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 will write more details about this method.

At the end of the first paragraph of section 7.3, we will add the following sentences:

“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. (2015 and 2017).”

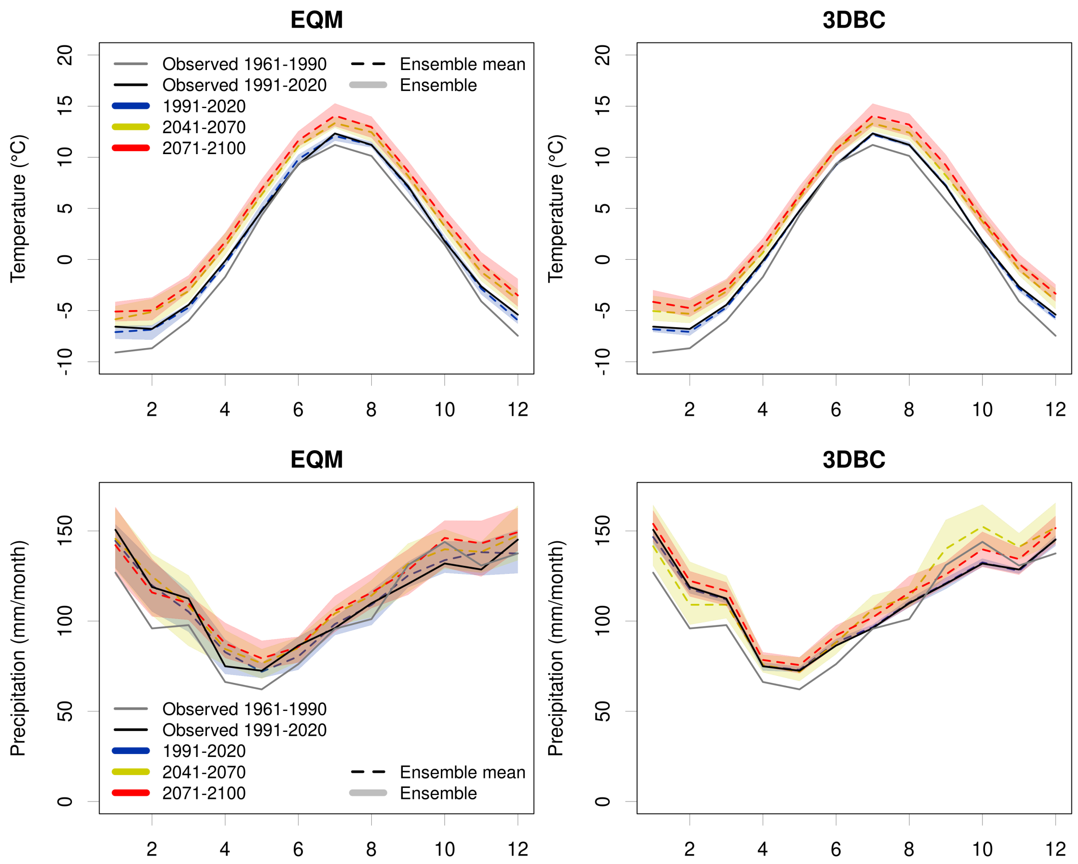
Also, I would like the authors to discuss their results in context of similar studies, either right when the results are presented or in the discussion. Have other studies also found such a large contribution of the BA effect for seasonal means?

Answer: Thanks a lot for the comment. We will include discussion on our results in the context of similar studies. Generally, studies on uncertainty attribution are mainly targeted at annual values rather than seasonal ones, e.g. Chae and Chung (2024) reported a larger variance contribution of BA methods than GCMs in precipitation and runoff projections while GCMs was found to play a larger role than BA methods in temperature projections. These findings are rather similar to ours, but the study was based on CMIP6 GCMs, different BA methods and confined to a small catchment in South Korea. Thus, the results are not necessarily comparable to ours. Other studies by Senatore et al. (2022) and Wu et al. (2022) demonstrated contradicting results with the model uncertainty dominating over the total uncertainty. Previous studies (Paz & Willems (2022); Lafferty & Sriver (2023)) confirmed that the contribution of uncertainty components might vary by region. Tong et al. (2021) and Zhang et al. (2024) have also concluded that the effects of BA methods are variable- and season-dependent when they investigated the seasonal climate change signals in temperature and precipitation over China and Queensland, Australia respectively. However, these studies have not carried out ANOVA or other similar analyses to quantify the contribution. Their results on winter (DJF) and summer (JJA) do not suggest larger uncertainty associated with BA methods than the variability within the model ensemble.

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 the comments. We will further discuss the limitations of 3DBC in the revised paper when we comment on the bump in Fig. 11. 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. We will add an extra line indicating observed 1961-1990 to the figure to highlight this.
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.



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 will add a paragraph on the selection of the CMIP6 data.

“There were (only) 14 RCM simulations based on CMIP6 available by June 2024, thus the selection criteria are different from the ones for the CMIP5-EURO-CORDEX projections. The main criterion for the selection is to include as many GCMs as possible while at the same time excluding model combinations that show fairly similar results in temperature and precipitation over Norway.”

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

Answer: we will adjust the text as below:

“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.”

LI257-258: 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 will correct this accordingly.

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

Answer: Thanks for the comment. We will reorganize and add a few sentences to the paragraph before section 5.1, including a short explanation why we ended up using the selected methods.

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. We realize that we can better explain the method we adopted by rewriting section 5.1 and introducing mathematical equations, which can help clarify how EQM was applied in this study.

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 will modify the sentence and further clarify how the extrapolation below 1st and beyond 99th quantiles were done.

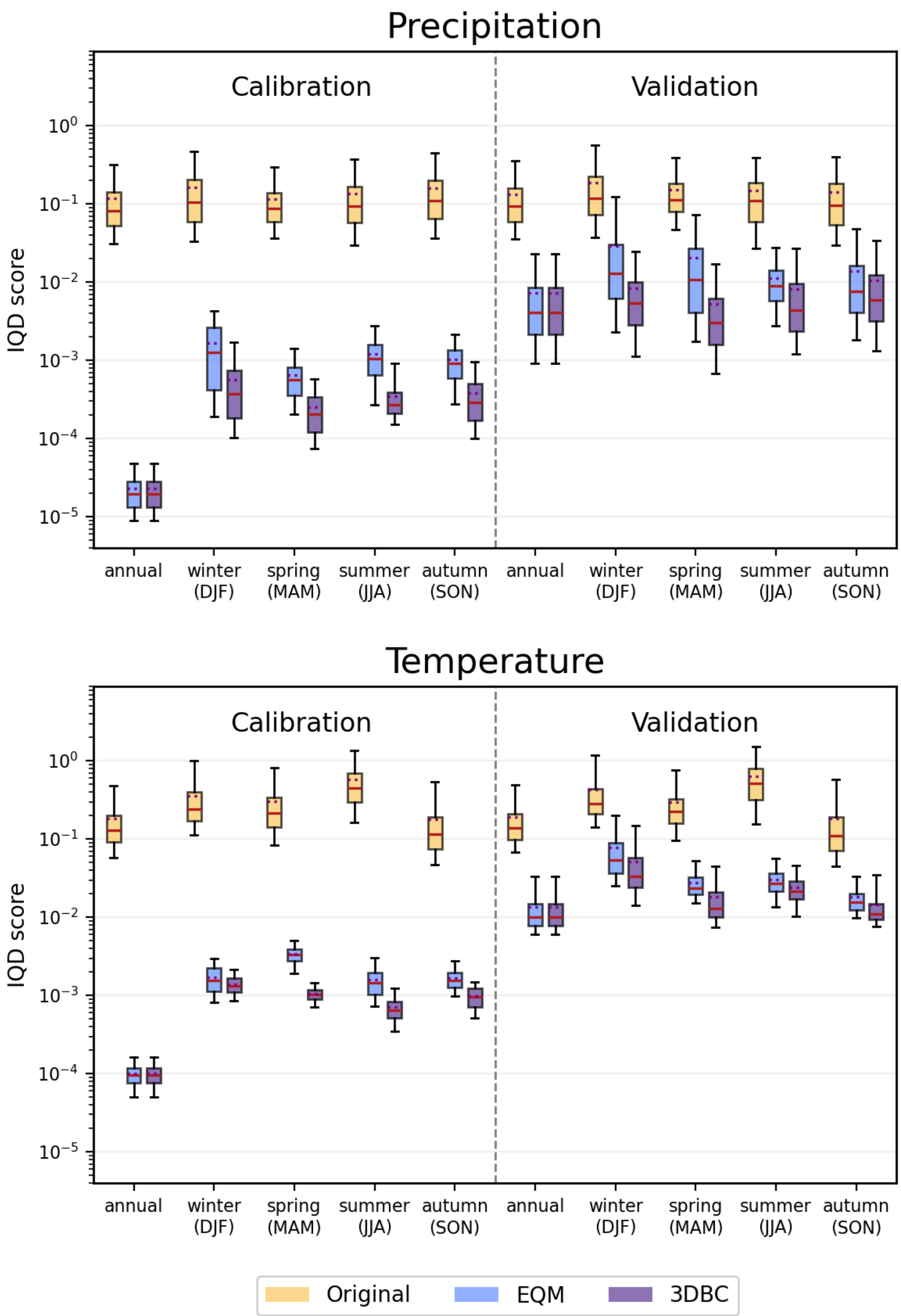
For values smaller than the 1st quantile and larger than the 99th quantile, constant linear extrapolation was performed based on the slopes derived from the 1st and 2nd quantiles and 98th and 99th quantiles respectively.

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. On individual GCM-RCM model level for temperature in the validation period, the IQD scores from particular RCMs (CCLM-4-8-17 and REMO2009/REMO2015) show that EQM actually provides marginally better results than 3DBC in autumn, which might indicate that the observed autumn temperature ranks between the calibration and validation periods are quite different. Otherwise, all the other individual results point to the same conclusion: 3DBC yields better bias-adjustment performance than EQM. See also the next comment about presenting the spread.

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 comments. In response to Reviewer 1, we have added a boxplot showing the ensemble spread of the IQD score, both annual and seasonal, instead of Table 2 (see the figure below). We will also extend Figure 4 to cover all four seasons as an example. In general, the rank correlation coefficients (ρ) are largest and positive in winter (warm days are wetter), followed by negative ρ in summer (warm days are dry, cold are wet). For spring and autumn, ρ is much smaller indicating a rather weak rank correlation between temperature and precipitation. We will add a figure showing the ensemble spread of ρ as well.



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 according to the comment from reviewer 1. In the new figure, one can clearly see the degree of the glacierization in the Engabreen 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.

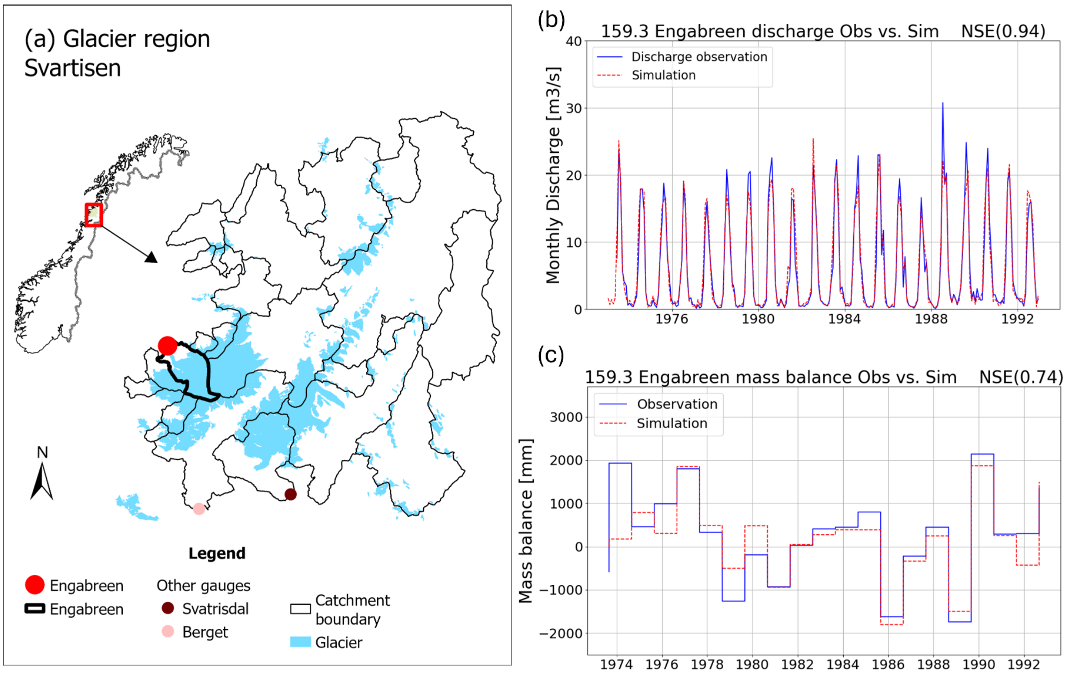
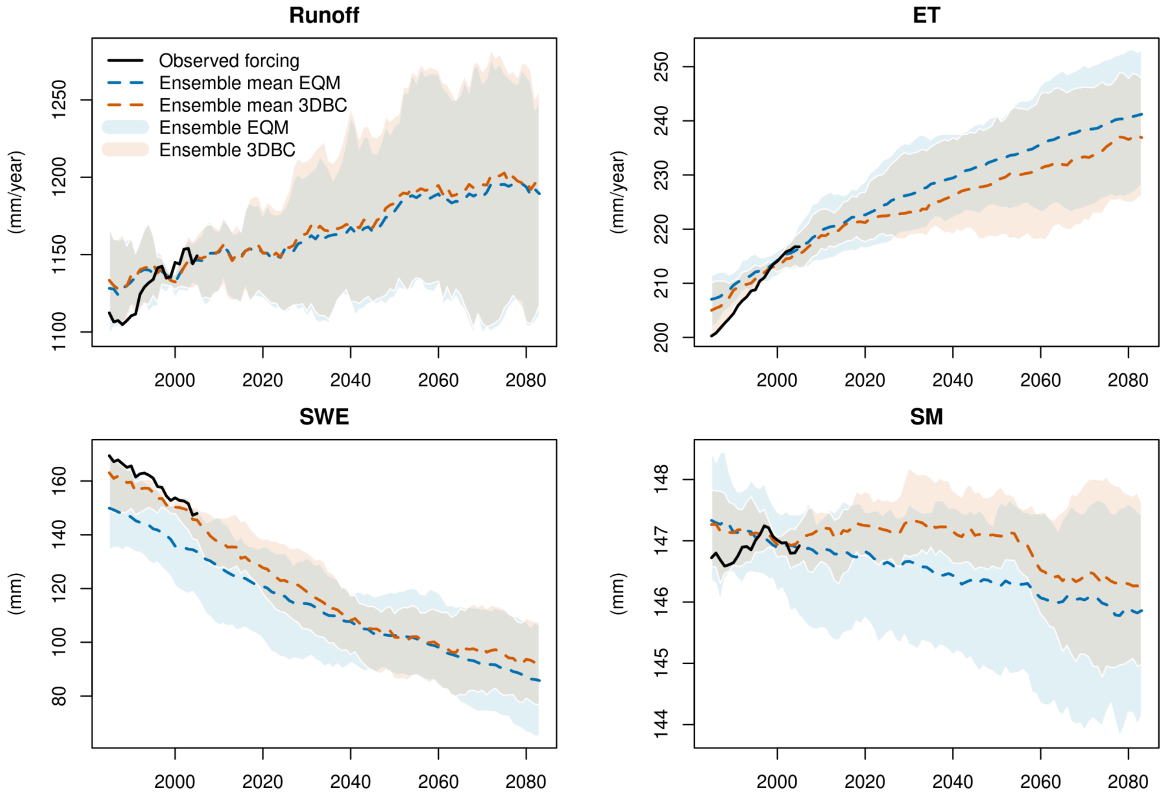


Figure 8: Is it correct to interpret that the ensemble spread would be exactly the same for the EQM and the 3DBC methods since the methods show equal spreads in annual change statistics (Fig. 5)? If so, I would suggest to add a note that BA method does not matter for the spread in the time series shown 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 will add an extra sentence stating this.

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 to separate them.



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 will change 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: The different assumptions in the two versions of the national projections poses a challenge in communicating the new results. The previous projections presented results as changes from 1971–2000 to 2071–2100, whereas the current projections present results as changes from 1991–2020 to 2071–2100. At a glance, the temperature increases and precipitation increases seem smaller than in CiN-2015, but this is a logical consequence of the fact that 20 years of climate change have already passed.

Decision-makers who have not yet used projections in their plans, should use the new projections. Decision-makers who have based their plans and computations on CiN-2015, however, are on the right track with climate adaptation and are not required to update their plans with new projections.

We will add the following paragraph to the paper:

The CiN-2025 projections generally display a smaller increase in temperature and precipitation than the ensemble used in CiN-2015, but the two versions of projections agree on the direction of change. This is mainly explained by the shorter period between reference and end of the century (1991–2020 to 2071–2100 used in CiN-2025 instead of 1971–2000 to 2071–2100 used in CiN-2015). In contrast, a larger increase in runoff is seen than previously, mainly due to the improved evapotranspiration routine. Overall, the ensemble spread is narrowed down, indicating a higher certainty of future projections.

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. Our answers are below:

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. In principle, we suggest using the full ensemble projections if possible, but we also provide general guidance on the methodology of selecting projections. For users who want to select a subset of climate models, they should firstly analyze the climate signals for their study area and periods 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. If the users only want to use the 3DBC-EQM adjusted projections that gives better dependence between variables, they should firstly analyze the seasonal trends for their study area and periods using both EQM and 3DBC-EQM projections. If the trends are similar, the 3DBC-EQM adjusted projections can be sufficient, otherwise, we strongly recommend using both bias-adjusted projections.

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 RCP:s and SSP:s.

Answer: Thanks for the suggestion. A reference (Benestad et al., 2008) about empirical-statistical downscaling and a brief description of the method will be added. However, ESD was not part of the complete modelling chain, i.e. it did not feed into hydrological and index calculations.

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. In principle, we would suggest the users do the same seasonal trend analysis (Fig. 5) for their study region first. If the trends using EQM and 3DBC-EQM are similar, then the 3DBC-EQM adjusted projections may be sufficient, otherwise, we strongly recommend using both bias-adjusted projections to account for the uncertainty.

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., Huang, S., Aich, V., Yang, T., Wang, X., Krysanova, V., and Hattermann, F.: Multi-model climate impact assessment and intercomparison for three large-scale river basins on three continents Earth System Dynamics 6:17-43, 2015.

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.