<u>**Response to Referee Comment**</u>

We appreciate the time taken by the reviewer for the careful and thorough reading of this manuscript and for considering the remarks of the first reviewer. The additional clarifications and constructive suggestions will certainly help to improve the quality of our manuscript. The comments have been carefully considered and responded. Please find our response to each comment below.

<u>**Specific comments:**</u>

1. ***The setup for the EMQ is completely unclear. Standard EQM is local, i.e. it would apply a different correction for each location for which observations are available, in this case for each gridcell of the observational datasets. There is no explanation of how the corrections for the subclasses (elevation and slope) are obtained. Are the local corrections averaged, or is the precipitation averaged prior to fitting the EQM?***
This is obviously a key aspect of the method and it is surprising that it is not explained.
The statement that standard bias correction methods do not include the effect of topography is wrong, as the observations, which are the basis for the fitting, do include these effects. What is presumably meant is that standard bias correction does not include these effects explicitly, which means it cannot be applied when the topography changes

<u>**RESPONSE**</u>:

We agree that the setup of the bias correction remains unclear. Still we would like to point out that one strength of our method is that it is not local (the standard EQM described by the reviewer is a bias correction plus statistical downscaling). The simple reason is that a localized correction would fail in different states like the Last Glacial Maximum as valleys are filled with ice. To make the suggested method clearer, we have modified the manuscript as follows:

- Page 5 lines 31 – 32

…To combine all steps, the EQM is applied to each (sub-) group and each month of the year, separately. This results in a set of TFs for each (sub-) group and each month of the year. Thus…

…To combine all steps, the local intensity scaling method and the EQM are applied to each (sub-) group defined in the first step and to each month of the year, separately, by pooling all grid points that belong to each group and handling them as a single distribution of daily precipitation. This results in a set of TFs for each (sub-) group and each month of the year. For instance, when the correction is carried out using height-classes of 400 m, a TF is defined for each group, resulting in nine TFs for each month and in total 108 TFs throughout the year. Moreover, the correction is afterwards applied to the daily precipitation in every grid point using the TFs that are common to all elements within the same group (or sub-group) and month. Thus…

We also agree that the observational data sets implicitly include effects of topography. Changes regarding this point are presented in the following lines of the manuscript:

- Page 2 line 33

…correction methods do not consider orographic features that…

…correction methods do only implicitly consider orographic features that…

- Page 3 line 6

…time includes orographic characteristics…

…time explicitly includes orographic characteristics…

2. ***As already pointed out by the first reviewer, determining joint bias corrections for the subclasses defined by topography and slope only makes sense if the local bias corrections within a class are more similar than those between the classes. This needs to be shown***

<u>RESPONSE</u>:

We appreciate this comment and we agree that we missed to show clearly enough the argumentation for using different classes. As reviewer 1 asked a similar question we present there the same answer: We thank the reviewer for bringing to our attention that we missed to show clearly enough the orographic dependence of the biases. To clarify this, we have attached a figure that presents the monthly mean biases for each height-class before and after the correction (Fig. R1). Figure R1 illustrates an overestimation at high elevations and an underestimation at the lower ones during the colder months. Moreover, different levels of underestimation are observed across the height-classes during the warmer months. Thus, the splitting into different height-classes is appropriate to be used in the bias correction. Moreover, we would like to mention that we explicitly present the model biases within two classes in the Fig. 3 (of the manuscript), and implicitly for all the height classes in Fig. 4 and 5. Note that the biases within the classes are much smaller than between the classes. Therefore, we have included in the revised manuscript a more balanced discussion of how our approach is removing the biases.

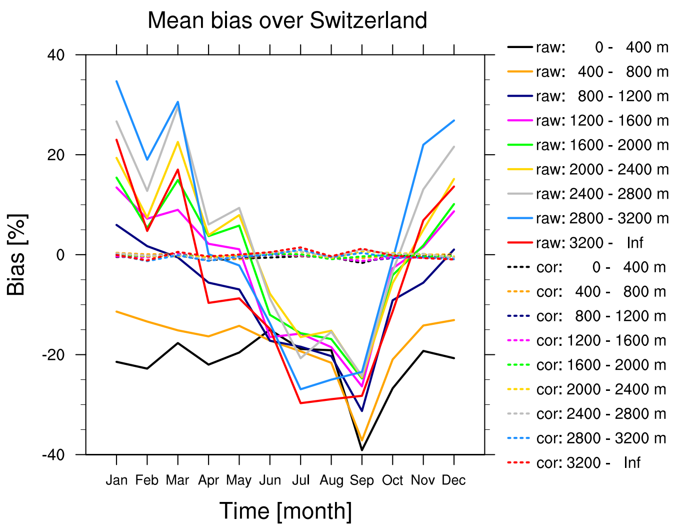

Figure R1. Mean bias over Switzerland for different height-classes.

3. ***The justification for the intended application is superficial and ignores key problems. In turn this means that the justification for the new approach itself is weak. As pointed out already by the first reviewer, many things in addition to the topography are different in a glacial climate, for instance the large-scale circulation or the moisture content. It is thus highly questionable whether applying a bias correction that is based on present climate, even if it explicitly accounts for topography, would yield meaningful results.***
This problem is closely related to the distinction of different types of errors and to the issue of propagation of GCM errors through dynamical downscaling. There are a few statements in the paper that mention that discrepancies of RCM simulations and observations might be caused by the driving GCM. However, there is no systematic discussion of what kind of errors bias correction could correct in a meaningful way. A discussion of these issues can be found for instance in
Maraun, et al., 2017: Towards process-informed bias correction of climate change simulations. Nature Climate Change, 7(11), 764-773
Maraun and Widmann, 2018: Statistical downscaling and bias correction in climate research. Cambridge University Press, ISBN 1107066050
Eden, J.M., Widmann, M., Grawe, D, and Rast. S., 2012: Reassessing the skill of GCM-simulated precipitation. J. Climate, 25(11), 3970-3984.

<u>RESPONSE</u>:

We appreciate that the reviewer brings up the point that it might be misleading about to what extent the presented bias-correction can be applied to other climate states. As already responded to reviewer 1, we would like to mention that the danger of correcting biases in a simulated climate with a method that has been trained with a climate that does not correspond to the simulated one is well-known in the statistical downscaling and correction methods. Statistical downscaling and correction methods suffer basically from the assumption of stationary biases, which implies that their algorithms trained with today's climate are considered to be also valid for different climate states. Thus, our work aims at presenting a new bias-correction that attempts to decrease this danger substantially by using orographic features as additional variables for the correction. Note, that the presented correction is obviously only applicable in regions where the topography is rather complex and where topography has certainly an influence on the local atmospheric circulation.

Moreover, precipitation biases are not only produced by initial and boundary conditions provided by the global climate models, but also by parametrisations, physical and numerical formulations that are described in both global and regional climate models. The main goal of the presented work is to correct wet or dry biases that come either from global or regional models or both. These biases can be produced by parametrisations and numerical formulations, but mainly by those that are associated with orography effects, namely, vertical motion and precipitation-related processes over complex terrain. To clarify this, we will present a discussion of the abilities and limitations of the presented correction in the results part of the revised manuscript. As responded to reviewer 1 we also include some general discussion on bias correction methods in the introduction part including the references suggested.

4. ***The fact that EQM leads to correct distributions for the fitting data is trivially true by construction. The informative part of the validation of statistical models is related to the aspects that are not trivially in agreement with observations. For each aspect of the validation it should be discussed to what extent a good skill can be expected by construction.***

For instance, given the unclear setup for fitting and application of the bias correction, it is not clear what causes the differences between observed and corrected distributions in Fig.3, or the differences in Fig. 4 and Fig. 5.
Some problems related to the validation of bias correction methods are discussed in Maraun, D. and M. Widmann, 2018, 'Cross-validation of bias-corrected climate simulations is misleading', HESS, 22(9), 4867-4873.

<u>RESPONSE</u>:

We thank the reviewer for this comment and agree that the validation may be poorly discussed. As noted by Bennett et al. (2014), the importance of cross-validation methods is that they can test the ability of bias-correction techniques on a different climate state. However, this might not be reasonable as the biases of the other climate state may not remain unchanged and the method's accomplishment relies on the biases caught during the period the method is trained on. We also recognise that recent studies by Maraun et al. (2017) and Maraun and Widmann (2018) have argued against carrying out a cross-validation for evaluating bias corrections. The authors remark that the observational and simulated data sets do not have a synchronised internal climate variability. Thus, this asynchronism in the internal climate variability may be one of the sources of the biases in free-running models.

Furthermore, as mentioned by Maraun and Widmann (2018), our cross-validation method does not compare the correction to the observations on the validation period (future or past climate state), which can produce false positive or true negative results due to internal variability in the model or observations, but the method assesses whether the statistical evolution of the model is kept.

Moreover, one of the reasons that may explain the remaining difference between the observational and the corrected data sets, as mentioned in the manuscript, can be traced back to the fact that some height classes sample over regions with slightly different biases. Hence, biases of one area can be diminished by the biases that are shared by the other areas. For instance, the strong negative biases observed in the Rhone Valley and Ticino are not fully corrected because the slight underestimation across the Swiss Plateau dominates the bias in this height-class.

Nevertheless, we agree that the evaluation and the argumentation for the remaining biases is not discussed clearly enough in the manuscript and that this should be better explained. Thus, we will include such an evaluation more explicitly in the results part of next version of the manuscript.

Bennett, James C., Michael R. Grose, Stuart P. Corney, Christopher J. White, Gregory K. Holz, Jack J. Katzfey, David A. Post, and Nathaniel L. Bindoff. 2014. 'Performance of an empirical bias-correction of a high-resolution climate dataset'. International Journal of Climatology 34 (7): 2189–2204. https://doi.org/10.1002/joc.3830.

5. ***It is not clear why the wet-day frequency is adjusted prior to the fitting of the EQM. If EQM is applied to the whole distribution including dry days, this adjustment is included in the EQM fitting. The justification might be linked to the unexplained details in the fitting setup.***

<u>RESPONSE</u>:

We thank the reviewer for highlighting this point and recognize that this adjustment may not be clear enough. We would like to mention that the adjustment does not mainly focus on the wet-day frequency, but the very low intensity values. As clarified already in the answer for reviewer 1, we agree that the argumentation for this adjustment may not be fully explained. To make this clear, we would like to mention that, in our study, we use an empirical quantile mapping technique (EQM) that differs from the parametric quantile mapping technique (QM). The reason of using an EQM is because this technique uses an empirical cumulative distribution function and does not fit any parametric distribution to the sample, i.e, (sub-) groups, as it is done in the QM. Therefore, we do not assume any known distribution either in our data sets or in the possible application to other climate states. However, the results of the EQM can become unrealistic if the very low intensity values are not adjusted previously. The reason for this is that the very small values can produce inappropriate TFs due to an important shift in the distribution, i.e., the quantiles.

To adjust these very low values, an additional parameter is included in the definition of days without precipitation that has been mentioned in the respond of the second major point of reviewer 1. The days without precipitation are now considered as censored values when they fall below a certain threshold. Many studies use a static threshold that is between 0.01 and 1.00 mm day^{-1}, whereas in our study, we calculate different thresholds to be consistent with the differentiate biases-treatment across the groups (or subgroups) and months of the year. The threshold is calculated using the local intensity scaling method and can vary between 0.001 and 1.00 mm day^{-1}. To clarify this, we have made some changes that are presented in the revised manuscript and also in response to the fifth major comment of reviewer 1.

6. ***Although it is mentioned that the errors in the observations should be taken into account when interpreting the results, there is no substantial effort to actually do this. For instance, it would be instructive to do a rough correction for the substantial undercatch of precipitation falling as snow, which strongly affects the high elevations, and assess to what extent the validation results are sensitive to this error.***

<u>RESPONSE</u>:

We appreciate this comment. We agree that we missed to show a wider discussion about the error in the observational data sets when interpreting the results of the correction method. As mentioned by (Isotta, 2014), the gridded observational data sets do not only present errors due to the interpolation methods, but they also show errors that may differ in quantity from one to other station (Sevruk, 1985; Richter, 1995) and are related to the “gauge undercatch”, whose magnitudes range from 5% over the flatland regions to 30% above 1500 m a.s.l.. Therefore, we will include a better discussion of these errors when analysing the correction, which will be presented in the results part of the revised manuscript.

Sevruk B. 1985. Systematischer Niederschlagmessfehler in der Schweiz. Der Niederschlag in der Schweiz, Beitr¨age zur. Geologischen Karte der Schweiz-Hydrologie 31: 65–75.

Richter D. 1995. Ergebnisse methodischer Untersuchungen zur Korrektur des systematischen Messfehlers des Hellmann-Niederschlagsmessers. Bericht Deutschen Wetterdienstes 194, 93 pp. (To be obtained from German Weather Service, Offenbach a.M., Germany.)

7. ***As the realization of internal variability is different the observations and in a free-running GCM (as opposed to a reanalysis) some differences between observations and simulations will be due to internal variability. This effect should be roughly quantified, for instance by showing fitting and validating the method for 10 or 15 year sub-periods (which would lead to 9 or 4 possible combinations of fitting and validation subperiods).***

RESPONSE:

We thank the reviewer for bringing to our attention the approach to quantify the biases that may be caused by differences between the internal variability of the observational data set and the simulated one. Furthermore, we would like to mention that correction methods are sensitive to the period the methods are trained on, and their accuracies would increase as more information from the observational data sets is taken into account (Lafon et al., 2013). Therefore, since the accuracy of our correction method needs to be kept as high as possible, we will carry out the suggestion made by the reviewer by splitting the data sets into two sub-periods. The outcome will be presented in results and conclusion part of the revised manuscript.

Once again, we would like to thank the reviewer for the time invested to review our paper so carefully and we are looking forward to meeting the expectations.

Best regards,
On behalf of the co-authors,

Patricio Velasquez