Dear Editor,

We thank you for handling our manuscript and we appreciate the effort that Referees have put into their assessment. Please enclosed you can find the revised version of the manuscript titled “Analysis of high streamflow extremes in climate change studies: How do we calibrate hydrological models?”, reference number hess-2021-467.

After having carefully read their comments, we believe we fully addressed each point as reported in the attached rebuttal document. Please also find enclosed a pdf document which details in track changes mode all the revisions we included into the revised manuscript.

We do believe that the revised manuscript improved significantly and meets the quality standards of the *HESS* journal.

Bruno Majone

on behalf of the authors

Address

Department of Civil, Environmental and Mechanical Engineering, University of Trento, via Mesiano 77, 38123 Trento, Italy

email: bruno.majone@unitn.it

Reply to Editor and Reviewers

We thank the Editor and the Referees for the valuable comments. Below we reply point to point and describe the modifications introduced in the revised version of the manuscript. Our replies are evidenced in blue and italic.

Reply to Editor

Dear authors,

Thank you for responding to the two reviews. You have responded to most comments carefully. Because some of the comments are substantial, your revised manuscript will be sent to the referees again.

Please revise the manuscript accordingly. I look forward to receiving your revised manuscript.

Sincerely,

Yi He, HESS Editor

Reply

We thank the Editor for his assessment and for the opportunity to submit a revised version of the manuscript. We took in great considerations all Referees' comments and in the revised manuscript we introduced the following modifications:

- *we streamlined the Introduction Section to better clarify the objectives of our work;*
- *we significantly revised the text and structure of the Methods section in order to incorporate all the suggestions provided by both Referees; in particular we introduced a new Section which details in a consistent manner the set-up adopted for the different simulations performed in the study;*
- *we included new analyses to discuss the potential overparameterization issue associated to the use of KS metric as requested by Referee 2;*
- *as requested by both Referees we streamlined the manuscript by removing all repetition and unnecessary details. Overall the manuscript did not change in length (few lines less than the original version) despite the inclusion of the additional text needed to clarify specific comments raised by Referees.*

Reply to Referee #1

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 took in great consideration this comment and in the revised manuscript: i) we restructured the manuscript by introducing all the suggested modifications; and ii) we revised the writing throughout all the document in order to remove unnecessary details and repetitions. All the modifications are highlighted in the enclosed pdf document detailing in track changes mode all the revisions we included into the revised manuscript.

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 $Q^M_{s,(i)}$ … indicates the ranked time series of the annual maxima $Q^M_{s,i}$ 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 clarified above, we deeply revised the manuscript, removing redundancies and unneeded information. In this respect, the

Methods section has been restructured and reduced in length. As ECDFs are computed after ordering the samples from the smaller to the larger, this information has been included in the revised manuscript at line 132.

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 simplified this part and clarified that p-value is used as a metric of coherence between observed and simulated maxima (see lines 152-154 of the revised manuscript).

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 clarified this aspect at lines 134-136 of the revised manuscript.

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 removed 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

Agreed. In the revised manuscript this unnecessary detail has been removed.

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. This information is now reported synthetically in Section 3.4 (see also reply to comment 7).

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 revised significantly both sections and moved this technical information into the new Section 3.4 which now details in a consistent manner the set-up adopted for the different simulations performed in this study. Furthermore, we inverted the ordering between the two sections as also requested in the ensuing comment.

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

We accepted the comment and reordered the Results and Discussion section in light of the Referee's comment. Furthermore, the content of old Section 4.2 has been also split into two sections (new Sections 4.1 and 4.2) in order to include the additional analysis that we performed to address comments raised by Referee n.2. Specifically, results highlighting the drawback when using CM forcing on hydrological models calibrated with ground data have been included into the new Section 4.2, while calibrations using CMs' outputs as input are now detailed in the new Section 4.3.

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 added this information in the revised manuscript at lines 101-105 and 259-261 of 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 6 of the revised manuscript). We clarified this methodological aspect at lines 264-266 of the revised manuscript.

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 streamlined the manuscript to better convey our results.

Reply to Referee #2

This paper calibrates a 12-parameter conceptual hydrologic model (HYPERStreamHS) for the 9850-km2 upper Adage River Basin (Italy), using observed data and bias-corrected data of three regional climate models (EURO-CORDEX), for the 1982-2010 reference period. The model is parameterized for the climate model data using (i) the Kolmogorov Smirnov (KS) statistic for the empirical distribution functions of annual extremes and (ii) flow duration curves. The KS test is subsequently used to test if the observed and simulated extremes are drawn from the same probability distribution. The paper also plots the parameter ranges of the 200 best solutions and models future streamflow extremes.

Reply

We thank the Referee for this careful summary of our work.

General comments

(1) The first key weakness of the research is that the authors calibrate 12 parameters of a conceptual hydrologic model using just 29 annual daily stream flow extremes. This is, of course, a terrible over-parameterization. The effect of over-parameterization on the streamflow simulations needs to be quantified.

Reply

We understand the concern by the Referee of a possible overparameterization. In the following we show that this is not the case in our work. In the original manuscript we considered the issue of overparameterization in two different ways:

1) *we investigated identifiability of model parameters for the case in which KS metric has been used (old Figure 5) on the basis of the best 200 retained simulations. Results indicate that, though not all the parameters are equally important, some of them present a very good identifiability, thus suggesting that overparameterization could be less than what could be suggested by the number of parameters;*
2) *Evaluation of parameterizations obtained by using KS as efficiency metric in terms of R_{FDC} led to satisfactorily results (in the range between 0.45 and 0.80) for all the*

investigated cases (old Table 2). Since R_{FDC} is a metric that uses the entire time series of observational data, we consider this as an additional evidence that the use of KS metric on annual maxima is not terribly biased towards over-parameterization.

However, we do agree with the Referee that this aspect deserves a supplement of analysis and discussion. In the revised manuscript we: i) expanded the discussion of the aforementioned points (see lines 360-364 and 447-457); and ii) we introduced new analyses supporting our conclusion that the use of KS metric does not lead to terribly overparameterized models (see lines 287-310), as it will be outlined afterwards.

To demonstrate the low level of overparameterization, similar to most applications, of our proposed approach we performed the following additional simulations. As a first step we evaluated the performance of the model in the time window 1950-1980, not used for calibration, at the gauging station of Ponte S. Lorenzo at Trento. The verification is done by using precipitation and air temperature provided by the ADIGE dataset. The parameters used were those obtained by calibrating with KS with the annual maxima of the time interval 1982-2010. Similarly, to the cases presented in the original manuscript the first two years of simulations have been used as spin-up period. Results are presented in Figure R1a which shows the simulated and observed ECDFs of annual streamflow maxima and the associated p-value of the Kolmogorov-Smirnov test. Reproduction of observed ECDF is satisfactorily, especially for high flow quantiles, with a KS value of 0.233 (it was 0.067 in calibration) and p-value = 0.372. In a strict statistical sense KS-ADIGE parameterization provides simulated samples of annual streamflow maxima belonging to the same population of observations also in the time window 1952-1980; the reduction of p-value from calibration to validation is not negligible but altogether rather common in hydrological models.

We also performed a spatial validation of the KS-ADIGE parameterization by simulating streamflow at the Bronzolo gauging station (about 6000 km^2) in the time window 1982-2010. These data were not used in the parameters calibration phase. Results presented in Figure R1b highlight an excellent reproduction of the observed ECDF of annual streamflow maxima with a KS = 0.133 and p-value = 0.951. This is in our opinion a noteworthy result which indicates that the KS-ADIGE parameterization is reliable and does not introduce distortion in the spatial distribution of hydrological processes, particularly for those associated to high streamflow events, i.e., runoff generation and streamflow concentration processes.

The reason of this good agreement obtained in both validations is in the fact that, although the model is calibrated to the annual maxima daily streamflows, it is applied to simulate the entire streamflow time series and therefore the maxima are reproduced correctly only if the interaction between the precipitation and streamflow relevant during high flow extremes is correctly reproduced. We commented this in the revised manuscript in addition to the loss of identifiability of parameters not relevant in the reproduction of high flow extremes such as parameters controlling snowmelting and groundwater contribution, the latter being relevant only for low flows (see lines 447-450 of the revised manuscript). This shows that we cannot exclude that additional analyses could be envisioned for limiting overparameterization (e.g., reduced number *of parameters, presence of constraints in the parameters range, etc.) in applications dealing with different hydrological models and different data availabilities (e.g. lower number of streamflow extremes). However, these evidences are in our view enough to consider the model reliable.*

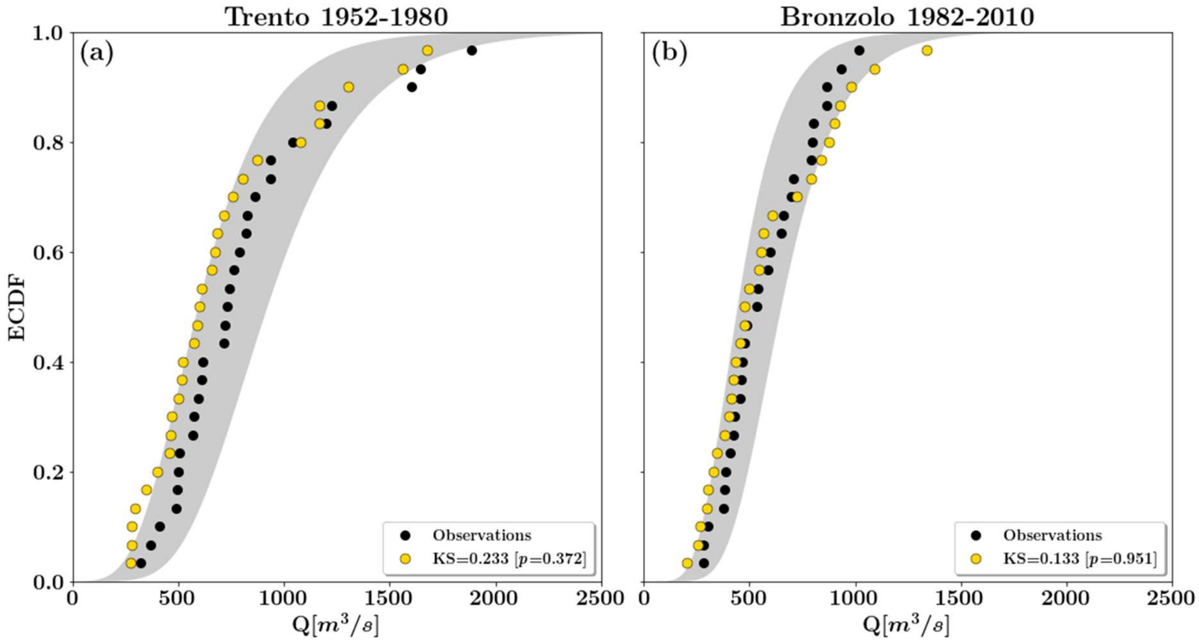

Figure R1. a) ECDFs of annual maximum daily streamflow at Trento gauging station in the period 1952-1980 and b) ECDFs of annual maximum daily streamflow at Bronzolo gauging station in the period 1982-2010, obtained by using as input the observational dataset ADIGE.

As a concluding comment, we remark that <u>our main objective is to present a methodological framework for improving high flow extreme estimation in climate change scenarios, not to propose a new multipurpose calibration procedure</u>. We investigated the overparameterization issue as requested by the Referee and we believe we provided arguments that our methodological approach does not lead to a "terribly overparameterized" model: a physiological deterioration of modeling performances in validation is indeed common to all hydrological models and applications.

(2) The second key weakness is that there is no evaluation (validation) of the parameterized models with an independent data series. How can we call this reliable and accurate? (Highlights, l.18-19)

Reply

Successful validation of the KS-ADIGE parameterization in the time window 1952-1980 was presented in the previous Reply. Additionally, we performed at the Trento gauging station the validation of parametrizations obtained by the NSE-ADIGE and R_{FDC}-ADIGE in the time window 1952-1980. These simulations led to NSE and R_{FDC} values (NSE = 0.803 and R_{FDC} = 0.804) which are only slightly lower than those obtained in calibration (NSE = 0.822 and R_{FDC} = 0.975), thus indicating once more the *reliability of the adopted modeling framework. In the revised manuscript we presented these additional results at lines 287-300.*

(3) Why not present the characteristics (figure, table) of the rainfall extremes of the observations and the climate models?

Reply

Evaluating extremes of precipitations is not the objective of this work and we are afraid that including them may cause a loss of focus, considering also that the comment n. (6) invites us to revise deeply the manuscript in order to avoid unnecessary details or discussions of relative importance with respect the main objective. The Euro-Cordex simulations have been widely studied in the literature and we feel that referencing these studies will suffice to describe the context and allow the interested reader to deepen this aspect not directly considered in our work. We believe that in this way we obtain a good compromise between the request of curtailing the paper and the need to introduce the additional results presented in the Replies 1 and 2.

(4) It is not surprising that the KS test for comparing the empirical distributions of observed and modeled annual flow extremes will give a better result for the model optimized for these extremes with the KS statistic than for the models optimized with the flow duration curves or with the NSE. However, we can also understand that the KS test has its limitations (see Figure 3), so please present in this light.

Reply

Our objective is to show that calibration should be tailored to the application and in particular if the objective are the maxima, minimizing KS (i.e. imposing that experimental and numerical ECDFs are as close as possible) is a better strategy that imposing correspondence between the FDCs or between chronological time series. In Figure 3 of the original manuscript we showed that low statistical coherence (i.e., small p-values of the Kolmogorov-Smirnov test) can be achieved when employing parameterizations obtained with a calibration approach directly targeting the statistics of extremes, but still using observational data as input, i.e., KS-ADIGE. We agree with the Referee that each metric has its own limitations and trade-offs and in the revised manuscript this point is commented at lines 278-285.

(5) The paper is written in a wild wild way. We find Methods in the Introduction, Methods in the Results and Discussion, Introduction in the Results and Discussion, no specific research objectives in the Introduction, inexact language, superfluous text and many repetitions.

Reply

We considered with extreme attention this comment and we restructured significantly the manuscript by introducing all the suggested modifications. All the modifications are highlighted in the enclosed pdf document detailing in track changes mode all the revisions we included into the revised manuscript.

(6) In summary, the paper needs to be completely restructured and rewritten in a concise and quantitative manner. Uncertainties stemming from the two key weaknesses (1 and 2 above) need to be quantitatively addressed, metrics and p-values of section 4.1 and 4.2 should be summarized together in one clear table. Expressions such as statistical coherence, forward simulations, extrapolations, 100% confidence bands (!?) need to be defined in the Methods and possibly reworded.

Reply

We already clarified the modifications that we introduced to address major comments (1), (2) and (5). Concerning the remaining issues, we summarized in two Tables all the presented metrics and p-values (see the revised version of old Table 2 now Table 4, and the new Table 3), as well as we revised significantly the description of the methodological section in the view of the Referee's suggestions.

Specific comments (non exhaustive)

l.8: error prone RC: Please quantify. The majority of your models are accepted, according to your KS p-value.

Reply

Here we were referring exclusively to the parameterizations derived from calibrations conducted using observed meteorological data. Though it is true that many of these models pass the Kolmogorov-Smirnov test (assuming as customary a level of significance of 0.05), their p-values are low and this is reflected in the estimation of high flow quantiles as depicted in Figure 4 of the original manuscript. In the revised manuscript we revised the sentence at lines 8-10 of the abstract to be more explicit.

l.39 Much less?

Reply

Thank you for noticing this. The sentence has been modified in the revised manuscript as a consequence of the restructuring made to the Introduction section.

l.57: iii) due to the impossibility of obtaining totally unbiased climate simulations there is no a-priori guarantee that simulations fed by climate models produce samples (e.g. time series of simulated annual streamflow maximum) that are statistically coherent with observations. RC: Your approach cannot address this problem either.

Reply

What we meant here is that climate model simulations are also affected by bias and they may differ from observed meteorological data. In this respect, we showed that calibrating the model by using observed meteorological data as input forcing can lead to a biased evaluation of the probability distribution of streamflow extremes when climate models are used. Instead, calibrating by using climate models and observed streamflow lead to a better reproduction of observed extremes. In some way this approach can be considered as an additional bias correction removing the biases that the standard procedure was not successful to remove. We do agree with the Referee that interpretation of point iii) may be misleading and we removed it in the revised manuscript.

l.61: by directly targeting. RC: non-scientific language

Reply

Agreed. We modified the writing, clarifying that calibration is conducted by maximizing the probability that the modeled and observed streamflow extremes belong to the same population (see lines 53-55 of the revised manuscript).

l.64-73: These are Methods

Reply

We moved most of this information in the Methods section maintaining here only the parts needed to highlight the novelty aspects of our work.

l.66: are constrained to maximize the chances?

Reply

We removed the mentioned sentence during the revision of Introduction section.

l.67: Statistical coherence RC: Please define statistical coherence or use another expression.

Reply

In our view statistical coherence implies that two samples are characterized by a large probability of belonging to the same population. In particular, we used the p-value associated to the well-known Kolmogorov-Smirnov test as a metric of coherence between observed and simulated streamflow maxima. On the light of Referee's comment, we modified the Methods sections to make this definition explicit (see lines 152-154 of the revised manuscript).

l.75: Do we really need six references for "goal-oriented"?

Reply

Agreed. We reduced the number of cited papers to 3.

l.102: Section 2.2 RC: It would make more sense to present this after Section 2.4

Reply

We revised significantly the Methods section to incorporate this and other comments raised by both Referees. In particular, we accepted this specific comment and the revised version of Section 2.2 (now Section 2.3, starting from line 144 of the revised manuscript) has been moved after the section presenting the calibration metrics (now Section 2.2).

l.111: A similar definition has been introduced for observed streamflow. RC What writing style is this?!

Reply

Agreed. The sentence has been removed.

l.113-119: State your null hypothesis and condense this text.

Reply
This section has been significantly condensed (see lines 145-154 of the revised manuscript). The null hypothesis is presented at the beginning of the section at lines 146-147.

l.120-125: Does this need a numbered Section?
Reply
Agreed. The text present in the old Section 2.3 has been incorporated into the revised version of Section 2.2 at lines 135-137.

l.121: daily average RC: average daily
Reply
Thanks for noticing this, we will modify accordingly.

l.135-158: The efficiency criteria are without the max and min.
Reply
Though we agree with the Referee that efficiency metrics are typically shown without this notation, we believe that in our work this additional remark is needed to clarify that the efficiency metric KS (eq. 3 in the revised manuscript) is designed to minimize the two-sample Kolmogorov-Smirnov statistic (D_n), with this latter being defined as the maximum absolute distance between observed and simulated ECDFs of annual streamflow maxima. The additional advantage is that we highlight how, contrary to NSE and R_{FDC}, KS metric should be minimized.

l.138: sensitive?
Reply
Here we meant that the metric considers the chronological time series of simulated and observed daily streamflow. We modified the writing.

l.146: repetition
Reply
Thanks for noticing this, we removed the repetition accordingly.

l.162-166: RC: Please condense.
Reply
The Methods section, including this portion of the text, has been significantly restructured and condensed following the suggestion provided by both Referees.

l.171: adaptation?
Reply
Here we meant statistical inference using a Gumbel distribution. The sentence has been modified also in the view of the ensuing comment.

l.172: for comparison purposes in order to extrapolate RC: Now what is it?

Reply

The term extrapolation refers to the common practice of estimating high flow quantiles for a return period beyond the available number of simulation years, procedure that cannot be done on the basis of the ECDFs and that necessarily is to be performed by means of a statistical inference of a theoretical distribution. This is needed in impact studies concerning the effect of climate change on high flow extremes. We clarified the meaning of this term at lines 159-161 of the revised manuscript.

l.181: portion?

Reply

Here we meant upper part of the Adige river basin. We modified accordingly.

l.288: parametric errors, RC: Without comma and what do you mean? All models are simplifications of reality.

What we meant here is that since model predictive errors are always present we cannot expect that a model reproducing streamflow time series properly is also expected to provide a good reproduction of high flow quantiles. During the revision of the manuscript the sentence has been removed.

l.233: provide an assessment?

Reply

Agreed. We modified the writing with "provide projections".

l.244: the correction used in the reference period 1989-2010 is extended to the period 1980-2010 RC: This is not clear. Is this done by you and if so how?

Reply

Thanks for noticing this. Climate models' biases have been corrected (by comparison with observations) as part of the Euro-Cordex experiment. The correction obtained during the overlapping period between climate models' outputs and observational data (i.e. 1989-2010) is extended to the previous 9 years in order to produce bias corrected scenarios for the entire control period 1980-2010. Since the adopted Euro-Cordex scenarios are the standard used in climate change impact studies in the revised manuscript we reduced this part of the text (see lines 224-231).

l.253-257: RC: Methods

Reply

Agreed. We moved this part into the new Section 3.4 where the set-up adopted for the different simulations performed in this study is presented in a consistent manner.

l.260: On the other hand. RC: Which other?

Reply

Agreed. We modified the writing.

l.278: cast?

Reply
We removed this term during the proofreading of the manuscript.

l.279-284: RC: Introduction
Reply
Agreed. We moved this part to the Introduction section.

l.311: coined here as Hydrological Calibration on Extremes (HyCoX) RC: Repetition
Reply
Thanks for noticing this. We removed the repetition.

l.326-335: RC: Methods and Introduction
Reply
Agreed. We moved this part into the new Section 3.4 where the set-up adopted for the different simulations performed in this study is presented in a consistent manner.

l.370: Fig 3 RC: It will be easier to follow if all metrics and p-values are presented together in Table 2.
Reply
Thanks for the suggestion. In the revised manuscript we summarized all the metrics and p-value in the new Tables 3 and 4.

l.373: Forward?
Reply
Here we meant simulations with parameterizations derived from a given calibration experiment. In the revised manuscript the title of Section 4.2 now reads as "Simulations using parameterizations derived from calibrations with observed ground data".

l.377: the 90% confidence interval of the observed ECDF. RC: of the fitted extreme value distribution function?
Reply
Thanks for noticing this. As spotted by the Referee, we evaluated confidence intervals by means of a parametric bootstrap. We will clarify this aspect throughout all the revised manuscript.

l.407: Extrapolation?
Reply
We already replied in a previous comment.

l.418: 100% confidence bands?!
Reply
Here we consider 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. We modified the writing in the revised manuscript to clarify this aspect (see lines 264-267).

l.429: Furthermore, we verified a-posteriori that the optimal parameters are inside the range of variation. RC: The methods are unclear. Is this “range of variation” (please use a better expression) for the 40,000 simulations? How can the optimal parameter fall outside the range?

Reply

Agreed. We used the term “parameter range” in place of “range of variation” and we clarified that this is valid for all the 40,000 simulations. We also revised significatively the Methods section to implement all the useful suggestions provided by the Referee (see lines 259-267). Notice here that 40,000 simulations are simply the maximum number of particle positions, given by the number of particles times the maximum number steps used in the particle swarming algorithm.

Concerning the last part Referee comment, we were referring to the preliminary analyses conducted to identify parameters range. We agree that the sentence was inaccurate and thus it has been removed during the revision.

l.440: The differences observed in the optimal value of model parameters are due to structural errors in the GCMs and RCMs. RC: Really? And now we use these errors to make an erroneous hydrologic model, without any independent model validation. One can understand that there are two modelling approaches each with assumptions and uncertainties. So please stick to quantitative evidence.

Reply

We accepted the suggestion and in the revised manuscript we reformulated the sentence by removing the reference to GCMs and RCMs epistemic uncertainty (see lines 458-459). Considering the aspect of model validation, we believe we provided in the first two replies evidences that our proposed methodological framework is accurate and reliable.

l.446: Furthermore, our approach provides an answer to the need of reducing uncertainty in climate change impact assessments. RC: Please quantify your uncertainty reduction.

Reply

What we have shown is that using in calibration the same climate model later used in the projection ensures statistical coherence, which is lost when calibration is performed by using observed meteorological input (given that projections should be done by using the climate models). In this respect, our goal-oriented calibration framework aims at improving the estimation of extremes by directly calibrating the selected hydrological model to the quantities of interest. In the revised manuscript we removed the sentence to avoid any misunderstanding.

l.452: Marked dashes?

Reply

We modified the term with “bold horizontal dashes”.

l.490-594: RC: Please be concise. Answer your research objectives, which you should have stated in the Introduction.

Reply

Agreed. We reduced significantly the Conclusion section as well as we modified the Introduction section to explicitly mention the research objectives (see lines 68-72 of the revised manuscript).
