, 2016.

van Pelt, W., Oerlemans, J., Reijmer, C., Pohjola, V., Pettersson, R., and van Angelen, J.: Simulating melt, runoff and refreezing on Nordenskiöldbreen, Svalbard, using a coupled snow and energy balance model, The Cryosphere, 6, 641–659, https://doi.org/10.5194/tc-6-641-2012, 2012.

Portoghese, I., Bruno, E., Guyennon, N., and Iacobellis, V.: Stochastic bias-correction of daily rainfall scenarios for hydrological applications, 11, 2497–2509, https://doi.org/10.5194/nhess-11-2497-2011, 2011.

Schwalm, C. R., Glendon, S., and Duffy, P.: RCP8.5 tracks cumulative CO2 emissions, Proceedings of the National Academy of Sciences, 117, 19656–19657, https://doi.org/10.1073/pnas.2007117117, 2020.

Yang, W., Andréasson, J., Phil Graham, L., Olsson, J., Rosberg, J., and Wetterhall, F.: Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies, Hydrology Research, 41, 211–229, https://doi.org/10.2166/nh.2010.004, 2010.