Comment 1:

This work aims to disentangle the long-standing question of hydrological sensitivities to climate change. This is done by feeding climate forcing into hydrological simulations. Specifically, the climate forcing is perturbed in order to assess the sensitivity of the hydrological response in a large mountainous basin. Of particular interest is the use of a detailed hydrological model capable of modelling not only streamflow but also snow cover and glacier mass balance, both key elements for medium-term water management and water resources optimisation. The authors point out that the novelty of the work reside in both the assessment of "patterns and drivers of local hydrological sensitivities" and to the case study. This is also due to the fact that there are contrasting results from similar analyses in the same region of interest. However, in my opinion, one of the most interesting features of this work is the adoption of a sort of multi-objective calibration based on streamflow, snow cover area, glacier mass balance and stream water isotopes, alongside with a spatial validation of the calibration. Therefore, I believe that this paper is worth publishing in HESS after answering the following questions.

Response:

Many thanks for your appreciation on our work. We will revise the paper thoroughly according to your comments.

Comment 2:

The model calibration is one of the weakest parts of the papers. There is no reference to the algorithm used as well as all the details regarding its applications which, in my humble opinion, should not only be mandatory to ensure the replicability of the paper, but are also important to understand the calibration results. For example, if the PSO scheme is adopted, the sensitivity analysis of the hydrological model parameters is not possible. Conversely, a Monte Carlo procedure could help to define the confidence band of the hydrological results. This should be explained according to the desired research objectives.

Response:

Many thanks for your comment. The calibration procedure was indeed described simplistically in the manuscript, because the main aim of the paper was to explore the hydrological sensitivity to perturbed climate, and only one set of parameters were adopted to derive the results. In specify, we adopted an automatic algorithm, the Python Surrogate Optimization Toolbox (pySOT, Eriksson et al., 2019) for model calibration. The pySOT algorithm uses radial basis functions (RBFs) as surrogate models to approximate the simulations, reducing the time for each model run. The symmetric Latin hypercube design (SLHD) method was used to generate parameter values, allowing an arbitrary number of design points. In each optimization run, the procedure stopped when a maximum number of allowed function evaluations was reached, which was set as 3000. In this study, we repeated the pySOT algorithm for 100 times, and a final parameter set was selected from the calibrated parameter sets manually based on the overall performance on multiple objectives. We acknowledge that the parameter calibration and uncertainty analysis are rather weak in this study, which is due to, as you said, the desired research objectives. We will add the details of calibration procedure in the Method section, and address the shortage in the Limitation section.

Comment 3:

I have some doubts about the objective functions. If I understand correctly, the model seems to have been calibrated using a single objective function composed of four equally weighted functions (NSE for streamflow at one site; NSE for isotope; RMSE for SCA; and RMSE for GMB). In this way, however, the authors mix different types of metrics. I try to be clear: the NSE varies between minus infinity and one, while the RMSE varies from 0 to plus infinity (theoretically, of course). As a consequence, different metrics have different impacts on the aggregated objective function. I therefore believe that the NSE (or RMSE) must be used for all the objects under consideration. Otherwise, one factor could be weighted more than the others. I would suggest to have a look at "Madsen, Henrik. (2003). Parameter estimation in distributed hydrological catchment modelling using multiobjective automatic calibration. Progress in Water Resources. 26. 205-216. 10.1016/S0309-1708(02)00092-1.,Section 2.2, Equation 3.

I also disagree with the definition of multi-object calibration. Basically, the authors use a single-objective calibration. In other words, they choose a specific solution in the multi-objective space. I believe that this strategy is fine if the aim of the paper is to focus solely on the uncertainty of climate forcing, although a more precise definition of the objective function and calibration is required, in my opinion.

Response:

Many thanks for your comments. The main reason that we mixed different types of metrics are as followed: (1) We found that NSE is not suitable for evaluation of objectives with strong essentially fluctuation (Schaefli et a., 2007), such as SCA and GMB. The NSE could be very low even if the simulation looks good. Meanwhile, error indexes such as RMSE and MAE were widely used for SCA/GMB simulation in previous studies (e.g., He et al., 2019; He et al., 2021; Lyu et al., 2023). (2) Although different metrics have different dimensional unit and range, previous results indicated that the values of these metrics are of the same order of magnitude when the model performance is acceptable (He et al., 2019; Ala-aho et al., 2017; Nan et al., 2021). (3) We did not simply adopt the parameter set which produced the best integrated objective function, but recorded all the parameters produced during the calibration procedure and manually selected one parameter set with best overall performance. Thus we believe that the choice of objective function had little influence on the main findings of this study. Nonetheless, we will address the potential influence of objective function in the limitation section.

We agree with you that the expression “multi-objective calibration” is inappropriate. What we mean here is that the model was calibrated toward datasets related to multiple objectives, but in the calibration practice a single objective function was adopted. We will make the expression more accurate in the revised manuscript.

Comment 4:

Model calibration and evaluation section. I propose not only to carry out a spatial validation of the hydrological model, but also a temporal validation with a calibration period and then a test period. The hydrological response could then be carried out taking into account all 15 years from 2001 to 2015. At this stage it is indeed important to ensure the reliability of the hydrological model.

Response:

Many thanks for your suggestion. The main reason for such calibration scheme was the large basin area. In our previous studies in this region (Nan et al., 2021, 2022), we found that the model performance during validation period was highly correlated with that of calibration period. In specify, when the NSE during calibration period was in the range of 0.86~0.94, the NSE during validation period was also at a high level within the range of 0.77~0.92. This could be partly due to the strong linearity of precipitation-discharge relation in the large basin. But the model performance at internal stations had large uncertainties when the discharge at outlet station was simulated well. Consequently, for simplicity, we did not divide the simulation period into calibration and validation period, but only used the discharge data of internal stations to validate the model. We will justify these in the revised manuscript.

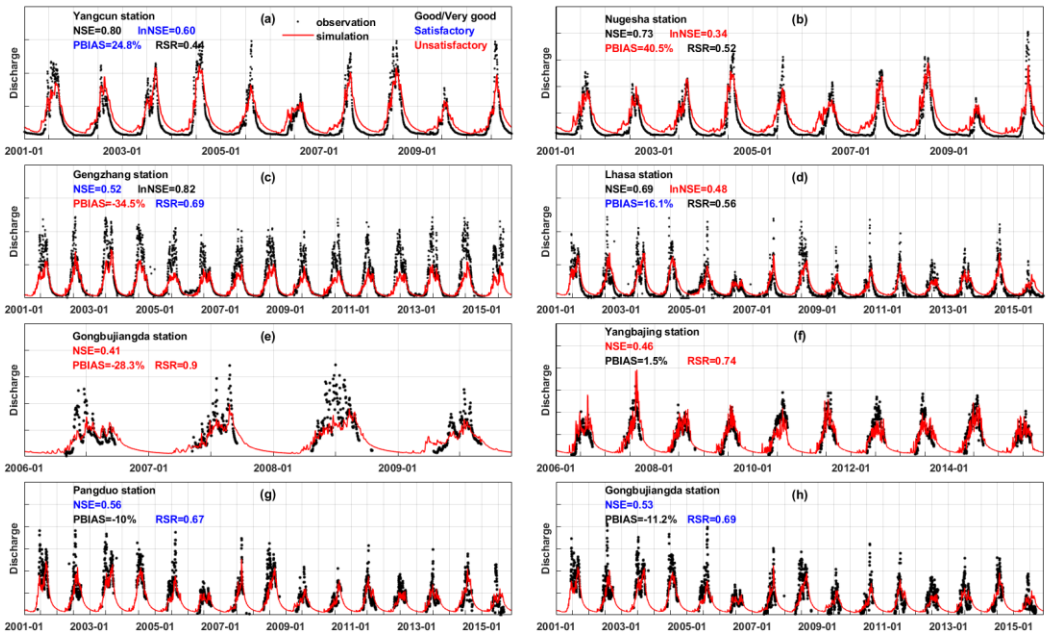
Comment 5:

Line 303: I propose to support the sentence "but were at acceptable levels with "D.N. Moriasi, J.G. Arnold, M.W. Van Liew, R.L. Bingner, R.D. Harmel, T.L. Veith Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations".

The limitation of hydrological performance in terms of either maximum flow or low flow is consistent with several studies showing how the use of one metric in calibration can lead to less than ideal results for other metrics, as each is sensitive to particular characteristics of the time series and has its own limitations and trade-offs. See for example "Schaefli, B. and Gupta, H. V.: Do Nash values have value?, Hydrol. Process., 21, 2075-2080, 2007; "Gupta, H. et al.: Decomposition of mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377, 80-91, 2009. "Mcmillan, et al. Five guidelines for the selection of hydrological signatures. Hydrol. Process., 31, 4757-4761, 2017. "; "Fenicia, F., et al.: Signature-domain calibration of hydrological models using approximate Bayesian computation: Empirical analysis of basic properties. Water Resour. Res., 54, 3958-3987." and "Majone, B. et al. Analysis of high streamflow extremes in climate change studies: How to calibrate hydrological models? Hydrol. Earth Syst. Sci. 2022, 26, 3863-3883". . I suggest that this consideration be taken into account when rewriting the section about the evaluation of model performance".

Response:

Thank you very much for providing these valuable references, which are really helpful for our deeper understanding on the calibration function selection issue. We have calculated the metrics PBIAS and RSR, and evaluated whether model performance is acceptable following the guidelines of Moriasi et al. (2007) (Figure 3 has been revised as below). We will also cite these useful literatures properly when describing the model performance evaluation in the revised manuscript.



Comment 6:

A few words should be devoted to the description of the concentration ratio and the concentration period, as is the case for NSE, LnNSE, RMSE ecc ecc.

Response:

Many thanks for your suggestion. We will add description and equations of CR and CP in the revised manuscript.

Comment 7:

For the sake of clarity, I suggest deleting Figures 5a and 5b. They do not add any additional information respect to Figures 5c and 5d.

Response:

Many thanks for your suggestion. We will delete Figures 5a and 5b in the revised manuscript.

Comment 8:

Figure 7: I do not understand figures from 7e to 7h and figures from 7m to 7p. It is not clear that the sum of the components is 100%. I suggest using a different type of figure style.

Response:

We are sorry that we make you confused about the figures. Figures 7e-h and m-p present the relative contribution of each runoff component in the total runoff during different season in different climate perturbation scenarios. We believe that the pie graph might be a better figure style which can make it clearer that the contributions of each component add up to 100%. However, five pie graphs would be needed for each subplot if this figure style is adopted, leading to too many subfigures. What's more, another reviewer required us to add the error bar to denote the standard deviation, which is difficult to present by other styles. Consequently, we prefer reserving the current Figure 7. Nonetheless, we will add a table listing the numbers in Figure 7 in the Supplementary Information to make it clearer.

Comment 9:

I suggest using the correlation values (r) in the text to identify and comment on positive\negative correlations and strong weak correlations (see Ratner, Bruce The correlation coefficient: Its values range between +1/-1, or do they not?). The p-value only indicates that there is a relationship between two groups.

Response:

Many thanks for your comments. We have calculated the correlation coefficient, and will add them in the text when describing the positive\negative correlations.

Comment 10:

Limitations section: I suggest adding that future work should address the problem of sensitivity analysis of hydrological models, multi-objective calibration and goal-oriented calibration.

Response:

Many thanks for your comments. We will address these issues in the limitation section of revised manuscript.

References

Ala-Aho, P., Tetzlaff, D., McNamara, J. P., Laudon, H., & Soulsby, C. (2017). Using isotopes to constrain water flux and age estimates in snow-influenced catchments using the STARR (Spatially distributed Tracer-Aided Rainfall–Runoff) model. Hydrology and Earth System Sciences, 21(10), 5089-5110.

Eriksson, D., Bindel, D., & Shoemaker, C. A. (2019). pySOT and POAP: An event-driven asynchronous framework for surrogate optimization. arXiv preprint arXiv:1908.00420.

He, Z. H., Pomeroy, J. W., Fang, X., and Peterson, A.: Sensitivity analysis of hydrological processes to perturbed climate in a southern boreal forest basin, Journal of Hydrology, 601, 10.1016/j.jhydrol.2021.126706, 2021.

He, Z., Unger-Shayesteh, K., Vorogushyn, S., Weise, S. M., Kalashnikova, O., Gafurov, A., Duethmann, D., Barandun, M., and Merz, B.: Constraining hydrological model parameters using water isotopic compositions in a glacierized basin, Central Asia, Journal of Hydrology, 571, 332-348, 10.1016/j.jhydrol.2019.01.048, 2019.

Lyu, H., Tian, F., Zhang, K., & Nan, Y. (2023). Water-energy-food nexus in the Yarlung Tsangpo-Brahmaputra River Basin: Impact of mainstream hydropower development. Journal of Hydrology: Regional Studies, 45, 101293.

Nan, Y., He, Z., Tian, F., Wei, Z., and Tian, L.: Can we use precipitation isotope outputs of isotopic general circulation models to improve hydrological modeling in large mountainous catchments on the Tibetan Plateau?, Hydrology and Earth System Sciences, 25, 6151-6172, 10.5194/hess-25-6151-2021, 2021.

Nan, Y., He, Z., Tian, F., Wei, Z., and Tian, L.: Assessing the influence of water sampling strategy on the performance of tracer-aided hydrological modeling in a mountainous basin on the Tibetan Plateau, Hydrology and Earth System Sciences, 26, 4147-4167, 10.5194/hess-26-4147-2022, 2022.