

We thank the reviewers for their careful reading and valuable comments! Please see our point-by-point response below.

Reviewer 1:

L472-474: does that mean that a correction factor is used but not calibrated? If yes, which value was used?

Answer: We didn't use any correction factors in this study. We revised the sentences to avoid misunderstanding.

Line 486-487 in the track-changes manuscript:

"In addition, we didn't correct rainfall or snow as in other HBV applications, because it will lead to inconsistency between the climate and hydrological projections in terms of water balance."

Reviewer 2:

General comments:

I read both the first manuscript and the revised version and got the impression that the authors carefully addressed the reviewer's comments in a fully satisfying way. The authors changed the manuscript substantially which made it a lot clearer and even added some interesting new aspects. Since there have been substantial changes, I mainly comment on those parts where I found that the changes might need some further improvements. It is only minor changes and I suggest acceptance once those comments have been taken into account.

Detailed comments:

L265-268: It sounds like the last selection step was made based on biases in mean precipitation and temperature biases have not been weighed in. On line 272, the authors write though that performance is checked both with respect to precipitation and temperature. It would be good if the authors clarified this inconsistency. Either, temperature bias was not used as an exclusion criterion or the differences of the temperature biases were considered negligible.

Answer: Thanks for the comment. We added this sentence on line 266-267 in the track-changes manuscript: *"The differences between the temperature biases in the remaining model combinations were found negligible."*

L332-333: I agree that RCMs often show higher wet-day frequency than reference data, at least in the study region. However, generally speaking, there are regions and seasons in which climate models tend to show a dry frequency bias (e.g. mediterranean region, Sahel zone, parts of South America). I would suggest that the authors make the statement less general and/or more specific to their study region.

Answer: Thanks for the comment. The reviewer is correct, and we have therefore changed the statement so that the overestimation of rainy days by RCMs is specified for our study region. We updated the text with another, more fitting reference.

Line 333-334 in the track-changes manuscript:

"Wet-day correction has also been applied prior to bias-adjustment of precipitation. In our study region, RCMs generally provide more rainy days than observed ones (see e.g. Lhotka et al., 2024)."

Thus, for each grid cell, a threshold ..."

L339-340: Is this really true? It probably depends what the authors mean by applicability. If it is about less extrapolation in the future periods, I would doubt that the statement is true. The transformation in equation 2 is a division of the heavily skewed random variable by a less skewed random variable. In theory, this leads to an even more heavily skewed random variable. Even in the case of the division of two nicely behaving random normal variables, the result would follow a Cauchy-distribution which is much more heavy tailed. I don't know the exact nature of the distributions of the nominator and denominator in equation 2, but I would guess that anything deviating from a normal distribution would make the distribution of the ratio even more complicated.

In the end, this could be related to what I read in Bürger et al. (2013) where they see that detrended quantile mapping was amongst the methods showing the strongest outliers. Have the authors looked at the performance of the methods for the tails? Did they see outliers?

In any case, I would suggest that the authors rethink this statement and maybe even remove it as it might be misleading to state a potential advantage of the method that actually might turn out to be just the opposite, depending on how one interprets the statement.

Answer: Thanks a lot for the constructive comments. We understand now our original formulation L339-340 was not clear enough and might lead to misunderstanding. Equation 1 simply refers to a scaling factor, which is calculated based on the ratio between the accumulated monthly precipitation for the projection and training periods, separately for each month of the year. This scaling factor is not a random variable but a single number, and dividing by it in Equation 2 will not change the characteristics of the original distribution. The idea is to scale the range of the future projection closer to the historical one so that we can reduce the necessity of extrapolation outside the historical range and use interpolation instead which is much more robust. To all our understanding, the statement is thus true and we want to keep it. However, we have restructured the fourth paragraph in section 5.1 as follows to make it clearer:

Line 341-350 in the track-changes manuscript:

"Scaling the range of future projection to align closer to the historical one improves the applicability of the transfer function by reducing the need to extrapolate beyond historical values, while the daily variability about the monthly mean remains unchanged. The trend was reimposed after the bias-adjustment of the residuals. For all the variables other than temperatures, trend removal and reimposition were performed multiplicatively, i.e. variability and residuals were considered as ratios rather than differences. The procedure was applied separately for twelve months. For example, for precipitation for month i , the scaling factor δP_i , which also represents the relative precipitation trend, is defined as:"

When it comes to outliers, we have not looked at them specifically. But we did check the performance of EQM in terms of some ETCCDI climate indices. The results (not shown) revealed that EQM provided comparable extreme precipitation amounts (e.g. R99pTOT) in the validation period (1960-1984) as observed and QDM (Quantile Delta Mapping), and we did not note any particular outliers.

L341: I was confused why residuals was written in quotes. It sort of implies to me that something else than residuals is meant. I suggest to either explain what is meant or to use another terminology. Maybe, the term anomaly could be used instead?

Answer: Thanks a lot for the comment. Residual usually refers to the difference between two values, and not related to division of such values. That was why residuals were written in quotes in

L341 because the term was also meant to include the variables using multiplicative trend removal and reimposition, which involves quotient/ratio between two values. The quotes are removed and it is further clarified in the text (see the answer above).

L346: The subscript j is defined as day in projection period. That is confusing to me as the subscript i might vary depending on j . Shouldn't it rather be 'day belonging to current month i in projection period'?

Answer: Thanks a lot for the comment. We have rewritten the equations to make them clearer. Equation 2 is formulated as follows and similar changes are made for equations 3, 4 and 5 to be consistent as well.

Line 353-358 in the track-changes manuscript:

"The de-trended daily precipitation in the projection period for a specific calendar month i (=Jan., Feb., ..., Dec.), $\hat{P}_i^{prj}(t)$, is:

$$\hat{P}_i^{prj}(t) = P_i^{prj}(t)/\delta P_i \quad (2)$$

where $t = 1, 2, \dots, N_i$ is the time step, and N_i the number of days belonging to the calendar month i in the projection period. $P_i^{prj}(t)$ denotes the original RCM daily precipitation time series for month i in the projection period."

L349: Use of 'normalized' is misleading (also on line 346). The distribution of the 'residuals' is still highly non-normal. I suggest to use the term 'detrended' as introduced earlier in the manuscript.

Answer: Thanks for the comment. We agree that using 'normalized' in this context might be misleading. We have changed the term to 'de-trended' instead in L353, L359, L360, L364 and L369 in the track-changes manuscript, respectively.

L365-366: I'm aware that this procedure is used for many bias-adjustments methods that require to be applied to a well-defined future time slice rather than the full transient projection time-series. The issue is that this can lead to jumps between the 10-year time slices. My expectation would be that the jumps at the transition between the decadal periods are much smaller than the day-to-day variability. When looking at overall 30-year statistics, such jumps most likely are negligible and the chosen method is a pragmatic way to handle a basic limitation of the method. However, when using the bias-adjusted and stitched-together time series as input to impact models which require transient time series, it might be more tricky to really see how it affects the outcome. I guess the transition happens at the end of year, right in the middle of the snow accumulation period. A sudden jump in temperature might lead to a clear difference in snow accumulation before and after the transition.

I don't want the authors to change anything in their method and I'm well aware that such pragmatic choices have to be made. Further, I do not really know how large the impact of those jumps is. I would like though that the authors briefly write about this limitation/property of the method if they agree with the comment. I could even see that the approach with the shifting 30-year window by 10-years in each step actually is a way forward to make the jumps smaller, rather than shifting the window by 30 years in each iteration. And maybe, this additional statement gives the authors even the chance to say what else they did or that one could do to reduce the jumps.

Answer: Thank you very much for these useful comments. We agree with the reviewer that our approach can lead to jumps at the transitions between the decadal periods, and such jumps are usually negligible if we look at the overall 30-year statistics. We have therefore added the following sentences to describe the limitation of the method:

Line 374-378 in the track-changes manuscript:

“This procedure can better preserve the decadal trend but at the same time might introduce discontinuities in the time series at the start of each decadal period. The impact of such jumps in e.g. temperature might affect the snow accumulation in hydrological modelling at the transitions between the decadal periods. On the other hand, such jumps are negligible when the focus is on overall 30-year statistics. The impact of this approach is also expected to be smaller than shifting the entire 30-year window in each iteration. This is a pragmatic way to handle a basic limitation of the method.”

Figure 5: I find the different color scales in the 4 rows rather confusing and make it hard to compare the seasons. My suggestion would be to use the same color scale for all 4 rows. I leave it to the authors if they want to make this change. There is nothing wrong with the way it is shown now, and they are free to make the suggested change in case they agree with my comment.

Answer: We did not find a suitable color palette that could show the details of all the seasonal results and satisfy the colour blindness requirement of the journal at the same time. The main focus lies on the comparison of the two methods within a season and not among the seasons. Thus we have chosen different color scales to highlight the details

Section 6.3

Thanks a lot for the helpful explanation about how DEW and distHBV were combined. I have a follow up question though. You write on lines 656-658 that runoff from distHBV gets replaced by the runoff from DEW for the glacierized grid cells. How is the shrinking glacier area handled? In the future, some formerly glacierized grid cells will have become non-glacierized, i.e. they do not get replaced by DEW anymore. Nevertheless, distHBV might still think that it is glacierized. According to figure 9, there are clearly almost no grid cells with overestimated runoff in panel c). Are formerly glacierized grid cells replaced by some other land use type in distHBV? Or was it handled in another way?

Answer: The mask of glacierized grid cells was selected in the historical period and it didn't change in the postprocessing procedure. Hence, the formerly glacierized grid cells were still replaced by the DEW results. We reformulated the sentences to make it clearer.

Line 566-571 in the track-changes manuscript:

“Note that DEW simulated the whole glacierized catchments but we only masked the glacierized grid cells in the historical period (e.g., the blue-color area within the catchments in Fig. 8a) for postprocessing. It is mainly because DEW uses a simpler potential evapotranspiration (PET) method and rougher landuse/soil classes for non-glacierized area than distHBV. The mask of the glacierized grid cells remained unchanged during the whole postprocessing procedure and the DEW outputs on the masked grid cells replaced the distHBV ones at each time step in both historical and future periods.”

L685-687: I do not follow the author's hypothesis that a reduced number of snow days would lead to a convergence of the ensemble spread and mean for the two different BA methods. Previously, it was argued that the different handling of correlation between precipitation and temperature leads

to the differences seen between the two BA methods. I find it not fully convincing that one here mainly attributes the convergence to the number of snow days instead of correlation pattern between precipitation and temperature. And as long as there are snow days, one should see similar differences if the difference in the correlation pattern still persists. Of course, I might miss an important point here and if so, I would like the authors to explain their hypothesis a bit more.

Answer: thank you very much for the comment. We agreed that the length of snow days may not necessarily be the major reason and our formulation was not clear enough. We have corrected the explanation in the revised manuscript.

Line 703-708 in the track-changes manuscript:

“However, it is interesting to see that the snow water equivalent projections do not differ substantially after 2040 between the two bias-adjustment methods. Although the correlation structures and temporal sequences in the 3DBC data for this period are still similar to the observed ones in all models, different climate change signals in the RCMs lead to a larger overall spread in the driving variables, particularly temperature (Fig. 10). Thus, the resulting snow water equivalent projections using the 3DBC data inherit the general climate change signals and become similar to the EQM projections.”

Grammar

L653: “soil moisture tend” instead of “soil moisture trend”

Answer: It has been corrected.

L663: “increases” instead of “increasing changes” to be consistent with the sentence before where the same term “increasing changes” has been replaced by “increase”

Answer: It has been corrected.

L839: “CiN-2025” instead of “CiN-2015” at the second appearance of CiN-2015.

Answer: It has been corrected.