Reviewer 2

General comments:

This manuscript focuses on evaluating the value of snow cover area, glacier mass balance, and isotopes in reducing uncertainty and equifinality of hydrological modeling in a large mountainous basin in the Tibetan Plateau. The Bayesian approach and GLUE method are adopted to investigate the research questions. The research topic aligns with the journal scope and the research findings are potentially useful for the readers. I have a few concerns regarding the modeling procedure, the details of the input data, and the interpretation of the results before the paper being accepted for publication.

Additionally, one thing I noticed here is that the time-series simulated and observed discharge does not have a y-axis (Fig.5), which is present on purpose due to data dissemination restrictions mentioned in the caption. However, this is not possible for readers to understand the model performance, and the magnitude of the simulated and observed discharge. A manuscript avoiding showing y-axis of time-series discharge plot in the results could potentially conflict with the basic principle of open science of HESS/Copernicus journals.

We thank the reviewer for raising this important point. We fully acknowledge that omitting the y-axis from the time-series discharge plots limits the reader's ability to assess the absolute magnitude of both simulated and observed flows. We also recognize that this may appear to be at odds with the open science principles promoted by HESS and Copernicus journals. However, the decision to exclude the y-axis was made to comply with data dissemination restrictions imposed by the data provider, who explicitly prohibits the publication of absolute discharge values. We are aware of the implications of this limitation and will revise the manuscript to address it as transparently as possible. Specifically, we will: (i) add a normalized y-axis to enable relative performance assessment; (ii) include a note explaining how access to the original data can be requested under the provider's terms; and (iii) explicitly discuss this limitation in the manuscript, including a clear justification and a reflection on its implications for reproducibility. We note that similar cases—where normalization or partial data concealment was necessary due to access restrictions—have been accepted in previous HESS publications, provided they were accompanied by proper documentation and acknowledgment of the limitations (e.g., Nan et al., 2021 and Nan and Tian 2024). We are, of course, open to following the editor's guidance on how to best address this issue while respecting both the journal's policies and the constraints imposed by the data provider.

Specific comments:

Modeling perspective:

The subsurface is overly-simplified represented in the model. The subsurface flow generates from the model is composed of the subsurface lateral flow ("interflow") in the unsaturated zone and the baseflow from groundwater to surface water in the saturated zone. These two subsurface flow components are simulated as a sum (L105 and Fig.1). It is thus not possible to conclude the role of groundwater in contributing to the streamflow and the groundwater- surface water interactions. The subsurface lateral flow can be high and not negligible in such large mountainous basin (>2*105 km2). It is recommended to be cautious in interpreting and concluding the result regarding the baseflow. All mentioning of groundwater baseflow in the manuscript actually refer to the subsurface flow, i.e. the sum of both unsaturated and saturated zone, e.g. on L134, it is subsurface flow, but not baseflow. The presented modeling approach is not able to investigate groundwater alone.

We thank the reviewer for this detailed and constructive comment. We agree that the representation of subsurface processes in our model is simplified, and we acknowledge that subsurface flow in the current setup comprises both lateral flow in the unsaturated zone (often referred to as interflow) and baseflow from the saturated zone. These two components are indeed treated as a single aggregated term in the model, and as such, it is not possible to explicitly separate or quantify the contribution of groundwater to streamflow, nor to analyze groundwater–surface water interactions in detail.
In light of the reviewer's observation, we will revise the manuscript to replace the term "baseflow" with "subsurface flow" throughout, when referring to the total contribution from both the unsaturated and saturated zones. This change will be made, for instance, in L134 and similar occurrences. We will also include a clarification early in the methods section to define explicitly what is meant by "subsurface flow" in the context of our model.
Moreover, we will address this structural simplification as a modeling limitation and clarify that the current modeling approach does not allow us to investigate groundwater dynamics in isolation. We fully agree that caution is required when interpreting model outputs in terms of baseflow or groundwater contributions, especially in a large mountainous basin where interflow can be substantial.
Finally, when referring to streamflow during dry periods, we propose using the term "low flow" to avoid potential misunderstandings and to emphasize that our analysis focuses on the overall behavior of the hydrograph during low-flow conditions, rather than on baseflow in the strict hydrogeological sense.

Regarding the modeling, are the spatial zones delineated the same for both the surface and subsurface? (this could potentially fragment the aquifers located at the boundaries). Is the subsurface flow allowed to cross the delineated boundaries? The conceptualization of the subsurface processes in the model potentially limits the ability of the model for investigating the surface-subsurface interactions. The model limitation should be clearly discussed in Section 4.3.
In the model, only the runoff concentration process through the river network is considered to occur crossing the boundaries of simulation units (i.e., REWs). The runoff generation processes (both surface and subsurface) only occur within the REW. The model actually only considers the shallow groundwater, which is frequently recharged by the infiltration, but does not consider the deep groundwater cycle. This is indeed a limitation in model, especially for the TP, where previous studies indicated deep interbasin groundwater pathways existed. We will discuss these in the limitation.
There are 4 discharge stations, but only the results at Nuxia Station are presented. The results for the other stations should be presented in the Supplementary Information.

The authors should also clarify if the conclusions achieved at Nuxia Station are held the same as the other three stations.

We thank the reviewer for pointing this out. Although there are four national discharge stations in the basin, the data are not publicly accessible and can only be requested from the provider, subject to approval and specific conditions. We were able to obtain data for only two additional stations—Yangcun and Nugesha—besides Nuxia. In the revised version of the manuscript, we will include the results for Yangcun and Nugesha in the Supplementary Information. We will also clarify that the conclusions drawn at Nuxia Station are consistent with those observed at these two additional stations, as shown in Figures 1 and 2.

Does glacier melt contribute to groundwater recharge? Or is it assumed that all glacier melt goes into streamflow? This assumption should be clear in the text as well.

Yes, we assume that glacier melt generates streamflow directly through the surface pathway, because of the low permeability of the glacier surface. We will clarify this in the revised manuscript.

The simple degree-day-factor methods are used to solve snowmelt and glacier melt. Glacier mass balance is estimated with a simple volume-area scaling factor approach. The limitations of these adopted simple approaches for solving snow and glacier melts should be discussed in terms of modeling limitations.

These methods are indeed rather simplified, and the adoption of a spatially uniform degree-day factor cannot reflect the spatial heterogeneity of the melting processes. We will discuss them in the limitation section. The reason why we adopted a simplified method is that the study basin is very large for a hydrological model, and we need to make the model adequately efficient for the subsequent GLUE analysis. Besides, although the degree-day factor method is simplified, it is a rather commonly used method, and performs effectively in snow and glacier simulation (especially when we are concerned about the characteristic at a large spatial scale).

L213: how are the Pareto fronts defined? Please justify this threshold used to show the Pareto fronts in Figure 3 and the conclusions obtained from this result relating to this threshold.

The Pareto fronts shown in Figure 3 (red points) are defined based on the standard multi-objective dominance criterion, where no other solution performs better across all objectives and strictly better in at least one. The blue dashed lines are not part of the Pareto front definition but are used to indicate performance thresholds (e.g., NSE > 0) that help distinguish between simulations with at least minimal predictive value and those that are clearly inadequate. In particular, an NSE value below zero indicates that the model performs worse than a simple baseline given by the mean of the observed data. The NSE = 0 line is therefore used as a permissive reference to highlight the broad performance landscape, including poor-performing solutions. A similar reasoning applies to the glacier mass balance indicator VE: values below zero indicate that the mean simulated mass balance deviates from the observed mean more than a null reference would (i.e., assuming zero glacier mass change), and are therefore also interpreted as indicative of non-behavioral model realizations. These thresholds support a transparent visualization of the trade-off space and enable a clear separation of solutions with meaningful predictive skill. The conclusions regarding trade-offs remain consistent across different thresholds, and the ones chosen here serve to support a meaningful interpretation of solution clusters. We acknowledge that this methodological choice and its implications were not clearly explained in the initial version of the manuscript. To improve clarity, we will add a dedicated paragraph to explicitly justify the threshold selection and discuss its influence on the interpretation of the results.

L248-249: Why does the KKA shows a noticeable convergence, but not KKD? They both are parameters that control the subsurface runoff outflow rates. Please clarify this point.

We acknowledge the reviewer's comment and agree that the contrast in convergence between KKA and KKD deserves clarification. Although both parameters influence the subsurface runoff outflow, their functional roles in the model differ significantly. KKA is an exponential coefficient, meaning that even small changes in its value can lead to large nonlinear variations in the outflow rate. This makes KKA highly sensitive and more easily constrained by the calibration targets. On the other hand, KKD acts as a linear coefficient, whose effect on the runoff response is more gradual and can be compensated by other interacting parameters. As a result, KKD tends to be less identifiable, leading to a flatter posterior distribution and a lack of noticeable convergence. We will clarify this point in the revised manuscript.

Data perspective:

Limited information is provided on the input data of this study. This could hamper the readers to interpret the results.

L79: The four river gauging are only given by names and no other information and data are available. It is recommended to provide details on the coordinates and elevations of the four river gauging stations in this mountainous basin, also their observed periods, frequency, and measurement method. Any observations errors/failures in the winter low flow and high flow periods? These details are important to interpret the observed and simulated discharge.

We appreciate the suggestion and will include a table summarizing the basic information of the hydrological stations utilized in our study. However, it's important to note that, despite being part of China's national monitoring network, detailed information—such as specific measurement methodologies and potential errors during low or high flow conditions—is not publicly accessible. Our access to these data was facilitated through personal connections, reflecting the broader challenges associated with water data availability in China. This situation aligns with the issues highlighted by Lin et al. (2023), who discuss the limited accessibility and usability of China's water data and advocate for the development of a national water data infrastructure to enhance data sharing and support effective water resource management.

Station	Coordinate	Elevation (m)	Streamflow	Isotope				
			Period	Period (in 2005)	Precipitation		Stream water	
					Sample number	$\overline{\delta^{18}O}$ (‰)	Sample number	$\overline{\delta^{18}O}$ (‰)
Nuxia	94.65°E, 29.47°N	3691	2001-2015	14 Mar-23 Oct	86	-10.33	34	-15.74
Yangcun	91.82°E, 29.27°N	4541	2001-2010	17 Mar-5 Oct	59	-13.14	30	-16.57
Nugesha	89.71°E, 29.32°N	4715	2001-2010	14 May-22 Oct	45	-14.29	25	-17.84
Lazi	87.58°E, 29.12°N	4889	/	6 Jun-22 Sep	42	-17.41	22	-16.52

Table 1: Data and sample information at four stations

Section 2.1: What is the modelling period? Please detail the start and end dates of the meteorological data sets and the modelling period. Also add details on which years of DEM, land use data, soil data, snow cover, and glacier data are used in this modelling study.

The modelling period is 2001.1.1-2015.12.31, which is also the start and end date of the meteorological data we have adopted.

Details of how each dataset used in the model:

- DEM is used for dividing the whole basin into several REWs, and calculating the REW attributions for model calculation (e.g., the basin area, the length and width of river channel, the slope of the hillslope).
- The vegetation data (NDVI and LAI) is used to determine the proportion of vegetation covered area (i.e., the vi-zone and vd-zone in the model) and the interception ability.
- The soil data is used to determine the soil properties not obtained by calibration, including saturated hydraulic conductivity, soil porosity, soil pore distribution index, field capacity, and air entry value, which are used for the simulations of infiltration, exfiltration and groundwater outflow processes.
- The snow cover data is used for the calibration of SCA simulation.
- The glacier cover area data is used for determination of the boundary of regions where glacier simulation is conducted, and for dividing the glacier simulation unit. The glacier elevation change data is used for the calibration of GMB simulation.

More detailed descriptions of each dataset and its role in the modeling framework are provided in the main text of the paper.

3. L79-93: Are the gridded meteorological satellite data corrected with in-situ station data? How are the different resolutions of various types of gridded spatial data used in the hydrological model? Please provide details on this.

The correction of satellite data with in-situ station data is not conducted in our study, but this process is included in the production process of some datasets, such as the precipitation and temperature data in the CMFD dataset (He et al., 2020: the first high-resolution meteorological forcing dataset for land process studies over China | Scientific Data). The only data correction conducted by ourselves is to correct the isoGSM precipitation isotope data by the measurement data, which will be explained in detail in the response to the next comment.

The simulation unit of the model is REW, the average area of which is ~700 km2, larger than the spatial resolution of all the gridded products. The areal averages of each factor are calculated in each REW, which are used as the input for the simulation.

We will explicitly state this in the revised version of the manuscript to improve clarity.

The description of the streamflow sampling is very vague, which is simply stated as "Grab samples of stream water were collected in 2005 at four stations..". Please provide details on how many samples and in which months the samples were collected. Do the authors have the precipitation (rainfall, snow) isotopes in the same year (2005) or in a different year (2008)? Using streamflow and precipitation isotope data of different years in the same model can be inappropriate.

We will provide a table to show the detailed information of the precipitation and river water samples, including the number of samples, the sampling period, and the isotopic characteristics. Precipitation samples were also collected in 2005 for the same period. The isotope data of these precipitation samples were used to correct the gridded output of the isotopic general circulation model isoGSM, to drive our model. We will add a description of this process in the data section of the revised manuscript.

How is the precipitation tracer estimated for rainfall and snow individually? This needs to be clarified in the manuscript.

As explained in the last response, the precipitation isotope produced by the isoGSM corrected based on observation precipitation isotope data was adopted as the model input. The isotope compositions of rainfall and snowfall were assumed to be the same, but that of snowpack and snowmelt in the catchment was simulated based on the balance equations of water and isotope mass, similarly as other water bodies.

We will clarify these aspects in the revised manuscript.

Interpretation of the results:

L256-259: The SCA shows a higher influence on the posterior distribution of T0 than the GMB, which does not show the strongest influence as the authors interpreted. Could the authors please clarify why they see GMB as the strongest from this result figure (Fig.4j)?

Thank you for pointing this out. You are absolutely right: the SCA shows a stronger influence on the posterior distribution of T_0 compared to the GMB, as is clearly evident in Figure 4j. We acknowledge this misinterpretation in our original statement and will correct the text accordingly in the revised manuscript to accurately reflect the results.

L306-308: The isotope data have increased uncertainty of the simulated glacier melt runoff (Fig.6d), but they are helpful to constrain other surface runoff components (rainfall runoff, snowmelt). Please clarify this result.

Glacier meltwater is assumed to generate runoff directly through surface pathway, so its contribution is not related to the partitioning between surface and subsurface runoff, for which isotope is helpful. Consequently, isotope cannot reduce the uncertainty of glacier meltwater.

L272-277: Including the isotope data leads to a decreased containing ratio. This means a significant under capture of the extremely low and high streamflow. Why including isotope data has decreased the streamflow simulation performance? Please clarify this result.

We thank the reviewer for this question. The observed reduction in the containing ratio when including isotope data is a consequence of the additional constraints introduced by the isotopic information. Isotope observations provide insights into flow paths and residence times, which reduce the set of parameter combinations that are simultaneously consistent with both hydrological and isotopic signatures. This leads to a more selective model response, where some simulations that were previously acceptable based on streamflow alone are now rejected due to inconsistencies with the isotope data. This effect is not limited to extreme flows, but applies more generally to the full range of simulated dynamics. It highlights the trade-offs that emerge in multi-objective calibration, where integrating multiple sources of information typically narrows the behavioral parameter space. The reduction in the containing ratio is particularly marked when compared to other observational targets such as snow cover area or glacier mass balance. While these variables mainly constrain the water balance and storage dynamics, isotope data provide independent information on flow partitioning and subsurface mixing processes. This imposes stronger internal constraints on the model structure and functioning, making it harder to compensate structural mismatches through parameter adjustment alone. As a result, fewer parameter sets are able to simultaneously satisfy both streamflow and isotope-based objectives. This highlights the added value of isotope data for model diagnosis and internal consistency, but also explains the greater selectivity they introduce during calibration.

Technical corrections:

L8: It would be helpful to mention which type of hydrological model the THREW-T is in the abstract. i.e. fully-distributed, semi-distributed, or conceptual?

L78: km2 should be straight upright, not italic. Please correct all formats of the units for similar cases.

L80, L82: Please add years between which the mean annual precipitation and mean annual temperature are calculated.

L100: distributed -> semi-distributed?

L104: what is bare zone? Bare soil, bare rock?

Figures 2 and 4: avoid using red and green colors together in the same figure to allow readers with colour vision deficiencies to correctly interpret your findings.

Table 1: table caption should be on top of the table.

L165: Please correct the formats of Equations 1-6 by following the journal guideline. e.g. the NSE should be straight upright, not italic. The text subscription should be straight upright as well.

L210-211: NSE, VE should be straight upright, not italic. Please check the format of all such mentioning.

L266-267, 281: Please remove the parentheses around the Section and Figure numbers, and correct all such mentioning in the manuscript.
Figure 4 caption i) covered area -> snow covered area.

Thank you for your careful review and helpful suggestions. We will address all the technical corrections you pointed out in the revised version of the manuscript.

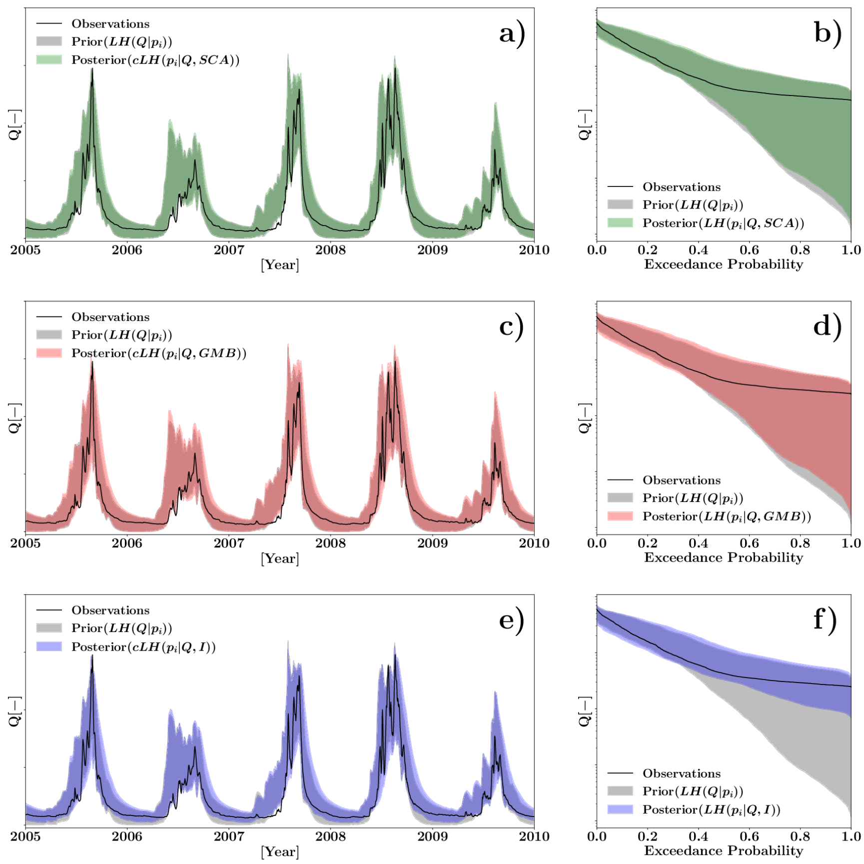

Fig. 1: The 5–95% percentile prior, conditioned solely on streamflow, and posterior predictive uncertainty ranges for streamflow, calculated under different conditions: snow cover area (SCA), glacier mass balance (GMB), and isotopes (I) at Yangcun gauging station . Left panels: daily streamflow time series for the period 2005–2010; right panels: flow duration curves for the entire period 2001–2015.

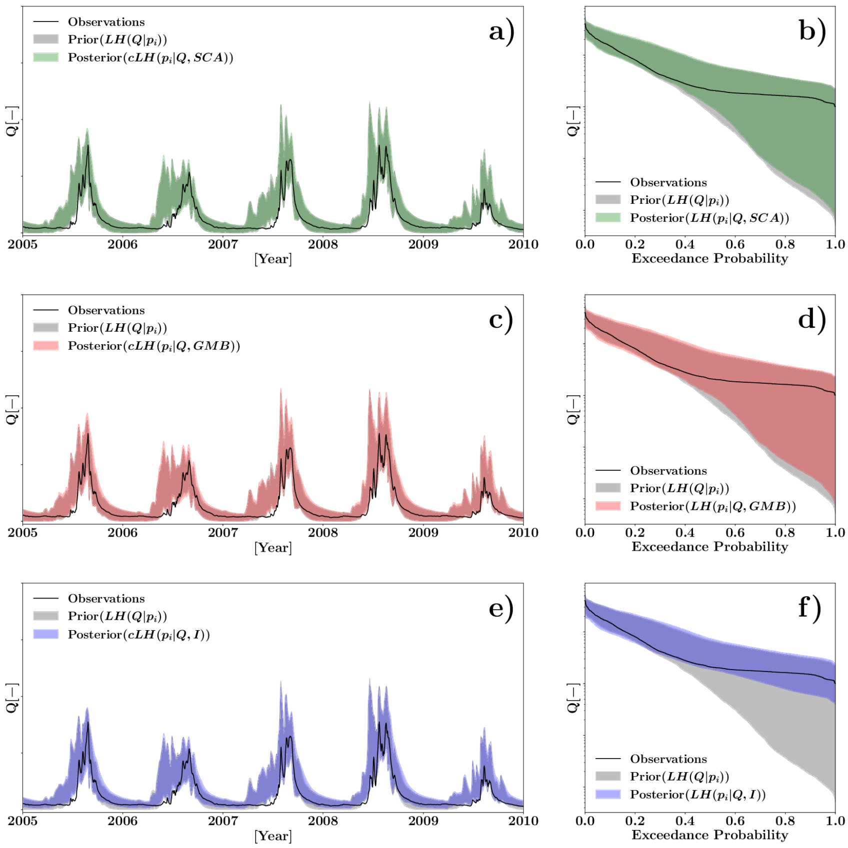

Fig. 2: The 5–95% percentile prior, conditioned solely on streamflow, and posterior predictive uncertainty ranges for streamflow, calculated under different conditions: snow cover area (SCA), glacier mass balance (GMB), and isotopes (I) at Nugesha gauging station . Left panels: daily streamflow time series for the period 2005–2010; right panels: flow duration curves for the entire period 2001–2015.

Bibliography:

He, J., Yang, K., Tang, W. et al. *The first high-resolution meteorological forcing dataset for land process studies over China.* Sci Data 7, 25 (2020). https://doi.org/10.1038/s41597-020-0369-y

Lin, J., Bryan, B.A., Zhou, X. *et al.* Making China's water data accessible, usable and shareable. *Nat Water* **1**, 328–335 (2023). https://doi.org/10.1038/s44221-023-00039-y

Nan, Y., & Tian, F. (2024). *Glaciers determine the sensitivity of hydrological processes to perturbed climate in a large mountainous basin on the Tibetan Plateau*. Hydrology and Earth System Sciences, 28(3), 669–689. https://doi.org/10.5194/hess-28-669-2024

Nan, Y., He, Z., Tian, F., Wei, Z., & Tian, L. (2021). *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(12), 6151–6172. https://doi.org/10.5194/hess-25-6151-2021