

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
<https://doi.org/10.5194/hess-2021-467-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on hess-2021-467

Anonymous Referee #1

Referee comment on "Analysis of high streamflow extremes in climate change studies: how do we calibrate hydrological models?" by Bruno Majone et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2021-467-RC1>, 2021

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:

- Unreliable distribution of simulated streamflow maxima when hydrological models are calibrated by optimizing metrics designed to reliably reproduce ordinary streamflow
- 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.

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.

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

- In Line 108 I read “where i is the position of Q_s ... and Q_o ... 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 $\delta_{i,s}$... indicates the ranked time series of the annual maxima $\delta_{i,s}$... 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 information since this does not affect the result of eq. (1), and this is reasonable.
- 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 $\delta_{i,s}^{1/4}$ 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.
- 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.
- Line 143-144 (just above eq.4) “simulated, $\delta_{i,s}$... and observed, $\delta_{i,o}$, 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)”
- 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?
- 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).
- 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.
- Maybe my previus 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).

- 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?

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.