Authors' response to the interactive comment of referee # 2 on hess-2018-317
"Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments" by Judith Meyer et al.

We thank the reviewer Sven Kotlarski for his positive evaluation of our manuscript and the helpful comments. Below we respond (in blue) to the reviewer comments (*in black*). We appreciate the efforts of the reviewer and are going to consider his valuable suggestions when revising our manuscript.

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 would not be suitable input to a hydrological model. We can follow the reasoning of the referee, yet, as there have been a few new studies now on the pure bias correction, we would like to keep focus on the hydrological application here. 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 case 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 by the referee above. Therefore we refer to our response to referee #1 regarding cross-validation (please see therein). 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).

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 intervariable 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 version of the manuscript. However, figures such as the one below might be added (as supplementary material) to the revised version, if considered helpful. 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 will extend the discussion a bit in this respect, but an ultimate answer may be beyond the scope of this study and would require separate investigations also based on an intercomparison for further observational datasets.

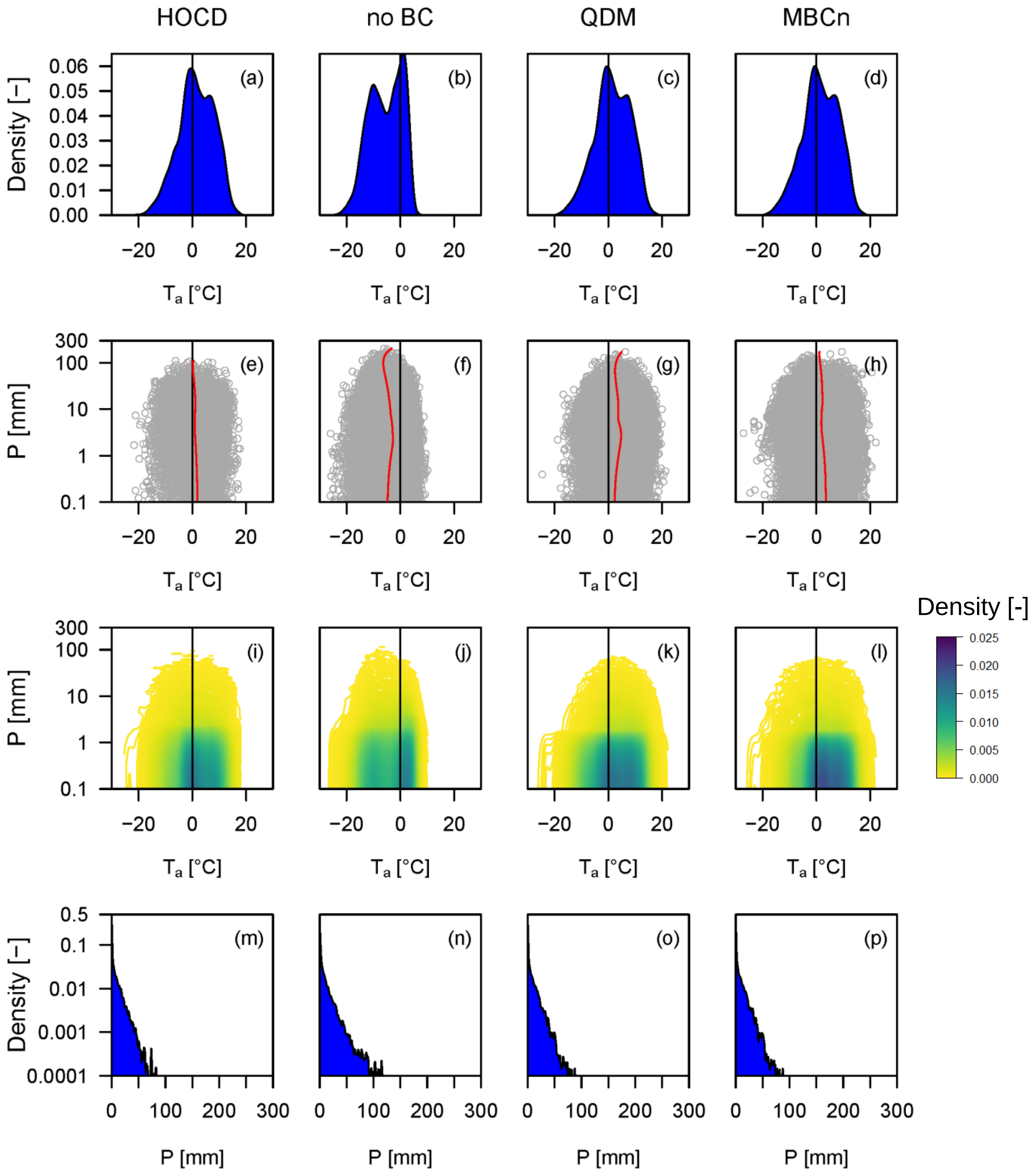

Figure A: Exemplary representation of T_a and P over the historical reference period 1977–2006 for the Schwarze Lütschine catchment according to the historically observed climate data (column HOCD) and uncorrected (column no BC), univariate-corrected (column QDM), and bivariate-corrected (column MBCn) 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).

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, we are aware of these studies and will integrate references to them in the revised version and extend the discussion. 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 will now reference the ESGF archive in relation to the download of the EURO-CORDEX data. In addition, we will expand 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 acknowledge the limitations of some of the individual EURO-CORDEX runs, which were pointed out to us during the CH2018 model selection process.

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 that this needs to be more precisely explained in the revised version. 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 calibration 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.

Please note: to address the request of referee #1 meanwhile we have already repeated the calibrations of the hydrological model with changed settings (slightly shortened calibration period to allow for a validation period and SWE calibration for extended elevation range) and carried out all model runs based on the results of this new (improved) calibrations. We will explain this in detail and update all result figures based on this re-calibration of the hydrological model in the revised version. In reference to the reviewer's comment above, please find below the new results for Figure 5. The additional calibration efforts 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 parametrisation of the hydrological model.

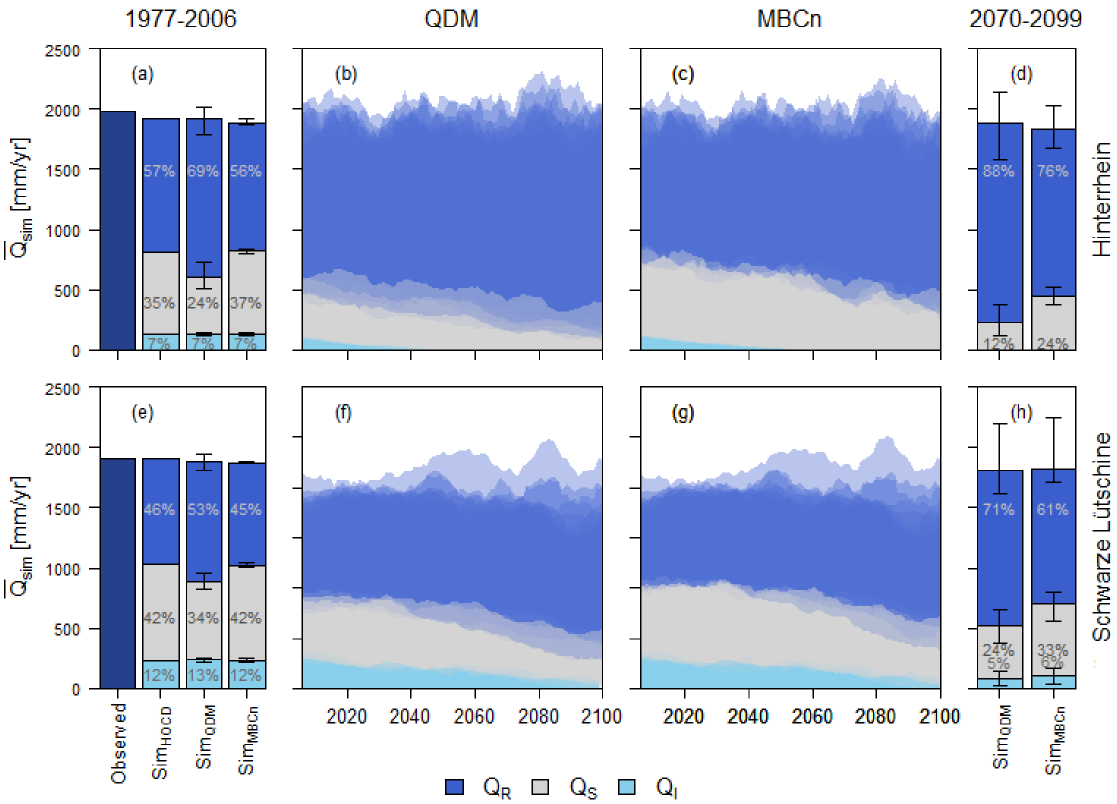

Figure B: As Figure 5 of the submitted manuscript version (see there for explanations) but showing results based on changed parametrisation after re-calibration of the hydrological model.

We thank the referee for noting the following minor issues, which will be corrected in the revised version:

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 ..."