Anonymous Referee #1

This article seeks to present a new method for bias correction of climate model outputs. Also, it seeks to present an intercomparison of bias correction between three available methods, one of which is the new method. Bias correction is a very important aspect of future planning for climate change given the strong limitations of raw model outputs for practical hydrological decision making. Thus, articles suggesting method improvements, or with insightful intercomparisons, should be encouraged.

However, in this case, I fear a lot of work is required to raise the material up to publication standard within HESS. I have a number of concerns, none of which are specifically about the technical details of the method, but some of which build on technical concerns raised by the reviewers of the original pre-print (HESS-2020-515). The concerns are listed below.

As referee #2 was also the more critical referee during the first submission, we want to emphasize here that his/her concerns have been addressed and solved.

1. Detrimental dual focus (aka what is the paper actually about?)

*In response to the reviewer comments, the authors have elected to increase the prominence of the method intercomparison and decrease the prominence of the new method. For example, the title is now 100% about the evaluation and doesn't mention the new method, and the abstract doesn't go into details about what is new in the method, rather, it just says "we present a method". In contrast, the body text follows the original manscript and is still focussed on the new method. So the paper has a dual focus. Unfortunately, it doesn't really work to have a paper about two things—each detracts from the other, leaving both unsatisfactory. So the authors need to decide: if the paper is about the new method, then the evaluation is subsidiary and should be specifically designed to justify why the new method is an improvement. Alternatively, if the paper is about the intercomparison/evaluation, the new method should *not* be mentioned at all.*

We are aware that the focus of the manuscript was not clear enough in the first submission and have worked a lot on this issue. The goal of this paper is to find a bias correction method that fulfils three well-founded demands that a highly relevant for climate impact scientists. Conversely, we also discuss the shortcomings of two other methods, which are representatives for larger groups of methods.

Also, in the existing literature it is not pointed out clearly that our demand (2) depends on how the quantile correction is applied. If the quantiles are calculated only with the historical values like in traditional QM, which means the bias adjustment is fixed to constant values of the variables, the raw CCS is changed. This we already mention, but it will be made clearer in the introduction, in the text for Fig. 3 and in a new subsection 3.1. as mentioned in our response for 3) of referee #2. We consider this aspect as valuable contribution to the community.

To meet your concerns and in agreement with the comments from the other referee, we move some parts of 3.1 to a new subsection (See 3.) from referee #2 where we discuss the conceptional differences between QM and EQA and all methods, that are in the same groups according to Table 1. So in the end, the description of EQA is strongly shortened and only serves to document our method for the readers so that it can be reproduced. This should also reduce the impression that we intend to introduce a completely new method.

2. Is the method an advance?

If the authors elect to focus on the method, the next question is whether the method is an advance. Prompted by reviewer 1, the authors admit that "Equation 2 in Li et al. (2010) is the correct mathematical *description for our method when correcting temperature" and that "for precipitation, our method is almost equivalent to PresRAT". However, "the important difference to these methods is that our method is stricly empiric/nonparametric". Given this description and perusing the HESS manuscript types at https://www.hydrology-and-earth-system-sciences.net/about/manuscript_types.html, I suggest that this kind of change is more appropriate for a "Technical Note" and does not constitute a theoretical or practical advance worthy of a research article. If this path was taken, the material would need to be significantly reduced to fit with the requirement of "a few pages".*

The bias adjustment method itself is an empiric application of existing methods to meet the three specified demands. This may not be sufficient for being called a new method, thus we avoided to call it "new". However, we decided to use a new name for it, as it is not completely identical to any other method:

For additive bias adjustment for e.g. temperature, EQA is equivalent to EDCDFm and QDM (Cannon et al., 2015), with the emphasis that we calculate purely empirical CDFs. For precipitation (multiplicative bias adjustment), EQA is equivalent to PresRAT with the difference that EQA is empirical.

A novelty is the combination of the above said, the fundamental discussion about them, the focus on complex terrain with strongly varying meteorological fields and that we compare the methods with synthetical model data to show how the methods deal with extremely dry models.

3. Are the methods being compared the right methods to compare?

The answer to this question depends on the choices made at point 1. If the author focus is to introduce a method that builds upon existing methods (as per point 2 above), then the most obvious point of comparison is the parent methods. Thus, the comparison should include EDCDFm, PresRAT, and any other methods the new method inherits from. This is the only way to determine whether the new method helps or hinders, and in which context this occurs. On the other hand, if the authors elect to focus on method evaluation/intercomparison, then I think some changes will be required to make the evaluation more novel. This will require reviewing what other literature tries to do this (with which I confess I am not fully familiar) and then ensuring there is something different (and substantial) about this intercomparison - perhaps it's the number of methods being compared, or that their types have never been compared, or perhaps it's the study area that's novel (Austria). In any case, it seems likely that more than two methods will need to be compared, which translates to more work for the authors I'm afraid.

The choice of methods depends on what is intended to show. We wanted to show the shortcomings of widely used methods. Thus, our selection had two reasons:

- We arranged bias adjustment methods in four groups in Table 1. We chose one method out of three of these groups. The forth group (trend altering and parametric) was not relevant for this comparison, as it unites negative features of two other groups, so we skipped this group.
- The chosen representatives of the groups should be relevant for the impact research community of the alpine area, which means that is should have been implemented (several times) in Austria or Europe. This applies to QM and SDM. SDM was used for a couple of projects in Austria, where the data is publicly available for climate impact studies.

EQA, QM and SDM all represent their groups in Table 1.

4. What literature review is required and how should the study motivation be framed?

In the case where a slight tweak is made to existing methods, there is no need for a lengthy literature review espousing the benefits of the parent methods. We can assume that the case was argued back when the *original methods were published, and thus all that is required is a short summary of the benefits (or, if debated, of either side of the debate). This is why a "Technical Note" can be (indeed, must be) so short. On the other hand, if the focus is on intercomparison, then the authors must provide a general review of all bias correction methods, a review of other intercomparison studies, and a justification for what this study is adding to those existing studies. Unfortunately, the current introduction does none of these things well. It is quite long and lacks narrative and structure.*

There is a structure in the introduction, which was also suggested by the second reviewer from the first submission. Every paragraph has a specific topic/focus. As it is not possible to make headlines in the introduction, these are not directly visible. The introduction should serve the following purposes:

1. It introduces our three demands and justifies them.
2. It introduces existing methods that may be relevant for meeting the three demands.
3. It should give the reader all the necessary background information to understand Table 1. (parametric/empirical, CCS, stationarity of biases)
4. It briefly discusses comparison studies.

However, we agree that the introduction is quite long and there is still room for the improvement of the structure. If we are offered the chance to stay in the reviewing process, the introduction will be rearranged to serve the purposes in the right order. We will shorten the introduction, e.g. by moving lines 96-107 to the end, where it could be added to a new “discussion” section. These lines were strongly influenced by referee #1 of the first submission, who wanted more bias adjustment papers with hydrological applications in mind. Also, we will move the three demands more towards the beginning of introduction, as they are the base for the whole following literature review. Additionally, single sentences can also be removed here and there without omitting too much information.

I could go into greater detail on some further technical aspects of the paper but I feel there is not much point until these big, overarching questions are settled, since they will impact many aspects of the manuscript. I would be happy to review future versions and give more detailed comments then.

In general, this manuscript has the feel of funded applied research being turned into a paper as an afterthought. Funded applied research is very worthwhile in and of itself but to add something to the academic literature there must be something substantive that is new, novel or insightful, and not every practical project has these aspects.

You are right in this aspect that we are partly paid by funded applied research projects. We have already implemented our method in projects. However, a literature screening did not explain satisfactorily the skills and limitations of the methods and which are suitable for our needs. This is why we decided this is worth a publication.

I feel that I should end on some positives. Notwithstanding my criticisms of structure, the article is quite well written, with a good standard of English, well presented (particularly the figures) and with good attention to detail. I wish the authors well in to alter/add/refine the material so that it does eventually appear in the published literature.

Thank you for pointing out the positive aspects in your opinion.