Reply to Referee #1

We thank the Referee for the valuable and insightful comments. Below we reply point to point and describe the modifications that we plan to introduce in the revised version of the manuscript. Our replies are evidenced in blue and italic.

The manuscript investigates different settings for calibration of hydrological models. Specifically, the authors focus on the reliability in representing streamflow extremes and analyze two major issues that arise when climate model forcing are used as input of hydrological models to assess possible change of discharge maxima, namely:

1. Unreliable distribution of simulated streamflow maxima when hydrological models are calibrated by optimizing metrics designed to reliably reproduce ordinary streamflow
2. Errors and biases attributable to the use of climate model forcing are used as input of hydrological models (when previously calibrated using ground data).

Although both the above issues are not new, they are often neglected in climate change studies. Here the authors present a tailored calibration approach to tackle both issues, i.e. providing a good and reliable representation of streamflow maxima when using climate model forcing as input of hydrological models.

The proposed approach is applied to a set of climate model outputs, as well as to ground data, to emphasize with exhaustive examples the magnitude of errors related to two issues.

In light of the above considerations, it is my opinion that the material and methods presented in the paper can be useful and of interest for HESS readers and, more in general, scientists interested in the field.

Reply

We thank the Referee for the overall positive assessment of our study and the encouraging comment.

However, despite I recognize the potential interest of the paper, I have a major concern related to writing, since there are some parts of the manuscript that are unclear, sometimes there is unnecessary information or excessive repetition of information. Moreover, the text can also be better organized. For these reasons, I cannot recommend the publication of the actual manuscript, but I am confident that the material and results can be presented in an effective and informative exposition, if ALL the authors dedicate the due amount of time to proofread and revise the manuscript.

Reply

We take in great consideration this comment and in the revised manuscript we plan: i) to restructure the manuscript by introducing all the suggested modifications; and ii) to revise the writing throughout all the document in order to remove unnecessary details and repetitions.

In the following I will provide some (not exhaustive examples related to my concerns.

1. In Line 108 I read "where i is the position of Qs … and Qo… in the ranked samples of the simulated (s) and observed (o) annual streamflow maxima, respectively,

....” and a few lines below (lines 110-112), the same information is repeated "As customary in statistics $Q_{s,(i)}^M$ … indicates the ranked time series of the annual maxima $Q_{s,i}^M$ of simulated streamflow. A similar definition has been introduced for observed streamflow." As a reviewer, I guess that authors assumed: i) that the reader knows what is a ranked variable, ii) that the reader know how to extract annual maxima from continuous time series (see my comment 5). Moreover, I notice also that no comment is provided on the rank ordering (i.e. decreasing or increasing), so I guess that the authors do not provide this in formation since this does not affect the result of eq. (1), and this is reasonable.

Reply

We agree with the reviewer and, as we already said above, our plan is to make an accurate revision of the manuscript, removing redundancies and unneeded information. As ECDFs are computed after ordering the data from the smaller to the larger, this information will be added to the revised manuscript.

2. Then I continue my reading and in lines 113-119 I find an explanation of the p-value (e.g. "The p-value is the probability of rejecting the null hypothesis when it is true. It can also be defined as the smallest significance level α_s at which the null hypothesis would be rejected"). So, the reader should know the meaning of ranked variables, but he should probably ignore the meaning of p-value???? Maybe that a statement that p-values associated to the Kolmogorov-Smirnov statistic is used as a metric of coherence between observed and simulated maxima would suffice.

Reply

Agreed. We propose to simplify this part and clarify that p-value is used as a metric of coherence between observed and simulated maxima.

3. Eq. (2) in Line 124 provides the Weibull plotting position formula that is introduced in line 121 by the sentence "The daily average annual streamflow maxima are extracted from the chronological daily time series ….", but authors forgot to state that this formula is not valid for chronological maxima, but for ranked records with increasing order.

Reply

Thanks for spotting this. We will amend the text accordingly.

4. Line 143-144 (just above eq.4) "simulated, $\hat{Q}_{s,(i)}(\boldsymbol{\theta})$, and observed, $\hat{Q}_{o,(i)}$, flow duration curves (i.e., the ranked streamflow values this time in descending order)" and just after eq. 4 a repetition in line 146 "(ranked from the larger to the smaller value)"

Reply

Thanks for evidencing this. We will remove the repetition.

5. Lines 160-163. Here it is explained with a confusing and wrong notation how annual maxima are extracted from a time series. Why this information is provided here and not before (see my comment 1)? Does the reader need this information?

Reply

We agree that this description should be simplified and provided when the extraction of annual maxima is introduced. We will modify the revised manuscript accordingly.

6. Lines 253-254 and line 257 provide the same information (line 257 report a reference to eqs. 4 and 5). The two text can be merged (or reference to equation should be provided first).

Reply

Thanks for noticing this. We will merge the two sentences.

7. Section 4.2 provide much more details (e.g. on calibration process and confidence bands) than previous Section 4.1. …. Again: usually the due information should be provide that first time is needed.

Reply

In the revised manuscript we plan to revise significantly both sections and to move part of the information directly in the Methods section.

8. Maybe my previous comment n.7 on the way of writing can be skipped. Indeed I do not understand the choice to present first in Section 4.1 the calibration with CM and then in Section 4.2 the calibration with ground data (with more details). I would suggest to exchange the order of the two sections to show first the drawback when using CM forcing on hydrological models calibrated with ground data (i.e. the actual content of Section 4.2) and then the improvement when calibrating the hydrological model with the same forcing used for simulation (i.e. the actual content of Section 4.1).

Reply

Given the relevant modifications that we plan to introduce in the manuscript text and structure it is hard at this stage to provide a definitive answer to this Referee's request. We will certainly take under serious consideration this comment during our review.

9. Lines 415-420. It is not clear how calibration is performed. I guess that parameters are randomly selected according to uniform distributions (in a 12-dimensional parameter space) bounded by the ranges in Table 3, but it is only my guess. The sampling rule should be clarified. It should be better clarified that 40000 have been consequently run and the best 200 ones retained (as I guess). What does "we considered the 100% confidence bands resulting from the retained solutions" means? I guess the maximum and minimum value of each parameter in the 200 retained simulations … is it?

Reply

Parameter estimation is performed by means of the genetic Particle Swarming Optimization algorithm and 40,000 was the maximum number of positions the particles

assume in exploring the parameters space (given by the product of the number of particles and the maximum steps of the PSO algorithm). We will add this information in the revised manuscript.

Furthermore, the Referee is right in stating that we considered as a metric of uncertainty for the calibrated parameter the range between the maximum and minimum value of each parameter in the 200 retained simulations (see Figure 5). We propose to revise this part of the manuscript to better specify the adopted methodologies

The above points are not exhaustive, but I am confident that if the authors devote the due time, they can properly revise the whole manuscript to effectively convey their results.

Reply

We thank the Referee for the insightful comments. In the revised manuscript we will streamline the manuscript to better convey our results.