Authors' response to editor's and referees' comments on hess-2018-317
"Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments" by Judith Meyer et al.

Dear Editor, dear Referees,

Thank you again for the evaluation of our manuscript and the many helpful comments.

Please find below our point-by-point reply in blue to all editor and reviewer comments *(in black)* followed by the track-change version of the revised manuscript.

Best regards,
Irene Kohn, on behalf of all authors

Editor comments
Both reviewers indicate that it is an interesting paper but requires some redesign. I agree with this assessment. For example, in the intro, the authors mention the objectives of the study but do not mention the scientific hypothesis on this this study is funded. In fact, this word is not even mentioned in the manuscript. It is also important to make clear why the state of the art leads to formulate such hypothesis. Otherwise a manuscript becomes a case study. This could be done based on the proposed objectives.

We revised the Introduction section and clarified our objectives and the underlying hypothesis.

Technically, I would like to see a proper comparison with existing methods such as that of Hempen et al.: Hempel, S., K. Frieler, L. Warszawski, J. Schewe, and F. Piontek (2013), A trend-preserving bias correction: the ISI-MIP approach, Earth Syst. Dynam., 4(2), 219–236, doi:10.5194/esd-4-219-2013.

In the submitted version of the manuscript we looked mostly for comparable studies that investigated consequences of a bi- or multivariate bias correction in hydrological impact modelling. We also agree with the suggestion of referee #2 that we should broaden our scope by considering the findings of other references that discussed inter-variable aspects in (univariate) bias correction methods also outside a hydrological modelling context. Consequently, we added such references in the Intro and Discussion Section. However, there are many more publications of further bias correction methods similar to the one mentioned above and also more recent ones. We don't agree that an even more comprehensive review of existing bias correction methods in general needs to be part of this manuscript, in which we exemplarily compared hydrological impacts of a uni- and a bivariate bias correction method in two alpine catchments. A more comprehensive discussion on (recent) bias correction approaches, including among others reference to Hempel et al. 2013, is presented in the original publications on the bias correction methods applied by us: Cannon et al. (2015) and Cannon (2018). The issue of preserving projected trends, being the main focus of Hempel et al., 2013, is treated in more depth in Cannon et al. (2015). However, **a reference to Hempel et al. (2013) was added in Section 3.1**.

As mentioned by the both reviewers, a proper model validation and crossvaligdation should be carried out. Nowhere in the paper a validation period is mentioned. Model performances against observation are not presented in any form. This is a must in this kind of studies.

Calibration of the hydrological model was carried out again with a slightly shortened period to allow for a short validation period. An additional table summarising model performance measures for the calibration and validation period is included in the revised manuscript and graphs showing model output vs. observation based data on streamflow, snow cover, and glaciers were added as supplementary material.

Comments Referee 1

The manuscript "Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments" by Meyer et al. investigates the effect of univariate versus (BC) multivariate bias correction (MBC) on the representation of snow, ice and rain representation in a climate change impact assessment approach. MBC has the advantage to control for variable interdependency (T and P namely) that in turn influences rain-snowfall fractioning. They used BC and MBC to bias correct and downscale 10 GCM-RCM combination using quantile mapping and drove the hydrological model HBV-light in a transient setting (1976-2099). They analyzed the effect of the different methods for snow-water-equivalent, icemelt, streamflow amounts and its composition (ice, snow, rain) over time. The paper adds to the ongoing discussion on the possible effects of intervariable dependency, especially by adding the information about the effect of streamflow composition. The paper is scientifically interesting, original, overall very well writing and certainly within the scope of the journal and of interest to the readers. Texts and figures are widely clear and lead the way for reasonable conclusions.

Beside some minor comments that I will list below, I have two more substantial concerns that refer to the design of the study.

In summary, as we understand, these more substantial concerns are: a) hydrological modelling results for a validation period are missing, and b) the use of a constant temperature lapse rate in HBV. We provide detailed explanations and suggestions for compromises below. But summarized in brief, these decisions need to be seen in context with the aim of the study to test principal effects of uni- vs. multi-variate bias correction in such an environment. The objective of the study is not to present a best possible hydrological impact assessment for the sake of a case study (in line with the Editor's comment, this would not be a suitable study for HESS) and results for the two catchments should not be taken as that, i.e. interpreted quantitatively as future reality. Instead, the case study catchments (rather than e.g. a completely hypothetical catchment, which would be an alternative) allow demonstrating and discussing the potential effects; i.e. the objective is to compare simulation results for the same boundary conditions in terms of hydrological modelling only differing in the T_a and P input series corrected with a uni- vs. a multivariate method. In the revised Introduction section we stated this more clearly.

The use of the set lapse rate is motivated by the small scale of the catchments and the fact that this is the commonly used approach in HBV applications. We argue that both do not affect our conclusions regarding the specific aims of this study, in which focus is on the comparison of systematic effects of bias correction methods, to keep the paper concise, and modelling with a conceptual model that is as parsimonious as possible. Note that Figure 2 (P at T > or ≤ 0 °C) compares results of QDM vs. MBCn independent of the hydrological modelling and related assumptions. Subsequent results of hydrological modelling are all reasonable consequences of that. Changes in the hydrological model parametrization or incorporation of changed lapse rates may have led to slightly changed simulation results but we are convinced that that would not lead to different main conclusions in the comparison of the simulations driven by univariate and multivariate corrected *P*- and *T*-input.

1. The hydrological model was calibrated for the entire reference time period, e.g. 1976–2006, (page 7, line 16 ff) against streamflow, snow (SWE and snow-covered area), and glacier volume.

Three things puzzles me here:

1.1 There is no validation period!

1.2 Model performances against observation are not presented at all, neither as statistical measure nor in the graphs. I am aware that only differences between input data sets are analyzed in this study, still the basic performance measures are needed to frame the results. E.g. if the representation snow melt is not well captured (what I do not assume here) but the streamflow is (hence snow melt insensitive) than the snow sensitivity to changes in the input data might also be underrated.

The reason for not reserving a validation period was to make full use of available observation-based data, in particular on glacier change. As indicated by the referee, we think that details of model performance and hydrological model validation are not of highest priority to compare the results of univariate and bivariate bias correction of precipitation and temperature in hydrological modelling of alpine catchments in a general way, as we intend with our study. However, we fully agree that model validation is an important component of hydrological modelling studies in general and that some information on model performance should be added as general information for the reader.

1.1.Validation period: In our case the historical reference period 1976–2006, which was used for model calibration, was chosen because (i) the used HYRAS climate data product unfortunately covers only the period ≤ 2006, and (ii) model warmup started in the year 1973, the simulation in 1976 and needs to go until at least 2003 in order to capture the climatology and to make full use of reliable data for the initial conditions and development of the catchments' glaciers. We consider the use of observation-based glacier data for both, model initialization and calibration, crucial. Hence, we used an estimate of glacier volume for the year 2003 based on glacier area (by Paul et al. 2010) in model calibration. Consequently, a calibration period of at least 1973/1976 to 2003 has been a result of the available glacier data for the years of 1973 and 2003. Our model calibration strategy was developed within a multi-catchment modelling study (Stahl et al. 2017), for which a consistent calibration scheme and consistent reference datasets for 49 alpine catchments (including their glaciers) were needed. We are not aware of any additional glacier observation-based data covering all glacierized parts of the modelled catchments Hinterrhein and Schwarze Lütschine that could have been used additionally in this study.

We agree that validation normally should be done but it was not used here because the study rather focused on the relative performances of the bias correction approaches than on the absolute model performance. However, to address the requests for a validation period, **we repeated model calibration for a period until the year 2003 only, which allows at least three years (2004–2006) to be used for model validation. Consequently, Figures 3–6 were exchanged with results based on the hydrological model parametrization based on this new calibration and text referring to the figures was adjusted where needed.**

1.2 Model performance: Performance is usually far from perfect. But considering that we deal with high-elevation catchments, incorporate streamflow, snow characteristics, and glacier data in calibration we consider it an acceptable real-world representation for the purpose of the study. **We included quantitative results on model performance for the calibration and the new validation period 2004–2006 and validation period based on the components of the multi-criteria objective function as additional Table** (Table 3) in the revised manuscript. **Obs. vs. sim. result graphs were**

added separately as supplementary material in order not to complicate the previously existing results graphs in which the focus should be kept on the comparison of QDM- and MBCn- based results.

1.3 this concern about validation and performance measures also extends to both QM approaches (was a cross-validation framework used? provision of verification statistics is needed) –> more emphasis should be laid on the presentation of validation and the introduction of a validation period/validation framework is required

The BC methods are optimized to the best possible transformation. We think, a randomized cross-validation would not provide any targeted and usable information regarding our aims of testing systematic effects of QDM vs MBCn. Because the univariate and multivariate bias correction algorithms are applied in an asynchronous fashion to freely running climate simulations – adjusting the marginal/joint distributions – it is almost guaranteed that they will perform well in terms of cross-validated measures of distributional fit. This is by construction. See also the critique by Maraun & Widmann in HESSD: https://www.hydrol-earth-syst-sci-discuss.net/hess-2018-151/.

Cross-validation does make sense when performance, especially for aspects not explicitly adjusted, is measured in a setting where climate model simulations are synchronized with the real-world climate state, for example in climate prediction or perfect boundary condition (e.g., reanalysis-driven) setups. We note that such reanalysis-driven evaluations have been performed in Cannon (2018) for the two algorithms used in this study. This was done over a large continental domain for a more complicated multivariate index (fire weather), that combines, in a nonlinear fashion, the current and lagged effects of temperature, precipitation, wind, and humidity. **We discussed these issues and past validation efforts in an additional paragraph in the revised Discussion section**.

2. If I understand correctly, the combination of climate model data and the hydrological model is as follows:
- The quantile mapping is performed between climate model output and the average catchment value for T and P
- This mean value is interpolated within the catchment by a lapse rate that is fixed for each day of the year (extracted from the reference period)

This approach might be needed in HBV-light, but is based on the assumption that the lapse rate is not changing over time and is independent of certain events. This is a very strong assumption that is disproven by numerous study showing e.g. elevation depended warming, for instance. I assume, and to my own experience, that the slope of the lapse rate is quite sensitive to SWE simulations. Hence, this strong assumption likely influences the robustness of your results. Furthermore, you cannot control for this in the calibration of the model, as you limited the evaluation of model performances of SWE and snow-covered area to 2000 – 2500 m asl (page 7, line 21), which is exactly the catchment mean elevation for which the lapse rate is of minor effect. To me the fixed lapse rate is a very critical points in the study and need to be solved. A possible workaround of this issue would be to perform a quantile mapping that is not based on catchment mean values, but for each grid cell of the HYRAS data set. This procedure is currently done for the new CH2018 climate scenarios by MeteoSwiss. Doing so, you can extract the lapse rate for each day separately and use this dynamic lapse rate. With this procedure you would not have to make this strong assumption of a static lapse rate and have much more reliable results.

Firstly, for clarification: calibration in terms of snow covered area fraction was based on the entire catchment only the calibration of SWE was limited to the elevation range 2000–2500 m a.s.l. Phrasing of the sentences referring to the calibration of snow characteristics in the submitted version is not wrong but a bit misleading. When we re-calibrated the model for the revisions we used a larger elevation range for the SWE calibration (see also separate comment on SWE calibration below).

We agree that lapse rates and their potential future changes are important for hydrological modelling of alpine catchments in particular and that hydrological impact assessments should attempt to incorporate them adequately. However, this is not easy and a topic related to high uncertainties anyway. The use of a linear gradient is a key feature of HBV. That is already a – if you like, strong, – simplification as one of many common in hydrological modelling.

We considered the possible ways to derive bias corrected lapse rates from GCM–RCM output:

- Referee's suggestion to apply BC on a grid-to-grid base to HYRAS: this would in fact be an application of BC to a grid distinctly higher in resolution (1x1 km) that is, as we mentioned in the paper, also criticized widely and would call for an explicit downscaling step, i.e. it would require aggregation and re-gridding of HYRAS. Consequently, results would also be strongly influenced by the underlying background climatology of this HYRAS interpolation product (also available only for recent conditions). We are not convinced that this is an appropriate/better solution.

- Theoretically, of course it is feasible to extract lapse rates from the GCM–RCM output and incorporate in the bias correction, including in case of the multivariate MBCn to apply it to 4 variables, not only to T and P catchment mean but additionally to lapse rates. Uncertainties in the extraction of lapse rates from 12x12 km CORDEX grid for mesoscale catchments (54 km^2 the smaller one) and their subsequent bias correction are expected to be substantial (even higher than those in case of the catchment averages), too. In addition, we think that evidence of trends in projected temperature lapse rates at the catchment scale is limited.

➔ Considering our aims and setting (model, data, objectives) **we don't think that extra efforts vs. expected added value at the expense of more assumptions may improve the confidence in the conclusion of the study**. We think it would rather complicate the interpretation of the results in terms of effects of the QDM vs. MBCn on T_a and P. We prefer simple, comparable boundary conditions for our modelling experiment, and kept lapse rates unchanged/unaffected. As stated above, we stressed more clearly that our study should be taken as that, a modelling experiment (and not a perfect impact assessment). **Taking the long-term seasonal average lapse rate pattern from observed data as stationary was one of our assumptions to facilitate the separation of the effect of incorporating or ignoring interdependencies of catchment averages of T_a and P in bias correction in the interpretation of results**.

I think that these proposed changes are accomplishable in a reasonable time. Therefore, if those and the following comments are addressed, I am happy to comment on the manuscript again and likely recommend a publication. I am looking forward to the revised version.

To summarise briefly: we re-calibrated for 1976–2003 with validation for 2004–2006, added information on performance, and we reconsidered the SWE calibration, i.e. extended it to a larger elevation range. While not crucial for our study's main objective, those revisions were straightforward and time efforts justifiable. But from our side and with the study's objective in mind we do not see the need for cross-validation of bias correction and argue that fixed lapse rates can be used (see above).

Specific comments:

Page 1, line 23ff. I suggest to state this result ("for the historical..:") prior to the effect on the future as this ensures an improved bias correction for MBC.

Thanks for this suggestion. We rephrased those sentences slightly.

Page 2, line 24: This publication might also be of interest (I am not an author): Wilcke,Renate Anna Irma; Mendlik, Thomas; Gobiet, Andreas (2013): Multi-variable error correction of regional climate models. In: Climatic Change 120 (4), S. 871–887. DOI:10.1007/s10584-013-0845-x.

Thanks for this suggestion. As also suggested by referee #2 we added references to this study in the Intro and Discussion sections.

Page 4 line 2: Only the "Unterer Grindelwald"-glacier is big (∼ 6.biggest in Switzerland). It is a glaciated catchment, but covered by smaller glaciers. Please, rephrase. Therefor,also the following sentence needs to be rephrased.

We agree, thanks. The sentences were corrected.

Page 5, line 7ff: Please highlight that you apply catchment averages and that these averages are the "Target" in the quantile mapping approach (If I understood you correctly)

Yes, we work generally with series of catchment averages also as "target" data as stated already at the beginning of the data section Page 4 line 16: "The resulting time series of catchment mean precipitation and temperature were used as input for the calibration of the glacio-hydrological model and as historically observed climate data (HOCD) for the bias correction." **We tried to stress this even more clearly in the revised manuscript by adding 'catchment mean' explicitly at several places, where it may be important to remind the reader on that**.

Page 5, line 10: This is a very unusual time period as it crosses to climate normal periods. Do you have any reason of this time window. It hampers comparability to other climate change impact assessment studies.

As already explained above, **the use of the period 1976–2006 was motivated by the modelling of glacier change**. Glacier data for the catchments (all glacierized parts) were available for the years 1973 (glacier area based on the inventory by Müller et al. 1976 / Maisch et al. (2000) and modelled ice thickness data provided by Matthias Huss) and 2003 (inventory by Paul et al. 2010). In our opinion, the initialization of the glaciers represents a major source of uncertainty for modelling glacierized catchments. Hence, we wanted to start the simulations (warmup) at a year for which glacier- area and -volume distributions are relatively well constrained by observation based data. Furthermore, we think it is important to incorporate glacier data in model calibration, what we realized by using glacier volume estimates based on glacier area data for the year 2003 (see also comment above). The used

HYRAS climate data product was updated recently but had not been available for periods beyond 2006 at the time of the study. Moreover, we would like to stress again that our study should be mainly considered as a general test to reveal and discuss principal "effects of univariate and multivariate bias correction" using two alpine catchments for demonstration rather than a specific climate impact assessment for the two case study catchments. We think, we clearly focused our interpretation of results on the comparison of univariate and bivariate bias correction; it is only limited possible to discuss the presented simulation results quantitatively in terms of projected impacts and to compare them with more comprehensive climate impact assessments.

Page 5, line 13. Which gauging station was used. I am only aware of the FOEN station in Lütschine-Gsteig, and the Weisse Lütschine, Zweilütschinen. Did you use differences of these stations?

Yes, exactly we used the records from the stations Lütschine–Gsteig and Weisse Lütschinen–Zweilütschinen to reconstruct streamflow 1973–2006 for the outlet of the study catchment (in mm/day) at the location of the station Schwarze Lütschine – Gündlischwand operated by the Canton of Bern. Observed stream flow at the station Schwarze Lütschine – Gündlischwand that was available for the period 1992–1999 was used to validate the reconstructed streamflow time series. **We explained this in the revised version**.

Page 6, line 6: This is phrased wrongly. Univariate QDM cannot be both widely accepted (and used since several years) and developed by Cannon et al. 2015.

We agree that this was not precisely phrased and revised this paragraph.

Page 6, line 9: detrending of a time series is problematic, as assumptions about the kind of trend are necessary. Can you please add information about the way the trend is treated and comment on possible effects.

The approach used for QDM follows the quantile delta change (Olsson et al. 2009, Atmos. Res.) and quantile perturbation (Willems and Vrac 2011, J. Hydr.) methods and is explained and discussed in detail in Cannon et al. (2015). Strictly speaking, the use of "detrending" in the submitted version was inaccurate. The detrended quantile mapping (DQM) algorithm in Cannon et al. (2015) *does* use a linear detrending step, but this is not the case with the QDM (and, by extension, MBCn methods). QDM adjusts all quantiles in the projection period by first removing the projected change signal (e.g., if the climate model projects a 20% increase in the 90th percentile of precipitation in the future period, this change is removed, etc.), then it applies standard quantile mapping, and finally the projected change signal is reintroduced. **We revised Section 3.1 to be more precise with respect to the preservation of the climate models' change signal and added reference to Hempel et al. (2013) as suggested by the editor.** We do not think that any further detailed explanations and discussion are needed within the scope of this study, given that this is included in the original paper on the method

Page 6, chapter 3.1: additional information about the validation procedure should be given as well as information about the "target value" (catchment averages). Please, see our general comment on cross-validation above. The last sentence of this section already referred to the catchment averages being subject to bias correction. This sentence was slightly rephrased to place a bit more emphasis on that.

Page 6, line 28: Please, quantify the difference by adding grid cell size and range of catchment area.

Information on catchment sizes (54 km² and 180 km²) and the resolution of the climate model data (0.11°: ~12x12km) and the HYRAS data (1x1km) used to derive the catchment mean values are given in the corresponding subsections but were now repeated here (in brackets) and the sentence slightly rephrased.

Page 7, line 16: so no separation of calibration and validation time period ! see general comments

Please, see response to general comment above, too. The **original period was split to obtain a model validation, result graphs updated, and information on model performance added**.

Page 7, line 21-22: is it correct that you not only evaluate but calibrate your model against this elevation limitation? Please rephrase. See below for a combined reply to this and the next comment.

Page 7, line 21-22: I disagree with the statement that only the area 2000-2500 is crucial as in my experience it is also very important for streamflow how much of the entire catchment is covered by snow – and hence contribute to snow melt. Please,comment.

See also response to general comment above. For model calibration, originally, we used: a) snow covered area fraction of the entire catchment!, b) mean SWE for the elevation range 2000–2500 m a.s.l. Both were derived from a gridded SWE climatology product from the Swiss SLF institute. Due to discrepancies in the data product's resolution and the high detail of elevation-and-aspect zones for glacierized and non-glacierized catchment part for the catchment model, we think a comparison of mean values calculated for the entire catchment or a (not too limited) elevation range most appropriate. The reference data product from the SLF is known to be based on relatively few stations for elevations above 2500 m a.s.l. (in the domain of the study); thus, we constrained the use of the data to elevations below 2500 m a.s.l.. Selection of the elevation range of 2000–2500 m a.s.l. was based on an analysis of the SLF product for SWE statistics over all elevation zones of 49 alpine catchments (alpine catchments modelled in Stahl et al. 2017). Focussing on all those catchments, we considered the elevation range 2000–2500 m a.s.l. particularly important. However, **we agree that an extension of elevations in the catchments SWE calibration is useful for the catchments in this study. Hence, we extended the elevation range used for calibration of SWE to elevations ≤ 2500 m a.s.l. without a lower limit**.

Page 8, Figure 2: What is also striking is that noBC is performing better than QDM for rainfall. Can you add on this?

This is not the case. If understood correctly, your statement is "true" if one looks at the median of the boxplots for rain for days with $T_a > 0$ °C only (left part of left panel plots for historical period) ignoring the considerable bias for rain for days with $T_a \leq 0$ °C (right part of left panel plots for historical period). Looking at total precipitation independent of T_a, considerable biases towards too large precipitation sums are evident in the noBC data from all GCM–RCM combinations. Both methods (univariate QDM and MBCn) correct for this to a similar degree, with marginal distributions the same for both methods, by definition, as also stated in the manuscript. See also the additional figures showing T_a and P distributions provided as supplementary material in response to the request by referee 2.

Page 10, Figure 4: Maybe an error occurred, as the boxplots in the lower panel of the Schwarze Lütschine graph is missing.

No, that is not an error but on purpose. As stated in the main text, for the Schwarze Lütschine catchments bias-corrected data from only few GCM–RCMs resulted in complete glacier retreat (volume = 0) by 2099. For those cases circles for the simulated year of complete glacier disappearance are plotted, whereas for most cases simulations resulted in a glacier volume ≥ 0 in 2099, thus no boxplot can be plotted. **We remarked on that in the figure caption**.

Page 14, line 22: Please rephrase: It depends not on the bias correction but more specifically on the consideration of intervariable dependencies.

Sentence was rephrased.

Page 14, line 24: Are the found glacier retreat comparable to other findings?

As stated above, we think one should discuss our simulations results with a focus on differences in the simulations (effects) as a result of the univariate vs. bivariate bias correction of temperature and precipitation rather than taking our results as specific future projections for the two catchments. We are not aware of projections specifically for the glaciers in our catchments that could be directly compared. Roughly, projections for mid-sized glaciers in the Aare basin might be compared with our simulations for the glaciers in the Schwarze Lütschine catchments and projections for small glaciers in the Alpine Rhine basin may be taken for comparison with our results for the glaciers in the Hinterrhein catchments. Results and discussion of projected glacier retreat and disappearance dates for small Swiss glaciers can be found in Huss & Fischer (2016, frontiers in earth science, https://doi.org/10.3389/feart.2016.00034); general results of projections for Swiss glaciers are presented, for instance, in the CCHydro report (FOEN, 2012: Effects of Climate Change on Water Resources and Waters. Synthesis report on "Climate Change and Hydrology in Switzerland" (CCHydro) project. Federal Office for the Environment, Bern. Umwelt-Wissen No 1217). In general, as far as a comparison is possible, the projected glacier retreat in our study compares well with results for projected glacier area change presented in those publications. However, note that i) it is generally difficult to compare the response and evolution of individual glaciers, ii) in our study **we consider the difference between the simulations driven by QDM- and MBCn-corrected data of much higher relevance than the exact date of glacier disappearance in comparison to other studies**.

Page 15, line 8: Much more critical to me is the assumption of a fixed lapse rate, even more under climate change conditions

Please, see general comments above. We agree that the question how to derive lapse rates for climate impact projections adequately is a difficult one and an important challenge requiring research efforts. However, **we restricted our study to demonstrating the effect of incorporating or ignoring interdependencies of catchment averages of T_a and P in bias correction**.

Comments Referee 2

The work by Meyer et al. presents an inter-comparison of a univariate and a multivariate bias correction (BC) method in terms of hydrological climate impact scenarios in two catchments of the Swiss Alps. For this purpose, daily temperature and precipitation amounts as simulated by ten EURO-CORDEX RCM experiments are bias-corrected toward observed catchment mean values and then fed into the HBV-light hydrological model. For BC the QDM and MBCn methods are employed, the latter taking explicitly into account variable interdependencies. The study finds important differences in the simulated streamflow for a historical period between QDM- and MBCn-based setups.

In general, shows MBCn shows a better performance. The main reason is an underestimation of snowfall amounts in the QDM-based setups (equivalent to a smaller snowfall fraction of total precipitation) which translates into smaller SWE amounts and an overestimation of winter streamflow while the spring meltwater peak is underestimated. The differences in the snowfall amounts between the two BC approaches furthermore translate into differences in future climate change signals of SWE, glacier coverage and, finally, streamflow. Qualitatively, the differences between the BC approaches are obtained for all ten climate model chains investigated, indicating a robust finding that seems to be valid for any GCM-RCM chain. In general, the paper is of very high quality and nicely highlights an important potential deficiency of bias-corrected climate scenarios in the Alpine region. It comes at a perfect time, as several recently released national reference scenarios are based on univariate BC approaches similar to QDM (e.g., Austria: ÖKS15, Switzerland: CH2018). As such, the study is certainly relevant for the journal's readership. Its setup is sound and convincing, the results are presented in an appropriate manner and the conclusions are well-based on the results obtained. There are no language issues except for the mixed use of past and present tense in the presentation of the results, which should be revised. There are a few minor issues that should be corrected for as well as two major remarks (see below). However, I'd leave it up to the authors to consider these major comments or not. I believe a consideration would further improve the quality of the paper, but the study is sound and convincing even in its current state. My recommendation is therefore to return the paper to the authors for minor revisions.

Congratulations for this nice piece of work! Sven Kotlarski

MAJOR ISSUES

Cross validation: Similarly to the point raised by a previous reviewer, I believe that a proper cross validation framework would be helpful. MBCn is a more complex method than QDM, and there's an increased danger of overfitting. As MBCn explicitly corrects for biased inter-variable dependencies, snowfall amounts (if derived by a fixed temperature threshold) are well represented by definition. Being aware of the criticism by Maraun & Widmann, cross validation still makes sense in a split sample framework, e.g. by separating the historical period into the 15 coldest/warmest/driest/wettest years and the 15 warmest/coldest/wettest/driest years and using these subsets for calibration and verification, respectively. In case this splitting cannot be handled by HBV-light because the transient character is lost, one could carry out a cross validation for at least one ERA-Interim EURO-CORDEX experiment (these experiments are available as well and are in basic temporal correspondence with the observations). In general a cross validation would make the point stronger that a multivariate BC is superior for the example presented.

Yes, indeed a split sample test would be problematic as input to a hydrological model as there might be rather long memory effects. We can follow your reasoning, yet, as there have been a number of studies on the bias correction recently, **we would like to keep focus here on the hydrological application**. In general, a lot of aspects have to be considered in the selection and application of a bias correction method for a given purpose. With our study we mainly want to point to the potentially significant consequences in terms of snowfall fraction from a hydrological (modeller's) perspective, as also acknowledged above. Therefore we refer to our response to referee #1 regarding cross-validation (above). **For this study with a bivariate application case of MBCn, we consider it sufficient to briefly address this in the revision and refer again to past cross-validation evaluation efforts presented in the original paper about the MBCn method (Cannon, 2018). Accordingly, a paragraph on these aspects was added in the Discussion Section.**

Reason for underestimated snowfall amounts by QDM: I understand that the paper puts an emphasis on the hydrological consequences of the two different BC methods. These effects are very well and convincingly presented. However, the question WHY QDM shows these deficiencies is not ultimately answered. The reason is to be found in the T-P relationship of the QDM data, and probably already appears in the raw RCM data. To analyze this further, 2D histograms would be extremely helpful and also illustrative.

Below we add as an example a graph (for only one raw RCM data set, one RCP, one catchment) that shows distributions and bivariate probability density plots of P and T_a in order to compare our HOCD, uncorrected, QDM-corrected, and MBCn-corrected data. Differences (biases) between the historical reference data (HOCD) and the uncorrected RCM data are evident. However, differences regarding the T_a–P inter-variable relationship between QDM- and MBCn-corrected data are present (see e.g. local regression line in plots e–h) but more difficult to recognise in such kinds of plots. Hence, we included only the corresponding precipitation sums for days below and above 0 °C as shown in Figure 2 in the submitted and revised version of the manuscript. However, **figures as the one below but for all GCM–RCM combinations, both RCPs, and both catchments were added as supplement**. Furthermore, we agree that it is of high interest to discuss and understand the causes of the T_a–P intervariable-dependence-bias resulting in differences in temperature-threshold defined snowfall fractions better. **We extended the discussion a bit in this respect (first paragraph of Discussion section), but think an ultimate answer is beyond the scope of this study and would require separate investigations also based on an intercomparison for further observational datasets**.

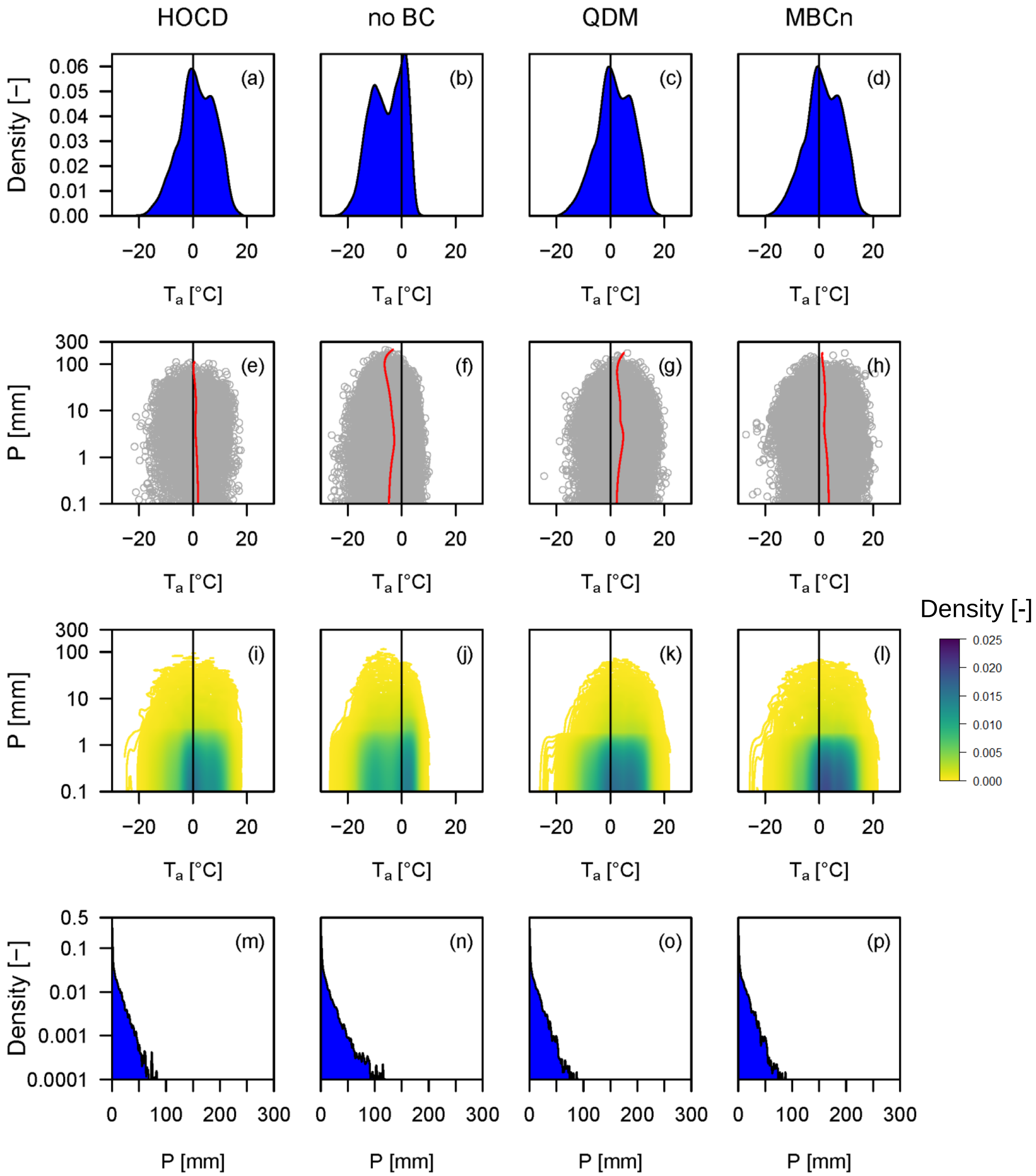

Figure A: Exemplary representation of T_a and P over the historical reference period 1977–2006 for the Schwarze Lütschine catchment according to the historically observed climate data (column HOCD) and uncorrected (column no BC), univariate-corrected (column QDM), and bivariate-corrected (column MBCn) climate model data from one GCM–RCM combination (ECEARTH–RACMO22E) for RCP 8.5. Top and bottom panel show marginal distributions of T_a and P, respectively; centre panels show bivariate plots for T_a and P with local regression lines (plots e–h) and density allocation (plots i–l). See Supplement for figures for all applied climate model projection datasets and for both catchments.

MINOR ISSUES

Introduction and conclusions: The literature review should account for the studies by Wilcke et al. (Climatic Change, 2013) and Ivanov & Kotlarski (Int. J. Climatol., 2017). Inter-variable dependencies in standard QM have already been analysed in there. One of the results was that QM does not distort inter-variable dependencies as long as they are approximately represented by the raw RCM data. The results of the present work therefore indicate some distorted inter-dependencies already in the RCM raw output (which could be better described if my major comment #2 would be considered). These issues should also be discussed in the discussion/conclusions.

Thank you, **references to those and further studies were added in the revised Intro and Discussion sections**. See also comment above.

p2 l4: The CORDEX data are actually not available from the CH2018 archive. The respective website only explains the selection of EURO-CORDEX models for the CH2018 Swiss climate scenarios. In the present study, EURO-CORDEX data were probably obtained from the ESGF archive.

Thank you for this remark. We agree that the CH2018 archive is not the most appropriate reference. **We now reference the ESGF archive** in relation to the download of the EURO-CORDEX data. In addition, we expanded our acknowledgements section to include the efforts of Dr. Urs Beyerle and to follow CORDEX terms of use (see point c):

https://www.hzg.de/imperia/md/assets/clm/cordex_terms_of_use.pdf

Table 2: Just a note: The two runs driven by CNRM-CM5 are critical as the driving GCM CNRM-CM5 has an inconsistency in the historical period. It is fine to use them for the present work, but in future works they might have to be removed. More information is available from the new EURO-CORDEX errata page available from www.eurocordex.net.

Thank you for pointing this out. We added a remark on that in the manuscript (footer, Table 2).

Chapter 3.1: The description of the QM methods is incomplete in the sense that it is not clear if the correction has been carried out for the bulk series (all days independent of the time of year) or depending on the time of year (e.g., seasonal or DOY dependence). This information is critical, as a bulk correction could be responsible for the deficiencies of QDM in my opinion. I believe the authors employed a seasonally dependent BC, but this needs to be better explained (even if reference to Cannon et al. is provided).

We agree with this critique. **The issue is now more precisely explained in the revised version of Section 3.1**. Yes, we applied bias correction in a seasonally dependent fashion. Specifically, bias corrections were applied over 3 x 10 = 30-year sliding windows. This involved replacing the central 10-years and sliding forward 10-years for each 30-yr window, until the end of the projection period is reached. Within each window – to ensure an unbiased seasonal cycle – bias corrections were applied separately for each calendar month.

p11 l30-31. Do you have any explanation for the higher mean streamflow amounts for QDM? Are differences in ETP involved?

Potential evapotranspiration (ET) was kept the same for all model runs but actual ET simulations can vary depending on water availability and presence of snow cover. However, this is not a main driver for the observed differences in total streamflow. The slightly higher mean streamflow for QDM compared to MBCn is mostly the case for the Hinterrhein catchment and might be partly explained by one HBV-light model parameter, the so-called snowfall correction factor that can potentially tackle snowfall undercatch measurement errors (by parameter values > 1.0) as well as snow sublimation losses (by parameter values < 1.0). For the Hinterrhein catchment the previous calibration (submitted manuscript version) of this parameter resulted in a value of 0.81, meaning that the model reduces any snow input by 19%. Since snow makes up the largest fraction of precipitation input in the alpine study catchments this snow-specific reduction due to the calibration parameter might finally result in a lower simulated streamflow for the MBCn-based data, for which the snow fraction is higher. In addition, slightly higher ice melt runoff simulations contribute to the higher mean streamflow amounts for QDM compared to MBCn over the historical reference period in case of the Schwarze Lütschine catchment. However, please note that **the additional calibration efforts (to address the requests by referee #1) led to overall less pronounced differences in terms of total streamflow and also a better agreement with observed streamflow, while the systematic differences in the snow and rain component fractions are not affected by the changed parametrization of the hydrological model (see Figure 5 in the revised manuscript)**.

Finally, we thank both referees for noting the following minor issues, which were corrected:

p2 l20: "... which correct for biases in the data's entire distribution..."

p6 l20-21: "... until the multivariate distributions of bias-corrected and observed data match."

p14 l24: "...disappearance vary by over a decade ..."

p15 l31: "...empirical-statistical bias correction methods ..."

Page 6, line 22: "Climate model data" instead of data.

Page 15, line 34: is there a type? "re bias "

Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments

Judith Meyer[1], Irene Kohn[1], Kerstin Stahl[1], Kirsti Hakala[2], Jan Seibert[2], and Alex J. Cannon[3]

[1]Faculty of Environment and Natural Resources, University of Freiburg, Freiburg, 79098, Germany
[2]Department of Geography, University of Zurich, Zurich, 8057, Switzerland
[3]Climate Research Division, Environment and Climate Change Canada, BC V8W 2Y2, Victoria, Canada

Correspondence to: Irene Kohn (irene.kohn@hydrology.uni-freiburg.de)

Abstract. Alpine catchments show a high sensitivity to climate variation as they include the elevation range of the snow line. Therefore, the correct representation of climate variables and their interdependence is crucial when describing or predicting hydrological processes. When using climate model simulations in hydrological impact studies, forcing meteorological data are usually downscaled and bias corrected, most often by univariate approaches such as quantile mapping of individual variables~~. However, univariate correction~~ neglecting~~s~~ the relationships that exist between climate variables. In this study we test the hypothesis that the explicit consideration of the relation between air temperature and precipitation will affect hydrological impact modelling in a snow-dominated mountain environment. ~~g~~Glacio-hydrological simulations were performed for two partly glacierized alpine catchments using a recently developed multivariate bias correction method to post-process EURO-CORDEX regional climate model outputs between 1976 and ~~2100~~2099. These simulations were compared to those obtained by using the common univariate quantile mapping for bias correction. As both methods correct each climate variable's distribution in the same way, the marginal distributions of the individual variables show no differences. Yet, regarding the interdependence of precipitation and air temperature, clear differences are notable in the studied catchments. Simultaneous correction based on the multivariate approach lead to more precipitation below air temperatures of 0 °C and therefore more simulated snowfall than with the data of the univariate approach. This difference translated to considerable consequences for the hydrological responses of the catchments. The multivariate bias correction forced simulations showed distinctly different results for projected snow cover characteristics, snowmelt-driven streamflow components, and expected glacier disappearance dates~~ in the future~~. ~~For the historical period~~ In all aspects – the fraction of precipitation above and below 0 °C, the simulated snow water equivalents, glacier volumes, and the streamflow regime – simulations resulting from the multivariate-corrected data corresponded better with reference data than the results of univariate bias correction. Differences in simulated total streamflow due to the different bias correction approaches may be considered negligible given the generally large spread of the projections, but systematic differences in the seasonally delayed streamflow components from snowmelt in particular will matter from a planning perspective. While this study does not allow concluding definitively that multivariate bias correction approaches are generally preferable, it clearly demonstrates

that incorporating or ignoring inter-variable relationships between air temperature and precipitation data can impact the conclusions drawn in hydrological climate change impact studies in snow-dominated environments.

1 Introduction

~~As the Earth's~~ With global ~~climate~~ changes, hydrological processes in high elevation regions have been significantly impacted (Messerli et al., 2004). In the European Alps, ~~an~~ the observed increase in air temperature ~~was witnessed during the last century~~ is a trend that is expected to continue in the future. Future ~~trends in~~ precipitation changes are less clear, ~~however, a~~ with an expected slight increase in winter precipitation ~~is expected~~ (Gobiet et al., 2014; Kotlarski et al., 2016). The hydrology of alpine catchments is especially sensitive to these changing climate ~~parameters~~ variables (Köplin et al., 2010). High elevations in the Alps are still characterized by snow cover and the existence of glaciers. However, rising air temperatures and a consequent upward shift of the zero-degree isotherm has led to a decrease in snow accumulation and an increase in glacier melt (Pellicciotti et al., 2010). Due to shrinking glacier areas, the glacial influence in the streamflow regimes has decreased. This is especially notable during late summer when water from ice melt can constitute a notable percentage of total streamflow. With progressive glacier retreat, the ice melt contribution to streamflow is expected to decrease (Jansson et al., 2003; Hock, 2005; Moore et al., 2009; Huss and Hock, 2018). The interdependence of air temperature and precipitation is particularly important for hydrological systems as it determines the physical state of precipitation. Bosshard et al. (2014) showed that an air temperature dependent shift from snowfall to rain has notable effects on catchment water storage and seasonal water availability in such an environment. A correct representation of climate variables and their interdependence is therefore essential in hydrological simulations of glacierized catchments.

In hydrological climate change impact studies, post-processing of climate model data has become a standard procedure. Despite continuous progress, raw outputs from regional climate models differ largely from observational reference data due to both spatial mismatches and systematic biases. Therefore, climate model outputs are downscaled and biases are adjusted statistically before being used in hydrological simulations (Ehret et al., 2012; Maraun, 2016; Teutschbein and Seibert, 2012). Many empirical statistical techniques have been developed to post-process climate model outputs for these purposes. For hydrological impact studies quantile mapping approaches, which correct for biases in the data's entire distribution, have often been recommended (Teutschbein and Seibert, 2012; Gudmundsson et al., 2012; Chen et al., 2013). However, these approaches correct the climate variables independently from one another. The interdependence of key climate variables, such as air temperature and precipitation, can be especially important when modelling snow-dominated catchments due to the aforementioned threshold effects of the transition of rain to snowfall or the conditions required for snow and ice melt. Studies that analyzed inter-variable aspects of bias correction showed that univariate quantile mapping retains the inter-variable dependencies as represented by the raw climate model output data (Wilcke et al., 2013; Ivanov and Kotlarski, 2017). But, these may not correspond to the local interdependencies in observations. To account for interdependencies, multivariate bias correction approaches have been developed that allow for the preservation of the interdependence of climate variables as represented by the target observation data throughout the bias correction process (Li et al., 2014; Cannon, 2016, 2018; Mehrotra and Sharma, 2016, 2015). A correction procedure that preserves the climate variables' interdependence may be considered more appropriate for subsequent impact analyses, such as the application of a calibrated hydrological model using multiple variables, than univariate techniques that ignore biases in inter-variable relationships (Cannon, 2018).

While many studies have evaluated bias correction methods in terms of their effects on the actual variables of precipitation and air temperature themselves, studies that use impact models to investigate the consequence of bias correction in the modelled impacts are still rare. So far, there have been only a few studies (Räty et al., 2018; Chen et al., 2018) that investigated the effect of using a multivariate bias correction technique on hydrological projections. Chen et al. (2018) found that jointly corrected precipitation and air temperature data better modelled eleven out of twelve catchments in the calibration period than the meteorological data that was corrected ~~based on~~with a univariate method. An advantage of using a bivariate bias correction approach was not evident for the coldest snow-dominated catchment of the sample though. ~~According to~~ Hydrological simulations by Räty et al. (2018) ~~their hydrological simulations~~ generally did not substantially benefit from bivariate bias correction approaches, ~~whereas~~but when looking more specifically, simulations of high flows and snow water equivalents in snow-influenced catchments improved slightly ~~in comparison to simulations using univariate corrected climate model data~~.

In this study we investigate the hypothesis that the explicit consideration of the relation between air temperature and precipitation in bias correction will affect hydrological impact modelling in snow- and glacier melt dominated environments. Here, dependencies are known to matter most as they have cumulative effects over a season through snow storage and at multi-year time scales through the glacier mass balance. The ~~objective~~ approach of this study was therefore to conduct climate impact modelling experiments that allow ~~compare~~ comparison of the effects of univariate and multivariate bias correction of precipitation and air temperature input ~~in~~on the hydrological ~~impact modelling of~~ change in alpine catchments. ~~This was done by systematically comparing hydrological simulations driven by climate data corrected with the two different bias correction methods.~~ The model experiments ~~was~~ were conducted for two meso-scale partly glacierized catchments in the Swiss Alps, for which snow accumulation, glacier mass balance, and streamflow were simulated from 1976 to ~~2100~~2099.

2 Study catchments and data

2.1 Study area

Two partly glacierized meso-scale catchments in the Swiss Alps, in the headwater of the Rhine River, were examined in this study: the Hinterrhein catchment and the larger Schwarze Lütschine catchment (Fig. 1, Table 1~~Table 1~~). Based on the dataset by Freudiger et al. (2018), used in this study, around the year 1900 glacier coverage was approximately 32% of the Hinterrhein catchment area and around 25% of the Schwarze Lütschine catchment area. Glaciers in both catchments retreated considerably during the 20th century. The Hinterrhein catchment is characterized by small, scattered glaciers, which by 1973 lost around half their area, leading to a glacier coverage of only 7% in 2010 (Table 1~~Table 1~~). In ~~T~~the Schwarze

Lütschine catchment~~, in contrast, holds some of the largest glaciers in the Swiss Alps such as the Grindelwald glacier. Consequently,~~ losses in relative glacier area have been smaller. This difference in glacier coverage is related to elevation with considerably higher maximum elevations in the Schwarze Lütschine catchment compared to the Hinterrhein catchment (Table 1~~Table 1~~).

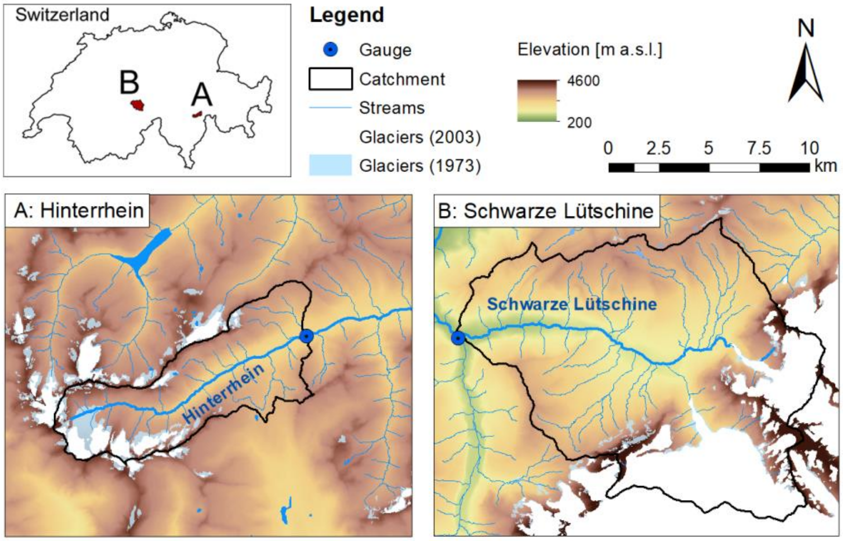

Figure 1: Map of the two study catchments and their location in Switzerland: Hinterrhein (A) and Schwarze Lütschine (B).

Table 1: Catchment characteristics including glacier cover information.

	Area	Elevation			Glacier cover*					
		mean	min	max	1973		2003		2010	
	[km²]	[m a.s.l.]			[km²]	[%]	[km²]	[%]	[km²]	[%]
Hinterrhein	53.9	2357	1587	3387	9.1	17.8	4.7	8.7	3.8	7.1
Schwarze Lütschine	179.9	2059	648	4086	37.0	23.5	34.4	19.1	29.7	16.5

* Based on glacier inventories by Müller et al. (1976) / Maisch et al. (2000) for 1973, Paul et al. (2011) for 2003, and Fischer et al.(2014) for 2010.

2.2 Data and data preparation

The application of bias correction algorithms to climate model outputs is generally based on three datasets: historical observations as reference (also called ‘target’) data, historical climate model simulations, and the corresponding climate model projections. In the present study the historical reference data for the study catchments were derived from an observation based interpolation product, i.e. the 1x1 km² gridded daily air temperature and precipitation datasets from the HYRAS product (Rauthe et al., 2013; Frick et al., 2014). Area-weighted mean values of precipitation and air temperature were extracted for the study catchments. The extracted catchment mean precipitation time series were corrected for undercatch based on the method by Sevruk (1989) and were then further adjusted through ~~a~~ validation with long-term annual mean precipitation sums resulting from a water balance approach (for details see Stahl et al., 2017). The resulting time series of catchment mean precipitation and air temperature were used as input for the calibration of the glacio-hydrological model and as historically observed climate data (HOCD) for the bias correction.

The climate model datasets were obtained from the Coordinated Regional Climate Downscaling Experiment (CORDEX, www.cordex.org) via the ~~CH2018~~ Earth System Grid Federation (ESGF) archive (http://www.cordex.org/data-access/esgf/~~http://www.ch2018.ch/en/home-2/~~). CORDEX is a collaborative effort within the climate modelling community where general circulation models (GCMs) are downscaled using regional climate models (RCMs). Since all catchments in this study are located in Switzerland, GCM–RCMs were selected from the European domain of the CORDEX project (EURO-CORDEX, http://www.euro-cordex.net/). EURO-CORDEX provides simulations at 0.11° (~12.5 km horizontal resolution) and 0.44° (~50 km horizontal resolution). Given that the catchments used in this study are situated in the Alpine domain, only the higher resolution 0.11° simulations were used. Two Representative Concentration Pathways (RCPs) were selected for this study: RCP 4.5 represents an intermediate mitigation scenario, where greenhouse gas (GHG) emissions will peak around 2040 and then steadily decrease, and RCP 8.5 represents a more pessimistic scenario, which assumes that GHG emissions will continue to increase throughout the 21st century (Meinshausen et al., 2011).

Precipitation (P) and air temperature (T_a) data were provided by the ten GCM–RCMs shown in Table 2 for the time period 1970–~~2100~~2099. For each catchment, raw GCM–RCM data were extracted using an area-weighted method as shown in Hakala et al. ~~(in review)~~(2018). Based on the areal fraction of an RCM grid cell overlying a particular catchment, five RCM grid cells contribute to each catchment. All GCM–RCMs used in this study utilize a Gregorian calendar.

Table 2: GCM–RCM combinations from the EURO-CORDEX initiative used in this study.

Driving GCM	RCM	RCM institution
CNRM-CM5-LR [1)]	CCLM4-8-17	Climate Limited-area Modelling Community
CNRM-CM5 [1)]	RCA4	Swedish Meteorological and Hydrological Institute
EC-EARTH [2)]	CCLM4-8-17	Climate Limited-area Modelling Community
EC-EARTH [2)]	HIRHAM5 [5)]	Danish Meteorological Institute
EC-EARTH [2)]	RACMO22E [5)]	Royal Netherlands Meteorological Institute
EC-EARTH [2)]	RCA4	Swedish Meteorological and Hydrological Institute
IPSL-CM5A-MR [3)]	WRF331F	Laboratoire des Sciences du Climat et de l'Environnement
IPSL-CM5A-MR [3)]	RCA4	Swedish Meteorological and Hydrological Institute
MPI-ESM-LR [4)]	CCLM4-8-17	Climate Limited-area Modelling Community
MPI-ESM-LR [4)]	RCA4	Swedish Meteorological & Hydrological Institute

GCM institutions: [1)] CNRM-CERFACS (Centre National de Recherches Météorologiques-Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique); note that a warning concerning an inconsistency in the historical run of CNRM-CM5 has been issued on the CORDEX errata page (https://www.euro-cordex.net/078730/index.php.en) after data had been downloaded and selected for this study, [2)] EC-Earth consortium, [3)] IPSL (Institut Pierre-Simon Laplace), [4)] MPI-M (Max Planck Institute for Meteorology)

[5)] CORDEX errata page (https://www.euro-cordex.net/078730/index.php.en) notes snow accumulation issues for these RCM runs.

The application of the hydrological model requires catchment mean time series of P and T_a. These were subjected to bias correction. Further data used as model input and for model calibration were not directly bias corrected. Daily potential evapotranspiration was calculated with an air temperature based approach provided by Oudin et al. (2005). Catchment specific air temperature lapse rates were determined based on daily values from the HYRAS product. Based on the reference period from 1976–2006 a mean for each day of the year was calculated and smoothed using an 11-day moving average. A mean precipitation gradient (in % per 100 m a.s.l.) was determined from the corrected HYRAS data and applied as constant value in all simulations.

Daily streamflow data for model calibration were provided by the Swiss Federal Office for the Environment (FOEN) and the "Amt für Wasser und Abfall des Kantons Bern". The available streamflow record for the station Gündlischwand (operated by the Cantone of Berne) at the outlet of the Schwarze Lütschine study catchment covered only the period 1992–1999. By using the record of a downstream station of the Lütschine River (station Gsteig) and subtracting the streamflow of its other major headwater tributary (record from the station Zweilütschinen of the Weisse Lütschine) the streamflow for the Schwarze Lütschine study catchment could be reconstructed for the entire simulation period. This reconstructed streamflow time series was validated with the available streamflow data from the station Gündlischwand for the subperiod 1992–1999 and then used for model calibration. Snow water equivalent (SWE) and snow cover data were derived from a snow map (interpolated grid) product by the OSHD-SLF (2013). The glacier area was assessed based on glacier inventory data by Müller et al. (1976) and Maisch et al. (2000) for the state in the year 1973, by Paul et al. (2011) for the state in 2003, and by Fischer et al. (2014) for the year 2010 (see Table 1~~Table 1~~). Estimates of glacier volume were derived based on gridded ice thickness data available for the years 1973 and 2010, which were computed using the approach by Huss and Farinotti (2012) and provided by Matthias Huss. Glacier volume for the year 2003 was estimated based on the glacier cover according to Paul et al. (2011) and glacier volume–area scaling. The glacier volume estimate for 1973 was used for model initialization. The estimate for 2003 was incorporated in the model calibration for the period 1976–2006. The estimate for 2010 was not directly used in the calibration but served the validation of model simulations beyond the year 2006.

3 Methods

3.1 Bias correction of climate data

Depending on the GCM–RCM combination, raw climate variables (noBC) of the control period (1976–2006) differ from the reference data (HOCD). To correct these biases, two different bias correction methods were applied to each climate model's T_a and P series: a univariate quantile mapping technique – Quantile Delta Mapping (QDM) – and a ~~M~~multivariate ~~B~~bias ~~C~~correction approach (MBCn). ~~Univariate QDM was used because of its often and widely accepted application.~~ Quantile mapping is based on a transfer function that transforms the cumulative distribution(s) of the modelled data to match the distribution(s) of the observed series. The obtained transfer function is then applied to all climate model data, historical and projected. Thus it corrects systematic distributional biases relative to historical observations and preserves model-projected relative changes. Quantile Delta Mapping (QDM) ~~QDM~~ is a variant of quantile mapping ~~approach~~ by Cannon et al. (2015) that was designed to avoid artificial deterioration of trends arising as a statistical artefact of standard quantile mapping. QDM corrects systematic distributional biases relative to historical observations and preserves model-projected changes in quantiles in the projection period. For a given time slice, ~~Therefore,~~ the climate model's change signal (Δ) – relative change for precipitation and absolute change for air temperature – is ~~extracted~~removed from all projected future quantiles in a first step. ~~The q~~Quantile mapping is then applied~~ to the detrended series,~~ before the projected ~~trends~~ changes in quantiles are reintroduced to the bias corrected model output. ~~Quantile mapping is based on a transfer function that transforms the cumulative distribution(s) of the modelled data to match the distribution(s) of the observed series. The obtained transfer function is then applied to all climate model data, historical and projected. Thus it corrects systematic distributional biases relative to historical observations and preserves model-projected relative changes (Cannon et al., 2015).~~

The MBCn multivariate bias correction algorithm by Cannon (2018) is based on the N-dimensional probability density function transform. This approach was originally developed for image processing (Pitié et al., 2007) but has been converted for post-processing climate model data. MBCn combines QDM and random orthogonal rotations to match the multivariate distributions of climate model data and observed data. In the MBCn approach, a random orthogonal rotation of the data points is applied before QDM. This exposes QDM to a linear combination of the original variables, which is then used to correct the marginal distributions of the rotated data. The QDM-corrected dataset is then rotated back and convergence to the observed multivariate distribution is checked. These steps are conducted iteratively until the multivariate distributions of ~~model~~ bias corrected climate model data and observed climate data match. In this~~ case~~ study, 100 iterations were conducted.

Both QDM and MBCn were applied in a seasonally dependent fashion. Specifically, bias corrections were applied over 30-year sliding windows. This involved replacing the central 10-years and sliding forward 10-years for each 30-yr window, until the end of the projection period was reached. Within each window – to ensure an unbiased seasonal cycle – bias corrections were applied separately for each calendar month. The combination of change-preservation by QDM, which is also a core component of MBCn, with sliding windows ensures that projected trends from the underlying climate model are largely preserved. This follows the general approach and recommendation of Hempel et al. (2013) concerning trend preservation of post-processed climate model output for impact modelling.

Climate model ~~D~~data is often simultaneously bias corrected and downscaled as the reference data stems from stations or higher resolution observations in comparison to the coarse grid resolution of RCMs. Undesirable effects in downscaling to finer scales have been one of the major limitations of current bias correction methods (Maraun, 2013; Ehret et al., 2012; Maraun et al., 2017). Such artefacts can occur especially in complex terrain and if the scale gap between climate model outputs and impact model data is considerable. In general, bias correction based on spatial resolutions that differ substantially should be avoided or handled with great care. ~~Since the hydrological model simulations in this study are~~ In this study the discrepancy in resolution is assumed acceptable as the bias correction was based on spatially aggregated mean climate variables for the meso-scale catchments (54 km² and 180 km²) with ~~the discrepancy in resolution~~ the original resolution of the underlying gridded datasets (GCM–RCM data: 0.11°, historical HYRAS data: 1 km) becoming of secondary importance~~is assumed acceptable~~.

3.2 Hydrological model simulations

The HBV model (Bergström, 1976; Lindström et al., 1997) is a semi-distributed bucket-type runoff model. Here the software implementation HBV-light (Seibert and Vis, 2012) was used, which recently has been extended to represent coupled glacio-hydrological processes of partly glacierized catchments (Seibert et al., 2018). This version of the HBV model also allows tracking the different components of streamflow resulting from rainfall (Q_R), snowmelt (Q_S), and glacier ice melt (Q_I) (Weiler et al., 2018; Seibert et al., 2018). The HBV model requires daily precipitation, air temperature, and potential evapotranspiration data as input to simulate daily runoff. In addition, linear gradients of air temperature and precipitation are needed for the interpolation over elevation zones. A general description of the basic model structure and the process conceptualization of the HBV model are found elsewhere (e.g., Lindström et al., 1997; Seibert and Vis, 2012; Seibert et al., 2018). Snow and ice accumulation and melt are based on a widely used air temperature index approach using a threshold air temperature as a model parameter to differentiate between precipitation falling as snow and rain as well as to simulate melt of snow and ice ~~melt~~ by additionally using a degree-day factor. Differences in the melt of glacier ice compared to snow are represented by another model parameter. The influence of differences in aspect on snow and ice melt was taken into account by distinguishing three aspect classes and applying an additional aspect factor parameter (Hagg et al., 2007; Hottelet et al., 1993). The latest version of the HBV-light software with the implementation of the coupled glacio-hydrological processes and the adjustment of glacier geometry to glacier mass changes based on the Δh-parametrization by Huss et al. (2010) is explained in detail in Seibert et al. (2018). It should be noted that with the implementation in HBV-light only one glacier per catchment or subcatchment can be represented. Hence, glacier cover areas in each of the two case study catchments were aggregated and simulated as one 'virtual' model glacier.

The model was calibrated for the ~~reference~~ period from 1976–200~~6~~3, preceded by a 3-year warm-up period, by optimizing a weighted objective function, giving special attention to streamflow dynamics (50%), snow simulation (25%), and glacier volume change (25%). The Lindström measure (Lindström, 1997) was used for the streamflow's general dynamic and volume errors, while the Nash–Sutcliffe efficiency (Nash and Sutcliffe, 1970) was computed based on logarithmically-transformed streamflow. Additionally the Nash–Sutcliffe efficiency was computed for the streamflow only during the summer months from June to September. ~~For evaluation of~~ To calibrate the snow simulations the snow covered area fraction of the catchment ~~was used~~ as well as the mean SWE of the elevation range ~~between 2000~~ < 2500 m a.s.l. were used.~~,~~ ~~which~~ Elevations below 2500 m a.s.l. ~~is~~ represent the crucial ~~elevation~~ range for the snow line and in this range the gridded SWE interpolation used as reference data is well-founded on station data. ~~The g~~Glacier volume was considered in the calibration process using glacier volume estimates for the years 1973 and 2003. The automated multi-criteria calibration was based on a genetic algorithm for parameter optimization (see Seibert, 2000). A 3-year model validation period (2003/10/01–2006/12/31) completed the historical reference period 1977–2006. Resulting performance measures for the calibration and validation period are summarized in Table 3 (see Supplement for additional figures comparing simulated variables and reference data). The retreat of the glaciers required all experiments to be run in a transient mode, i.e. the model was forced with climate model scenario data for the period from October 1976 to September 2099.

Table 3: Model performance criteria for the calibration (1976/-10/-01–2003/-09/-30) and validation (2003/10/01–2006/12/31) of the hydrological model formulated (see footers) that the ideal value for a perfect fit is 1.0.

Model performance criteria	Weight in calibration	Hinterrhein		Schwarze Lütschine	
		Calibration	Validation	Calibration	Validation
Nash–Sutcliffe efficiency (R_{eff}) [5] for streamflow	-	0.773	0.763	0.910	0.880
Kling–Gupta efficiency[6] for streamflow	-	0.861	0.877	0.934	0.898
Volume error (V) [7] for streamflow	-	0.972	0.962	1.000	0.965
Lindström measure [8] for streamflow	0.20	0.770	0.759	0.910	0.877
R_{eff} [5] for log transformed streamflow	0.15	0.840	0.648	0.908	0.749
R_{eff} [5] for streamflow in Jun–Sep	0.15	0.684	0.711	0.795	0.749
Root mean square error for snow covered area fraction [9]	0.10	0.856	0.761	0.863	0.803
Mean absolute normalized error (MANE) for SWE [10]	0.20	0.642	0.557	0.757	0.553
Glacier volume change objective function [11]	0.20	0.999998	-	0.999994	-

Formulation of model performance criteria:

[5] $R_{eff} = 1 - \frac{\sum(Q_{obs}-Q_{sim})^2}{(Q_{obs}-\bar{Q}_{obs})^2}$ where Q_{obs} and Q_{sim}, respectively, are observed and simulated streamflow [mm/day]

[6] see Gupta et al. (2009)

[7] $1 - V$ with $V = \frac{\sum|(Q_{obs}-Q_{sim})|}{\sum(Q_{obs})}$ where Q_{obs} and Q_{sim}, respectively, are observed and simulated streamflow [mm/day]

[8] $1 - R_{eff} - 0.1\,V$ with R_{eff} [5] and V [6] see Lindström et al. (1997)

[9] $1 - \sqrt{\frac{1}{n}(C_{sim} - C_{ref})^2}$ with C_{ref}: snow covered catchment area fraction (C) [-] as per gridded SWE reference data; C_{sim}: simulated C; n: number of time steps

[10] $1 - \frac{\sum|(S_{ref}-S_{sim})|}{\sum S_{obs}}$ *with* S [mm]: mean SWE for elevation range below 2500 m a.s.l. where S_{ref} is derived from SWE reference data and S_{sim} is simulated

[11] $1 - \frac{|\Delta W_{sim}-\Delta W_{obs}|}{\Delta W_{obs}}$ with ΔW [mm]: change of glacier ice volume in water equivalent between the years 1973 and 2003, where ΔW_{obs} corresponds to an estimate based on observed glacier area and ΔW_{sim} is simulated

3.3 Data analysis

Effects of the bias correction approaches on the hydrological simulation were based on comparisons of the simulation results for the historical reference period 1976–2006 using P and T_a time series derived from the HYRAS datasets as input (Sim_{HOCD}) and simulations forced with P and T_a series from the output of the ten different GCM–RCMs for the two different RCP scenarios, each uncorrected (Sim_{noBC}) and bias corrected based on QDM (Sim_{QDM}) and on MBCn (Sim_{MBCn}). In total, this le~~a~~d~~s~~ to 61 hydrological model runs (1 Sim_{HOCD}, 20 Sim_{noBC}, 20 Sim_{QDM}, and 20 Sim_{MBCn}) per catchment. In a first step (Results Section 4.1), the different P and T_a series were evaluated for the amount of precipitation occurring at air temperatures above and below of 0 °C due to the importance for the simulation of snow accumulation and melt processes~~, since within HBV-light, as in many other hydrological models, the air temperature determines the state of precipitation~~.

Furthermore, the simulation results were assessed in terms of ~~the~~ SWE, glacier ice volume (V_I) evolution (Results Section 4.2), and eventually streamflow with its three individual components Q_R, Q_S, and Q_I (Results Section 4.3).

4 Results

4.1 Climate variables bias correction

The two applied bias correction methods le~~a~~d to differences concerning the interdependence of P and T_a. The distribution of annual precipitation sums during air temperatures above and below 0 °C of the entire ensemble is represented in Fig. 2, while results for the individual GCM–RCM output series are provided in the Supplement. Generally, the uncorrected climate model data (noBC) have a wider variability than the reference data (HOCD). Particularly for the Schwarze Lütschine the uncorrected data yielded precipitation amounts remarkably higher than historically observed. However, differences also existed between the correction methods. For both catchments precipitation falling above air temperatures of 0 °C was overestimated with QDM. Accordingly, precipitation falling below air temperatures of 0 °C was underestimated in the univariate bias corrected data. MBCn appears to have better reproduced the historical reference data in this respect.

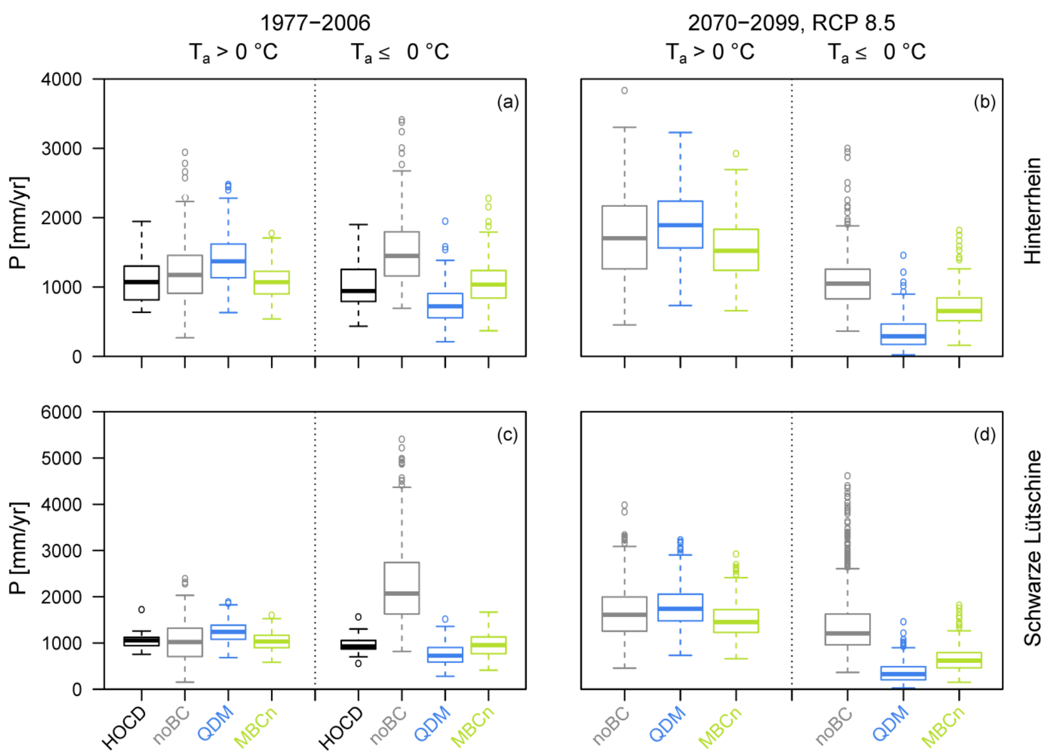

Figure 2: Annual precipitation sums for days with air temperatures above or below 0 °C.

4.2 Hydrological model simulations – cryosphere

Application of the climate scenarios clearly revealed a decreasing role of snow for both study catchments. Figure 3~~Figure 3~~ illustrates a distinctly smaller snow accumulation in the course of a year simulated for the period 2070–2099 compared to the historical reference period (1977–2006) and a more complete melt during the summer. This extended the snow free period during the summer in the Hinterrhein catchment. The spread between the simulations divergesd for the ~~future~~ simulations of future conditions. In the Schwarze Lütschine catchment with its higher maximum elevations all effects weare comparable, yet a permanent snow cover ~~was~~ remained still present based on most scenarios. As expected, simulations based on the RCP 4.5 scenario (not shown) led to a clear but less severe decrease in mean SWE than for the RCP 8.5 scenario.

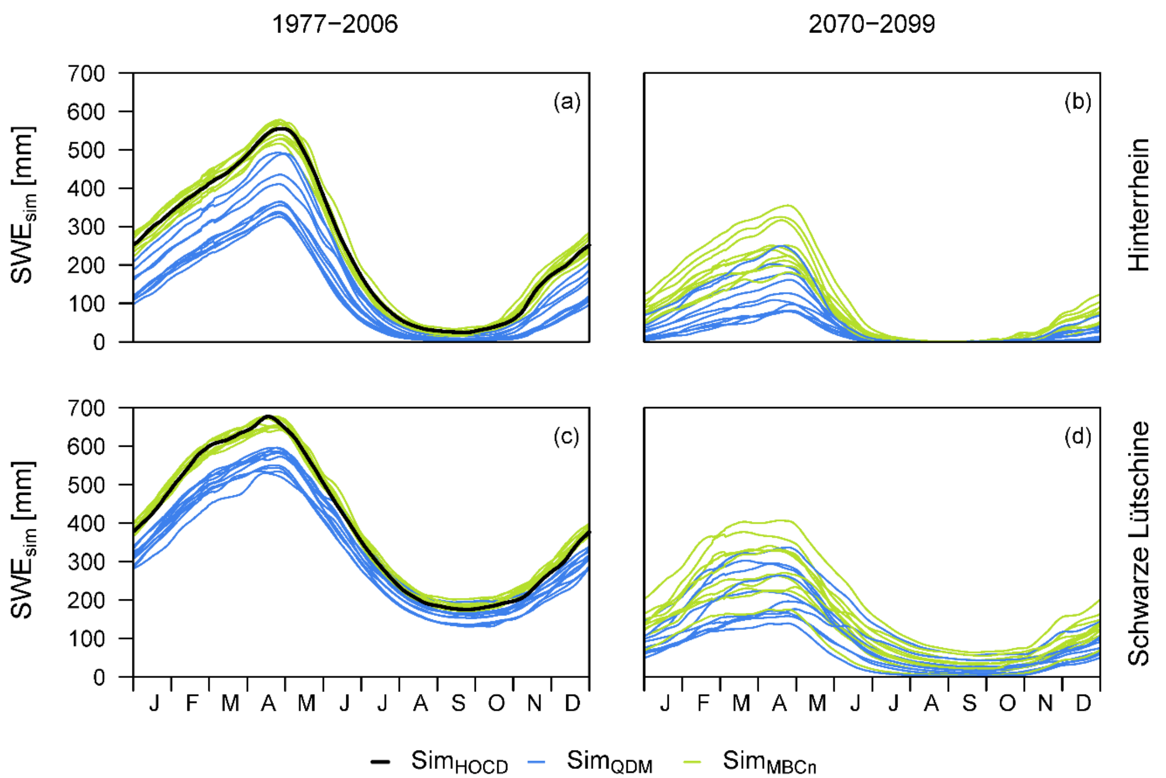

Figure 3: Mean annual SWE regime, calculated using the 11-day moving average of daily simulated SWE (catchment mean) for a~~)~~ and c~~)~~ the historical reference period and b~~)~~ and d~~)~~ at the end of the scenario period based on the RCP 8.5 scenario.

The differences in the interdependence of precipitation and air temperature resulting from the application of QDM versus MBCn to the GCM–RCM data can be seen in the simulated SWE (Fig. 3). The state of precipitation defined by the calibrated threshold air temperature parameter TT (Schwarze Lütschine TT = -0.~~47~~29 °C; Hinterrhein TT = -0.73 °C) influenced the snow accumulation and therefore led to differences in the annual SWE regime (Fig. 3). As MBCn-corrected

GCM–RCM data caused more precipitation to fall as snow, the accumulated catchment mean SWE in spring was simulated to be up to around 100–200 mm higher ~~during snow accumulation~~ in the historical reference period compared to simulations based on QDM-corrected forcing data ~~for both study catchments~~. Simulated SWE based on the two different bias correction methods differed notably. Comparing the results with the reference simulation (Fig. 3~~Sim_{HOCD}~~) indicates that MBCn performed better. The systematic difference in simulated SWE resulting from the bias correction methods was a bit less clear for the Schwarze Lütschine catchment in the scenario period, yet overall the differing tendencies between QDM- and MBCn-corrected data were considerable.

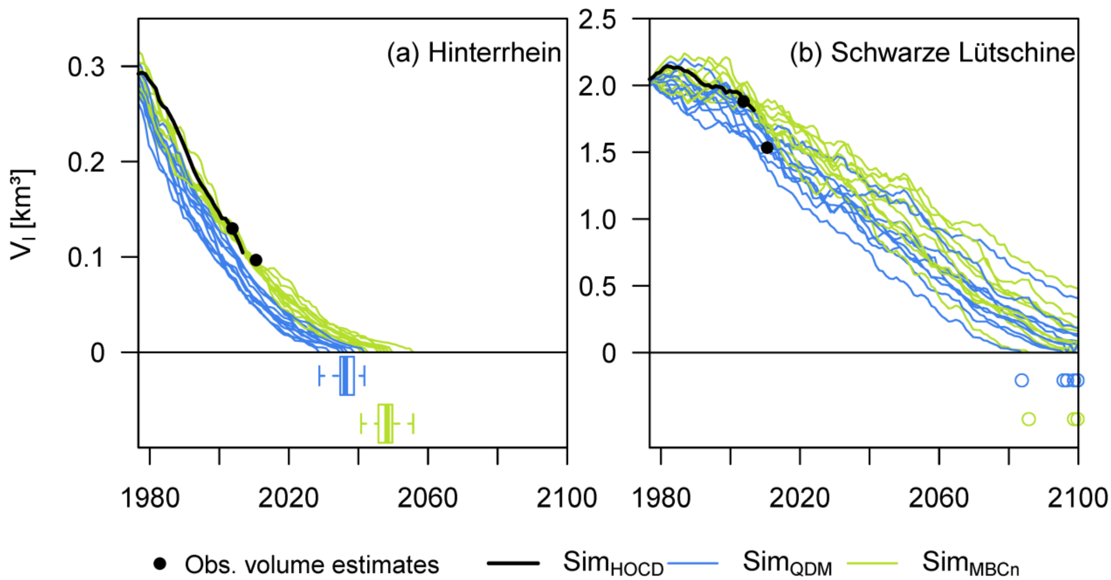

Figure 4: Simulated glacier ice volume from 1977 to ~~2100~~2099 using the RCP 8.5 scenario forcing in the two catchments (a, b). In the lower part of the graphs the boxes in the left figure and the dots in both figures indicate the simulated years of the complete glacier ice melt. For the Schwarze Lütschine only 5 (3) out of the 10 Sim_{QDM} (Sim_{MBCn}) simulations led to complete glacier melt by the end of 2099, not allowing to show any boxplots. Filled black circles are glacier volume estimates based on observed glacier area data in 2003 and 2010.

For the period 1976 to ~~2100~~ 2099 the glacier volume was simulated to decrease in both catchments. In the Hinterrhein catchment, glaciers diminished continuously from the beginning of the simulation period ~~for both, the RCP 4.5 and the RCP 8.5 scenario,~~ and were simulated to have disappeared between ~~2030~~2028 and 2055 under the RCP 8.5 scenario depending on the GCM–RCMs and the applied bias correction method (Fig. 4). In the Schwarze Lütschine catchment, data from a few GCM–RCMs resulted in an increase in simulated glacier volume in the 1970s and 1980s, which is in line with the historical reference simulation (Sim_{HOCD}). In the following years, glacier volume decreased continuously. In contrast to the Hinterrhein catchment, glaciers were not simulated to have disappeared by the end of ~~2100~~2099 based on the RCP 4.5 scenario (not shown). However, in the simulations the glacier volume diminished to on average roughly a third of its initial size at the beginning of the simulation period. The RCP 8.5 scenario from a few certain GCM–RCM combinations even led to complete glacier disappearance in the Schwarze Lütschine catchment within the 21st century.

Focusing on systematic differences between simulations using data corrected based on QDM and MBCn, the simulations of glacier volume showed similar tendencies as were found for SWE. For both catchments, but again more clearly for the Hinterrhein catchment, MBCn-corrected GCM–RCM data resulted in a slower decline in glacier volume in comparison to simulations based on QDM-corrected data. All projections led to complete glacier disappearance in the Hinterrhein catchment by about the year 2050 with a clear tendency towards earlier dates for QDM-based simulations (2028~~6~~–2041~~39~~, mean: 2036~~3~~) compared to MBCn-based simulations (2040~~38~~–2055~~1~~, mean: 2047~~4~~). For the Schwarze Lütschine catchment the range of QDM- and MBCn-based glacier volume simulations overlapped~~s~~ largely as simulations in general diverged considerably. However, for each individual GCM–RCM dataset, glacier melt was simulated to be faster using the QDM-corrected data compared to the MBCn-corrected data. ~~In contrast to the inconclusive results for the SWE regimes, t~~The less intense decline in glacier volumes resulting from MBCn-corrected forcing data appeared to correspond better with the reference simulation (Sim_{HOCD}) in the initial phase of the historical period and with the observation-based glacier volume estimates for the year 2003 (and also for the year 2010 in case of the Hinterrhein catchment). MBCn thus led to more realistic results for the historical reference period.

4.3 Hydrological model simulations – streamflow

Time changes of annual variables and mean monthly hydrological regimes were assessed for streamflow Q and for the individual streamflow components, i.e. the rain component Q_R, the snowmelt component Q_S, and the ice melt component Q_I. Mean annual streamflow of the study catchments ~~appeared to stay relatively unchanged~~ showed a small decrease over the entire simulation period from 1976 to ~~2100~~2099 for most simulations, while for some a slight increase was noticed (Fig. 5). However, the simulations based on different GCM–RCM outputs diverged over time. ~~Streamflow slightly decreased in some simulations and increased in others.~~ While – on average – the total annual streamflow stayed largely unchanged, its composition changed clearly. The streamflow component from glacier ice melt decreased slowly over time as the glaciers retreated. Likewise, the snowmelt component of streamflow decreased over time. On average, for the RCP 4.5 scenario's MBCn-corrected data these decreases ~~a~~were around ~~6~~14% in the Hinterrhein and 16~~0~~% in the Schwarze Lütschine~~;~~ for the RCP 8.5 scenario's QDM-corrected data they ~~are more than 60~~were around 53% in the Hinterrhein and ~~45~~33% in the Schwarze Lütschine.

The streamflow simulations ~~also~~ reflected the changes from the different bias correction methods found for the cryosphere. Simulations based on QDM-corrected data ~~mostly show higher~~ led to slightly different total streamflow than MBCn-corrected data (Fig. 5 a, d, e). These differences ~~are even~~ were much more pronounced regarding the individual streamflow components. Modelling based on QDM-corrected climate data led to an approximately 10% higher rain component of streamflow Q_R in comparison to MBCn-corrected simulations. The snowmelt component of streamflow Q_S varies proportionally, being notably smaller ~~for models~~ when using QDM-corrected GCM–RCM data. Comparing the means of the ice melt components of streamflow Q_I for the 30-year periods ~~in~~at the beginning and ~~in~~at the end of the entire simulation period showed no differences from the bias correction methods for the Hinterrhein catchment and differences in the range of only 1% for the Schwarze Lütschine catchment.

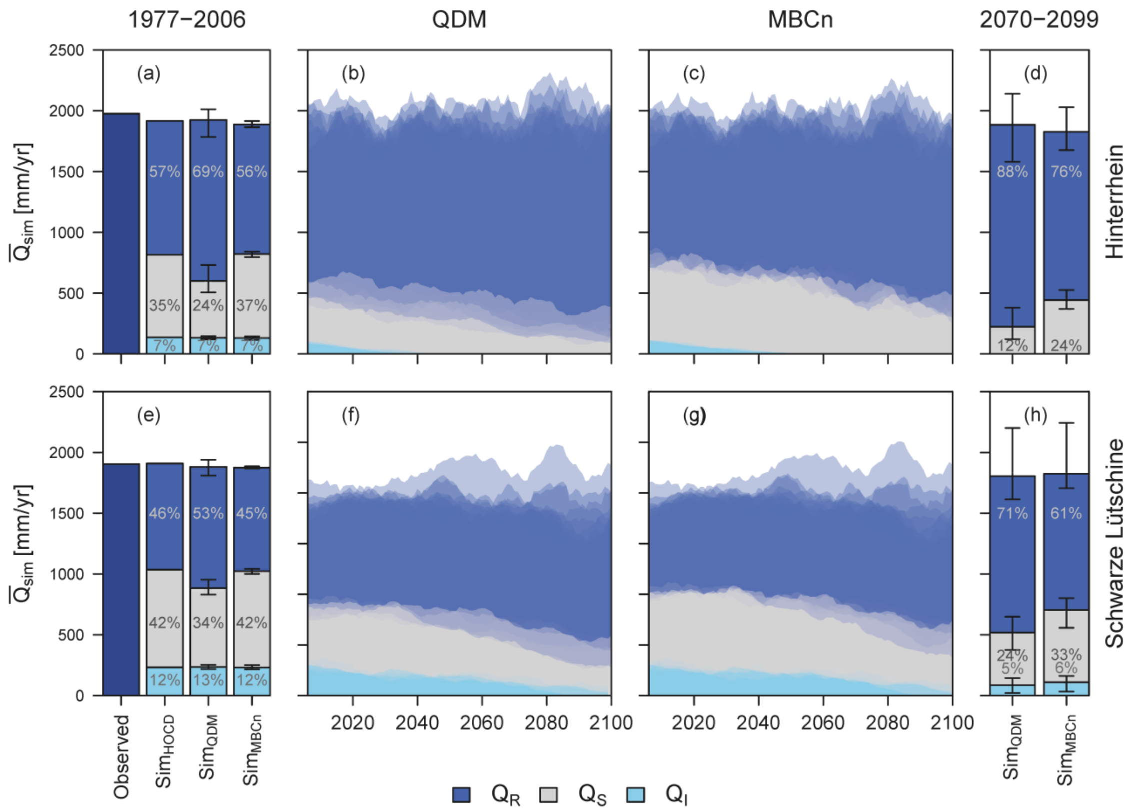

Figure 5: Observed total streamflow and simulated streamflow components for the historical reference period and for the different simulations under the RCP 8.5 scenario. Stacked bar plots show mean values over the historical reference period (~~plot~~ a, ~~and~~ e) and for the period 2070–2099 (~~plot~~ d, ~~and~~ h), stacked bar plots for $\mathrm{Sim_{QDM}}$ and $\mathrm{Sim_{MBCn}}$ show ensemble mean with ensemble spread (error bars). Simulation results over the scenario period 2006–2099 (~~plots~~ b, c, f, ~~and~~ g) are shown as semi-transparent polygons for each GCM–RCM combination.

Simulated streamflow and its components, Q_I, Q_S, and Q_R, also changed seasonally (Fig 6). In the historical reference period (1977–2006), the two catchments had a nivo-glacial streamflow regime peaking in the summer due to snow and ice melt and with little streamflow during winter. According to the projections the streamflow peak in early summer remaineds a dominant characteristic until the end of the simulation period. Yet, for the Hinterrhein catchment, the peak's timing was simulated to shift causing streamflow to concentrate in May and the peak to become much narrower than in the past. For the

Schwarze Lütschine catchment the simulations for the RCP 8.5 scenario resulted in very variable summer streamflow regimes for 2070–2099 and a tendency towards a lower summer streamflow peak than in the past. In the reference period, the glaciers' influence showed during late summer, where it extended the melt peak into autumn. This effect was simulated to diminish ~~and~~ with then decreased total streamflow in late summer. During autumn and winter, simulated streamflow for 2070–2099 ~~i~~was nearly double~~d~~ the level of the historical period ~~based on~~ mainly due to an increase in the rainfall component of streamflow. Despite similar tendencies of reduced Q_S in the future, differences arising from the different bias correction methods are notable. Q_S was more prominent in all regimes based on MBCn-corrected GCM–RCM outputs, which simulated higher peaks during the snowmelt season and a generally higher fraction during the rest of the year, especially for the future periods. Accordingly, QDM-corrected data supported a larger Q_R component beyond the summer. As a consequence, during low flow periods in winter, QDM-corrected forcing data overestimated the streamflow in the historical reference period. ~~Whereas~~ In contrast, ~~during the summer month~~ QDM-corrected forced simulations tended to slightly underestimate the streamflow during the spring and summer months, as Q_S was underestimated. Generally, MBCn-corrected data matched more closely with the reference simulations based on observed data.

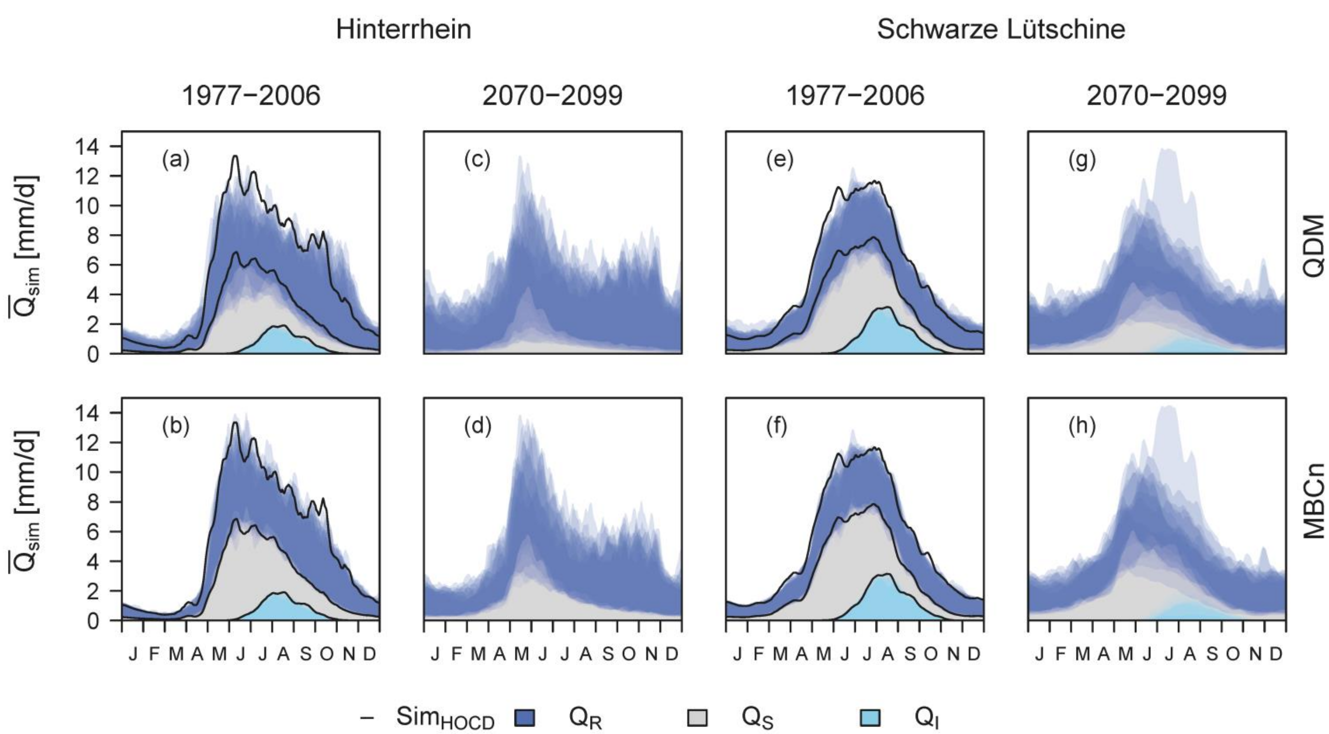

Figure 6: Streamflow regimes based on 11-day moving averages of daily streamflow during 30-year periods in the historical reference period and as projected for the period 2070–2099 under the RCP 8.5 scenario for the two catchments. Simulation results for each ensemble member are shown as semi-transparent polygons. For the historical reference period also the results of the simulations based on the historical reference *P* and T_a time series are shown (black lines).

5 Discussion

Both bias correction methods employed ~~with~~in this study, univariate QDM (Cannon et al., 2015) and multivariate MBCn (Cannon, 2018), are based on the same quantile mapping approach and by definition the marginal distributions of the corrected P and T_a series are the same as those of the historical reference data. However, the bias correction methods do result in differences in terms of P and T_a interdependency (see marginal and joint distributions of P and T_a series in the Supplement). Preserving the ranks of the climate model simulations, univariate bias correction approaches retain the inter-variable dependencies as represented in the raw climate model output (Vrac, 2018), as also demonstrated and discussed in previous studies using univariate quantile mapping methods (Wilcke et al., 2013; Ivanov and Kotlarski, 2017). However, often observed inter-variable dependencies are misrepresented in climate model simulations and hence biases therein are also retained by univariate methods (Wilcke et al., 2013; Gennaretti et al., 2015; Zscheischler et al., 2019). Such biases in the interdependency representation were also found in this study for P–T_a interdependency of raw climate model output from ten GCM–RCM simulations compared to the used historical observational dataset (see also Supplement). In snow-dominated environments, the representation of precipitation–temperature interdependence is important for hydrological modelling but also for many other aspects impacted strongly by snow cover extent and duration (Gennaretti et al., 2015). Further studies that compare P–T_a representation in climate model output and multiple observational datasets are needed to explore the causes of differences between climate model output and reference data such as those found here.

As air temperature determines the distinction between liquid precipitation and snow, differences in the climate variables' interdependence can lead to differences in simulated snowfall (Fig. 2), and consequently in snow accumulation and the catchments' seasonal water storage (Fig. 3–6). For the MBCn-corrected data in this study there was clearly more precipitation at air temperatures below 0 °C in comparison to the QDM-corrected data, resulting in more precipitation falling as snow, being stored, and accumulated than for univariate bias corrected forcing data. In glacierized catchments the higher amounts of snow from MBCn compared to QDM also affected the glaciers with higher winter mass balances and a later start of the melt season in spring/summer. The existence or non-existence of water storages in the form of snow and ice as well as the liquid precipitation directly contributing to streamflow had~~ve~~ notable influences on the streamflow composition and regime. For instance, the larger fraction of liquid precipitation at the cost of snow simulated with QDM-corrected data led to a systematic overestimation of streamflow during the winter months in the historical reference period. This error was not present in simulations based on MBCn-corrected P and T_a forcing.

It bears noting that results from QDM and MBCn in the historical reference period are, as for example also in Zscheischler et al. (2019), evaluated without cross-validation. However, because the univariate and multivariate bias correction algorithms are applied in an asynchronous fashion to freely running climate simulations – adjusting the marginal/joint distributions – it is, by construction, almost guaranteed that they will perform well in terms of cross-validated measures of distributional fit. Cross-validation does make sense when performance – especially for aspects not explicitly adjusted – is measured in a setting where climate model simulations are synchronized with the real-world climate state, for example in climate prediction or perfect boundary condition (e.g., reanalysis-driven) setups. We note that such reanalysis-driven cross-validation experiments have been performed in Cannon (2018) for the two algorithms used in this study. This was done over a large continental domain for a complicated multivariate fire weather index that combines, in a nonlinear fashion, the current and lagged effects of air temperature, precipitation, wind, and humidity. Hence, it is expected that results reported here are robust and would be similar in an out-of-sample evaluation.

There have long been concerns over climate change impacts on mountain water towers. Many climate impact studies for ~~alpine/~~snow-dominated catchments agree that due to continued warming, a decrease in snow cover characteristics and time-shifted snowmelt contributions to streamflow are to be expected under climate change scenarios (e.g. Barnett et al., 2005; Farinotti et al., 2012; Köplin et al., 2014; Addor et al., 2014; Milano et al., 2015; Coppola et al., 2016; Jenicek et al., 2018; Hanzer et al., 2018). In fact, the shift and ~~loss~~ decrease of the snowmelt peak ~~is~~ are one of the most robust results of such studies. In this study we showed that the ~~magnitude of decrease in the~~ snow component strongly depends not only on the GCM–RCM outputs but also on ~~the~~ whether the bias correction method applied incorporates inter-variable dependence of P and T_a or not. The simulated glacier volume ~~also~~ showed a clearly decreasing trend over the scenario period. However, net mass balances and hence rates of glacier ice melt and the mean timing of the final glacier disappearance vary~~ies~~ by over a decade in the Hinterrhein catchment. While the ensemble covers a wide range, the bias correction approach makes a difference for each GCM–RCM forcing. The changes ~~of~~ in snow accumulation and glacier melt then propagate into changes of streamflow regimes. In future projections, snowmelt peaks tend to occur earlier and with a more concentrated melt season. A ~~delayed~~ potential effect ~~visible only from~~ of this storage shift on streamflow however is potentially relevant year-round as could be visualized by the specific streamflow component modelling~~ is the potential contribution of stored water to streamflow year-round~~. The simulations suggest that ~~this~~ the melt contribution to streamflow depends on ~~the chosen bias correction method and hence~~ the interdependence of air temperature and precipitation and hence the chosen bias correction method. Furthermore, streamflow during the late summer decreases as the release of stored water from glaciers, which makes up a notable percentage of streamflow during the late summer, will have diminished. ~~Rate and timing of all of these effects are influenced by the bias correction method applied to the GCM–RCM data.~~ These systematic differences in hydrological impact scenarios originating from the applied univariate or multivariate bias correction method such as those found here, e.g. differences in glacier disappearance dates or differences in seasonal (summer vs. winter) water availability, may appear negligible given the overall large uncertainties of climate impact modelling yet may still be ~~of~~ relevant~~ce~~ for some specific adaptation management questions. The timing of ‘peak water’ occurrence or complete disappearance of glaciers may be relevant for the planning horizon of hydropower schemes (Hänggi and Weingartner, 2012; Schaefli~~, 2015~~ et al., 2019). The ~~E~~earlier recession of the melt peak may sooner or later affect early-summer flood hazard or increase the hazard of late-summer low flow due to the loss of ice and snow components of streamflow (Beaulieu et al., 2012; Godsey et al., 2014) requiring the planning of respective measures.

These ~~study's~~ results also require discussion of implications on common conceptual hydrological modelling concepts that are needed to simplify meteorological and hydrological complexity. The use of a threshold air temperature for the distinction of precipitation in snow and rainfall is a key concept of the HBV model and many other hydrological models. Hence, it may be expected that the simulations of the snow-dominated catchments respond particularly sensitive to changes~~/~~ and biases in P–T_a interdependencies. The question is the degree to which this may influence the hydrological variables discussed above. So far, few studies have evaluated multivariate-corrected GCM–RCM data in hydrological modelling. Chen et al. (2018) found that the joint bias correction of precipitation and air temperature led to a much better performance in terms of hydrological modelling for all their study basins located in various climates except for the coldest Canadian basin. In contrast, an overall additional benefit of using bivariate bias correction methods for hydrological impact projections was not evident in results by Räty et al. (2018) when compared to using a univariate quantile mapping applied as a delta change method, i.e. retaining present-day correlation structures. However, their analysis ~~of SWE simulations~~ indicated that the selection of the bias correction method was most important and the added value of using multivariate approaches most clearly found for ~~this hydrological variable~~ SWE simulations, supporting the findings of this study. Based on these case studies, it may be assumed that simulations with any hydrological model that include calibration over a historical reference period will be somewhat affected by a biased representation of inter-variable dependence of its input variables in GCM–RCM outputs. Further studies are needed to investigate other effects of multivariate bias correction for other types of climatological input variables, hydrological models, catchment types, and dominating processes.

This study demonstrates the importance of considering the representation of the interdependence of precipitation and air temperature in the specific case of hydrological impact modelling of snow and glacier dominated catchments. As shown, in the representation of the climate variables' interdependence, the multivariate bias correction approach leads to results closer to the climatological historical reference data as well as partly to hydrological simulations closer to the historical reference simulations as for instance for the simulated glacier volumes. Cannon (2016, 2018) also demonstrated better results for multivariate-corrected data in other examples, including fire weather indices and atmospheric river detection. In practice, some kind of bias correction is needed for many impact studies, although it is known that recent literature is rich in controversial debate of its use and major limitations of the application of empirical-statistical bias correction methods (e.g. Ehret et al., 2012; Addor and Seibert, 2014; Maraun, 2013, 2016; Clark et al., 2016; Maraun et al., 2017; Casanueva et al., 2018; Zscheischler et al., 2019). Some of the fundamental issues, the details of which ~~will be~~ are beyond the scope of this study, are shared with univariate bias correction, for example, the question of stationarity (regarding biases in marginal distributions). In addition, joint correction is often based on the assumption that the structure of the bias in variables' interdependence is stationary, i.e. the same for control as for projections. This is not strictly true for MBCn, which allows the multivariate distribution to evolve in the projection period. However, the extent to which model projected changes in dependence structure are preserved by MBCn have yet to be evaluated closely. More generally, whether the preservation of inter-variable dependence structures is a robust assumption or dependence structures should evolve from the reference to the future period are still open questions for multivariate bias correction methods development (Vrac, 2018). Furthermore, the correction of the multivariate dependence structure will necessarily affect the time sequencing of the climate model variables (Cannon, 2016), which can lead to modification of temporal autocorrelation. Maraun (2016) cautions that modifications of spatial, temporal or multi-variable interdependence may break the consistency with the driving climate model and many others have argued for the least possible transformation of GCM–RCM outputs for this reason. This study does not address these fundamental questions and critiques nor does it generally recommend or not recommend the use of multivariate bias correction methods. The objective of the study was to compare the differences resulting from univariate vs. multivariate methods. We demonstrated a case in which biases in inter-variable dependencies can affect hydrological simulations considerably. This is~~ notable~~ important, particularly as it is common practice to use hydrological models calibrated to climatic conditions represented by historical climate variable series. In the same way that the use of several climate and hydrological models is recommended, the incorporation of uncorrected, univariate-, and multivariate-corrected scenario data in the ensemble may be considered as one part of a transparent and honest communication of the full range of uncertainties.

6 Conclusions

~~In t~~This study systematically tested the effects of ~~a univariate and a ~~multivariate bias correction of projected air temperature and precipitation versus a traditional univariate bias correction on hydrological impact modelling in alpine environments~~were compared~~. Jointly corrected air temperature and precipitation series simulated more snowfall and consequently up to 50% more snow accumulation than univariate-corrected GCM–RCM data. Subsequently, glacier volume was simulated to decrease by up to a decade slower under multivariate-corrected scenarios. These differences also impact the simulations of streamflow and its components with higher snowmelt components and accordingly smaller rainfall components under multivariate-corrected scenarios compared to univariate-corrected scenarios. These are relevant systematic differences despite variations of the GCM–RCM ensemble. ~~The interdependence of air temperature and precipitation is hence of such importance that multivariate correction methods perform more accurately compared to univariate quantile mapping approaches. ~~The~~y~~ ~~of~~ choice between a univariate and a multivariate bias correction approach may therefore have implications for future water resources planning, as the snow component presents an important seasonal storage, and for the protection against hydrological hazards such as a higher vulnerability to drought.

~~The study therefore demonstrates a case, where the interdependence of air temperature and precipitation is of such importance that multivariate correction methods perform more accurately compared to univariate quantile mapping approaches. The results achieved by these two different~~Beyond the specific case this study suggests that the effect of bias correction methods may be generaliz~~able~~ed for ~~other ~~catchments that include the elevation range of the snow line.~~ Especially in alpine catchments,~~ Mountain hydrology modelling relies on the correct representation of the interdependence of air temperature and precipitation ~~plays ~~due to a crucial role ~~in hydrological modelling that uses a ~~of threshold air temperature concepts for the distinction of liquid and solid precipitation. This study ~~can therefore be regarded as~~makes an argument for the explicit consideration of interdependencies of climate variables by using multivariate bias correction methods in hydrological climate change impact studies in snow-dominated catchments. But also many other threshold effects drive relevant climate impacts and are parameterized in many models or indices. ~~It supports a call to study ~~The study provides a strong incentive to test similar effects in hydrological systems and their model representations that may be dominated by other climate variable interdependencies.

Code availability

An R package (R Core Team 2015) including the MBCn and the QDM algorithm is available for download from https://CRAN.R-project.org/package=MBC. The HBV-light software is freely available for download from https://www.geo.uzh.ch/en/units/h2k/Services/HBV-Model.html.

Data availability

EURO-CORDEX data can be accessed via different European datanodes, available at https://www.euro-cordex.net/060378/index.php.en. The HYRAS interpolation product used to derive the historical reference climate time series was made available by the German Weather Service (DWD) and the German Federal Institute of Hydrology (BfG). Streamflow time series were provided by the Swiss Federal Office for the Environment (FOEN) and the Amt für Wasser und Abfall des Kantons Bern. Snow data of the "SLF-Schneekartenserie Winter 1972-2012" used for model calibration are available upon request by the WSL Institute for Snow and Avalanche Research (SLF). Glacier ice thickness data were provided by Matthias Huss, other glacier data are available according to the given references.

Author contribution

JM, IK, KS, and JS designed the study. JM carried out bias correction, modelling, and all analyses and wrote the first draft. IK calibrated the hydrological model and prepared snow, glacier, and hydrological data. KH prepared the EURO-CORDEX data for the catchments. AC provided and helped with his bias correction scripts. All co-authors contributed to and edited the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

Work for this study was based on data acquired and methods developed within the project 'The snow and glacier melt components of the streamflow of the River Rhine and its tributaries considering the influence of climate change' (ASG-Rhein, see Stahl et al., 2017) funded by the International Commission for the Hydrology of the Rhine basin (CHR). We thank Dr. Urs Beyerle for his assistance with the retrieval of EURO-CORDEX data and further thank all data providers (see ‘Data availability’).

References

Addor, N., Rössler, O., Köplin, N., Huss, M., Weingartner, R., and Seibert, J.: Robust changes and sources of uncertainty in the projected hydrological regimes of Swiss catchments, Water Resour. Res., 50, 7541–7562, doi:10.1002/2014WR015549, 2014.

Addor, N. and Seibert, J.: Bias correction for hydrological impact studies - beyond the daily perspective, Hydrol. Process., 28, 4823–4828, doi:10.1002/hyp.10238, 2014.

Barnett, T. P., Adam, J. C., and Lettenmaier, D. P.: Potential impacts of a warming climate on water availability in snow-dominated regions, Nature, 438, 303–309, doi:10.1038/nature04141, 2005.

Beaulieu, M., Schreier, H., and Jost, G.: A shifting hydrological regime: A field investigation of snowmelt runoff processes and their connection to summer base flow, Sunshine Coast, British Columbia, Hydrol. Process., 26, 2672–2682, doi:10.1002/hyp.9404, 2012.

Bergström, S.: Development and application of a conceptual runoff model for Scandinavian catchments, SMHI, Norrköping, 1976.

Bosshard, T., Kotlarski, S., Zappa, M., and Schär, C.: Hydrological Climate-Impact Projections for the Rhine River: GCM–RCM Uncertainty and Separate Temperature and Precipitation Effects. J. Hydrometeor, 15, 697–713, doi:10.1175/JHM-D-12-098.1, 2014.

Cannon, A. J.: Multivariate Bias Correction of Climate Model Output: Matching Marginal Distributions and Intervariable Dependence Structure, J. Climate, 29, 7045–7064, doi:10.1175/JCLI-D-15-0679.1, 2016.

Cannon, A. J.: Multivariate quantile mapping bias correction: An N-dimensional probability density function transform for climate model simulations of multiple variables, Clim Dyn, 50, 31–49, doi:10.1007/s00382-017-3580-6, 2018.

Cannon, A. J., Sobie, S. R., and Murdock, T. Q.: Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes?, J. Climate, 28, 6938–6959, doi:10.1175/JCLI-D-14-00754.1, 2015.

Casanueva, A., Bedia, J., Herrera, S., Fernández, J., and Gutiérrez, J. M.: Direct and component-wise bias correction of multi-variate climate indices: The percentile adjustment function diagnostic tool, Climatic Change, 121, 2075, doi:10.1007/s10584-018-2167-5, 2018.

Chen, J., Brissette, F. P., Chaumont, D., and Braun, M.: Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America, Water Resour. Res., 49, 4187–4205, doi:10.1002/wrcr.20331, 2013.

Chen, J., Li, C., Brissette, F. P., Chen, H., Wang, M., and Essou, G. R.C.: Impacts of correcting the inter-variable correlation of climate model outputs on hydrological modeling, Journal of Hydrology, 560, 326–341, doi:10.1016/j.jhydrol.2018.03.040, 2018.

Clark, M. P., Wilby, R. L., Gutmann, E. D., Vano, J. A., Gangopadhyay, S., Wood, A. W., Fowler, H. J., Prudhomme, C., Arnold, J. R., and Brekke, L. D.: Characterizing Uncertainty of the Hydrologic Impacts of Climate Change, Curr Clim Change Rep, 2, 55–64, doi:10.1007/s40641-016-0034-x, 2016.

Coppola, E., Raffaele, F., and Giorgi, F.: Impact of climate change on snow melt driven runoff timing over the Alpine region, Clim Dyn, 76, 503, doi:10.1007/s00382-016-3331-0, 2016.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions "Should we apply bias correction to global and regional climate model data?" Hydrol. Earth Syst. Sci., 16, 3391–3404, doi:10.5194/hess-16-3391-2012, 2012.

Farinotti, D., Usselmann, S., Huss, M., Bauder, A., and Funk, M.: Runoff evolution in the Swiss Alps: Projections for selected high-alpine catchments based on ENSEMBLES scenarios, Hydrol. Process., 26, 1909–1924, doi:10.1002/hyp.8276, 2012.

Fischer, M., Huss, M., Barboux, C., and Hoelzle, M.: The New Swiss Glacier Inventory SGI2010: Relevance of Using High-Resolution Source Data in Areas Dominated by Very Small Glaciers, Arctic, Antarctic, and Alpine Research, 46, 933–945, doi:10.1657/1938-4246-46.4.933, 2014.

Freudiger, D., Mennekes, D., Seibert, J., and Weiler, M.: Historical glacier outlines from digitized topographic maps of the Swiss Alps, Earth Syst. Sci. Data (Earth System Science Data), 10, 805–814, doi:10.5194/essd-10-805-2018, 2018.

Frick, C., Steiner, H., Mazurkiewicz, A., Riediger, U., Rauthe, M., Reich, T., and Gratzki, A.: Central European high-resolution gridded daily data sets (HYRAS): Mean temperature and relative humidity, metz, 23, 15–32, doi:10.1127/0941-2948/2014/0560, 2014.

Gennaretti, F., Sangelantoni, L., and Grenier, P.: Toward daily climate scenarios for Canadian Arctic coastal zones with more realistic temperature-precipitation interdependence, J. Geophys. Res. Atmos., 120, 11,862–11,877, doi:10.1002/2015JD023890, 2015.

Gobiet, A., Kotlarski, S., Beniston, M., Heinrich, G., Rajczak, J., and Stoffel, M.: 21st century climate change in the European Alps—A review, Science of The Total Environment, 493, 1138–1151, doi:10.1016/j.scitotenv.2013.07.050, 2014.

Godsey, S. E., Kirchner, J. W., and Tague, C. L.: Effects of changes in winter snowpacks on summer low flows. Case studies in the Sierra Nevada, California, USA, Hydrol. Process., 28, 5048–5064, doi:10.1002/hyp.9943, 2014.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T.: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations - a comparison of methods, Hydrol. Earth Syst. Sci., 16, 3383–3390, doi:10.5194/hess-16-3383-2012, 2012.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, Guillermo, F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

Hagg, W., Braun, L. N., Kuhn, M., and Nesgaard, T. I.: Modelling of hydrological response to climate change in glacierized Central Asian catchments, Journal of Hydrology, 332, 40–53, doi:10.1016/j.jhydrol.2006.06.021, 2007.

Hakala, K., Addor, N., and Seibert, J.: Hydrological modeling to evaluate climate model simulations and their bias correction~~: Manuscript re-submitted to the~~ J~~.ournal of~~ Hydrometeor~~ology after minor revisions~~, 19(8), 1321–337, doi: 10.1175/JHM-D-17-0189.1, 2018.

Hänggi, P. and Weingartner, R.: Variations in Discharge Volumes for Hydropower Generation in Switzerland, Water Resour Manage, 26, 1231–1252, doi:10.1007/s11269-011-9956-1, 2012.

Hanzer, F., Förster, K., Nemec, J., and Strasser, U.: Projected cryospheric and hydrological impacts of 21st century climate change in the Ötztal Alps (Austria) simulated using a physically based approach, 22, 1593–1614, 2018.

Hock, R.: Glacier melt: A review of processes and their modelling, Progress in Physical Geography, 29, 362–391, doi:10.1191/0309133305pp453ra, 2005.

Hottelet, C., Braun, L. N., Leibundgut, C., and Rieg, A.: Simulation of snowpack and discharge in an alpine karst basin, IAHS Publications, 218, 249–260, 1993.

Huss, M. and Farinotti, D.: Distributed ice thickness and volume of all glaciers around the globe, J. Geophys. Res., 117, n/a-n/a, doi:10.1029/2012JF002523, 2012.

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

Huss, M., Jouvet, G., Farinotti, D., and Bauder, A.: Future high-mountain hydrology: A new parameterization of glacier retreat, Hydrol. Earth Syst. Sci., 14, 815–829, doi:10.5194/hess-14-815-2010, 2010.

Ivanov, M. A. and Kotlarski, S.: Assessing distribution-based climate model bias correction methods over an alpine domain: Added value and limitations, Int. J. Climatol., 37, 2633–2653, doi:10.1002/joc.4870, 2017.

Jansson, P., Hock, R., and Schneider, T.: The concept of glacier storage: A review, Journal of Hydrology, 282, 116–129, doi:10.1016/S0022-1694(03)00258-0, 2003.

Jenicek, M., Seibert, J., and Staudinger, M.: Modeling of Future Changes in Seasonal Snowpack and Impacts on Summer Low Flows in Alpine Catchments, Water Resour. Res., 54, 538–556, doi:10.1002/2017WR021648, 2018.

Köplin, N., Rößler, O., Schädler, B., and Weingartner, R.: Robust estimates of climate-induced hydrological change in a temperate mountainous region, Climatic Change, 122, 171–184, doi:10.1007/s10584-013-1015-x, 2014.

Köplin, N., Viviroli, D., Schädler, B., and Weingartner, R.: How does climate change affect mesoscale catchments in Switzerland? - a framework for a comprehensive assessment, Adv. Geosci., 27, 111–119, doi:10.5194/adgeo-27-111-2010, 2010.

Kotlarski, S., Zubler, E., Fischer, A., Winter, K. J. P., Gobiet, A., and Liniger, M. A.: Patterns of 21st Century Climate Change in the European Alps: The CORDEX RCM ensembles, EGU General Assembly Conference Abstracts, 18, 2016.

Li, C., Sinha, E., Horton, D. E., Diffenbaugh, N. S., and Michalak, A. M.: Joint bias correction of temperature and precipitation in climate model simulations, J. Geophys. Res. Atmos., 119, 13, doi:10.1002/2014JD022514, 2014.

Lindström, G.: A simple automatic calibration routine for the HBV model, Hydrol Res, 28, 153–168, 1997.

Lindström, G., Johansson, B., Persson, M., Gardelin, M., and Bergström, S.: Development and test of the distributed HBV-96 hydrological model, Journal of Hydrology, 201, 272–288, doi:10.1016/S0022-1694(97)00041-3, 1997.

Maisch, M., Wipf, A., Denneler, B., Battaglia, J., and Benz, C.: Die Gletscher der Schweizer Alpen: Gletscherhochstand 1850. Aktuelle Vergletscherung, Gletscherschwundszenarien, vdf Hochschulverlag, Zurich, 2000.

Maraun, D.: Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue, J. Climate, 26, 2137–2143, doi:10.1175/JCLI-D-12-00821.1, 2013.

Maraun, D.: Bias Correcting Climate Change Simulations - a Critical Review, Curr Clim Change Rep, 2, 211–220, doi:10.1007/s40641-016-0050-x, 2016.

Maraun, D., and Widmann, M.: Cross-validation of bias-corrected climate simulations is misleading. Hydrol. Earth Syst. Sci., 22(9), 4867-4873, 2018.

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., Hagemann, S., Richter, I., Soares, P. M. M., Hall, A., and Mearns, L. O.: Towards process-informed bias correction of climate change simulations, Nature Climate change, 7, 664–773, doi:10.1038/nclimate3418, 2017.

Mehrotra, R. and Sharma, A.: Correcting for systematic biases in multiple raw GCM variables across a range of timescales, Journal of Hydrology, 520, 214–223, doi:10.1016/j.jhydrol.2014.11.037, 2015.

Mehrotra, R. and Sharma, A.: A Multivariate Quantile-Matching Bias Correction Approach with Auto- and Cross-Dependence across Multiple Time Scales: Implications for Downscaling, J. Climate, 29, 3519–3539, doi:10.1175/JCLI-D-15-0356.1, 2016.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J.-F., Matsumoto, K., Montzka, S. A., Raper, S. C. B., Riahi, K., Thomson, A., Velders, G. J. M., and van Vuuren, D. P.P.: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, Climatic Change, 109, 213–241, doi:10.1007/s10584-011-0156-z, 2011.

Messerli, B., Viviroli, D., and Weingartner, R.: Mountains of the World: Vulnerable Water Towers for the 21st Century, Ambio, 29–34, 2004.

Milano, M., Reynard, E., Bosshard, N., and Weingartner, R.: Simulating future trends in hydrological regimes in Western Switzerland, Journal of Hydrology: Regional Studies, 4, 748–761, doi:10.1016/j.ejrh.2015.10.010, 2015.

Moore, R. D., Fleming, S. W., Menounos, B., Wheate, R., Fountain, A., Stahl, K., Holm, K., and Jakob, M.: Glacier change in western North America: Influences on hydrology, geomorphic hazards and water quality, Hydrol. Process., 23, 42–61, doi:10.1002/hyp.7162, 2009.

Müller, F., Caflish, T., and and Müller, G.: Firn und Eis der Schweizer Firn und Eis der Schweizer Alpen, Gletscherinventar, vdf Hochschulverlag, Zurich, 1976.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10, 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

OSHD-SLF: Datenblatt zur SWE-Kartenserie 1972-2012 // KHR-Projekt. Stand 30.11.2013., Operationeller Schneehydrologischer Dienst des SLF (OSHD-SLF), Davos, 2013.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall–runoff model? Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling, Journal of Hydrology, 303, 290–306, doi:10.1016/j.jhydrol.2004.08.026, 2005.

Paul, F., Frey, H., and Le Bris, R.: A new glacier inventory for the European Alps from Landsat TM scenes of 2003: challenges and results, Ann. Glaciol., 52, 144–152, doi:10.3189/172756411799096295, 2011.

Pellicciotti, F., Bauder, A., and Parola, M.: Effect of glaciers on streamflow trends in the Swiss Alps, Water Resour. Res., 46, 145, doi:10.1029/2009WR009039, 2010.

Pitié, F., Kokaram, A. C., and Dahyot, R.: Automated colour grading using colour distribution transfer, Computer Vision and Image Understanding, 107, 123–137, doi:10.1016/j.cviu.2006.11.011, 2007.

Räty, O., Räisänen, J., Bosshard, T., and Donnelly, C.: Intercomparison of Univariate and Joint Bias Correction Methods in Changing Climate From a Hydrological Perspective, Climate, 6, 33, doi:10.3390/cli6020033, 2018.

Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., and Gratzki, A.: A Central European precipitation climatology – Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS), metz, 22, 235–256, doi:10.1127/0941-2948/2013/0436, 2013.

Schaefli, B., Manso, P., Fischer, M., Huss, M., and Farinotti, D.: The role of glacier retreat for Swiss hydropower production, Renewable Energy, 132, 615–627, doi:10.1016/j.renene.2018.07.104, 2019.

~~Schaefli, B.: Projecting hydropower production under future climates: A guide for decision-makers and modelers to interpret and design climate change impact assessments, WIREs Water, 2, 271–289, doi:10.1002/wat2.1083, 2015.~~

Seibert, J.: Multi-criteria calibration of a conceptual runoff model using a genetic algorithm, Hydrol. Earth Syst. Sci., 4, 215–224, doi:10.5194/hess-4-215-2000, 2000.

Seibert, J. and Vis, M. J. P.: Teaching hydrological modeling with a user-friendly catchment-runoff-model software package, Hydrol. Earth Syst. Sci., 16, 3315–3325, doi:10.5194/hess-16-3315-2012, 2012.

Seibert, J., Vis, M. J. P., Kohn, I., Weiler, M., and Stahl, K.: Technical note: Representing glacier geometry changes in a semi-distributed hydrological model, Hydrol. Earth Syst. Sci., 22, 2211–2224, doi:10.5194/hess-22-2211-2018, 2018.

Sevruk, B.: Reliability of precipitation measurement, in: Precipitation measurements: WMO/IAHS/ETH Workshop on Precipitation Measurements, Sevruk, B. (Ed.), St. Moritz, 3-7 December 1989, Zurich, 13–19, 1989.

Stahl, K., Weiler, M., Freudiger, D., Kohn, I., Seibert, J., Vis, M., Gerlinger, K., and Böhm, M.: The snow and glacier melt components of streamflow of the river Rhine and its tributaries considering the influence of climate change. Final report to the International Commission for the Hydrology of the Rhine Basin (CHR): www.chr-khr.org/en/publications (last access: 5 June 2018), 2017.

Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, Journal of Hydrology, 456-457, 12–29, doi:10.1016/j.jhydrol.2012.05.052, 2012.

Vrac, M.: Multivariate bias adjustment of high-dimensional climate simulations: The Rank Resampling for Distributions and Dependences (R^2D^2) bias correction, Hydrol. Earth Syst. Sci., 22, 3175–3196, doi:10.5194/hess-22-3175-2018, 2018.

Weiler, M., Seibert, J., and Stahl, K.: Magic components—why quantifying rain, snowmelt, and icemelt in river discharge is not easy, Hydrological Processes, 32, 160–166, doi:10.1002/hyp.11361, 2018.

Wilcke, R. A. I., Mendlik, T., and Gobiet, A.: Multi-variable error correction of regional climate models, Climatic Change, 120, 871–887, doi:10.1007/s10584-013-0845-x, 2013.

Zscheischler, J., Fischer, E. M., and Lange, S.: The effect of univariate bias adjustment on multivariate hazard estimates. Earth System Dynamics, 10(1), 31-43, 2019.