

Dear editor and referee #3,

We thank referee #3 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 #3 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 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 general comments (written in light blue and *italic* font) of referee #3 are presented below.

*1. 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.

We agree that it is important that bias correction methods ensure the consistency between the observed and modelled SMB, and yet we disagree that this should be achieved at a cost of non-physical bias correction methods (or “tricks” after the referee’s terminology). Local bias-correction procedures through massive calibration of model parameters should be only applied when the modeled climate has attained a generally realistic state, i.e., close to its observed state, over regional scales. Hence, one should be careful when choosing local bias-correction methods to avoid introducing model calibrations that mask large-scale model biases in favor of matching local observations on the scales of glaciers or valleys. This is why our first and foremost concern in this study is to identify and quantify regional biases that leak into our SMB model domains and how successfully or unsuccessfully they are corrected through existing bias correction techniques.

When citing Huss and Hock (2015) referee #3 seems to use bias-corrections and calibration procedures interchangeably with the downscaling procedures. We are unsure which of the three terms is really meant here, but the procedures both in Huss and Hock, (2015) and Compagno et al., (2021) are based on a heavy data modification and parameter tuning including bias-correction for air temperature input prior to the calibration procedure and further adjustments to the precipitation and air temperature input again during the calibration to match the modeled glacier-wide SMB with measurements. The bias-correction method is a so-called delta change (DC) approach (Hay et al., 2000), which cannot deal with covariance and variability of the weather variables (Yang et al., 2010). We are not criticizing this procedure as it seems to be a common practice in many existing glacier SMB studies (e.g., Huss and Hock, 2018; Zekollari et al., 2019; Frei et al., 2018; Compagno et al.,

2021), but emphasize that our goal in this study is to answer how appropriate these existing bias-corrected climate data are for glacier impact studies and what kind of uncertainties different bias-correction methods bring to our glacier mass evolution study. We will work to make the purpose of this study more transparent.

Regarding the conclusion drawn by Compagno et al., (2021) that the impact of the climate product used on the modelled future glacier evolution is very limited, we would like to point out that the resolution of their simulations is much coarser (monthly temporal resolution and a spatial resolution of several kilometers depending on the climate forcing) than the resolution we use in our study (100 meters). Their focus is on regional mass change in the future, whereas our aim is to provide high resolution model results for impact studies in glacier-covered mountain regions. These differences define dissimilar demands on the bias correction, calibration and evaluation methods that we are testing and presenting in this article.

Regarding downscaling, we have presented the downscaling model component, MicroMet, in Sect. 2.2.1. It interpolates coarser-resolution RCM and reanalysis outputs onto our model grid through a distance-dependent weighting function and adjusts the interpolated data to compensate for the topography mismatch using an air temperature lapse rate factor and a precipitation adjustment factor. These two factors are generated from long-term observational data and presented in Sect. 2.6; we will however reiterate this part in Sect. 2.2.1 to make it clearer.

Finally, our calibration procedure is presented in Sect. 2.6, where we have tuned the snowfall fraction scheme and melting snow albedo. The results of the calibration of the benchmark simulation (driven by NORA10) are presented in Fig. A1 (a) in Appendix A, and the modeled SMB from the benchmark simulation does match observations considering our 100m resolution. We did not correct the climate forcing for bias ourselves but instead used climate forcings that have already been bias-corrected through more sophisticated procedures beforehand and are openly accessible – the fact that will be detailed in the text to avoid any confusion during the next round of reviews.

2. 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):

o 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.

We agree with the referee that validation of CORDEX climate products against NORA10 should be done over a common period of 2000-2014, and we will update relevant sections of the manuscript accordingly. However, we have chosen to maintain our evaluation of the overall performance of different bias correction methods over the entire period of 2000-2020 in order to have more robust conclusions about potential long-term impacts of errors in bias-corrected climate products and departures between different RCP-model projections (see response to the 3rd comment) on the glacier SMB modeling and glacier impact studies in Norway. We believe it is not justified to limit such evaluation to 2000-2014, since *in situ* observations are available over a longer interval, and our aim is to address the origins and effects of departures between different climate model products (with different bias-corrections and RCP scenarios) in the most comprehensive manner.

3. 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.

The climate forcings under different scenarios do diverge from 2005 (e.g. Schwalm et al., 2020). We therefore feel that there is more to say about how fast and to which extent climate models respond to RCP forcings and what such departures between climate model products mean for regional glaciological studies. We have therefore included this analysis in our revised study plan.

We believe it is useful and even necessary to look at how much different scenarios cause the climate to diverge from the ‘real’ (=observed) present-day state in a control period, what impacts bias-correction have on this divergence (Yang et al., 2010) and what kind of uncertainty this will bring to future projections. We will clarify and discuss the choices of different bias-corrected data and scenarios more in the revised manual script.

Regarding the limited skill of the CORDEX data to reproduce year-to-year variability, we agree that we should build our study upon the analysis of statistical quantities such as multi-annual means, seasonal distribution, etc., in the CORDEX simulation chain. However, it is the year-to-year variability that activates feedback mechanisms that lead to a multidecadal glacier commitment to a continuous retreat or gradual stabilization; hence, our focus on the glacier SMB and impact studies necessitates an assessment of both errors arising from the lack of year-to-year variability and multiannual climatic parameters, both of which play significant roles in the long-term glacier behavior. The bias-correction methods used to correct daily NORA10 and CORDEX data are applied to the time series and are meant to preserve the variability described by different climatic conditions generated by RCM projections and maximize the utilization of RCM outputs to obtain more realistic input data for impact studies such as hydrological modeling (e.g. Yang et al., 2010). It is however unclear to which degree such variability can be preserved by the aforementioned bias correction methods, since a detailed quantification of their effects on climate products has not been performed until now – the knowledge gap that we would like to fill.

4. 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 GCMRCM couple is used.

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.

5. 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.

We agree with the referee that there is a difference between “evaluation” and “validation”, and we tried to be careful when using such terminologies in our manuscript. Our opinion is that if the outputs of models are compared to a comprehensive set of observations, we should use the word “validation”, e.g., validate the model results of our benchmark simulation against glacier-wide surface mass balance measurements. However, this process can be also termed “evaluation” when it comes to the relative skills of the model runs. We will leave it to the editor to decide which of the two terms we should prioritize in our study.

Definition of “model validation”: Model validation is the process by which model outputs are (systematically) compared to independent real-world observations to judge the quantitative and qualitative correspondence with reality (<https://www.sciencedirect.com/topics/earth-and-planetary-sciences/model-validation>).

The specific response to the specific and technical comments of referee #3 are followed. **Due to the substantial re-framing of the manual script some of these comments might not apply to the new manual script.**

• l. 16: 'the variables mentioned above': not entirely clear which variables this refers to.

We refer to the ‘drainage basin-wide air temperature, precipitation, and glacier-wide climate mass balance are then validated against observations’ mentioned in l. 14-15. We will make it more clear in the revised manual script.

• 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.

The response is presented under the 1st general comment.

We agree that it is important that bias correction methods ensure the consistency between the observed and modelled SMB, and yet we disagree that this should be achieved at a cost of non-physical bias correction methods (or “tricks” after the referee’s terminology). Local bias-correction procedures through massive calibration of model parameters should be only applied when the modeled climate has attained a generally realistic state, i.e., close to its observed state, over regional scales. Hence, one should be careful when choosing local bias-correction methods to avoid introducing model calibrations that mask large-scale model biases in favor of matching local observations on the scales of glaciers or valleys. This is why our first and foremost concern in this study is to identify and quantify regional biases that leak into our SMB model domains and how successfully or unsuccessfully they are corrected through existing bias correction techniques.

When citing Huss and Hock (2015) referee #3 seems to use bias-corrections and calibration procedures interchangeably with the downscaling procedures. We are unsure which of the three terms is really meant here, but the procedures both in Huss and Hock, (2015) and Compagno et al., (2021) are based on a heavy data modification and parameter tuning including bias-correction for air temperature input prior to the calibration procedure and further adjustments to the precipitation and air temperature input again during the calibration to match the modeled glacier-wide SMB with measurements. The bias-correction method is a so-called delta change (DC) approach (Hay et al., 2000), which cannot deal with covariance and variability of the weather variables (Yang et al., 2010). We are not criticizing this procedure as it seems to be a common practice in many existing glacier SMB studies (e.g., Huss and Hock, 2018; Zekollari et al., 2019; Frei et al., 2018; Compagno et al., 2021), but emphasize that our goal in this study is to answer how appropriate these existing bias-corrected climate data are for glacier impact studies and what kind of uncertainties different bias-

correction methods bring to our glacier mass evolution study. We will work to make the purpose of this study more transparent.

Regarding the conclusion drawn by Compagno et al., (2021) that the impact of the climate product used on the modelled future glacier evolution is very limited, we would like to point out that the resolution of their simulations is much coarser (monthly temporal resolution and a spatial resolution of several kilometers depending on the climate forcing) than the resolution we use in our study (100 meters). Their focus is on regional mass change in the future, whereas our aim is to provide high resolution model results for impact studies in glacier-covered mountain regions. These differences define dissimilar demands on the bias correction, calibration and evaluation methods that we are testing and presenting in this article.

Regarding downscaling, we have presented the downscaling model component, MicroMet, in Sect. 2.2.1. It interpolates coarser-resolution RCM and reanalysis outputs onto our model grid through a distance-dependent weighting function and adjusts the interpolated data to compensate for the topography mismatch using an air temperature lapse rate factor and a precipitation adjustment factor. These two factors are generated from long-term observational data and presented in Sect. 2.6; we will however reiterate this part in Sect. 2.2.1 to make it clearer.

Finally, our calibration procedure is presented in Sect. 2.6, where we have tuned the snowfall fraction scheme and melting snow albedo. The results of the calibration of the benchmark simulation (driven by NORA10) are presented in Fig. A1 (a) in Appendix A, and the modeled SMB from the benchmark simulation does match observations considering our 100m resolution. We did not correct the climate forcing for bias ourselves but instead used climate forcings that have already been bias-corrected through more sophisticated procedures beforehand and are openly accessible – the fact that will be detailed in the text to avoid any confusion during the next round of reviews.

• *1.31-32: suggest adding a reference to the landmark paper by Immerzeel et al. (2020)*

We agree. We will add the reference to similar sentence in the revised manual script.

• *1.36-37: '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.*

For glacier-covered watersheds in western Norway, the runoff is projected to increase to 30-40% in 2021-2050 compared with 1961-1990 (Jóhannesson et al., 2012). And according to IPCC AR5, the 'peak meltwater' for glaciers in Norway and Iceland is projected to be in mid- to late- century (Cisneros et al., 2014). We will use these references in Sect.1 in the revised manual script.

• *1.42: global trend in glacier retreat. Here makes sense to refer to recent study by Hugonnet et al. (2021).*

Thanks for pointing out. We will use this reference in relevant texts if needed.

• *1.45-46: advance of glaciers and link to climatic conditions: possibly refer to study by Trachsel and Nesje (2015) here.*

Thanks for pointing out. We will use this reference in relevant texts if needed.

• *1. 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.*

Thanks for pointing out. We will use this reference in relevant texts if needed.

• *1.49: 'Regional Climate Models'*

We will correct the 'Reginal' to 'Regional'.

• *l.50: 'constraints from observations availability' à 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)*

We meant that the outputs must be constrained (validated) by observations via parameter tuning but not that the observation is a constraint. We will change the sentence from 'Their performance depends on constraints from observations ...' to 'Their performance depends on calibration...'

• *l.56: '...and glacier dynamics': true. But you do not include this in your study, do you?*

True. We will delete 'glacier dynamics'.

• *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?*

Agreed. We will change 'national scale' to 'local scale'.

• *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...*

Agreed. And see the response to the second major comment.

We agree with the referee that validation of CORDEX climate products against NORA10 should be done over a common period of 2000-2014, and we will update relevant sections of the manuscript accordingly. However, we have chosen to maintain our evaluation of the overall performance of different bias correction methods over the entire period of 2000-2020 in order to have more robust conclusions about potential long-term impacts of errors in bias-corrected climate products and departures between different RCP-model projections (see response to the 3rd comment) on the glacier SMB modeling and glacier impact studies in Norway. We believe it is not justified to limit such evaluation to 2000-2014, since *in situ* observations are available over a longer interval, and our aim is to address the origins and effects of departures between different climate model products (with different bias-corrections and RCP scenarios) in the most comprehensive manner.

• *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?*

'Full surface energy balance' refers to the full energy balance equation used in the model, which is presented in Sect. 2.2.2.

• *l. 107: five submodules à four submodules? (or maybe I am missing one..)*

Thanks for pointing out. It is four submodules not five.

• *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)*

Yes, the glacier extent is fixed. But the glacier thickness will decrease when the snow on glacial ice disappears, and energy is available for melting in eq. 1. We will improve the description of the submodules and how we calculated the climate mass balance in Sect. 2.2.

• *l. 130: sublimation. Must be very limited here? If not, could this be quantified? (e.g. vs. melt)*

The model takes into account mass loss due to sublimation. It might be limited or negligible in a regional scale study. But we will not quantify sublimation in this study as it is not our research focus here. We are just simply describing what the model can do.

• *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)

Yes, we agree that the glacier in Norway has retreated in the past 20 years. In our study, our model resolution is 100m. Most of the glaciers in west Norway retreated a few hundreds of meters in 2000-2018 (the longest retreat is 838m; Andreassen et al., 2020). We consider the disappearance of ice on these limited amount of grid cells not crucial for our model period. But it will become crucial when doing multi-decadal or century scale projections, which is what Laurent et al. (2020) and Muelchi et al. (2021) have done in their studies. We have discussed this as another constraint for achieving better future projection in Sect. 5. Glacier retreat will be counted in when doing future projections in a later stage. In this study, we focus solely on quantifying the uncertainties brought by bias-corrected climate forcing.

• 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.

We did not correct the raw CORDEX data ourselves. We will make this clear in the revised manual script.

• 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...

We have explained the reason in the response to the referee’s 4th general comments above. We will make clear why we choose this CORDEX data in the revised manual script.

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.

• 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.

We thank the referee for this comment.

• 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?

The calibration is done via tuning the snowfall fraction scheme and melt snow albedo (Sect. 2.6). We did not modify the climate forcing themselves to achieve the best agreement with the observation as

what has been done in Huss and Hock, (2015) as our research focuses are different. But we will make this clear in the revised manual script.

• *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”?*

Thanks for pointing out. The validation was done in Feiccabrino et al (2012). We will change the sentence to ‘which were validated...’

• *l.239: why not compare over the same time period?*

We have explained the reason in the response to the referee’s 2nd general comments above.

We agree with the referee that validation of CORDEX climate products against NORA10 should be done over a common period of 2000-2014, and we will update relevant sections of the manuscript accordingly. However, we have chosen to maintain our evaluation of the overall performance of different bias correction methods over the entire period of 2000-2020 in order to have more robust conclusions about potential long-term impacts of errors in bias-corrected climate products and departures between different RCP-model projections (see response to the 3rd comment) on the glacier SMB modeling and glacier impact studies in Norway. We believe it is not justified to limit such evaluation to 2000-2014, since *in situ* observations are available over a longer interval, and our aim is to address the origins and effects of departures between different climate model products (with different bias-corrections and RCP scenarios) in the most comprehensive manner.

• *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?*

We will use the same color scheme with the same extent for the same variable, i.e., temperature and precipitation, regardless of the season.

• *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.*

We have explained the reason in the response to the referee’s 3rd general comments above.

The climate forcings under different scenarios do diverge from 2005 (e.g. Schwalm et al., 2020), and it is interesting to demonstrate and attribute this phenomenon. We therefore feel that there is more to say about how fast and to which extent climate models respond to RCP forcings and what such departures between climate model products mean for regional glaciological studies. We have therefore included this analysis in our revised study plan.

We believe it is useful and even necessary to look at how much different scenarios cause the climate to diverge from the ‘real’ (=observed) present-day state in a control period, what impacts bias-correction have on this divergence (Yang et al., 2010) and what kind of uncertainty this will bring to future projections. We will clarify and discuss the choices of different bias-corrected data and scenarios more in the revised manual script.

• *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)*

What we mean here is that even though we did not include glacier retreat in our model study, this very negative mass balance captured by the simulation could indicate that Nigardsbreen could have retreated, which aligns with the observation. But we admit that this statement is misleading. Since we will reframe the manual script, this sentence might not be included in the revised version.

• 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)?

We realize that l. 303 should be rewritten. We meant that compared to 2000 the 5-year running mean in 2014 has increased. We will re-write Sect. 3.3. And this statement will most likely not be included.

Hugonnet et al. (2021) only provides the source data of mass change rate for the entire Scandinavia (-26.7 Gt yr⁻¹). But in l. 303 we are talking about the mass changes in individual drainage basin shown in Fig. 1, which could not be compared to Hugonnet et al. (2021) due to their coarser resolution. According to Fig.11 in Andreassen et al., 2020, the mass change rate of some glaciers in our study area are positive in certain years. We will plot the observations against our results of the benchmark simulation in the validation section, Sect. 3.1.

• When describing these variations, does NAO play a role? (Marzeion and Nesje, 2012; Trachsel and Nesje, 2015)

Yes, NAO influence the mass balance. And it has been discussed in other articles (e.g. Andreassen et al., 2020). But we cannot verify that in our study. And we will not focus on this particular aspect in the revised manual script neither.

• 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?

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.

Sect. 3.5 will be drastically re-written to compare some statistical quantities between the benchmark and CORDEX simulations and analysis the uncertainties.

• 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?

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.

• l. 383: good agreement. Not a negative bias because keep glaciers constant in time? See comment above (l. 298).

We thank the comment. We will address the limitations in our study in the revised manual script in more details.

• 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.

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

• I.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.

We have explained the reason in the response to the referee’s 4th general comments above. We will make clear why we choose this COREDEX data in the revised manual script.

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.

• I.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.*

We believe it is useful and even necessary to look at how much different scenarios cause the climate to diverge from the ‘real’ (=observed) present-day state in a control period, what impacts bias-correction have on this divergence (Yang et al., 2010) and what kind of uncertainty this will bring to future projections. We will clarify and discuss the choices of different bias-corrected data and scenarios more in the revised manual script.

• I.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.

Agreed. This will be done in a separate study in a later stage.

• I.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)*

Thanks for the suggestion. We will cite the work in relevant text in the revised manual script.

• I.479-480: *link with NAO?*

Yes, NAO influence the mass balance. And it has been discussed in other articles (e.g. Andreassen et al., 2020). But we cannot verify that in our study. And we will not focus on this particular aspect in the revised manual script neither.

• I.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. COREDEX.*

The entire Sect.4 will be drastically changed. 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 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.

The response to the comment of referee #3 on the figures are followed. **Due to the substantial re-framing of the manual script some of these comments might not apply to the new figures in the revised manual script:**

o 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.

We will improve the figures.

o 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 directly in a presentation for instance.

We will improve the figures according to the suggestions.

Reference

Andreassen, L. M., Elvehøy, H., Kjølmoen, B., and Belart, J. M. C.: Glacier change in Norway since the 1960s – an overview of mass balance, area, length and surface elevation changes, 1–16, <https://doi.org/10.1017/jog.2020.10>, 2020.

Cisneros, J., Taikan, O., Nigel, A., Gerardo, B., Graham, C., Petra, D., Tong, J., Shadrack, M., and Zbigniew, K.: Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, in: *Climate Change 2014: Impacts, Adaptation, and Vulnerability*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 229–269, 2014.

Compagno, L., Zekollari, H., Huss, M., and Farinotti, D.: Limited impact of climate forcing products on future glacier evolution in Scandinavia and Iceland, 67, 727–743, <https://doi.org/10.1017/jog.2021.24>, 2021.

Frei, P., Kotlarski, S., Liniger, M. A., and Schär, C.: Future snowfall in the Alps: projections based on the EURO-CORDEX regional climate models, 12, 1–24, <https://doi.org/10.5194/tc-12-1-2018>, 2018.

Hay, L. E., Wilby, R. L., and Leavesley, G. H.: A COMPARISON OF DELTA CHANGE AND DOWNSCALED GCM SCENARIOS FOR THREE MOUNTAINOUS BASINS IN THE UNITED STATES1, *JAWRA Journal of the American Water Resources Association*, 36, 387–397, <https://doi.org/10.1111/j.1752-1688.2000.tb04276.x>, 2000.

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.: A new model for global glacier change and sea-level rise, 3, <https://doi.org/10.3389/feart.2015.00054>, 2015.

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.

Jóhannesson, T., Aðalgeirsdóttir, G., Ahlstrøm, A., Andreassen, L., Beldring, S., Björnsson, H., Crochet, P., Einarsson, B., Elvehøy, H., Guðmundsson, S., Hock, R., Machguth, H., Melvold, K., Pálsson, F., Radić, V., Sigurðsson, O., and Thorsteinsson, T.: Hydropower, snow and ice, in: *Climate Change and Energy Systems Impacts, Risks and Adaptation in the Nordic and Baltic countries*, Nordic Council of Ministers, Copenhagen, Denmark, 227, 2012.

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., Pohjola, V., Pettersson, R., Marchenko, S., Kohler, J., Luks, B., Hagen, J. O., Schuler, T., Dunse, T., Noël, B., and Reijmer, C.: A long-term dataset of climatic mass balance, snow conditions, and runoff in Svalbard (1957–2018), 13, 2259–2280, <https://doi.org/10.5194/tc-13-2259-2019>, 2019.

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.

Zekollari, H., Huss, M., and Farinotti, D.: Modelling the future evolution of glaciers in the European Alps under the EURO-CORDEX RCM ensemble, 13, 1125–1146, <https://doi.org/10.5194/tc-13-1125-2019>, 2019.