Reviewer1:

This study presents the outcomes of bias-correcting and downscaling a climate model ensemble for hydrological applications in Norway. Three emission scenarios, ten GCM/RCM combinations, and two bias corrections were considered. First, the results are evaluated based on temperature and precipitation simulations in the historical and future periods. Then, the simulations and projections of a hydrological and a glacier model are analyzed. Overall, this study is very comprehensive and well presented. As it will probably serve as a basis for many climate and hydrological modeling applications in Norway, I believe it is relevant to GMD. However, I have a number of comments regarding some of the methodological choices, as well as the need to clarify some of the presented results. Since the list of comments is extensive and may require significant effort, I recommend making major revisions before accepting the study.

**Answer: we thank the reviewer for the valuable and constructive comments. Please see the detailed answers below.**

Major comments

- I understand that, for the sake of brevity, the authors chose to present only one scenario, the "intermediate" one. However, since the SSP scenario is one of the novelties of this project compared to the previous related project, it might be worthwhile to present only the results for this scenario or to add some of the key results for this scenario in the SI.

**Answer**: We are happy to inform you that we got permission to present our SSP3-7.0 results from the three RCM groups during the review process. As the raw EURO-CORDEX CMIP6 data has not been published yet, we added the key results for this scenario in the supplementary materials (Fig. S4-S19). In addition, we added new sections 9.2 and 9.3 in the discussion section to discuss the results of this scenario and the differences of results in the old and new national reports.

***"9.2 Comparison of results between the old and new national reports***

*The improved modelling chain generated updated climate and hydrological projections for Norway, which resulted in slightly different climate change signals and climate impacts compared to the analysis in the old national report. Under the RCP4.5 scenario, the projections for the old and new national reports agree on the direction of change, but CiN-2025 projections display a smaller increase in annual temperature (ensemble mean of 2.0 °C) and precipitation (ensemble mean of 6%) than the ensemble means in CiN-2015 (2.7 °C and 8% increase in temperature and precipitation respectively) at the end of the century. In addition, the ensemble spread in CiN-2025 is narrower than in CiN-2015, indicating more robust climate change signals. However, these differences are caused not only by the new selection of climate models, but also by the selection of the reference period. In CiN-2015, 1971–2000 was used as the reference period while in CiN-2015, it is 1991–2020. As temperature has already risen considerably in recent decades in Norway, annual mean temperature is higher in 1991–2020 than in 1971–2000, and the differences in temperature between the periods 2071–2100 and 1991–2020 are consequently more moderate than those between 2071–2100 and 1971–2000. In contrast, a larger increase in runoff is seen in CiN-2025 projections than in the previous one, mainly due to the improved evapotranspiration routine in the hydrological model (Huang et al., 2026).*

*Another major difference between the old and new national report is that CiN-2025 selected SSP3-7.0 as the high-emission scenario, which assumes lower emission than RCP8.5 used in CiN-2015. Under SSP3-7.0, the ensemble mean increases in annual mean temperature, precipitation and runoff are 3.4 °C, 11% and 10% in 2071-2100 relative to 1991-2020, respectively (Fig. S4, S5 and S16 in Supplementary materials). These increases are also smaller than the ones in 2071-2100 relative to 1971-2000 under the RCP8.5 scenario, shown in the old national report. Hence, users who have made computations based on the CiN-2015 projections, should notice these differences and justify whether their computations should be updated or not.*

***9.3 The effects of bias-adjustment methods under the high-emission scenario SSP3-7.0***

*This paper comprehensively compared the two bias-adjustment methods applied to EURO-CORDEX RCP4.5 simulations and found that the two methods can lead to considerable differences in seasonal changes and snow simulations under the moderate-emission scenario. However, the impact of the bias-adjustment methods may not only vary between climate models and future periods but also between emission scenarios. Thus, results for the same comparison between the two bias-adjustment methods under SSP3-7.0 are included in the Supplementary material (Figures S6-S19). Although the magnitudes of the projected changes between SSP3-7.0 and RCP4.5 differ, the general effects and differences between the two bias-adjustment methods are similar. One interesting aspect is a generally better agreement of the SSP3-7.0 ensemble mean temperature and precipitation with the observed values than in the RCP4.5 simulations during the reference period (Fig. S4).*

*For hydrological projections, 3DBC still provides better historical simulations than EQM under the SSP3-7.0 scenario, but the difference in future projections varies between the bias-adjustment methods and seasons. Considerable differences of ensemble mean changes in runoff are found in all seasons in the near future and in winter in the far future. The ensemble spreads of monthly projections for snow water equivalent and runoff are similar between 3DBC and EQM for almost all months, indicating that 3DBC does not help to reduce the projection uncertainty substantially under extremely warm conditions."*

- From the text (e.g., lines 68 and 375), it is often unclear that univariate correction does not bias or modify the dependence structure of the projections. On the contrary, as mentioned in line 372, univariate quantile mapping does not modify the simulated dependence of the uncorrected simulations; rather, it leaves it uncorrected. This point should be clarified to avoid confusing readers.
Additionally, it would be interesting to discuss whether univariate corrections might be sufficient if the climate models initially simulate this dependence well compared to observations, for instance, to simulate SWE (Figure 17 for EQM; probably the climate simulations resulting in the upper part of the ensemble, blue lines). Since multivariate correction degrades the preservation of the climate signal and impact modelers may use only a subset of the climate model ensemble, such a discussion could guide the selection of simulations and their interpretation.

**Answer**: Part 1: We have changed and added some text to try to make this clearer:

line 69-*70: "As univariate bias-adjustment methods, this approach does not modify inter-variable dependency structures but keeps them as in the original model data which can be inaccurate."*

*Line 368-370 (Line 419-422 in the track changes document): "The bias-adjusted climate projections based on the univariate EQM approach show the same dependency structures as the uncorrected RCM simulations. To impose inter-variable, temporal and spatial dependency structures obtained from the reference datasets, an additional post-processing step has been applied."*

Part2: Figure 17 shows the average results of the whole Norway, but the dependence simulated by each climate model varies in space and time. Since our users may be only interested in part of Norway, the direct guidance of model selection based on the country-average results can be misleading. Hence, we added a discussion on how users can select climate models and bias correction methods in section 9.4.

*Line 884-896 (Line 1007-1019 in the track changes document): "In principle, we suggest using the full ensemble projections with both bias-adjustment methods to account for the uncertainty of the whole modelling chain. But in practice, users may want to select a subset of climate models and one bias-adjustment method to reduce the computational cost of further applications. As the users may be only interested in parts of Norway and the performance of climate models and bias-adjustment methods vary in space and time, we are not able to give a straightforward suggestion on the subset of climate models and bias-adjustment methods based on the national analysis. However, the methodology as well as the analysis in this paper provides examples of selecting models and bias-adjustment methods. In order to select a subset of climate models, the users can analyze the climate signals for their study area and periods as in Fig. 3 and then select the models based on the study purpose, e.g., studies aiming to assess the driest and warmest climate conditions or the wettest and coldest conditions in the near or far future. Based on the selected models, the users can further assess the seasonal trends for their study area and periods using both EQM and 3DBC projections as in Fig. 6. If the trends are comparable between the two bias-adjustment methods, the 3DBC adjusted projections can be preferred, especially when the study is focused on seasonal changes and snow processes. Otherwise, we strongly recommend to use the projections adjusted by EQM and 3DBC to account for the uncertainty of bias-adjustment methods."*

- It is my understanding that bias correction is performed at the target resolution (1 km) rather than the native resolution (~0.11°). While I understand this choice, which was also made in similar projects such as CH 2018, using statistical bias correction techniques to make such a "jump" between scales can result in simulation artifacts, such as the overestimation of extremes and the overcorrection of the drizzle effect for area means (Maraun et al., 2013). While I don't think it's reasonable to make a different choice, I believe it's important for the authors to discuss this topic briefly in the discussion section. For instance, they could highlight that hypotheses about changes in processes below 0.11° resolution might not be valid and shouldn't be overinterpreted. One example is the changes in the spatial variability of rainfall extremes.

**Answer**: We agree that this should be discussed further. We have added a few sentences to section 9.1 which emphasize the importance of not ‘zooming’ too much to the results with 1 x 1 km resolution as the uncertainty generally increases with decreasing grid size:

Line 806-813 *(Line 927-934 in the track changes document)*: *“In this study, RCM outputs have firstly been interpolated to the resolution of the observations (1 km) using the nearest-neighbour method and then bias-adjusted on that resolution. However, users should not overinterpret projected changes which are finer than the native RCM resolution (~12.5 km) as the resolving power of the RCM sets the natural lower limit on which local-scale physical processes can be considered. Maraun (2013) has shown that the quantile mapping method cannot resolve this scale mismatch. This effect might be less important for temperature than precipitation as temperature usually has a much higher spatial coherence while precipitation is more subjected to small-scale variability. Although generally both the EQM and the 3DBC methods in our setup do not bring in climate change signals below the RCM resolution, some artifacts might be introduced which do not represent true local, small-scale climate changes.”*

- I found the choice of calibration and evaluation periods, as well as the explanations of the calibration method for the hydrological model, to be unclear. Although the model calibration seems to be the result of another study using a specific technique, I think more information is needed. For example, it is unclear if each spatial unit has a different set of parameters or if the calibration is "lumped." On line 413, it is unclear if 10 or 44 parameters are being calibrated and, again, if it is one set of parameters per gauge station or per 1 km cell. Also unclear is why the model was calibrated between 2000 and 2007 when data has been available since 1981 (Vali3 covers 1981–2014). For such a project, it would be good practice to calibrate on the "cold" period and evaluate on a warmer period to assess the temporal generalizability of the simulation, i.e., whether the model residuals depend on forcings projected to change. Finally, Figure 6a shows that the performance of Vali1 is similar to or higher than that of Cali, which is counterintuitive to me because many studies report degraded performance outside the calibration period (e.g., Guo et al., 2020). Could the input product be of higher quality after 2007?

**Answer**: thank you for the constructive suggestion. We added more information on the description. For the whole Norway, there is only one calibration parameter set, including 44 parameters. There are six major soil types in Norway classified in this project and each soil type is characterized by six parameters associated with subsurface processes. The snow parameters are also distinguished by three land use classes (deciduous forest, coniferous forest and others). In total, there are 44 parameters, including all subsurface, snow, glacier and lake parameters. The parameters vary between 1x1km grid cells due to different combinations of soil and land use types.

*Section 6.1 Line 465-474 (*Line 546-556 in the track changes document)*:*

*“Different from the lumped version of HBV, the parameters associated with snow and subsurface processes in distHBV vary by land cover type (deciduous forest, coniferous forest and others) and soil-type (the five soil types based on the sediment map (Section 2.3) plus glacier bed), respectively. In total, there are 44 parameters for modelling mainland of Norway, including six snow parameters (two snow parameters times three land use classes), 36 soil parameters (six soil types times six subsurface parameters) and two parameters associated with lake and glacier processes. The parameters vary between grid cells due to different combinations of soil and land cover types within grid cells. Note that we didn’t distinguish the snow parameters for all land cover types because it will increase equifinality*

*risks due to too many calibration parameters and forest is one of the dominant land cover types in Norway (Huang et al., 2026). In addition, we didn't calibrate the rainfall or snow correction parameters as in other HBV applications, because it will lead to inconsistency between the climate and hydrological projections in terms of water balance."*

For the validation period, we fully agree that the model should be tested in a warm period, and we changed the validation period from 2011 to 2020. In Norway, temperature tended to increase significantly in the recent 50 years (Yang and Huang, 2023) and the last 10 years was the warmest period for most catchments. Compared to the calibration period (2000-2007), the average increase in mean temperature of all 123 catchments is about 0.43 degree in 2011 - 2020. Hence, we validated our model for the last 10 years to show the model performance for warmer conditions.

During the new validation period, the model performance in terms of NSE is slightly degraded and there are more gauges with absolution BIAS larger than 0.1. The old and new validation results show that model performance varies with time. It may be partly due to parameter transferability problems under different climate conditions and partly due to input data quality that also varies with time. The number of temperature stations has been constantly increasing since 2000, but the number of precipitation data was reduced in the 2000s and increased again in the 2010s (Lussana, 2019). As a result, the observations used to generate the gridded seNorge data vary in recent years.

*Section 6.1 line 489-506 (*Line 572-592 in the track changes document*):*

*"The model was validated against the discharge of the 85 calibration stations and additional 38 gauging stations from 2011 to 2020 to evaluate the temporal and spatial transferability of the model, respectively. The validation period (2011 – 2020) was selected because it is the warmest period for most catchments in the recent decades. Compared to the calibration period (2000 – 2007), the average increase in annual mean temperature of all 123 catchments is about 0.43 degrees in 2011 – 2020. Hence, the validation results show the model performance under warmer conditions.*

*Figure 7 shows the calibration and validation results in terms of NSE and BIAS. During the calibration period, about 50% and 29% of the catchments show good (NSE>0.65 and |BIAS|<0.1) and satisfactory (0.65>NSE>0.55 and 0.1<|BIAS|<0.15) results (Moriasi et al., 2007), respectively. The model generally underestimates discharge with the median bias of -5%, mainly due to underestimation of precipitation in seNorge2018 v20.05 data. The model performs similarly in terms of NSE in the validation period for the 85 gauging stations, with the median NSE degraded by only 0.01. The median bias is reduced by 0.025 in the validation period than in the calibration period but there are more catchments with |BIAS|>0.1. The model performance varies with time, partly due to parameter transferability problems under different climate conditions and partly due to the quality of seNorge2018 v20.05 dataset that also varies with time (Lussana et al., 2019; 2020). The validation results for the additional 38 gauging stations show robust spatial transferability of the model, with good or satisfactory (NSE>0.55 and |BIAS|<0.15) model performance for about 58% of the catchments. The model generally underestimates discharge with BIAS less than -0.1 for about half of the validation gauging stations. Such calibration and validation results are acceptable with consideration of the quality of the meteorological forcing data in such a mountainous region and simultaneous calibration for all catchments."*

- Initially, it was unclear to me whether the glacier model was coupled with distHBV. I now understand that the models run independently. This should be made clearer early in the paper. Also, in areas where the glacier model is used, is the runoff evaluated? How does it compare to runoff simulated by distHBV?

**Answer**: we revised some sentences in Section 3 to clarify it. In addition, we added a new method section 6.3 to demonstrate the postprocessing method and compare the runoff simulations between distHBV and DEW in a new figure Fig.9.

Section 3, line 221-222 *(*Line 223-224 in the track changes document): *"A postprocessing procedure was carried out to combine the distHBV and DEW outputs to generate final runoff projections for mainland Norway."*

Section 6.3 Line 550-561 *(*Line 655-666 in the track changes document):

### *6.3 Postprocessing of distHBV and DEW outputs*

*The final runoff projections for mainland Norway were produced by replacing distHBV outputs with the DEW ones. Note that DEW simulated the whole glacierized catchments (Fig. 8a) but only the outputs for the grid cells with glaciers were used to replace the distHBV results. It is mainly because DEW uses a simpler potential evapotranspiration (PET) method and rougher landuse/soil classes than distHBV. Figure 9 (a and b) shows the simulated runoff projections using DEW and distHBV for the glacier region Svartisen (Fig. 8a) on 31st August 2100 driven by the ecearth-r12i1p1-cclm climate projection. Without considering glacier retreat, distHBV projected high runoff (> 20 mm) for most grid cells where glaciers exist at present while DEW projected much lower runoff for these grid cells than distHBV, confirming that distHBV overestimates runoff under warming conditions. After we replaced the distHBV results with the DEW ones, the final output is more reasonable for this region than the distHBV one (Fig. 9c). Note that there are still single grid cells with high runoff in the final product because of different glacier masks used by DEW and distHBV. In addition, small glaciers outside the glacierized catchments (Fig. 8a) were not simulated by the DEW model and the results for these small glaciers cannot be corrected.*

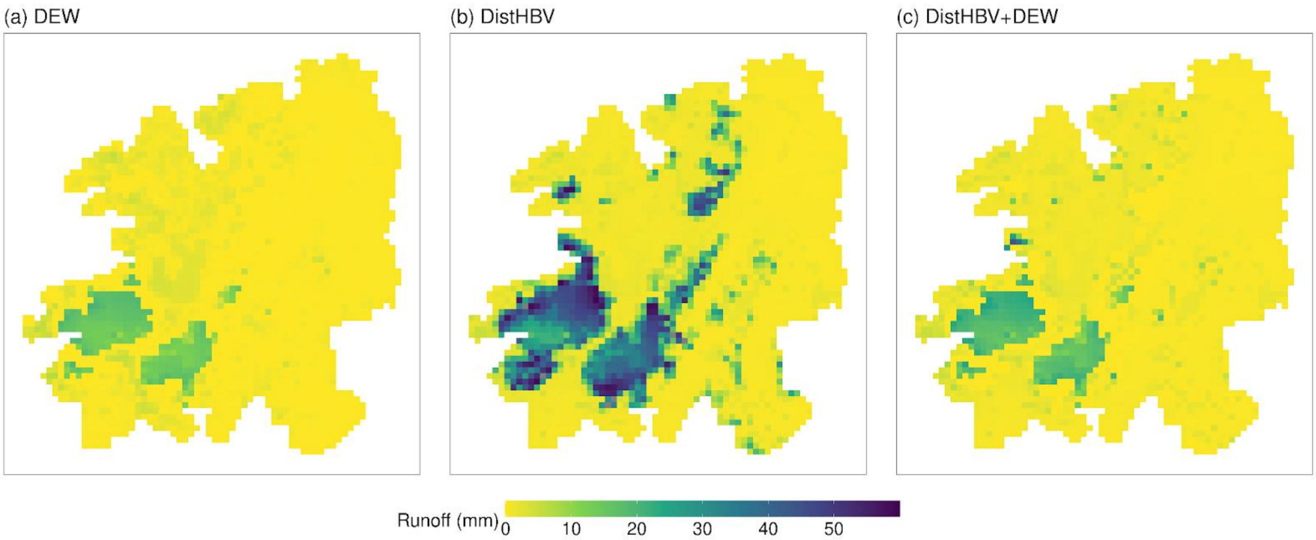


***Figure 9: Simulated runoff on 31st August 2100 for the glacier region Svartisen by DEW (a), distHBV (b) and the combination of distHBV and DEW (c) driven by the ecearth-r12i1p1-cclm climate projection.***

- I think it would be interesting to report on the methodology used for selecting the climate model. However, I think a few elements are missing from this section. For example, the second criterion is based on Table 6 of McSweeney et al. (2015). It should explicitly state that the table is used for model selection, not just methodology. More context should be provided for the last selection criterion. The term "visually striking" is vague and should be better defined. An example plot could be added to the supplementary information (SI). Overall, it is difficult to determine which models were excluded because of the second criterion and which were excluded because of the last criterion.

**Answer**: Thanks for pointing this out. We have revised the first part of section 4. Now, we explicitly state that Table 6 of McSweeney et al. (2015) is used for a GCM quality check, and we also explicitly state and provide a reason why model combinations were excluded. In addition, we revised Table 1 to show all 17 model combinations and which ones were excluded.

*Section 4, Line 246-268 (*Line 250-275 in the track changes document)*:*

*"Currently, the EURO-CORDEX CMIP5 projections comprise the largest high-resolution regional climate model ensemble for Europe and Norway with more than 30 simulations based on RCP2.6, more than 20 simulations based on RCP4.5 and more than 70 simulations based on RCP8.5. However, there are (only) 17 model combinations covering all three RCPs (Table 1). To be able to do a proper comparison between future projections of different RCPs, it is important to use identical model combinations for each RCP. These identical model combinations comprise five GCMs, namely CNRM_CM5, EC_EARTH, HadGEM2-ES, MPI-ESM-LR and NorESM1-M. Given time and computational constraints, we defined an upper limit of ten model combinations that are used as forcing data for the hydrological models, thus seven model combinations had to be excluded.*

***Table 1. Summary of the 17 GCM-RCM combinations available for RCP2.6, RCP4.5 and RCP8.5. Combinations in bold were selected for downscaling and bias-adjustment for the mainland of Norway. [1]: Original data has 360 days only. Additional days added. [2]: Leap-year days added. [3]: Spatial smoothing applied to tasmin, tasmax, tas and hurs.***

| ***Model combination name*** | ***GCM modelling institute*** | ***GCM*** | ***RCM modelling institute*** | ***RCM*** | ***Data coverage*** |
|---|---|---|---|---|---|
| ***cnrm-r1i1p1-aladin*** | ***CNRM-CERFACS*** | ***CNRM-CM5*** | ***CNRM*** | ***ALADIN63*** | ***1951–2100*** |
| ***ecearth-r12i1p1-rca[3]*** | ***ICHEC*** | ***EC-EARTH*** | ***SMHI*** | ***RCA4*** | ***1970–2100*** |
| ***ecearth-r12i1p1-cclm*** | ***ICHEC*** | ***EC-EARTH*** | ***BTU & KIT (CLMcom)*** | ***CCLM4-8-17*** | ***1950–2100*** |
| ***ecearth-r3i1p1-hirham[3]*** | ***ICHEC*** | ***EC-EARTH*** | ***DMI*** | ***HIRHAM5*** | ***1951–2100*** |
| ***hadgem-r1i1p1-rca[1,3]*** | ***MOHC*** | ***HadGEM2-ES*** | ***SMHI*** | ***RCA4*** | ***1970–2098*** |
| ***hadgem-r1i1p1-remo[1]*** | ***MOHC*** | ***HadGEM2-ES*** | ***GERICS*** | ***REMO2015*** | ***1950–2098*** |

| | | | | | |
|---|---|---|---|---|---|
| ***mpi-r1i1p1-cclm*** | ***MPI-M*** | ***MPI-ESM-LR*** | ***BTU (CLMcom)*** | ***CCLM4-8-17*** | ***1950–2100*** |
| ***mpi-r2i1p1-remo*** | ***MPI-M*** | ***MPI-ESM-LR*** | ***MPI-CSC*** | ***REMO2009*** | ***1951–2100*** |
| ***noresm-r1i1p1-rca***[2,3] | ***NCC*** | ***NorESM1-M*** | ***SMHI*** | ***RCA4*** | ***1970–2100*** |
| ***noresm-r1i1p1-remo*** | ***NCC*** | ***NorESM1-M*** | ***GERICS*** | ***REMO2015*** | ***1950–2100*** |
| *cnrm_r1i1p1_alaro* | *CNRM-CERFACS* | *CNRM-CM5* | *RMIB-UGent* | *ALARO-0* | *1950–2100* |
| *cnrm_r1i1p1_racmo* | *CNRM-CERFACS* | *CNRM-CM5* | *KNMI* | *RACMO22E* | *1950–2100* |
| *ecearth-r12i1p1_racmo* | *ICHEC* | *EC-EARTH* | *KNMI* | *RACMO22E* | *1950–2100* |
| *ecearth_r12i1p1_remo* | *ICHEC* | *EC-EARTH* | *GERICS* | *REMO2015* | *1950–2100* |
| *hadgem_r1i1p1_racmo*[1] | *MOHC* | *HadGEM2-ES* | *KNMI* | *RACMO22E* | *1950–2098* |
| *hadgem_r1i1p1_hirham*[1] | *MOHC* | *HadGEM2-ES* | *DMI* | *HIRHAM5* | *1951–2098* |
| *mpi_r1i1p1_remo* | *MPI-M* | *MPI-ESM-LR* | *MPI-CSC* | *REMO2009* | *1951–2100* |

*As a first quality check we used Table 6 in McSweeney et al. (2015) to see if the five GCMs perform satisfactorily in the representation of two out of the three physical phenomena consisting of i) annual temperature and precipitation cycles, ii) circulation and iii) storm tracks. This criterion did not lead to any exclusion of the 17 model combinations. As a next check we verify if the GCM-RCM combinations are ranked in the 'best half' for 24 variables and impact-based indices (Table 2 in Vautard et al., 2020) for the region of Scandinavia (Figure 12a in Vautard et al., 2020). This made us exclude cnrm _r1i1p1_alaro. Further, we excluded three simulations performed with the RCM RACMO22E which are affected by a bug in the snow albedo which again strongly affects the temperature signal above and around glaciers. This bug is documented in the EURO-CORDEX Errata table. Lastly, we checked the GCM-RCMs' performance with respect to the observed temperature and precipitation climate in Norway by using the seNorge v20.05 dataset as reference data. The largest precipitation biases (> 14 %) were found for the historical simulations with hadgem2_r1i1p1_hirham, mpi-r1i1p1_remo and ecearth-r12i1p1_remo, hence we excluded these simulations."*

- Quite often in the paper, descriptions of the main changes for certain variables are provided (e.g., Figures 8 and 9). While I find these results interesting, I'm not sure showing such results is the purpose of the paper. As I understand it, the paper presents the methodology and evaluation of the climate ensemble, the hydrological modeling setup, and the bias correction. In my opinion, replacing these parts of the analysis with more comparison plots that provide additional information about the climate model selection or the hydrological modeling setup would be more useful (see previous comments).

**Answer**: We agree with the reviewer but we also think it is necessary to give an overview of all projections before we compare the projections with different methods. Hence, we tried to keep these sections (7.1 and 8.1) as short as possible. Besides, we provided additional information as suggested. We revised Fig. 8 in hydrological modelling (Section 6) to better explain the model setup for glacier regions. We also added new figures (Fig. 9 and Fig. 18) in Section 6 and 8.2 to demonstrate the postprocessing procedure and additional comparison of the projections between the bias-adjustment methods.

Line 539-546 (Line 641-649 in the track changes document): *"Figure 8a shows one of the 12 glacier regions, called Svartisen, as an example. For this region, DEW was setup for all catchments where glacier melt contributes to river discharge. Among these catchments, only three catchments (Engabrevatn, Svartisdal and Berget) have discharge observations in good quality and only Engabrevatn has the measured mass balance data. Based on the data availability of both discharge and mass balance data, DEW was calibrated against the discharge of the three catchments and glacier mass balance in Engabrevatn for the period 1974–1993. The calibrated parameters were then transferred to other catchments of this region for hydrological projections. Figure 8b and 8c compare the observed and simulated discharge and mass balance for Engabrevatn in the calibration period. It shows that the model can well reproduce both monthly discharge and annual glacier mass balance with NSE larger than 0.7."*

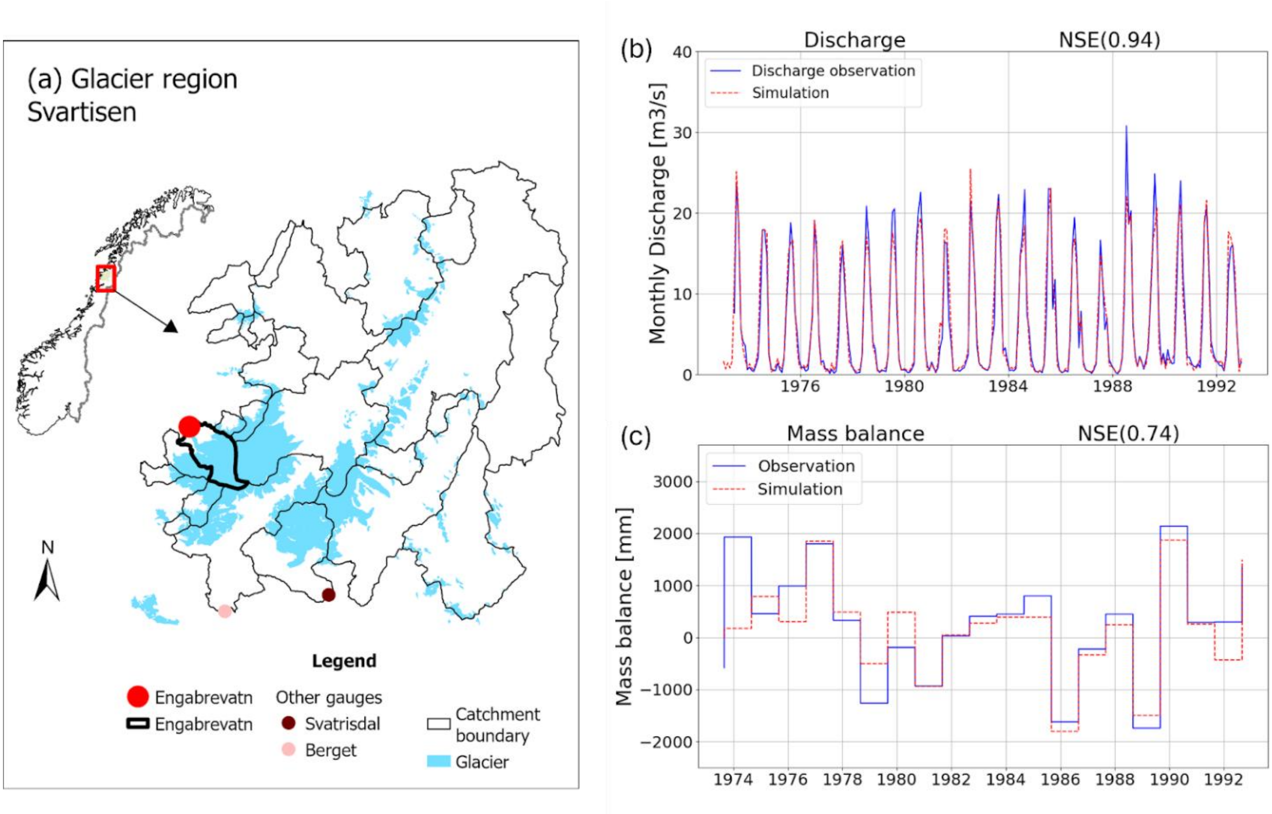


**Figure 8: The glacier region Svartisen (a), observed and simulated discharge (b) and annual mass balance (c) for the catchment Engabrevatn.**

Line 677-690 (Line 790-803 in the track changes document)*: "Figure 18 shows again the projected annual sum/mean of hydrological variables from 1971 to 2098 for mainland Norway, but separating the projections between the two bias-adjustment methods. The results show that the two bias-adjustment methods play a minor role on ensemble means as well as ensemble spread for runoff, evaporation and soil moisture, with the differences between the bias-adjustment methods less than 10 mm/year for runoff and evaporation and less than 1 mm for soil moisture. The ensemble mean of snow water equivalent using the 3DBC method has a better agreement with the results driven by observed forcing data than the ensemble mean using the EQM method, which always leads to underestimation of snow water equivalent in the historical period. In addition, the ensemble spread for snow water equivalent is narrower using 3DBC than EQM, especially before 2040, indicating lower uncertainty of projections using 3DBC. However, it is interesting to see that the snow water equivalent projections do not differ substantially after 2040 between the two bias-adjustment methods, probably due to less snow days in a warming climate. The minor impact of bias-adjustment methods on annual values also leads to similar spatial distributions of the changes in runoff, evaporation and soil moisture, but considerable differences of changes in snow water equivalent are found along the coast and northmost Norway between the bias-adjustment methods (Fig. S3 in Supplementary material)."*

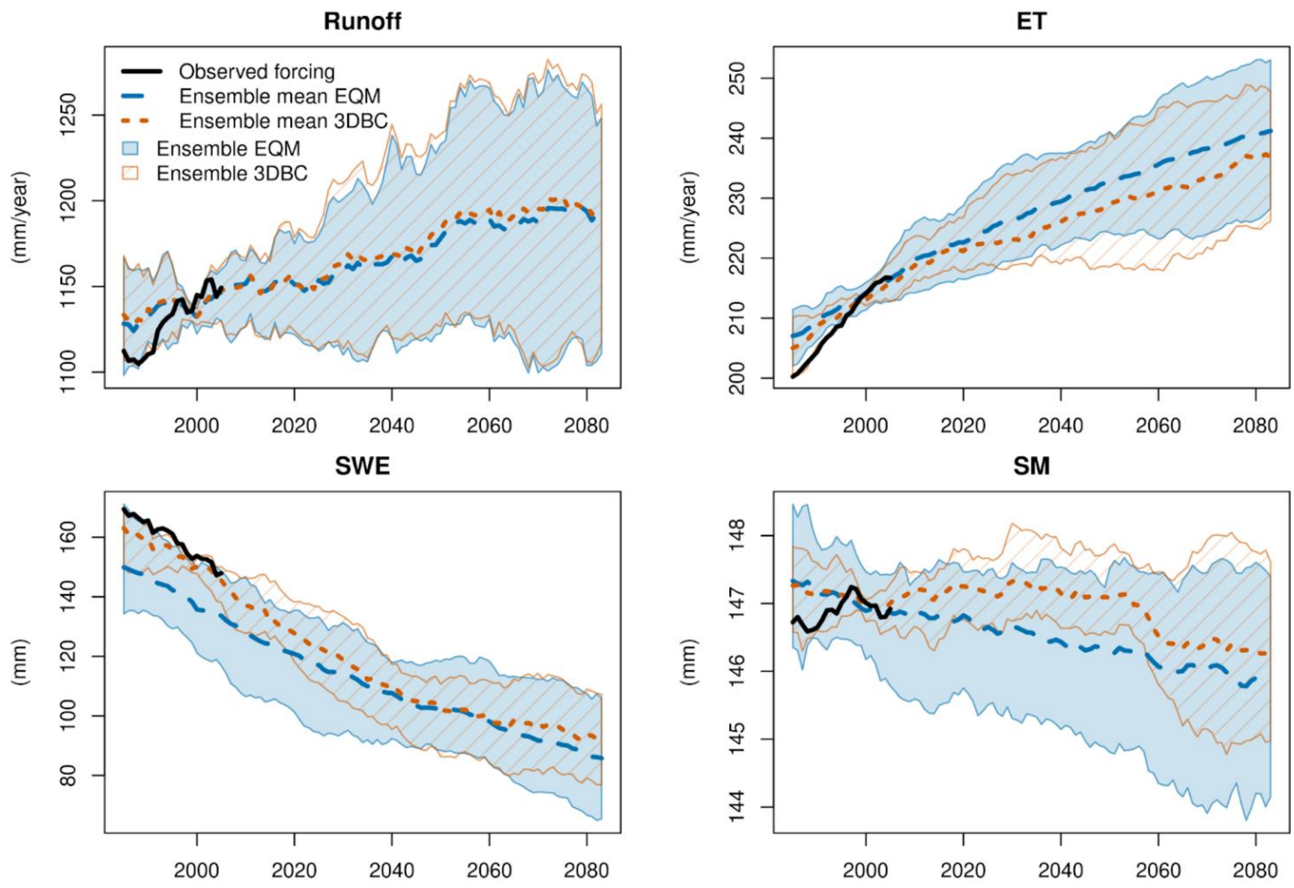


**Figure 18: The same as Fig. 16 but the projections using different bias-adjustment methods are separated.**

- I think a comparison of the results from the two bias correction methods would be more useful than figure 15, especially for the SWE results. This would provide more insight into the potential added value of multivariate correction over univariate correction for certain hydrological applications.

**Answer**: We have plotted a comparison of the results using the two bias-adjustment methods but we found that the differences are marginal, especially for runoff, ET and soil moisture, because they are changes in annual values. Hence, we didn't replace Fig. 15 with the new figures, but put the new figures in the Supplementary material (Fig. S3). We also add a description of the new figures in the first paragraph of Section 8.2 (see the answer above).

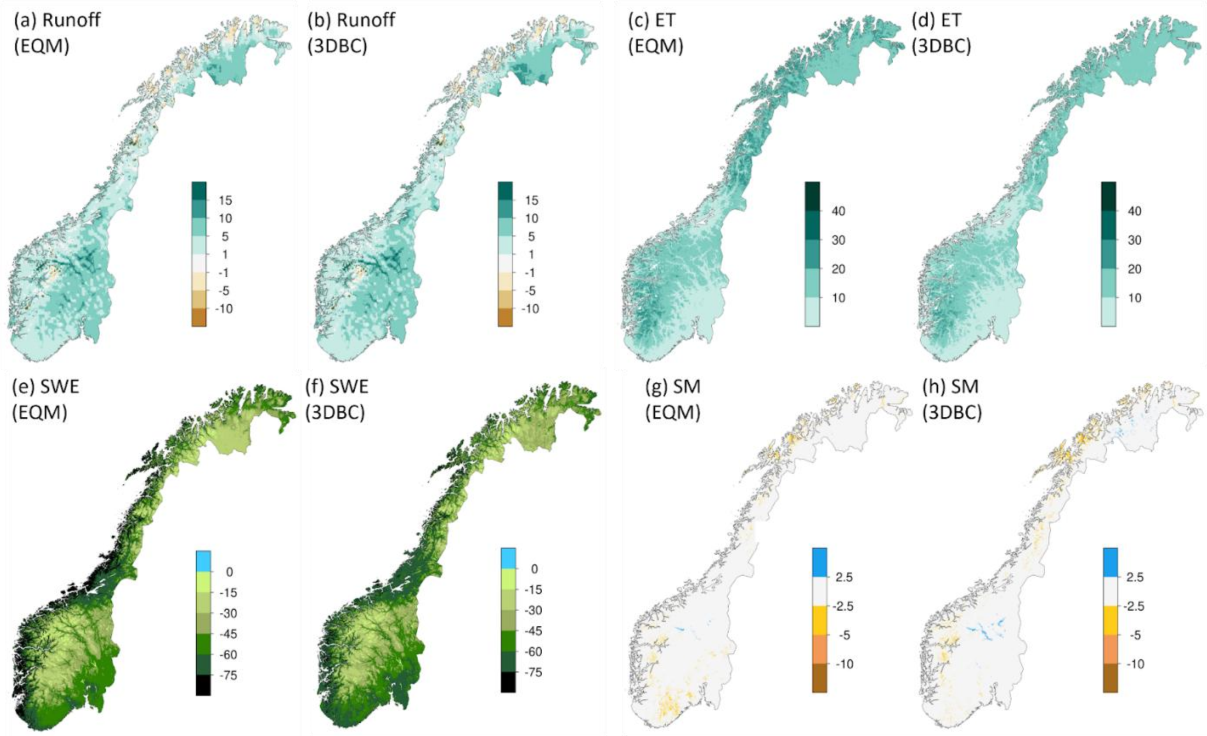


**Figure S3: ensemble mean changes (%) in annual runoff (a and b), evaporation (ET) (c and d), snow water equivalent (SWE) (e and f) and soil moisture (SM) (g and h) using two bias-adjustment methods in the scenario period 2071–2100 relative to the reference period 1991–2020 under the RCP4.5 scenario for mainland Norway.**

- For many applications, modelers may use only a subset of the provided projections. In light of the results presented in this analysis, I think a discussion about this is needed, as well as some guidance about good practices in that case. This would be very useful for impact modellers.

**Answer**: We added one paragraph in Section 9.4 on how users can select climate models and bias correction methods.

*Line 884-896 (*Line 1007-1019 in the track changes document)*:*

*"In principle, we suggest using the full ensemble projections with both bias-adjustment methods to account for the uncertainty of the whole modelling chain. But in practice, users may want to select a subset of climate models and one bias-adjustment method to reduce the computational cost of further applications. As the users may be only interested in parts of Norway and the performance of climate models and bias-adjustment methods vary in space and time, we are not able to give a*

*straightforward suggestion on the subset of climate models and bias-adjustment methods based on the national analysis. However, the methodology as well as the analysis in this paper provides examples of selecting models and bias-adjustment methods. In order to select a subset of climate models, the users can analyze the climate signals for their study area and periods as in Fig. 3 and then select the models based on the study purpose, e.g., studies aiming to assess the driest and warmest climate conditions or the wettest and coldest conditions in the near or far future. Based on the selected models, the users can further assess the seasonal trends for their study area and periods using both EQM and 3DBC projections as in Fig. 6. If the trends are comparable between the two bias-adjustment methods, the 3DBC adjusted projections can be preferred, especially when the study is focused on seasonal changes and snow processes. Otherwise, we strongly recommend to use the projections adjusted by EQM and 3DBC to account for the uncertainty of bias-adjustment methods."*

Minor comments

- L202: I think that a clear distinction between the period used for calibration of the bias correction and the period used for calibration of the hydrological model should be made explicitly somewhere.

**Answer**: We use the term 'training period' for bias-adjustment procedure (1985-2014), so that calibration period refers to the calibration of the hydrological model only.

- L277: I'm wondering whether this method might introduce an overestimation bias for precipitation which might then be corrected by the bias correction method. This could be avoided by considering that one day in the 360-day system is 24.3 hours. If this is too much work to update the simulations based on this, maybe just a quick test for a few cells might be useful to check the influence of artificially adding days in the projections.

**Answer**: Precipitation is given as flux, i.e. with units mm/s. To get the annual amount, this needs to be multiplied by 365(.25) * 24 * 3600, i.e. we need to shoot in extra days. Otherwise the annual amount would only be 360*24*3600*mean(daily). Whether this has been already considered in HadGEM, i.e. whether the conversion from daily amounts to fluxes assumes 24.3 h long days, is beyond our knowledge. However, as suggested by the reviewer we have tested the impact of shooting in extra days at a few grid cells from different parts of Norway. As expected, the effect is small relative to the existing biases in the original RCM data.

- L283: CH2018 is not properly referenced.

**Answer**: The report https://www.meteosvizzera.admin.ch/dam/jcr:6604da58-be19-4629-9dc0-a3d2352fbccb/CH2018_Technical_Report-compressed.pdf states that one should cite it like this (on page 3):

CH2018 (2018), CH2018 – Climate Scenarios for Switzerland, Technical Report, National Centre for Climate Services, Zurich, 271 pp. ISBN: 978-3-9525031-4-0

- L297: quickly explain which method was used for the wet-day correction.

**Answer**: To better describe the method used for wet-day correction, following sentences have been added to section 5.1.

Line 332-336 *(*Line 381-385 in the track changes document): *"Wet-day correction has also been applied prior to bias-adjustment of precipitation because RCMs generally provide more rainy days than the observed ones (Frei et al., 2003). For each grid cell, a threshold value is derived such that the wet-day frequency in modelled precipitation is equal to that in the corresponding reference data for the training period. All modelled precipitation values which are below the derived threshold value are then set to zero for both training and projection periods (Gudmundsson et al., 2012)."*

- L308: remove "which is assumed to be valid for use in the projection period" because it is mentioned earlier already.

**Answer**: We have removed it.

- L308: quickly explain why seasonal correction is often needed.

**Answer**: We have added some text on the monthly bias adjustments and why this is essential to produce realistic seasonal flow patterns and hydrological regimes in the subsequent hydrological modelling to section 5.1 first paragraph.

Line 322-324 *(*Line 371-373 in the track changes document)*: "With these monthly bias-adjustments we correct model biases which are varying throughout the year. This is essential for instance to produce realistic seasonal flow patterns and hydrological regimes in the subsequent hydrological modelling."*

- L317: explain why you apply this procedure -> for trend preservation?

**Answer**: Yes, it is done to better preserve the decadal trend. We have added this sentence to the last paragraph in section 5.1.

*Line 361-364 (*Line 410-413 in the track changes document)*:*

"The projection period starting from 2015 to 2100 was further divided into seven overlapping 30-year time slices. The first time slice, however, only covers 2015-2040, followed by 2021-2050, 2031-2060, etc. After the bias-adjustment of each time slice using the established monthly transfer functions, only the 10-year results in the middle of the period were being kept. **This procedure can better preserve the decadal trend**.*"*

- L337: Does this method affect temporal autocorrelation? (see François et al., 2020).

**Answer**: Yes, to make this clearer, we have added "temporal": "3DBC adjusts the ranks for future periods according to changes in the temporal auto-correlations". However, as also noted, in our implementation "adjustments in the variable auto-correlations for the future periods have a limited effect".

- L342: "ansatz" is a not very frequent term to describe "approach" -> I would use "approach" here

**Answer**: We use "approach" in the revised manuscript.

- Table 2: reporting performance results in a table makes the reading quite difficult in my opinion. I would replace by distributions, which would also allow to visualize more than the average over all grid cells.

**Answer**: Thanks for the constructive comment. A new boxplot, new Figure 4, is provided instead of Table 2 which shows the overall model ensemble results. For individual GCM-RCM results, we refer to the figure (Fig. S1) in Supplementary materials.

Line 398-407 *(*Line 450-461 in the track changes document):

*"IQD scores for precipitation and temperature are presented in Fig. 4. The results clearly demonstrate that both bias-adjustment approaches are far better at reproducing the full distributions of observed precipitation and temperature by several orders of magnitude than the original RCM outputs. As expected, the improvements are larger (smaller $d_{IQ}$) in the training period than the validation period. EQM and 3DBC have the same performance on annual results, but 3DBC generally performs better than EQM on seasonal results because 3DBC utilizes additional information about the intra-annual order of the observed time series in the post-processing. The only exceptions are the IQD scores from particular RCMs (CCLM4-8-17, REMO2015 and REMO2009), which show that EQM provides marginally better results than 3DBC in autumn (Fig. S1 in Supplementary materials). It might indicate that the observed ranks in autumn of the training and validation periods are quite different and that those models partly capture this change. Overall, 3DBC provides added value as compared to EQM when seasonal statistical properties are of importance."*

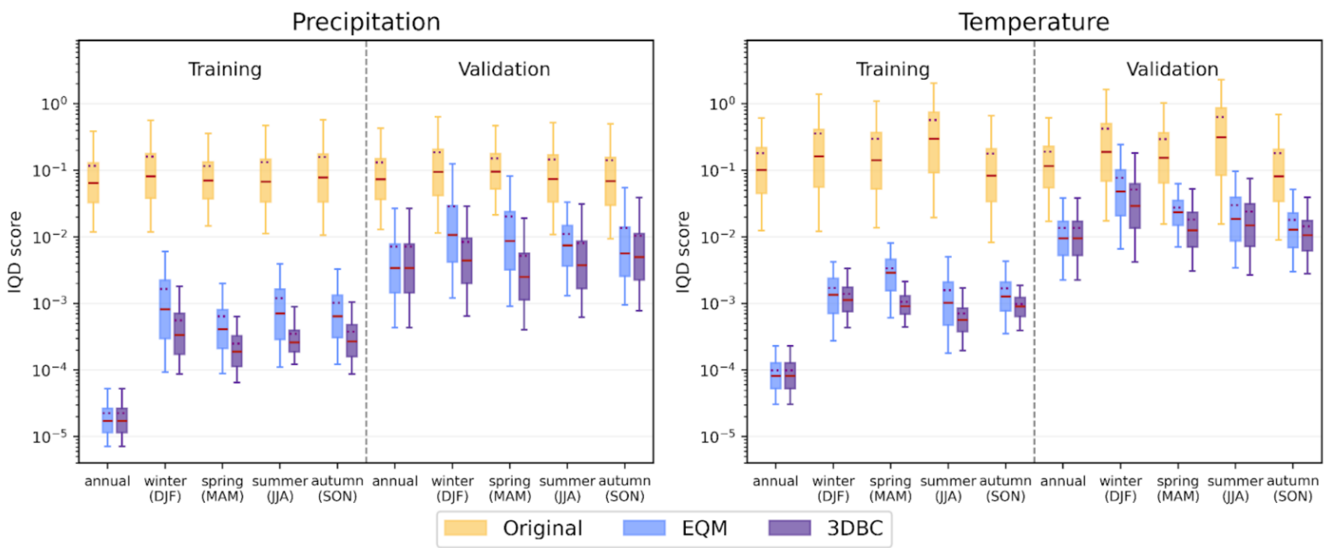


***Figure 4: Integrated quadratic distance (IQD) scores for precipitation (left panel) and temperature (right panel) based on the CMIP5 model ensemble. Bias-adjusted results from EQM and 3DBC in addition to the original model outputs are compared with the reference datasets seNorge2018 v20.05 over the training (1985–2014) and validation (1960/70–1984) periods. The red line on the box indicates the median value whilst the dotted line represents the mean. The lower and upper boundaries of the box are the 25th and 75th percentiles. The lower and upper ends of the whiskers refer to the 5th and 95th percentiles.***

- Fig 5: is there any dependence to elevation? This seems to be a common pattern found in recent studies (Matiu et al., 2024, Astagneau et al., 2025).

**Answer**: Thanks a lot for this constructive comment! The old Figure 5 (now Figure 6) has been updated presenting the climate change signals with respect to four elevation bands.

Line 434-450 *(*Line 501-511 in the track changes document):

*"The two bias-adjustment methods can lead to different climate change signals in the future periods (e.g. 2071-2100) relative to the reference period (1991-2020). Figure 6 shows the annual and seasonal changes grouped in four elevation bands (< 500, [500, 1000>, [1000,1500>, > 1500), including 52%, 32.7%, 13.6% and 1.7% of the grid cells in Norway, respectively. The result aligns with other recent studies demonstrating that the change signals are elevation-dependent (Astagneau et al.,2025, Matiu et al., 2024). The two bias-adjustment methods provide identical annual climate change signals, since 3DBC uses the same bias-adjusted results from EQM before further post-processing. EQM generally preserves the seasonal climate change signals from the original RCMs in terms of both mean and median changes as well as the spread of changes for all elevation bands. However, by reshuffling the chronological order intra-annually, 3DBC modifies the seasonal change signals from the original RCMs, leading to larger increases in precipitation and temperature in winter and smaller increases in spring than the original RCM outputs. In summer and autumn, the climate change signals can be underestimated or overestimated by 3DBC depending on the variables and elevations.*

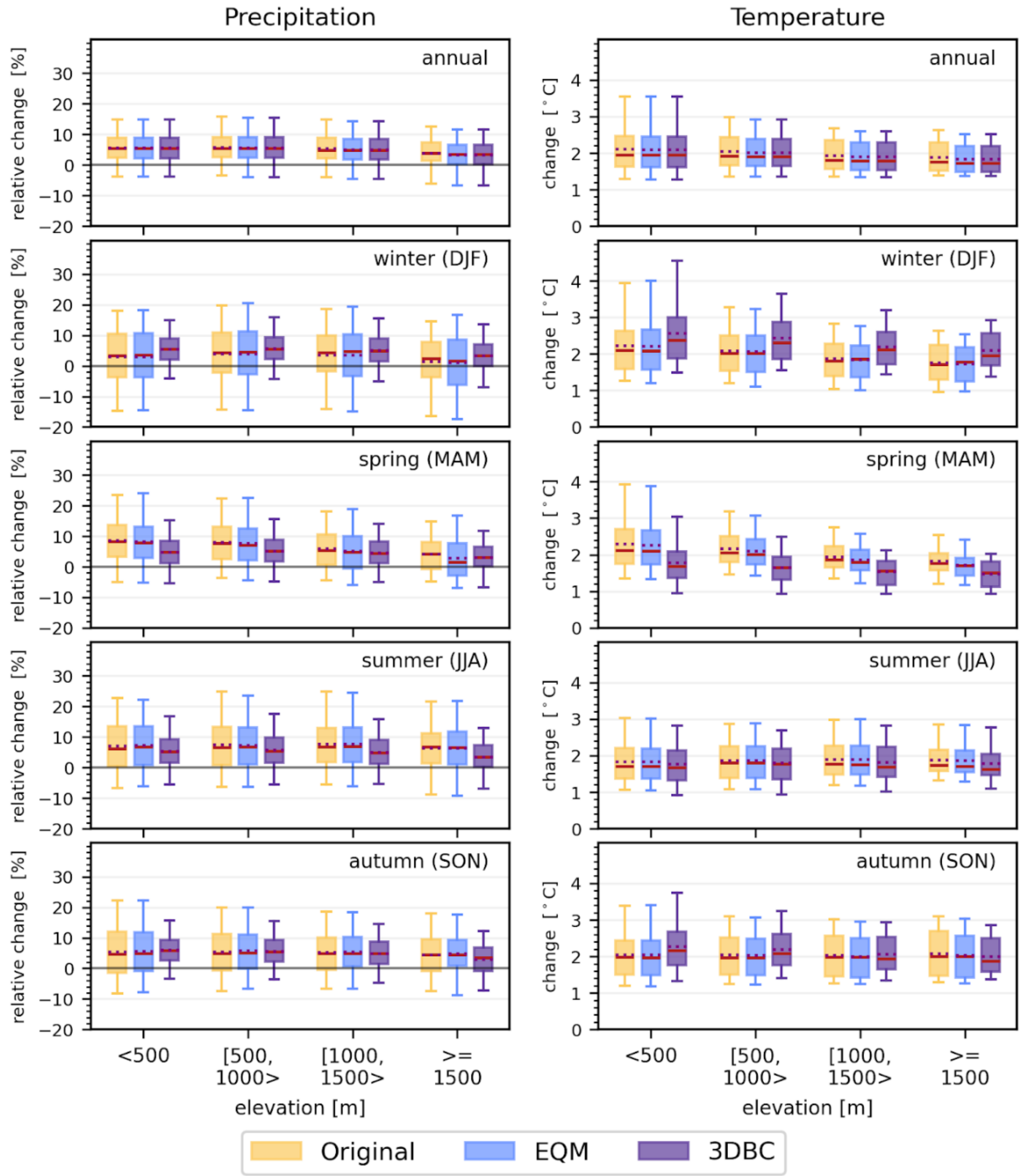


***Figure 6: Projected annual and seasonal changes in precipitation (relative change in %, left column) and temperature (change in °C, right column) from 1991–2020 to 2071–2100 for RCP4.5 in terms of elevation. Results grouped in four elevation bands from two bias-adjustment procedures, EQM and 3DBC, are compared to the original RCM projections. The red line on the box indicates the median value whilst the dotted line represents the mean. The lower and upper boundaries of the box are the 25th and 75th percentiles. The lower and upper ends of the whiskers refer to the 5th and 95th percentiles.***

- Table 3: The column "category" seems to be wrong: e.g. where is the snow routine category given that there is a snow melt temperature parameter? The resolution of this table is also not sharp.

**Answer**: We revised Table 3 (now it is Table 2).

**Table 2: list of calibration parameters. Note that the parameters associated with the snow and subsurface processes vary by land cover and soil type, respectively.**

| Associated process | Parameter | Explanation | Unit | Min | Max |
|---|---|---|---|---|---|
| Lake | KLAKE | Rating curve constant | - | 1.00E-04 | 0.1 |
| Snow and glacier | SMELT_T | Snow melt temperature | °C | -1 | 2 |
| | SMELTR | Temperature index for snow melt rate | m/°C | 1.00E-04 | 0.01 |
| | IMELTR | Ice melt rate for glaciers additional coefficient to SMELTR | - | 1 | 4 |
| Subsurface | FC | Field capacity | m | 1.00E-02 | 1 |
| | BETA | Shape coefficient of soil moisture | - | 1 | 5 |
| | KUZ | Upper zone recession coefficient | - | 1.00E-03 | 1 |
| | ALFA | Upper zone nonlinear drainage coefficient | - | 1 | 2 |
| | PERC | Percolation from upper zone to lower zone | - | 1.00E-03 | 0.5 |
| | KLZ | Lower recession coefficient | - | 1.00E-03 | 1 |

- L434: is there a calibrated snow correction factor in the hydrological model? This is a common parameter for HBV. If yes it would be interesting to reflect on this underestimation considering such a correction factor.

**Answer**: No, in this study, we didn't correct rainfall and snow data, because we wanted the simulated water balance to be strictly based on the input meteorological data. It allows consistent analysis between climatic and hydrological variables for the "Climate in Norway" report. We highlighted it in the manuscript.

*Line 472-474 (*Line 554-556 in the track changes document)*: "In addition, we didn't calibrate the rainfall or snow correction parameters as in other HBV applications, because it will lead to inconsistency between the climate and hydrological projections in terms of water balance."*

- L445: it is not clear what is used from DEW in terms of hydrological projections. For instance, for the Engabreen catchment, will the projected discharge be given by DEW or distHBV? Figure 2 does not provide this information but maybe I missed something in the text. It is also not clear from the data provided.

**Answer**: We added a new section 6.3 to show an example of a glacierized catchment and how we used the results from DEW for this catchment.

### *6.3 Postprocessing of distHBV and DEW outputs*

*The final runoff projections for mainland Norway were produced by replacing distHBV outputs with the DEW ones. Note that DEW simulated the whole glacierized catchments (Fig. 8a) but only the outputs for the grid cells with glaciers were used to replace the distHBV results. It is mainly because DEW uses a simpler potential evapotranspiration (PET) method and rougher landuse/soil classes than distHBV. Figure 9 (a and b) shows the simulated runoff projections using DEW and distHBV for the glacier region Svartisen (Fig. 8a) on 31st August 2100 driven by the ecearth-r12i1p1-cclm climate projection. Without considering glacier retreat, distHBV projected high runoff (> 20 mm) for most grid cells where glaciers exist at present while DEW projected much lower runoff for these grid cells than distHBV, confirming that distHBV overestimates runoff under warming conditions. After we replaced the distHBV results with the DEW ones, the final output is more reasonable for this region than the distHBV one (Fig. 9c). Note that there are still single grid cells with high runoff in the final product* *because of different glacier masks used by DEW and distHBV. In addition, small glaciers outside the glacierized catchments (Fig. 8a) were not simulated by the DEW model and the results for these small glaciers cannot be corrected.*

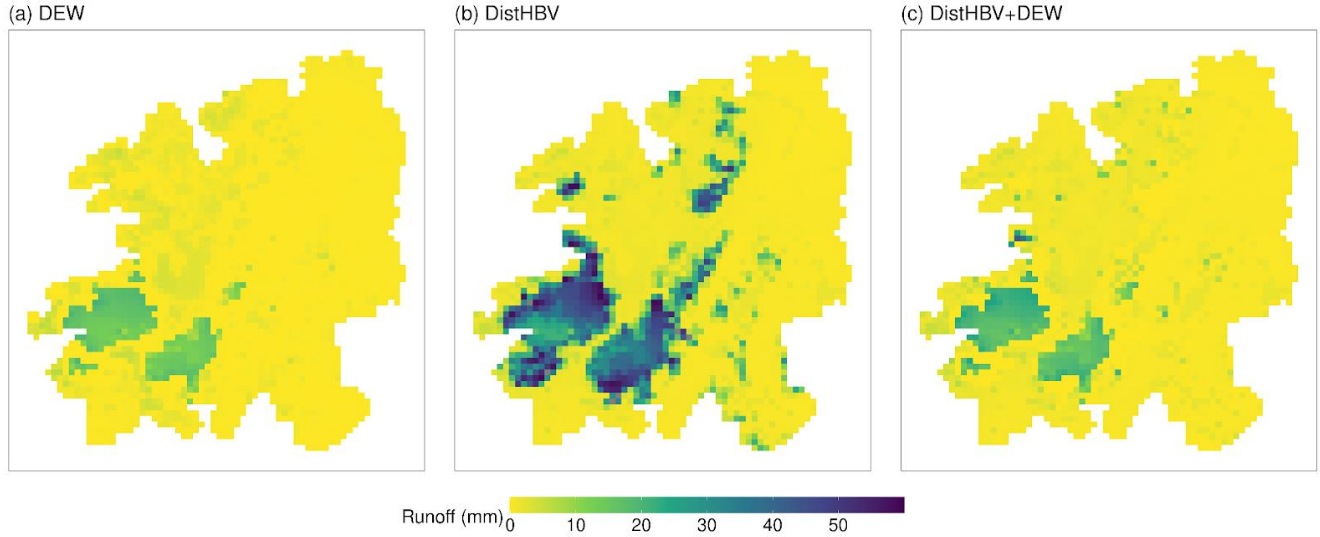


***Figure 9: Simulated runoff on 31st August 2100 for the glacier region Svartisen by DEW (a), distHBV (b) and the combination of distHBV and DEW (c) driven by the ecearth-r12i1p1-cclm climate projection.***

- Figure 8: it seems that the spread is reduced around the year 2000. I think this is because all models are calibrated independently and brought close to observations. I think something needs to be mentioned about this. It is also interesting to see that it is almost not the case for SWE and SM (Fig. 14).

**Answer**: Yes, the values are bias-adjusted such that they match the observed average in the training period 1985-2014. To make this more clear, we added a sentence noting this.

*Line 573-575 (*Line 680-681 in the track changes document): *"Note that all RCMs are bias-adjusted to match the observed values averaged over the training-period 1985-2014 and the spread in the ensemble equals zero for the middle of that period (year 2000)."*

*Line 655-656 (*Line 769-770 in the track changes document)*: "In addition, the spread of projections is reduced around the year 2000, as the bias-adjusted data matches the statistics of the observations better in the training period 1985-2014 than other periods."*

- 499: Why is EQM designed to preserve trends? Because of the detrending process and the sliding 30y window method? Since 3DBC acts as a postprocessor, why is the trend not preserved, at least for the marginals, not the dependence?

**Answer**: That is correct. The classical EQM method itself does not preserve trends. But the de-trended variant we have used in this study and the sliding 30-year window altogether contribute to better trend preservation. This feature has already been mentioned in the revised section 5.1 about EQM, the following sentences about 3DBC have been added to section 5.2:

*Line 381-385 (*Line 432-437 in the track changes document)*: "We thus adapted the 3DBC method to work within single years of the EQM data, an approach that maintains the climate change signals **and trends** from the RCMs (and EQM) on an annual scale. … **Since 3DBC reshuffles the bias-adjusted***

***times series resulting from EQM within a year, the marginal distributions at seasonal scale might be modified.****”*

- 7.3: I do not understand why this section appears after presenting the hydrological model. Change section order?

**Answer**: Section 7 includes all analysis on climate variables, including the uncertainty analysis for temperature and precipitation (7.3). Section 8 shows all analysis on hydrological variables.

- L525: why is bias correction uncertainty larger for precipitation? Because the initial climate bias is larger and therefore more difficult to correct? Or is this because of greater differences between correction techniques?

**Answer**: The uncertainty analysis quantifies the contributions to the uncertainty in the **changes**, not the actual values, i.e. the bias in the initial data does not influence this. The larger uncertainty is an effect of the two bias-adjustment methods resulting in different change signals. These differ more for precipitation than temperature, showing that results for temperature from a single bias-adjustment method are more robust than for precipitation. We have added a sentence on this to the manuscript in chapter 7.3.

*Line 628-630 (*Line 737-739 in the track changes document)*: “This is an effect of the two bias-adjustment methods resulting in seasonal change signals that differ more for precipitation than temperature, showing that results for temperature from a single bias-adjustment method are more robust than for precipitation. This is especially true for the near future.”*

- Figures 12, 13 and 19 are not sharp (image resolution).

**Answer**: Both the resolution and color of these figures (now they are Fig. 14, 15 and 22) have been improved.

- L547: why is this the case? A comment on this is needed.

Answer: The climate forcing data have been bias-adjusted to match the statistics of observations in the period 1985-2014, so the simulated hydrological variables using these forcing data show similar pattern of the simulated ones using the observed forcing.

*Line 655-657 (*Line 768-770 in the track changes document)*:*

*“In addition, the spread of projections is reduced around the year 2000, as the bias-adjusted data matches the statistics of the observations better in the training period 1985-2014 than other periods.”*

- L561: is this the only driver? Isn’t snow recharge changing antecedent conditions too?

**Answer**: Good point. It is indeed a driver.

*Line 669-670 (*Line 782-784 in the track changes document)*: “Soil moisture will decrease in most parts of the country due to the increase in evaporation and earlier snow melt, and moderate to strong decreases (<-5%) are mainly found in some southern areas and the coastal regions in the north.”*

- L683: the paper does not deal with this aspect -> I would remove.

**Answer**: We added a reference for this sentence.

“The Penman-Monteith method substantially improves evaporation estimates under climate scenarios by considering more climate variables and representing different land cover types (Huang et al., 2026),”

- L691: the link with the previous sentence is not obvious -> I would remove.

**Answer**: This sentence has been removed.

- L715: it was not clear to me whether the multivariate correction method is also used to correct these additional variables.

**Answer**: 3DBC method was used to further post-process all the atmospheric variables mentioned in this paper. We have added the following to line 227 *(*Line 230 in the track changes document)and 879 *(*Line 1002 in the track changes document)to make this clearer: “ ,**each bias-adjusted both with EQM and 3DBC** :”

Reviewer 2:

The study presents the method and some results of the new national climate projections for Norway. It uses a commonly employed setup for climate impact modelling in which emissions scenarios are used as input to global circulation models, which in turn are dynamically downscaled by regional climate model and bias-adjusted to eventually drive hydrological models in various catchments around Norway. Such modelling chains where run both for CMIP5-CORDEX and some available CMIP6-CORDEX data. All links in the modelling chain have undergone substantial developments since the last national assessment report in Norway and the authors specifically point out an additional multivariate bias-adjustment method, a more physical parameterization of the evapotranspiration by Penman-Monteith and an improved glacier model as new features. In the presentation of the results, the focus is put on bias evaluation before and after bias-adjustment in both a calibration and validation period, the differences in the results of the two bias-adjustment methods and various indices for climate change like annual cycles and 30-year moving means of hydroclimatic variables.

General comment

- The study nicely shows a typical example of a national hydroclimatic projection assessment modelling setup. It focusses on the modelling chain and particular elements of it while other aspects important for national climate projections – for e.g. the user perspective - are left untouched. This is a clear decision of the scope of the manuscript and in my point of view justified. The study mentions an important aspect that every similar study in a national context has to consider, i.e. the continuous development of the model chain from one assessment to the next one and the impact of the choices made in different modelling steps on the results. Those results are valuable to share with the scientific community. In general, the manuscript could benefit to be a bit more by embedding it better in existing literature.

  I recommend to publish the manuscript with major revisions, as some of the comments might require more work and might even affect the conclusions of the study.

  I enjoyed reading the article!

**Answer:** Thank you very much for your encouraging comments! Please see our answer to each comment below.

Major comments

- ANOVA analysis of the results:

  I very much appreciate the inclusion of an ANOVA to see how the variance in the climate change indices is decomposed into contributions from different effects. The setup is sound as all combinations of the modelling chain are available. However, since the sample sizes for the two effects differ so much and especially since one of the effects (BA) only has two samples, I would like to point out the issue of the biased variance estimator in ANOVA (see Dequé et al. 2007, section 3.3). In short, this issue leads to an underestimation of the variance and in case of only two samples, this underestimation is 50%, i.e. the actual variance calculated by an unbiased variance estimator would be double as large. Bosshard et al (2013) propose a subsampling procedure to albeit not removing, so at least to equalize the variance bias between the effects.

I suggest that the authors apply a similar subsampling scheme if they think this might be insightful, or at least discuss the issue in a way that the contribution of BA to the total variance is most likely larger than suggested by the ANOVA results.

**Answer:** Thank you very much for this comment! In line 516-517, we stated that "In this section, we analyse the contribution of these two uncertainty sources using the ANOVA method (Vetter et al., 2017)." The method used by Vetter et al. has already included the subsample scheme proposed by Bosshard et al. (2013) and it is well documented in Vetter et al. (2015) and Vetter et al. (2017). We didn't write about the ANOVA method used by Vetter et al. in detail because it was not the component of the modelling chain. Since the reviewer pointed this out, we wrote more details about this method.

*Line 614-618 (*Line 723-727 in the track changes document)*:*

*The ANOVA method provides not only variations in the impact on temperature/precipitation from these two major sources, but also their interaction term. To avoid the bias caused by different sample sizes of the sources, the ANOVA was implemented for a number of subsamples, each of which includes two climate models and two bias correction methods, and then the obtained estimates of subsamples were averaged. For more explanation of the method and equations, please refer to Vetter et al. (2017).*

- Also, I would like the authors to discuss their results in context of similar studies, either right when the results are presented or in the discussion. Have other studies also found such a large contribution of the BA effect for seasonal means?

**Answer:** Thanks a lot for the comment. We have included the following discussion on our results in the last paragraph in section 7.3.

*Line 638-643 (*Line 749-754 in the track changes document)*: "Our findings bring new insights into uncertainty attribution for seasonal projections, because most studies on uncertainty attribution are mainly targeted at annual values rather than seasonal ones, e.g. Paz & Willems (2022) and Lafferty & Sriver (2023). There are only a few uncertainty analyses for seasonal changes, but they did not find larger uncertainty associated with bias-adjustment methods than the variability within the model ensemble (Tong et al., 2021; Zhang et al., 2024). Our results highlight that bias-adjustment methods* *can be an important uncertainty source for seasonal projections and their seasonal effects should be considered in future studies."*

- Difference between the bias-adjustment methods:
  It sounds so appealing to use a multivariate BA-method that can reuse univariately bias-adjusted data and just correct the correlation in time and space! It would be great if the authors could discuss the limitations of the 3DBC method a bit more, just to help the reader to understand it better. I like the comparisons that are made throughout the manuscript. However, the probable reasons for the differences are discussed only very little. In particular, the bump around November in Fig. 11, bottom-right panel for the period 2041-2070, it would be very interesting to know how the authors reason why this happens and if they think that this could be problematic. Can it be related to the reference data and that they might have a high share of high precipitation days falling into this time of the year, and that this pattern, i.e. rank structure, is then imposed on all future periods? But why is it stronger in 2041-2070 than in 2071-2100 – I'm more used to gradual changes in the mean precipitation, meaning that the far future usually shows amplified changes with respect to the near future, except for small climate change signals that might be due to natural/internal variability. I'm not saying the results are wrong or impossible, but I would like to understand more about it and suggest that the authors discuss this and other differences more. For instance, it would be valuable for the reader to know if the modification of the climate change signal by 3DBC has been seen in other studies, too.

**Answer**: Thank you very much for this comment. The reviewer's hypothesis is correct. It is related to the reference data. In our implementation of 3DBC, we have used the rank structure from reference years 1961-1990 for 2041-2070 and the reference years 1991-2020 for 2071-2100. The mean annual cycle of the reference period is imprinted on the future period. And the autumn precipitation in the former normal period 1961-1990 is actually larger. So the larger bump in 2041-2070 than 2071-2100 originates from that.
The following sentences on the bump in Fig. 11 (Fig. 13 in the revised manuscript) are now included in the manuscript in chapter 7.2:

*Line 595-600 (*Line 701-707 in the track changes document)*:*
*For the near future (2041–2070), the difference in the changes from EQM and 3DBC are less systematic but 3DBC shows a pronounced increase in autumn precipitation which is absent in EQM (Fig, 13). This shift in the 3DBC results can be traced back to its implementation: the rank structure from the reference years 1961-1990 are used for the 2041-2070 period. Since the autumn precipitation in the period 1961-1990 has been large (Fig. 13a), this is imprinted on the mean annual cycle of the near future period. For the current climate (1991–2020), the 3DBC method results in climatologies that are similar for all models and thus a small ensemble-spread compared to the EQM data. This is especially true for precipitation (Fig. 13).*

We also added one line for the period 1961-1990 in Fig. 12 and 13.

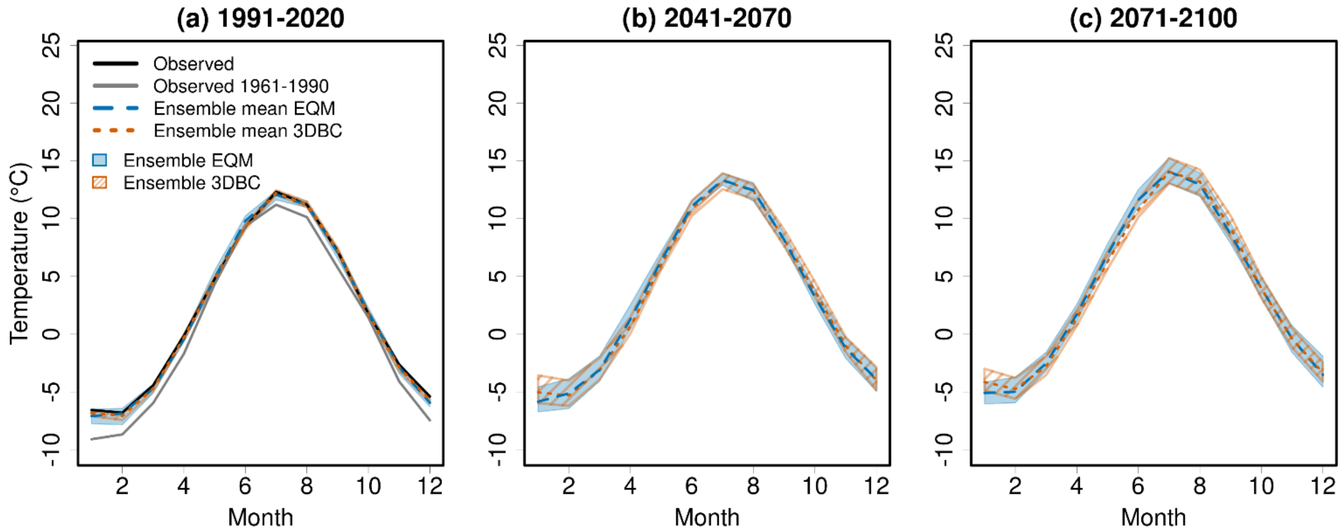


**Figure 12: 30-year mean monthly temperatures for Norway for different time periods using the EQM and 3DBC bias-adjusted climate projections under the RCP4.5 scenario. Black line: Observed temperature in 1991–2020. Grey line: Observed temperature in 1961-1990. Blue and orange lines: ensemble means of simulated temperature. Blue and orange striped areas: ensemble spread of 10 projections.**

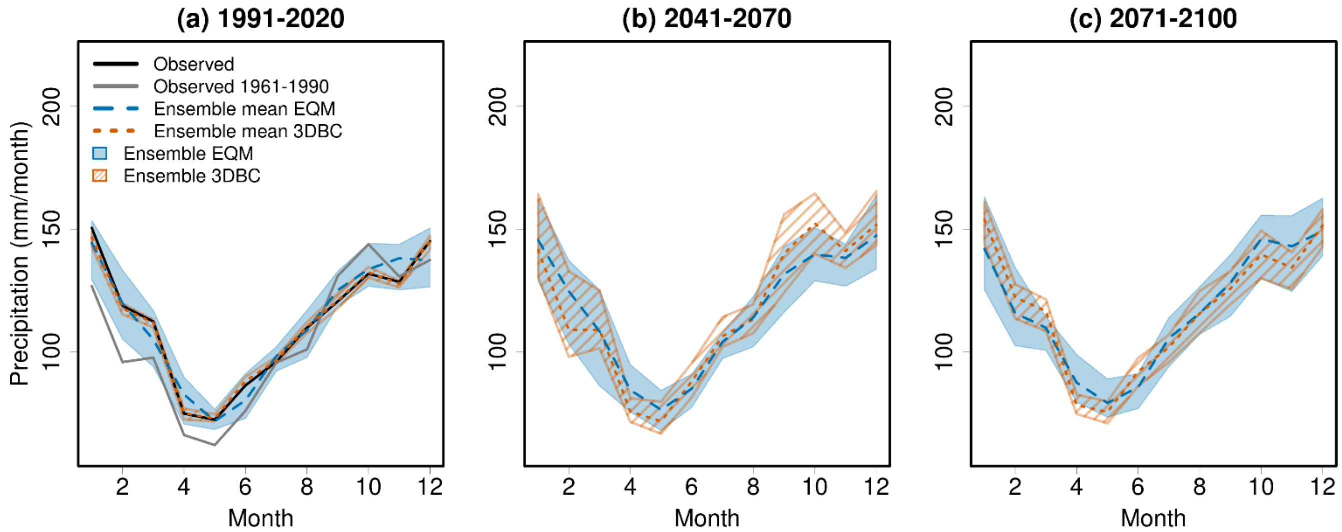


**Figure 13: The same as Fig. 12 but for precipitation.**

To our knowledge, only one study has used 3DBC in bias-adjustment (Nivron et al., 2024). The study compared the performance of six different BA methods, in which 3DBC was one of them, with their proposed machine learning model in terms of correcting daily maximum temperature and derived heatwave events. However, the study has not investigated the possible effects of 3DBC on climate change signals.

Nivron, O., Wischik, D.J., Vrac, M., Shuckburgh, E. and Archibald, A.T.: A temporal stochastic bias correction using a machine learning attention model. Environmental Data Science, 3: e36, 1–40, http://doi.org/10.1017/eds.2024.42, 2024.

- Specific comments

  Section 4: I could not find information about how the selection was done for the CMIP6 data and would ask the authors to include a description of it.

**Answer:** We added a paragraph on the selection of the CMIP6 downscalings. The selected models for CMIP6 are shown in Supplementary material.

Line 289-297 (Line 311-319 in the track changes document):
*"The selection criteria for the EURO-CORDEX CMIP6 projections are different from the ones for the EURO-CORDEX CMIP5 projections, because there were (only) 14 RCM simulations based on CMIP6 available by June 2024 (Table S1 in Supplementary material). Also, a selection of CMIP6 GCMs providing a satisfactory performance over Europe and covering a reasonable part of the climate change signal had already been carried out for the EURO-CORDEX CMIP6 simulations by Sobolowski et al. (2025). The main criteria for the selection here were thus to include as many GCMs as possible, a balanced RCM selection and excluding model combinations that show fairly similar results in temperature and precipitation over Norway. This led us to exclude the model combinations ecearthveg_r1i1p1f1_icon, ecearthveg_r1i1p1f1_racmo and miroc_r1i1p1f1_hclim. We further excluded noresm_r1i1p1f1_racmo due to a low climate sensitivity and small precipitation changes during the summer months."*

- L251: "2071-2100 minus 1991-2020" is not correct for the case of precipitation. Please adjust the text.

**Answer:** we have adjusted the text as below:

Line 270-271 (Line 277-278 in the track changes document): *"Projected changes in temperature and precipitation for mainland Norway by the end of the century relative to the reference period (1991-2020) under the RCP4.5 scenario."*

- Ll257-258: If it mattered at all, did you also have the requirement that a model combination had to have all variables available? If so, please state this in the list of criteria.

**Answer:** This did not really matter as we were interested in standard surface variables only. These are *pr, tas, tasmin, tasmax, sfcWind, rsds, rlds, hurs* and *ps* (Table 10.2.4 in Dyrrdal et., 2025; https://doi.org/10.60839/4rgq-nn84). All these variables are available (or can be calculated from the available variable in case of missing hurs) from the selected models.

- L 271: The given ranges of projected changes cannot really be compared to the one given on l 249, as the former is based on the 10 selected GCM-RCMs, while the latter is derived from the kernel-smoothed distribution. I suggest to have comparable metrics, either both taken from kernel-smoothed distribution or both taken directly from the ensemble of climate projections.

As it is now, it looks like the selection of an ensemble reduces the uncertainty range, while it does not look so when comparing the selected GCM-RCM points in Figure 3 to all available points.

**Answer:** Thanks a lot for this comment. We have removed the sentence on the kernel-smoothed distribution ranges and taken the comparable metrics directly the ensemble of climate projections. Note, we have rewritten the first part of section 4 to implement suggestions of reviewer 1.

*Line 275-277* (Line 296-298 in the track changes document)*: "Based on the selected ten GCM-RCM combinations (coloured dots in Fig. 3), the projected changes in temperature and precipitation in Norway range from 1.2 °C to 2.8 °C and from -1 % to 9 % in the future period 2071–2100 relative to the reference period 1991–2020."*

- Ll303-305: It would be interesting to learn which criteria the other non-selected methods failed on.

**Answer:** Thanks for the comment. We did not exclude other methods because they failed on any criteria but included EQM because it fulfills the criteria and is widely used. This is stated in the manuscript. However, we have added a few sentences to the paragraph right before section 5.1 to elaborate on the selected methods.

Line 311-317 (Line 338-344 in the track changes document): *"… In the end, the univariate bias-adjustment adopting de-trended empirical quantile mapping (EQM) approach (*Bürger et al., 2013*) was used to bias-adjust one climate variable at a time because the method meets all the aforementioned criteria and is widely used in adjusting climate model data.* ***EQM is effective in removing the model biases, preserving the trend and climate change signal moments (i.e. mean and standard deviation) and estimating extremes. As no univariate method can correct the possible biases in correlation among the atmospheric variables, all*** *the EQM results were further post-processed with the multivariate 3DBC approach (Mehrota and Sharma, 2019) to rectify inter-variable,* ***temporal*** *and spatial dependency structures."*

- Section 5.1: I think the description of the EQM method could be improved. It first reads like a standard EQM application, which has no dependency on future scenario periods but is only calibrated on historical model and reference data. Next, it is written that it nevertheless was applied in 30-year time slices. First after that, the reason for this is given, namely the handling of the mean trend. To me, it was confusing. First, I would have thought the daily data is used to calibrate EQM. But after the last section, it reads more like that the daily residuals are used. Are you doing a bias-adjustment on the residuals around the period means or on the daily data? And how do you handle the biases in the annual means, are those bias-adjusted as well?

Maybe, even a few concise mathematical equations could help to clarify the way EQM was applied.

**Answer:** Thanks very much for this constructive comment. In brief, it is indeed a de-trended variant of the quantile mapping method we have used. And it is correct that we have only bias-adjusted the daily residuals after removing trend (difference or relative difference in 30-year mean between the projection and training periods). After bias-adjustment, the trend is put back to the bias-adjusted residuals to obtain the proper bias-adjusted results. We have rewritten section 5.1 and introduced mathematical equations, which can help clarify how EQM was applied in this study.

Line 337-360 (Line 386-409 in the track changes document):

*To reduce the potential impact of over-adjustment (modifying the long-term linear trend) and extrapolation (model-projected values lying outside the range of the historical distribution), the long-term linear trend (usually 30-year) of the projected period was first removed from model projections. This shifting of the future distribution can better secure the applicability of the transfer function based on historical distribution. And the daily variability about the monthly mean remained unchanged. The trend was later reimposed after the bias-adjustment of the 'residuals'. For all the variables other than temperatures, trend removal and reimposition were performed multiplicatively. For example, relative trend for precipitation for month i,* $\delta P_i$*, is defined as:*

$$\delta P_i = \overline{P}_i^{prj} / \overline{P}_i^{tng} \quad (1)$$

*where* $\overline{P}_i^{prj}$ *and* $\overline{P}_i^{tng}$ *refer to mean monthly accumulated precipitation for month i for the projection and training periods, respectively. The de-trended (normalized) daily precipitation,* $\hat{P}_{ij}^{prj}$*, for month i and day number j in the projection period is:*

$$\hat{P}_{ij}^{prj} = P_{ij}^{prj} / \delta P_i \quad (2)$$

*where* $P_{ij}^{prj}$ *denotes the original RCM daily precipitation for month i and day number j in the projection period. EQM was then applied to the 'normalized' time series, and the precipitation trend for month i was then re-introduced to the bias-adjusted normalized results* $\breve{P}_{ij}^{prj}$ *for month i and day number j. The bias-adjusted precipitation,* $\tilde{P}_{ij}^{prj}$*, for month i and day number j can be obtained by:*

$$\tilde{P}_{ij}^{prj} = \breve{P}_{ij}^{prj} \cdot \delta P_i \quad (3)$$

*Similarly linear trend removal and reimposition for the projected values of temperature variables were done additively. Following similar notation, the temperature trend, $\delta T_i$, for month i simply equals to $\overline{T}_i^{prj} - \overline{T}_i^{tng}$, and the de-trended (residual) daily temperature, $\hat{T}_{ij}^{prj}$, can be derived from:*

$$\hat{T}_{ij}^{prj} = T_{ij}^{prj} - \delta T_i \qquad (4)$$

*where $T_{ij}^{prj}$ represents the original RCM daily temperature for month i and day number j in the projection period. The bias-adjusted temperature for month i and day number j, $\tilde{T}_{ij}^{prj}$, can be recovered by adding the temperature trend $\delta T_i$ to the bias-adjusted residual data $\check{T}_{ij}^{prj}$:*

$$\tilde{T}_{ij}^{prj} = \check{T}_{ij}^{prj} + \delta T_i \qquad (5)$$

- L315: I think I understand what 'constant linear extrapolation' means here. However, technically, it is not fully defined by only one point (1st or 99th quantile). Which slope did the constant linear extrapolation have?

**Answer:** Thanks very much for the comment. We have modified the sentence in the second paragraph in section 5.1 and further clarified how the extrapolation below 1st and beyond 99th quantiles were done.

Line 329-331 (Line 378-380 in the track changes document): *"For values smaller than the 1st quantile and larger than the 99th quantile, linear extrapolation was performed based on the slopes derived from the 1st and 2nd quantiles and from the 98th and 99th quantiles respectively."*

- L358: The average results might hide the spread between all the GCM-RCM combinations. Are some GCM-RCMs sticking out?

**Answer:** Thanks very much for the comment. The following sentences and Fig. S1 in Supplementary material have been added:

Line 401-406 (Line 454-460 in the track changes document)*: "EQM and 3DBC have the same performance on annual results, but 3DBC generally performs better than EQM on seasonal results because 3DBC utilizes additional information about the intra-annual order of the observed time series in the post-processing. The only exceptions are the IQD scores from particular RCMs (CCLM4-8-17, REMO2015, and REMO2009), which show that EQM provides marginally better results than 3DBC in autumn (Fig. S1 in Supplementary material). It might indicate that the observed ranks in autumn of the training and validation periods are quite different and that those models partly capture this change."*

- Table 2: In line with comment L358, please add GCM-RCM spread to the numbers given in the table. Also, since you chose to evaluation BA performance using IQD and rank-correlation, I suggest to ad the bias in rank-correlation to Table 2 as well. Figure 4 illustrates well the advantage of 3DBC. However, it only shows it for one GCM-RCM and one season. With numbers of the bias in rank-correlation for all seasons and all GCM-RCMs, one would get a more complete picture.

**Answer:** Thanks a lot for the constructive comments. In response to Reviewer 1, we have added a boxplot (new Figure 4) showing the ensemble spread of the IQD score, both annual and seasonal, instead of Table 2. We have also extended old Figure 4 (now Figure 5) to cover all four seasons as an example. In addition, a new figure showing the spread of ρ of each model combination can be found in the Supplementary materials (Fig. S2). And we have added the following comments in section 5.3.1:

Line 398-427 (Line 450-487 in the track changes document):

*IQD scores for precipitation and temperature are presented in Fig. 4. The results clearly demonstrate that both bias-adjustment approaches are far better at reproducing the full distributions of observed precipitation and temperature by several orders of magnitude than the original RCM outputs. As expected, the improvements are larger (smaller $d_{IQ}$) in the training period than the validation period. EQM and 3DBC have the same performance on annual results, but 3DBC generally performs better than EQM on seasonal results because 3DBC utilizes additional information about the intra-annual order of the observed time series in the post-processing. The only exceptions are the IQD scores from particular RCMs (CCLM4-8-17, REMO2015 and REMO2009), which show that EQM provides marginally better results than 3DBC in autumn (Fig. S1 in Supplementary materials). It might indicate that the observed ranks in autumn of the training and validation periods are quite different and that those models partly capture this change. Overall, 3DBC provides added value as compared to EQM when seasonal statistical properties are of importance.*

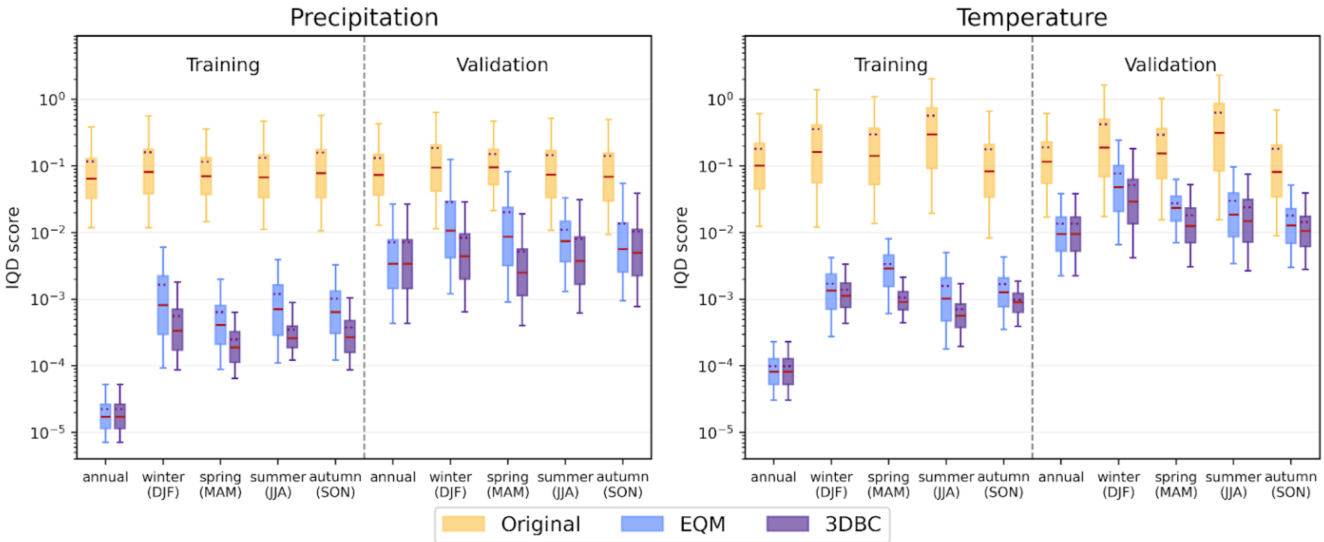


***Figure 4: Integrated quadratic distance (IQD) scores for precipitation (left panel) and temperature (right panel) based on the CMIP5 model ensemble. Bias-adjusted results from EQM and 3DBC in addition to the original model outputs are compared with the reference datasets seNorge2018 v20.05 over the training (1985–2014) and validation (1960/70–***

***1984) periods. The red line on the box indicates the median value whilst the dotted line represents the mean. The lower and upper boundaries of the box are the 25$^{th}$ and 75$^{th}$ percentiles. The lower and upper ends of the whiskers refer to the 5$^{th}$ and 95$^{th}$ percentiles.***

*Besides the seasonal statistics, 3DBC can simulate better spatial correlation structures between precipitation and temperature in the historical period than EQM, as it reorders the modelled ranks of precipitation and temperature based on observations while the univariate EQM method keeps the spatial rank correlation pattern from the RCM. Figure 5 shows an example of the spatial distribution of seasonal Spearman's rank correlation coefficient (ρ), calculated based on the bias-adjusted datasets from EQM and 3DBC and the reference datasets for training (1985–2014) and validation (1960–1984) periods for one RCM. In general, ρ are largest and positive in winter (warm days are wetter), followed by negative ρ in summer (warm days are dry, cold are wet). In spring and autumn, ρ is much smaller than in summer and winter, indicating a rather weak rank correlation between precipitation and temperature. The differences in the spatial correlation structure between these two methods are often most pronounced in winter and summer. EQM usually overestimates the positive rank correlations almost over the whole country in winter, whilst it underestimates negative dependencies in summer. And this spatial rank correlation pattern seems to be rather stable from one period to another. Aggregated results for each model combination are shown as boxplots (Fig. S2 in Supplementary materials), and they confirm that 3DBC performs better in recovering the inter-variable spatial dependency structure for all RCMs.*

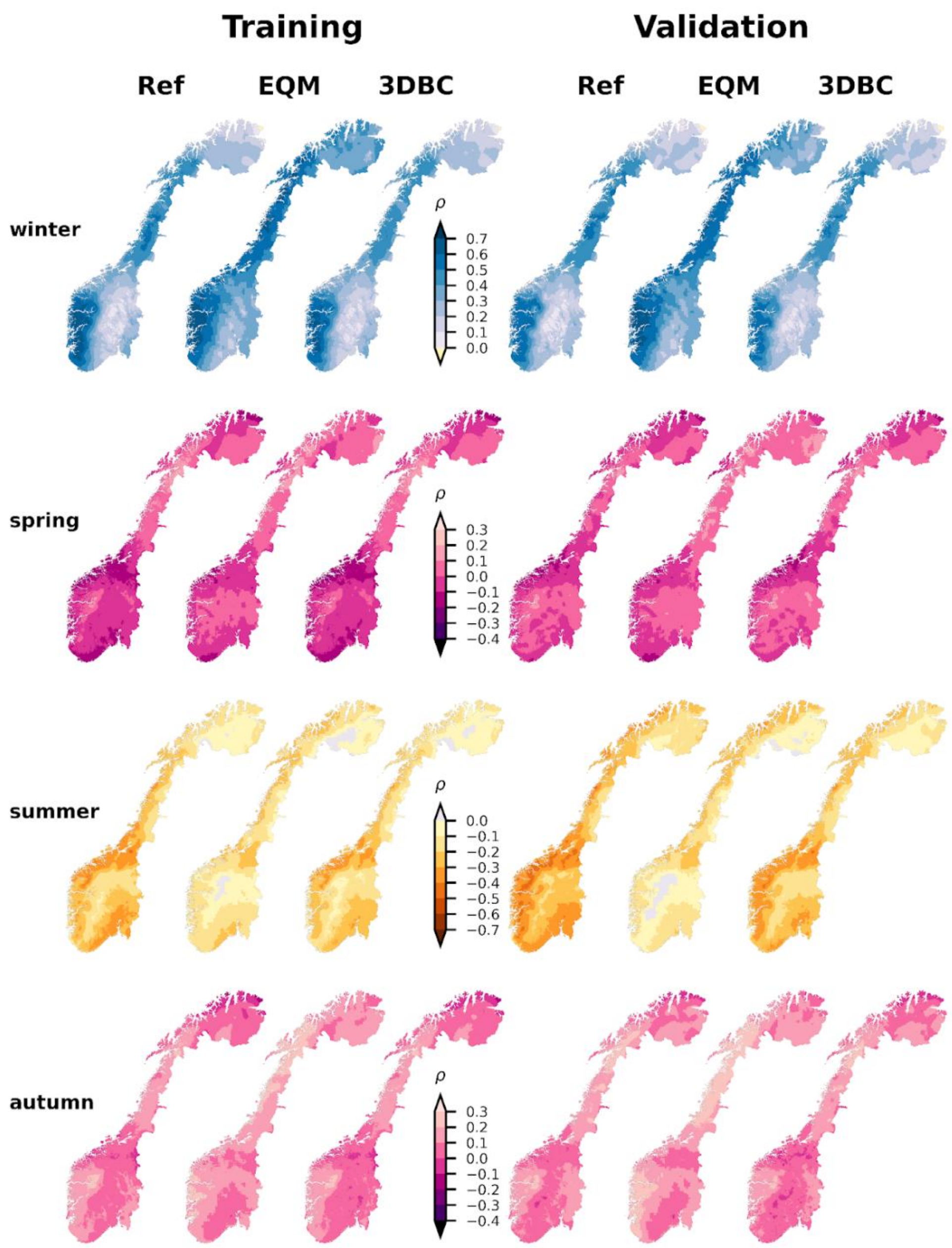


***Figure 5: Spatial distribution of Spearman's rank correlation coefficient ρ of daily precipitation and temperature in winter (DJF), spring (MAM), summer (JJA) and autumn (SON) for the two bias-adjustment methods. For training (1985–2014) and validation (1960–1984) periods, the two bias-adjusted datasets, EQM and 3DBC are based on historical run from CMIP5-based mpi-r1i1p1-cclm and compared with reference datasets seNorge2018 v20.05.***

- L472 and Figure 7: Please indicate the degree of the glacierization in the catchment, i.e. what percentage of the catchment area that is covered by the glacier. Also, it would be good to know which parts of the shown time series that belongs to the calibration and validation period.

**Answer:** We have revised Fig. 7 (Fig. 8 in the revised manuscript) according to the comment from reviewer 1. In the new figure, one can clearly see the degree of the glacierization in the Engabrevatn catchment (about 70% of the catchment is covered by glaciers). The time series belongs to the calibration period. For the validation period, there is only observed discharge available.

Line 539-546 (Line 641-648 in the track changes document):

*Figure 8a shows one of the 12 glacier regions, called Svartisen, as an example. For this region, DEW was setup for all catchments where glacier melt contributes to river discharge. Among these catchments, only three catchments (Engabrevatn, Svartisdal and Berget) have discharge observations in good quality and only Engabrevatn has the measured mass balance data. Based on the data availability of both discharge and mass balance data, DEW was calibrated against the discharge of the three catchments and glacier mass balance in Engabrevatn for the period 1974–1993. The calibrated parameters were then transferred to other catchments of this region for hydrological projections. Figure 8b and 8c compare the observed and simulated discharge and mass balance for Engabrevatn in the calibration period. It shows that the model can well reproduce both monthly discharge and annual glacier mass balance with NSE larger than 0.7.*

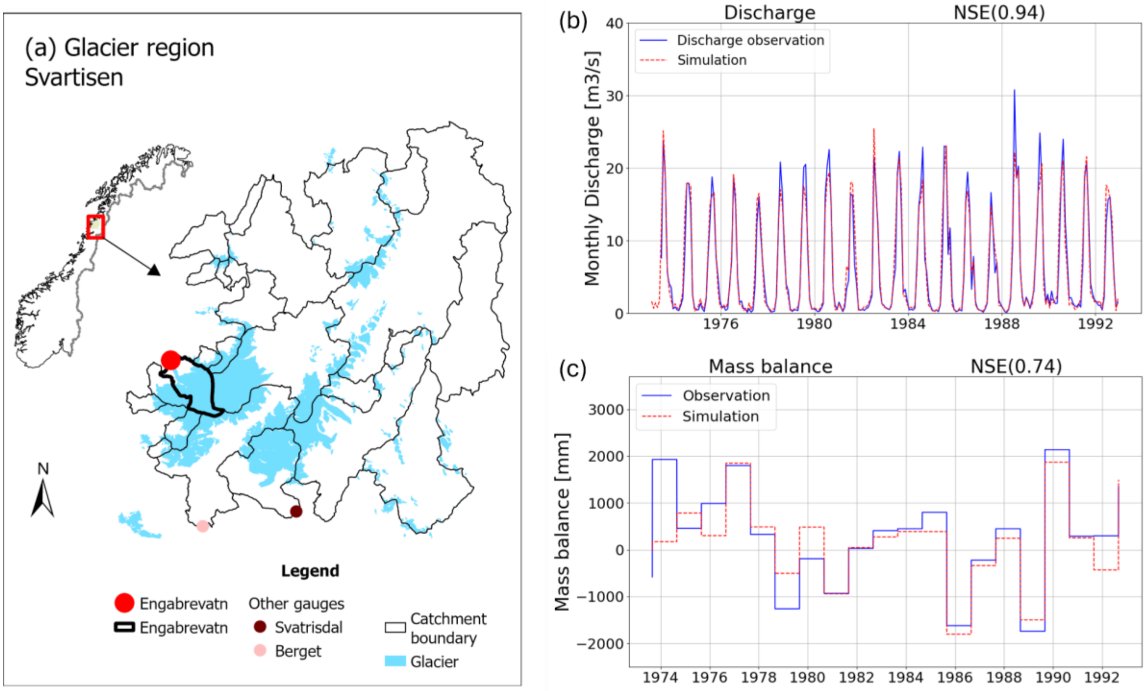


**Figure 8: The glacier region Svartisen (a), observed and simulated discharge (b) and annual mass balance (c) for the catchment Engabrevatn.**

- Figure 8: Is it correct to interpret that the ensemble spread would be exactly the same for the EQM and the 3DBC methods since the methods show equal spreads in annual change statistics (Fig. 5)? If so, I would suggest to add a note that BA method does not matter for the spread in the time series shown in Figure 8. If not, I would like to see the spread of the EQM and 3DBS sub-ensembles in Fig. 5 to see how their contribution to total spread changes over time.

**Answer:** Yes. The ensemble spread is exactly the same for both EQM and 3DBC methods as 3DBC has the same change statistics as EQM on an annual basis. We have added an extra sentence at the end of the first paragraph of section 7.1.

*Line 575-576* (Line 681-682 in the track changes document)*: "In addition, the ensemble spread is exactly the same for both EQM and 3DBC methods as 3DBC has the same change statistics as EQM on an annual basis."*

- Ll542-551 and Figure 14: I suggest to separately show the EQM and 3DBC ensembles in the different panels of Fig. 14 as it might be very interesting to see in connection to the text describing the figure. For e.g., you mention that the underestimated SWE is mainly seen in EQM. If you separated the two BA-ensembles, that would really give more weight to the statement. Also, you are showing earlier in the manuscript that 3DBC has lower uncertainty in the climate change signal. If so, that would even show in Figure 14 and it would be valuable to see how the uncertainty spread evolves over time. Also, one might even see in the figure whether one of the BA-ensembles is overconfident, i.e. that the observed data falls outside the uncertainty range too often.

**Answer:** Yes, we plot a new figure (Fig. 18) to separate them. We also added texts for this figure in the first paragraph of Section 8.2.

*Line 677-690* (Line 790-803 in the track changes document)*:*

*Figure 18 shows again the projected annual sum/mean of hydrological variables from 1971 to 2098 for mainland Norway, but separating the projections between the two bias-adjustment methods. The results show that the two bias-adjustment methods play a minor role on ensemble means as well as ensemble spread for runoff, evaporation and soil moisture, with the differences between the bias-adjustment methods less than 10 mm/year for runoff and evaporation and less than 1 mm for soil moisture. The ensemble mean of snow water equivalent using the 3DBC method has a better agreement with the results driven by observed forcing data than the ensemble mean using the EQM method, which always leads to underestimation of snow water equivalent in the historical period. In addition, the ensemble spread for snow water equivalent is narrower using 3DBC than EQM, especially before 2040, indicating lower uncertainty of projections using 3DBC. However, it is interesting to see that the snow water equivalent projections do not differ substantially after 2040 between the two bias-adjustment methods, probably due to less snow days in a warming climate. The minor impact of bias-adjustment methods on annual values also leads to similar spatial distributions of the changes in runoff, evaporation and soil moisture, but considerable differences of changes in snow water equivalent are found along the coast and northmost Norway between the bias-adjustment methods (Fig. S3 in Supplementary material).*

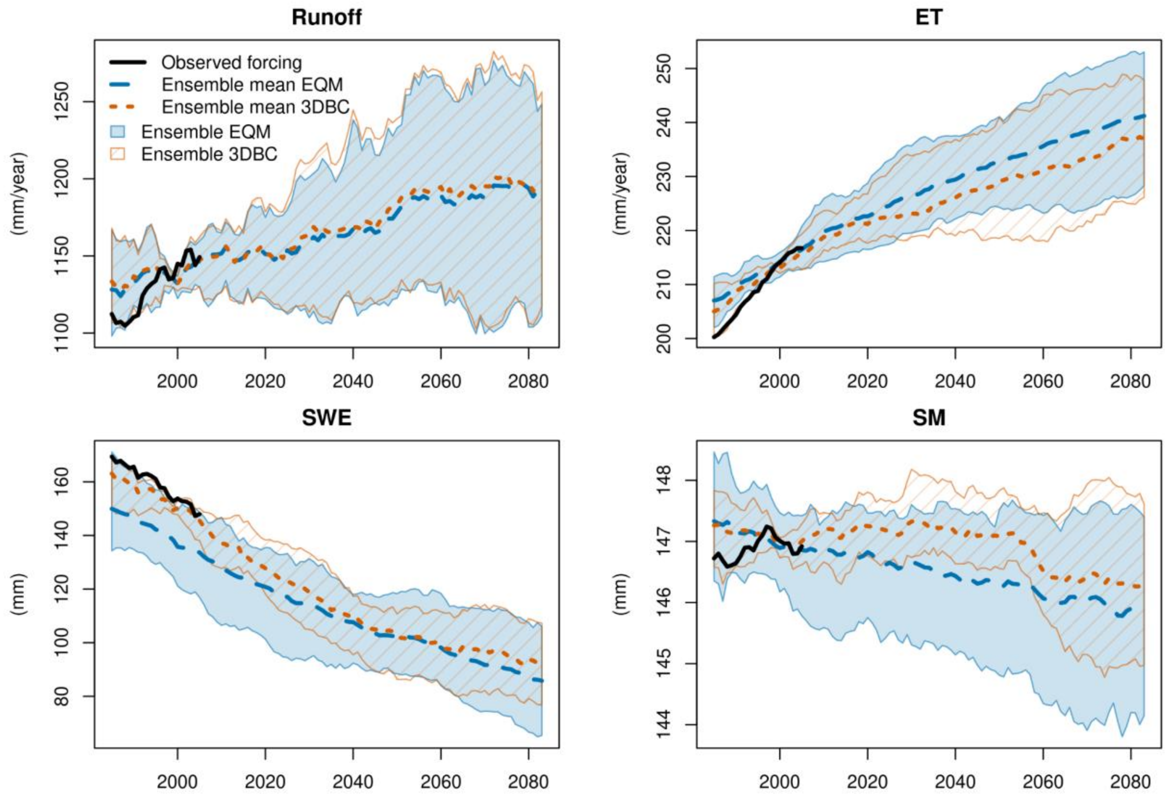


**Figure 18: The same as Fig. 16 but the projections using different bias-adjustment methods are separated.**

- L553, 554: I'm uncertain how to interpret the term 'increasing changes'. Strictly speaking, I would interpret it as the changes are getting larger over time. However, it refers to a time slice result without comparison over time. Should it be rather 'positive changes' or 'increase of …'?

**Answer:** We changed it to "increase in runoff".

- Section 9: An often-seen aspect in national climate assessment report is the updating of the results in more or less regular intervals. It would be great if the authors could share their thoughts and experiences with regard to that topic. For e.g., how are questions handled like 'How to the new results compare to the old ones?', 'How is the trust in the results affected by differences to previous results?' and 'Are the old results totally outdated or can the still be useful?'

**Answer:** We added a new sub-section 9.2 to discuss the differences in results in the old and new national reports:

***9.2 Comparison of results between the old and new national reports***

*The improved modelling chain generated updated climate and hydrological projections for Norway, which resulted in slightly different climate change signals and climate impacts compared to the analysis in the old national report. Under the RCP4.5 scenario, the projections for the old and new* *national reports agree on the direction of change, but CiN-2025 projections display a smaller increase in annual temperature (ensemble mean of 2.0 °C) and precipitation (ensemble mean of 6%) than the ensemble means in CiN-2015 (2.7 °C and 8% increase in temperature and precipitation respectively) at the end of the century. In addition, the ensemble spread in CiN-2025 is narrower than in CiN-2015, indicating more robust climate change signals. However, these differences are caused not only by the new selection of climate models, but also by the selection of the reference period. In CiN-2015, 1971–2000 was used as the reference period while in CiN-2015, it is 1991–2020. As temperature has already risen considerably in recent decades in Norway, annual mean temperature is higher in 1991–2020 than in 1971–2000, and the differences in temperature between the periods 2071–2100 and 1991–2020 are consequently more moderate than those between 2071–2100 and 1971–2000. In contrast, a larger increase in runoff is seen in CiN-2025 projections than in the previous one, mainly due to the improved evapotranspiration routine in the hydrological model (Huang et al., 2026).*

*Another major difference between the old and new national report is that CiN-2025 selected SSP3-7.0 as the high-emission scenario, which assumes lower emission than RCP8.5 used in CiN-2015. Under SSP3-7.0, the ensemble mean increases in annual mean temperature, precipitation and runoff are 3.4 °C, 11% and 10% in 2071-2100 relative to 1991-2020, respectively (Fig. S4, S5 and S16 in Supplementary materials). These increases are also smaller than the ones in 2071-2100 relative to 1971-2000 under the RCP8.5 scenario, shown in the old national report. Hence, users who have made computations based on the CiN-2015 projections, should notice these differences and justify whether their computations should be updated or not.*

- I also miss a discussion of the selection of the ensemble. There are several ways how to do model selection and it would be helpful for the reader that the chosen method is put into perspective.

**Answer:** Thank you very much for this comment! Reviewer 1 has also commented on this issue. We added one paragraph in Section 9.4:

*Line 884-896* (Line 1007-1019 in the track changes document)*: " In principle, we suggest using the full ensemble projections with both bias-adjustment methods to account for the uncertainty of the whole modelling chain. But in practice, users may want to select a subset of climate models and one bias-adjustment method to reduce the computational cost of further applications. As the users may be only interested in parts of Norway and the performance of climate models and bias-adjustment methods vary in space and time, we are not able to give a straightforward suggestion on the subset of climate models and bias-adjustment methods based on the national analysis. However, the methodology as well as the analysis in this paper provides examples of selecting models and bias-adjustment methods. In order to select a subset of climate models, the users can analyze the climate signals for their study area and periods as in Fig. 3 and then select the models based on the study purpose, e.g., studies aiming to assess the driest and warmest climate conditions or the wettest and coldest conditions in the near or far future. Based on the selected models, the users can further assess* *the seasonal trends for their study area and periods using both EQM and 3DBC projections as in Fig. 6. If the trends are comparable between the two bias-adjustment methods, the 3DBC adjusted projections can be preferred, especially when the study is focused on seasonal changes and snow processes. Otherwise, we strongly recommend to use the projections adjusted by EQM and 3DBC to account for the uncertainty of bias-adjustment methods."*

- ll657-658: Empirical-statistical downscaling is mentioned as another methodological approach. Please include a link to the study/report so that readers might read up on that or even better, describe it a bit more. It is probably of interest to the readers to learn that you have downscaled a really huge number of available CMIP5/CMIP6 GCMs for all available RCPs and SSPs.

**Answer:** Thanks for the suggestion. A recent reference (Benestad et al., 2025) including details on the empirical-statistical downscaling method has been added (chapter 9.1). However, ESD was not part of the complete modelling chain, i.e. it did not feed into hydrological and index calculations. We however now also refer more clearly to the CiN-2025 report which includes the ESD results:

*Line 779-782* (Line 899-902 in the track changes document)*: "In CiN-2025, this limitation is taken into account for temperature and precipitation using results from empirical-statistical downscaling (ESD) of the complete set of available GCMs. The ESD results are shown in the CiN-2025 report but not in this study as our focus is on the complete modelling chain. A detailed description of the ESD method used in CiN-2025 is given in Benestad et al. (2025)."*

- L680-681: I agree that users might have to select a BA method appropriate for their use (or use both). However, in the presented case here, I would find it difficult to recommend the users which one to use. I suggest that the authors give a few examples of applications in which one BA method should be preferred over the other.

**Answer:** Please see our reply to the previous comment and the new paragraph in Section 9.4:

*Line 884-896* (Line 1007-1019 in the track changes document)*: "In principle, we suggest using the full ensemble projections with both bias-adjustment methods to account for the uncertainty of the whole modelling chain. But in practice, users may want to select a subset of climate models and one bias-adjustment method to reduce the computational cost of further applications. As the users may be only interested in parts of Norway and the performance of climate models and bias-adjustment methods vary in space and time, we are not able to give a straightforward suggestion on the subset of climate models and bias-adjustment methods based on the national analysis. However, the methodology as well as the analysis in this paper provides examples of selecting models and bias-adjustment methods. In order to select a subset of climate models, the users can analyze the climate signals for their study area and periods as in Fig. 3 and then select the models based on the study purpose, e.g., studies aiming to assess the driest and warmest climate conditions or the wettest and coldest conditions in the near or far future. Based on the selected models, the users can further assess* *the seasonal trends for their study area and periods using both EQM and 3DBC projections as in Fig. 6. If the trends are comparable between the two bias-adjustment methods, the 3DBC adjusted projections can be preferred, especially when the study is focused on seasonal changes and snow processes. Otherwise, we strongly recommend to use the projections adjusted by EQM and 3DBC to account for the uncertainty of bias-adjustment methods."*

References

Benestad, R. E., Chen, D. and Hanssen-Bauer, I.: Empirical-statistical downscaling. World Scientific Publishing Company. https://doi.org/10.1142/6908, 2008.

Bosshard, T., M. Carambia, K. Goergen, S. Kotlarski, P. Krahe, M. Zappa, and Schär, C.: Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections. Water Resour. Res., 49, 1523–1536, https://doi.org/10.1029/2011WR011533, 2013.

Chae, S.T. and Chung, E.-S.: Significant contribution of bias correction methods to uncertainty in future runoff projections under CMIP6 climate change. Journal of Hydrology: Regional Studies 56, 101973, https://doi.org/10.1016/j.ejrh.2024.101973, 2024.

Déqué, M., Rowell, D.P., Lüthi, D. et al. An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. Climatic Change 81 (Suppl 1), 53–70, https://doi.org/10.1007/s10584-006-9228-x, 2007.

Nivron, O., Wischik, D.J., Vrac, M., Shuckburgh, E. and Archibald, A.T.: A temporal stochastic bias correction using a machine learning attention model. Environmental Data Science, 3: e36, 1–40, http://doi.org/10.1017/eds.2024.42, 2024.

Paz, S.M. and Willems, P.: Uncovering the strengths and weaknesses of an ensemble of quantile mapping methods for downscaling precipitation change in Southern Africa. J. Hydrol. -Reg. Stud. 41, 101104, https://doi.org/10.1016/j.ejrh.2022.101104, 2022.

Senatore, A., Fuoco, D., Maiolo, M., Mendicino, G., Smiatek, G. and Kunstmann, H.: Evaluating the uncertainty of climate model structure and bias correction on the hydrological impact of projected climate change in a Mediterranean catchment. J. Hydrol. Reg. Stud. 42, 101120, https://doi.org/10.1016/j.ejrh.2022.101120, 2022.

Vetter, T., Reinhardt, J., Flörke, M., van Griensven, A., Hattermann, F., Huang, S., Koch, H., Pechlivanidis, I. G., Plötner, S., Seidou, O., Su, B., Vervoort, R. W., and Krysanova, V.: Evaluation of sources of uncertainty in projected hydrological changes under climate change in 12 large-scale river basins, Climatic Change, 141, 419–433, https://doi.org/10.1007/s10584-016-1794-y, 2017.

Tong, Y., Gao, X., Han, Z., Xu, Y. and Giorgi, F.: Bias correction of temperature and precipitation over China for RCM simulations using the QM and QDM methods, Climate Dynamics 57, 1425-1443, https://doi.org/10.1007/s00382-020-05447-4, 2021.

Wu, Y., Miao, C., Fan, X., Gou, J., Zhang, Q. and Zheng, H.: Quantifying the uncertainty sources of future climate projections and narrowing uncertainties with bias correction techniques. Earth's Future 10 (11), e2022EF002963, https://doi.org/10.1029/2022EF002963, 2022.

H. Zhang, H., Chapman, S., Trancoso, R., Toombs, N. and Syktus, J.: Assessing the impact of bias correction approaches on climate extremes and the climate change signal, Meteorol. Appl., 31, 1-18, https://doi.org/10.1002/met.2204, 2024.