Dear Editor,

We thank you for the assessment and for the opportunity to submit a revised version of the manuscript. Enclosed please 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 your comments and those provided by Referee #1, 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 in 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,

As you will see from the reviewers' comments, one of the reviewers still has major concerns. In addition to that, I also have several comments listed below.

Reply

We thank the Editor for his assessment and for the opportunity to submit a revised version of the manuscript. We took into great consideration your comments and a detailed reply is provided below.

1. The manuscript needs proofreading to improve its English language. It is not well written in its current form and sometimes affects the understanding of the content.

Reply

We performed additional proofreading of the manuscript. All the modifications are highlighted in the enclosed pdf document detailing in track changes mode all the revisions included in the revised manuscript.

2.a The "goal-oriented" calibration framework named by the authors essentially correct the statistical distribution of the fitted annual maxima, but it can result in a poorly calibrated model.

Reply

We respectfully disagree with this comment. Although the model is calibrated against 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 analyzed in-depth this aspect during the first round of revisions to address concerns raised by Referee #2 about the reliability of our proposed approach. In particular, we showed that: i) calibration using KS as objective function leads to unbiased identification of model parameters (see results presented in Sect. 4.4); and ii) the values of R_{FDC} and NSE, computed a-posteriori with the parameters obtained as described in the point i), were satisfactorily in all investigated cases (either using ADIGE observational dataset or climate models' outputs) as evidenced by the values provided in Tables 3 and 4 of Sects. 4.1 and 4.3, respectively. Since R_{FDC} and NSE (in the case in which the ADIGE dataset is adopted as forcing) are metrics that use the entire time series of observational data, we considered this as an additional evidence that the use of KS metric does not lead to poorly calibrated models.

This latter conclusion was also enforced by the additional analyses we introduced in the first round of revisions, as requested by Referee #2. In particular, we obtained excellent performances in two validation runs conducted using the KS-ADIGE parameterization, namely: a first model evaluation in the time window 1950-1980, not used for calibration, at the gauging station of Ponte S. Lorenzo at Trento; and a spatial validation at the Bronzolo gauging station in the time window 1980-2010 (see results presented in Sect. 4.1).

2.b The CORDEX bias corrected outputs (P/T) should be compared with the observations to set the ground for comparison. The better match employing the KS parametrization shown in Figure 5 could be due to the wrong reasons.

Reply

Agreed. In the revised manuscript we introduced the requested comparison (see the new Figure 2) where we show clearly that climate models' outputs are not biased with respect to ADIGE observational dataset. Please refer to major comment (1) by Referee #2 for an extensive reply.

2.c The worse performance when employing the NSE-ADIGE parameterization could be due to insufficient model set up and calibration.

Reply

We respectfully disagree with this comment given that NSE-ADIGE parameterization provided an excellent value of NSE = 0.822 during calibration (see Table 3). Validations of this parameterization, performed with reference to the time frame 1950-1980 at the gauging station of Trento and at Bronzolo during the period 1980-2010, led to NSE values (NSE = 0.803 and NSE = 0.787, respectively, see Table 3) which are only slightly lower than those obtained in calibration, thus indicating that a proper set-up and calibration of the HYPERstreamHS hydrological model has been obtained (see results presented in Sect. 4.1).

The framework is meant to be used in impact studies in a warmer climate, but it is not guaranteed the calibrated model can outperform the models calibrated with a standard calibration procedure (if the model is well calibrated using the objective functions such as NSE or KGE).

Reply

We believe we provided sufficient evidence in the first version of the revised manuscript (see also our previous reply) that the standard calibration procedure (i.e., NSE-ADIGE parameterization) can be considered accurate and reliable. However, we also showed that tailoring the calibration procedure to the objective actually improves the quality of the simulations. In particular, if the objective is the distribution of the annual maximum streamflow, minimizing KS (i.e. imposing that experimental and simulated ECDFs are as close as possible) is a better strategy than imposing correspondence between the FDCs or between chronological time series of daily data. In simpler terms, the better reproduction of streamflow quantiles depicted in Figure 5 obtained when employing KS parametrizations cannot be attributed to poor calibration

of the model with a standard procedure but rather to the use of a metric specifically addressing high streamflow extremes.

2.d If the model is poorly calibrated and the water balance goes wrong over a long term, the simulated maxima can also go very wrong. If you claim it is only for matching the peak discharges, one can argue there is no need for using a hydrological model as the model does not correctly simulate the discharges other than the peaks.

Reply

Our previous replies highlighted that, although the model is calibrated against the annual maxima daily streamflows, it is also able to simulate the entire streamflow time series with good accuracy, although suboptimal with respect to the standard calibration. However, the high flow extremes are better reproduced with the proposed procedure. In particular, we showed that evaluation runs conducted using KS parameterizations led to satisfactorily values of R_{FDC} and NSE metrics. We do agree with the Editor that each efficiency metric has its own limitations and trade-offs and that calibration with high streamflow extremes may be suboptimal if the objective is to reproduce flow duration curves or the chronological time series of daily streamflow. This aspect is indeed already acknowledged in the revised manuscript at lines 301-306. In the light of the Editor's comment, we decided to include in the Conclusion section (see lines 544-547) an additional remark that multi-objective approaches could be envisioned to investigate the trade-off in model performance emerging from the use of multiple metrics including the one here proposed, though we consider this aspect beyond the scope of the present contribution.

3. It is unclear how you treat those cells that hardly have any catchment area in it. Please explain.

Reply

In the case of macrocells containing portions of neighbor river basins, the macrocell contributing area is reduced accordingly such as to guarantee that the entire modelled domain equals the drainage area of the investigated case study. A clarification sentence has been introduced in the revised manuscript at lines 256-258.

4. Please use consistent terminology throughout the manuscript. E.g. watershed or basin.

Reply

Thank you for noticing this. We opted for using the "basin" term throughout the manuscript.

5. Page 4: parameter range. Why the use of a division sign?

Reply

Thank you for noticing this. We now use the dash sign for the parameter ranges.

6. Page 5: the equation for NSE should have the max in it. The objective is to maximize the NSE but the NSE formulation presented in the manuscript is inappropriate. And it

should be referred to as Nash–Sutcliffe model efficiency instead of Nash-Sutcliffe index.

Reply

We are not sure to understand what the Editor means with "the NSE formulation presented in the manuscript is inappropriate". In our view eq. (1) reports the classical definition of NSE as a normalized statistic that determines the relative magnitude of the variance of the residuals compared to the variance of observations. We also note how the notation for observed and simulated streamflow time series within the different adopted metrics has been homogenized during the first round of revision to address specific comments by Referee #1. If Editor's comment concerns the use of "max" notation outside of the metrics' equation, our aim was to clarify that, contrary to NSE and R_{FDC}, KS should be minimized. To avoid any misunderstanding in the revised manuscript we removed the "min" and "max" notation from the equations presenting the efficiency metrics. Finally, following the Editor's suggestion, we now refer to Nash–Sutcliffe model efficiency throughout the manuscript.

7. Page 6: Weibull distribution was mentioned in Line 135, but later on you used Gumbel distribution. Please clarify.

Reply

The Extreme Value type I (Gumbel) distribution was used in this work to infer the theoretical probability distribution of simulated and observed annual streamflow maxima (see text reported in Sect. 2.4). The Weibull formulation is introduced in Eq. (4) solely to compute the Empirical Cumulative Distribution Functions (ECDFs) of the simulated and observed samples of daily average annual streamflow maxima ranked in ascending order. We modified the manuscript at lines 137-138 to better clarify this aspect.

8. Figure 1: River network is listed in the Legend, but it isn't on the map. Please correct.

Reply

Thank you for noticing this. The updated version of Figure 1 now correctly presents the river network.

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

Sincerely,

Yi He, HESS Editor

Reply

We thank the Editor for the encouraging comment.

Reply to Referee #2

Dear Authors,

Thanks for all the corrections. The manuscript has been much improved.

Reply

We thank the Referee for the positive comment.

(1) However, your repeated statements that the standard hydrologic model calibration procedure is biased (l.8, l.522, l.524-528) has no validity if you do not evaluate the climate data. How do we know that the issue is not that the climate model data are biased? I also asked this in my previous round. I copied this below with your reply. Your reply does not hold. You need to evaluate the climate data you use for your application, not generically refer to Euro-Cordex data evaluations. And comment (6), which you referred to, concerned the poor quality of the manuscript (poor structure, repetitions, superfluous text) and asked for concise and quantitative presentations. Absolutely no reason to not complete your scientific argument.

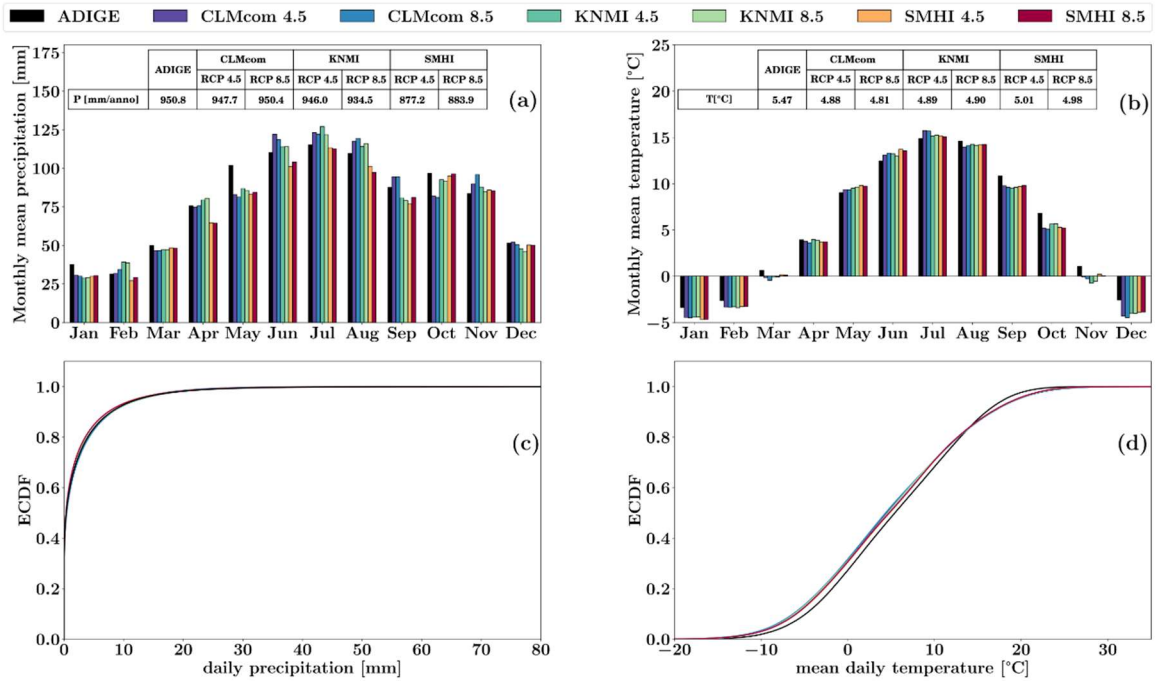
Copied from previous review round: "(3) Why not present the characteristics (figure, table) of the rainfall extremes of the observations and the climate models?

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."

Reply

We now better understand the Referee's concern about the presence of a possible bias in climate model forcing. We honestly did not understand this concern during the first round of revision and opted for not including such details to limit the manuscript's length. To address this comment, we now included in the revised manuscript the new Figure 2 (the Figure is also included below for easiness of reading) which details the comparison between basin-averaged precipitation and temperature forcing of the ADIGE observational dataset and of the six climate models. In our opinion, this comparison indicates that CMs' outputs for both variables are in good agreement with the observations during the reference period in terms of both seasonal dynamics (subplots a and b) and extremes (as highlighted in the basin-averaged daily ECDFs of subplots c and d). This corroborates our statements that improved representation of high streamflow extremes can be achieved using a goal-oriented approach and that standard calibration procedures may be suboptimal, provided that climate data are not biased as is the case in our case study.

We added these details in the revised version of the methodological Section 3.3 at lines 230-244.



	ADIGE	CLMcom RCP 4.5	CLMcom RCP 8.5	KNMI RCP 4.5	KNMI RCP 8.5	SMHI RCP 4.5	SMHI RCP 8.5
P [mm/anno]	950.8	947.7	950.4	946.0	934.5	877.2	883.9

	ADIGE	CLMcom RCP 4.5	CLMcom RCP 8.5	KNMI RCP 4.5	KNMI RCP 8.5	SMHI RCP 4.5	SMHI RCP 8.5
T[°C]	5.47	4.88	4.81	4.89	4.90	5.01	4.98

Figure R1: Annual cycle of basin-averaged monthly mean precipitation (a) and temperature (b) during the reference period 1980-2010 for both ADIGE and the 6 CMs used (different color bars). The associated annual averages are also shown in the insets. ECDFs of basin-averaged daily precipitation and temperature for the same datasets are presented in subplots (c) and (d), respectively.

Specific comments

l.14 distributed hydrologic model

RC: In l.93 we read: “12 parameters, which are assumed as spatially uniform,” This is obviously not a distributed model application, so please correct this.

Reply

We believe that the use of the term “distributed” is justified due to the following reasons: i) in the model we adopt macrocell-specific meteorological forcing; and ii) evapotranspiration, infiltration and runoff generation processes are spatially varying given that properties useful for the evaluation of these fluxes are also computed at the macrocell level. We acknowledge that this latter consideration was indeed missing and in the revised manuscript at lines 91-95 we included the following text: “Spatial heterogeneity of evapotranspiration, infiltration and runoff generation are accounted for by computing for each macrocell all relevant properties (e.g., maximum infiltration capacity, average elevation, soil type, crop coefficient etc.) based on available DEM and land-use/land-cover spatial maps”.

l.54: a physically-based hydrological model

RC: You model a 9850-km2 area with 12 parameters and you call this physically-based?

Reply

We classify as physically-based a model which provides a mechanistic representation, even if in a lumped formulation, of the different hydrological processes occurring within a river basin, as opposed to approaches which can be considered data-driven. The spatial variability of the relevant physical processes is guaranteed by adopting spatially-varying properties as discussed in the reply to the previous comment l.14. To avoid any misunderstanding, in the revised manuscript we removed the term "physically-based".

l.311: Table 3
RC: Why not report all metrics for the validation runs, instead of leaving empty spaces?
Reply
Agreed. In the revised version of Table 3, we included all the metrics for the validation runs.

l.513: a few climate models
RC: You lost count?
Reply
Thank you for noticing this. We now properly refer to six Climate Models.

l.514: distributed hydrologic model
RC: see l.14
Reply
Please refer to the reply provide to comment l.14.

l.518: We remark that such approach may lead to a suboptimal performance if the target is different from the one employed in this study, limitation that is outweighed by the improvements in representing high flow extremes in line with the goal-oriented framework pursued in this work.
RC: What a weird sentence. You reached your objective, why refer to this as outweighed?
Reply
Agreed. In the revised manuscript we reformulated the sentence along the lines suggested by Referee (see lines 543-544).