Anonymous Referee #2

Main comments

1. The study is well-written and concise. The model calibration and sensitivity analysis are detailed and well described However, there are some remaining concerns that the authors may account for before their manuscript can be considered for publication:

We sincerely thank the reviewer for carefully reviewing our manuscript and for the thoughtful, constructive feedback.

2. For the reader, who is not familiar with the author's preceding work, it is not clear how the model works. The schematic description in Fig 2 indicates that ET is taking place from the slow and fast karst groundwater storages, which would be quite unusual. To avoid misconception, please provide a complete model description in appendix (the table A1 is hardly understandable).

Reply: In this model, the karst critical zone in the hillslope was conceptualized as one reservoir, but the water stored in the reservoir was further sub-divided into upper active and lower passive storage zones (Fig 2) for the simulation of isotope ratios and estimation of water ages. This division follows our previous measurements of the vertical distribution of the rock fracturs/conduits along hillslopes where the large rock fracturs/conduits decrease exponentially in the vertical direction (Zhang et al., 2011).

The karst critical zone in the depression was conceptualized as two connected reservoirs, fast and slow flow, representing the solutional conduits in karst aquifers connecting with intergranular pores and fractures (often termed as matrix porosity).

The evapotranspiration could occur from the rich conduit/fracture areas by extended plant roots in the deep aquifer (Rong et al., 2011). Therefore, evapotranspiration is sourced from both the fast and slow reservoirs in the model.

We will revise the model descriptions in an appendix for clearer explanation of the module functions and meanings in the revised manuscript.

3. Some clarification on where the novelties of this work start is necessary. The authors inform the reader that in Zhang et al. (2017), the model was developed in previous work that used tracer data in addition to stream discharge to constrain the model structure, improve parameterization, and aid calibration. If this was done before, and the methods only describe how the isotope enabled model was parametrized and evaluated, what is the novelty of this particular study?

Reply: The model in the preceding work (Zhang et al., 2017), conceptualized the flow and the geochemical solute (Ca+Mg) routings using conceptualization of the dual flow system at the catchment scale. So, the original model had no basis for disaggregating the hydrological connectivity between different landscape units (e.g. "hillslope- to- depression- to- stream" in the study catchment). The hillslope-depression is a typical landform with variable hydrological connectivity in the karst catchments in southwest of China (Figure r1, Chen et al., 2018). Here, we improved our previous model structure by conceptualizing the hillslope and depression units (the improved part is in the red dotted box in Figure r2), and then use the hourly discharge and isotope values to calibrate the model. In addition, the new model has the parameters to represent passive storage inferred by isotope damping and the function of estimating the water ages from various landscape units in the catchment.

Although the tracer-aided model enhanced our understanding of the hydrological connectivity between different landscape units and the mixing processes, it increased the model parameters in the tracer modules. Therefore, we also evaluated the uncertainty of the simulation results including flow discharges, isotopic values, storages and ages at the different landscape units in this study.

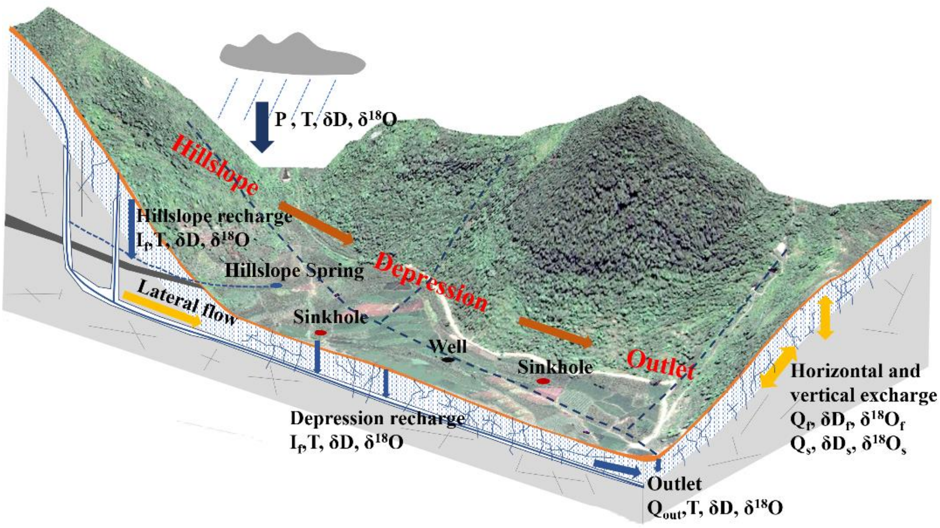


Figure r1 Sketch map of karst hydrological processes (Chen et al, 2018)

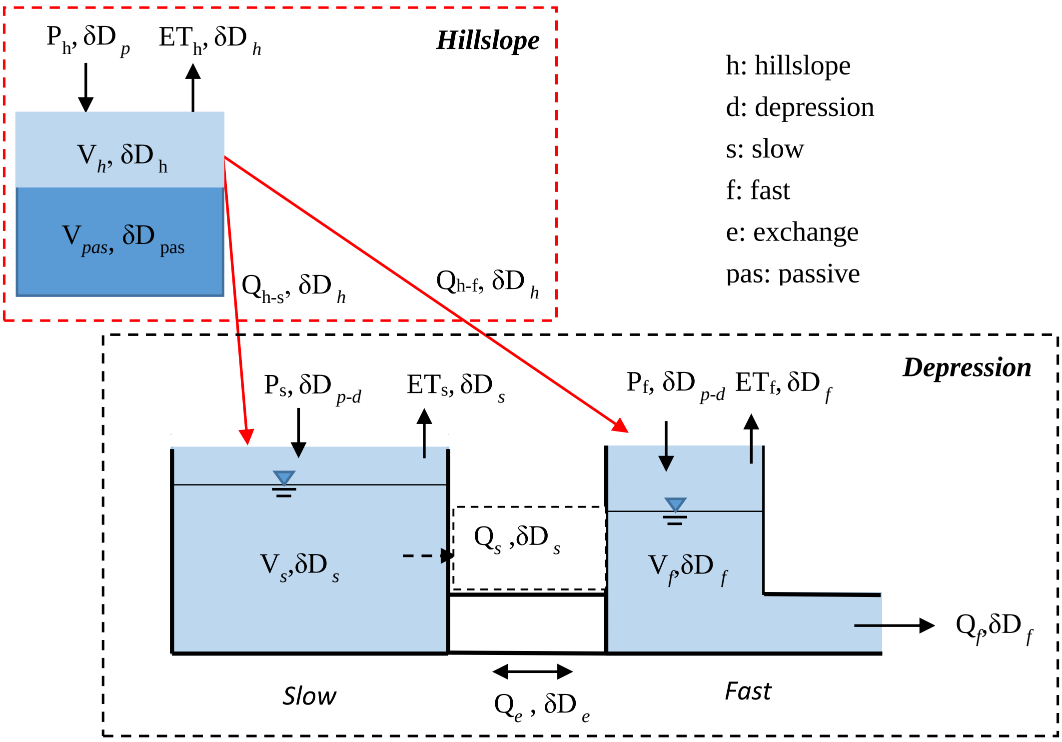


Figure r2 Structure of the improved model, and the improved part is in the red dotted box

4. Figure 4 shows that only 5 of 12 parameters are sensitive, which is quite a low number. Usually, discharge contains enough information to identify 4-6 parameters (Jakeman & Hornberger, 1993). Adding of additional information like isotopes should increase this number, if the model structure is well-chosen. To check the contribution of discharge data and isotopes, could the authors show the parameter sensitivities using discharge or water isotopes only?

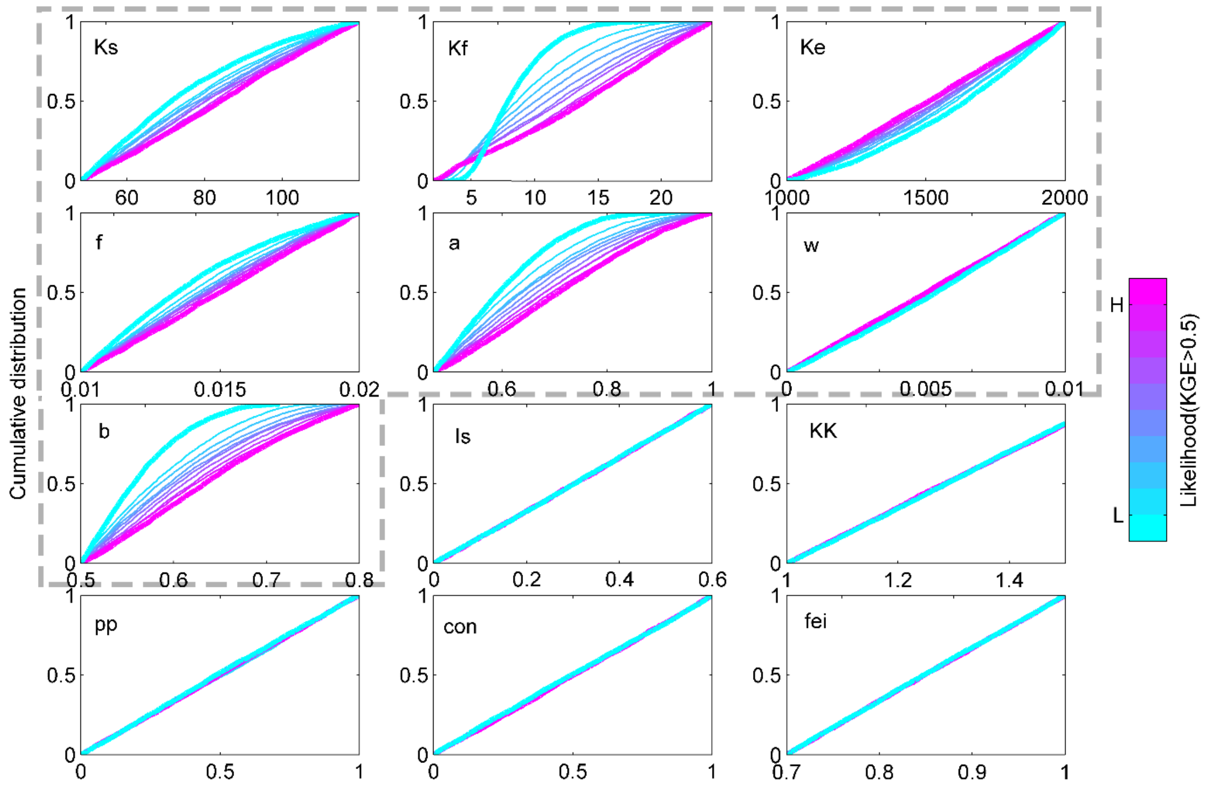
Reply: The trace-aided model includes 12 parameters, seven for flow routing (*Ks*, *Kf*, *Ke*, *f*, *a*, *w*, and *b)* and five for isotope ratios and water ages *(Is*, *KK*, *pp*, *con* and *fei*). So, the overall model increased by five parameters in the isotopic module.

We analyzed the parameter sensitivities using either the outlet discharge and/or water isotopes. Targeting the discharge, six parameters (except w) among the seven parameters in the flow routing module are sensitive and the parameters in the isotopic module are all insensitive (Fig r3 (a)). Targeting only isotopic values and both flow discharge and isotopic composition, the sensitive parameters are same, including *Kf*, *a*, and *b* in the flow routing module, and *Is* and *fei* in the tracer module) (Fig r3 (b) and (c)). Using both flow discharge and isotopic composition as the target, these parameters were more sensitive than those using only isotopic values (see the wide ranges of the cumulative distributions in Fig r3 (c)).

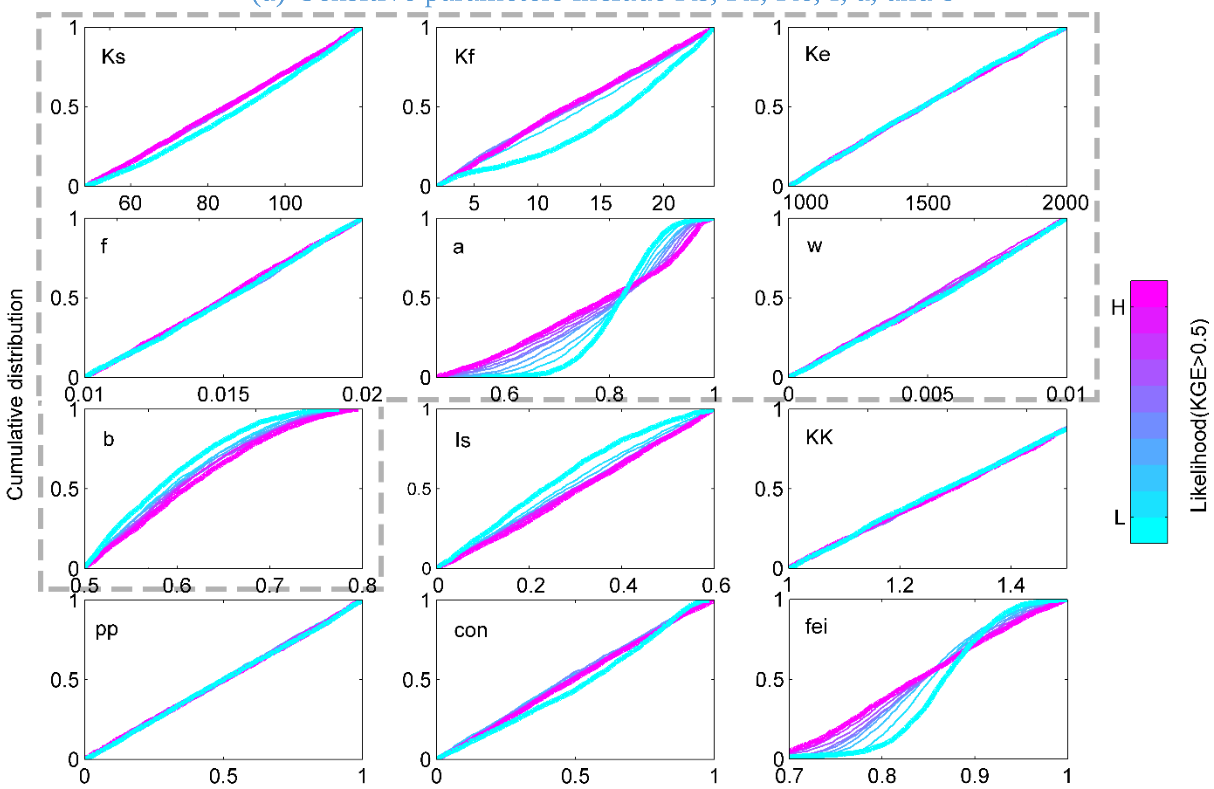
Interestingly, increasing the two sensitive parameters in the isotopic module (the coefficient for evaporation fractionation *Is* and the weighted isotope composition of rainfall input by the parameter *fei*) results in three parameters in the flow module becoming insensitive (slow reservoir constant (*Kf*), the exchange constant between the two reservoirs *Ke* and the ratio of porosity of the quick to slow flow reservoir *f*).

This can be explained as follows: the former two sensitive parameters in the isotopic module emphasize atmospheric effects on the outlet flow (being "old/new"). Larger *Is* indicates more evaporative effect on the stored water, leading to the stored and released water being older, particularly during the dry period. Larger *fei* indicates newer rainfall recharge (more negative isotopic values) into aquifer, leading to the stored and released water being newer during rainfall period. Alternatively, the latter three parameters in the flow module emphasize effects of fast (newer) and slow (older) flows in aquifer on the outlet flow (being "old/new"). More water release from the slow reservoir (larger *Kf*) and greater release of the slow reservoir into the fast reservoir (larger *Ke*) could lead to the released water being older in the dry season; a high proportion of the fast flow storage (larger *f*) and a greater exchange between the fast reservoir and the slow reservoir (larger *Ke*) could lead to the released water being newer in the wet season. Consequently, there is equifinality for these parameters in the trace-aided model, which can be overcome only when we have additional data to constrain some of the parameters, e.g. knowing the evaporative effect on water *Is* and the weighted isotope composition of rainfall input by the parameter *fei*.

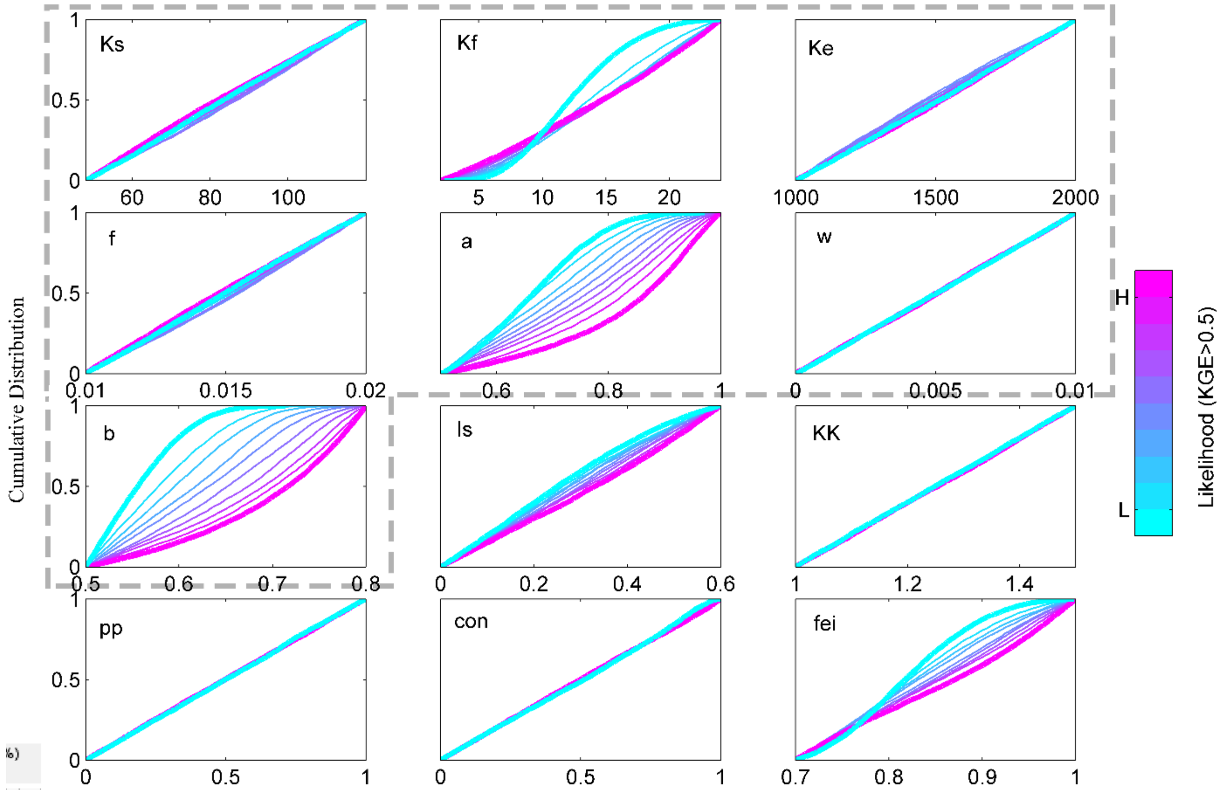
We will add the above reasoning in our discussion of the revised manuscript.



(a) Sensitive parameters include Ks, Kf, Ke, f, a, and b



(b) Sensitive parameters include Kf, a, b, Is and fei



(c) Sensitive parameters include Kf, a, b, Is and fei

Figure r3 Sensitivity of 10 model parameters using (a) flow, (b) isotope composition and (c) combined simulation of flow and isotopic composition. (The parameters inside the gray dotted box are for flow routing, and the outside parameters are for isotope routing.)

5. With a large fraction of the model parameters insensitive, how conclusive are the interpretations on the model internal dynamics that the authors use to explain connectivity and water age distribution in the system? In some of the figures, uncertainty ranges are provided and they are quite wide. In other figures (e.g., Fig. 5), only the mean is provided although the parameters controlling the observed processes ("w" in case of Fig. 5) are insensitive.

Reply: In the revised manuscript, we will describe the uncertainty of the modeled results for the various landscape units in the catchment. We reached the following conclusions:

The outlet hydrometric and isotope observations (consisting of mostly young and fast flows) were used as the calibration targets in this study. The outlet simulations had the least uncertainty, while uncertainty in the hillslope and depression units were highly related to their hydrological connectivity with the outlet. The simulated fast flows in the hillslope and depression units had lower uncertainty than the simulated slow flows in the depression since the two former units are highly connect with the outlet.

Although some parameters (e.g. w controlling hillslope flow dynamics) are insensitive, uncertainty bands of the hillslope flow (Fig r4) are narrow and the model captures quite well the hillslope seasonality and event-based dynamics through targeting the best matching of outlet discharges and isotopic values. This indicates that the hillslope dynamics are closely linked to the outlet dynamic patterns (with strong connectivity between them), which is consistent with the ranges of δD and $\delta^{18}O$ values at the hillslope spring being close to the ranges at the outlet discharge in Table 2.

For some of the other insensitive parameters (e.g. *Ke* that determines the exchange amount between the fast and slow flow reservoirs), uncertainty of the simulated exchange flux is much higher than that of the water fluxes from direct rainfall recharge and hillslope fluxes (see Figure r5), indicating weaker connectivity of the slow flow with the outlet flow.

In the revised manuscript, we will also present analysis of uncertainties of the modelled storages and ages (using the new capability of the tracer-aided model) at the various units. This shows that: the uncertainty increases with ages in comparison of the three water fluxes (Fig 9 in the original version of the manuscript), i.e. the narrowest uncertainty bands for the youngest hillslope flow and the widest for the oldest slow flow. Seasonal change in uncertainty also increases with age for the younger hillslope flow and the fast flow in depression. However, for the slow flow in the depression, change in the uncertainty decreases with ages, e.g. the bands tend to be wide for the younger water during rainfall season in Fig 9. The greater uncertainty as the slow flow becomes younger also reflects that the uncertainty is likely explained by the insensitive parameter of *Ke* (reducing effect of the frequent exchanges between fast and slow reservoirs on the outlet flow).

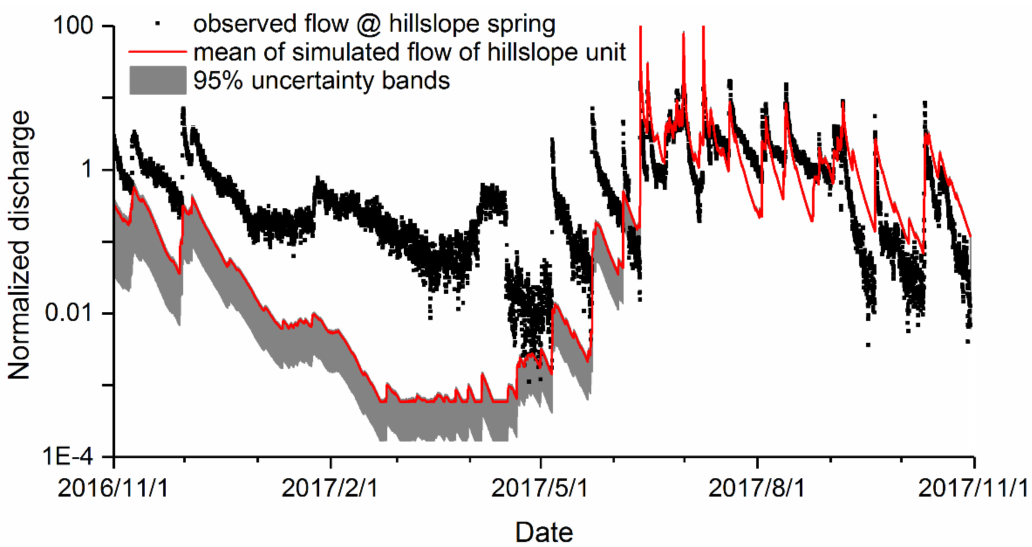


Figure r4 Observed discharge at hillslope spring against the simulated discharge of hillslope unit (values are normalized)

Detailed comments

Line 66: There are a few studies on water storage, flux and age dynamics using tracers in karst environments.

Reply: we will revise this expression.

Line 132: If this was done before, what is the novelty of this study?
Reply: See response to Q3 in the main comments above.

Line 145: Is there are distinction between soil/epikarst and groundwater? What controls matrix-conduit exchange?
Reply: We will revise the descriptions. In this model, we conceptualized the groundwater aquifer in the depression by a dual flow system (involving fast and slow flow reservoirs), and the groundwater aquifer in the hillslope by an upper active storage (mostly from epikarst) mixing with a lower passive storage since the rock fractures/conducts reduce with depth from the ground surface in the hillslope profile according to our previous investigations.

The exchange between matrix and conduit is controlled by the water storage (relate to water level) and the exchange constant between the two reservoirs (*Ke*) in each reservoir.

Line 222: This is not correct - please remove
Reply: We will revise this.

Line 260: typo
Reply: We will revise this.

Line 296: only 5 of 12 parameters are sensitive, which is quite a low number. Usually, discharge contains enough information to identify 4-6 parameters (Jakeman & Hornberger, 1993). Adding of additional information like isotopes should increase this number, if the model structure is well-chosen. To check this, could the authors show the parameter sensitivities using discharge only?
Reply: see the reply to Q4 in the main comments above.

Line 299: In the text, a rejection limit of 0.3 is mentioned. Please clarify
Reply: Two step calibrations were carried out in this study. First, 10^5 different parameter combinations were selected with the broad ranges of initial parameter values. And then, we obtained the narrower ranges of the parameters according to the best modelled results (meeting the KGE >0.3 criteria). For the second calibration, the narrowed ranges of the parameters were used as the initial ranges of the parameter to search the next best modelled results (KEG >0.5). We will revise the descriptions in the new manuscript accordingly.

Line 308: The parameter w is completely insensitive meaning that this storage's dynamics are not well identifiable, right? Please provide all behaviorals instead of the mean to show the precision of simulation of its discharge
Reply: See the reply to Q5 in main comments above.

Line 324: how much can you conclude from such wide uncertainty ranges?
Reply: the greater uncertainty of the modelled isotopic values in the depression arose from the insensitive parameters of *Ke* and *Ks* that affect the slow flow discharge and its exchange with the fast flow when the outlet hydrometric and isotope observations (consisting of mostly young and fast flows) used as metrics for the objective function for model calibration.

Here, the modelled isotope composition in the depression (see Figure 6b) refers to the release of water from the slow flow reservoir, representing a relatively constant source. The uncertainty bands can cover the limited variability of the measured values of δD at W1 and W5 (blue and yellow points in Figure 6b) where the aquifer has much lower permeability (W5) and is confined (W1) (cf the geophysical survey reported by Chen et al, 2018). This means that our tracer-aided model capture the slow flow dynamics in the depression even though the uncertainty is large.

The highly negative values of δD at W3 and W4 (red and black points in Figure 6b) are mostly below the uncertainty bands. This means that the stored water at W3 and W4 was younger than water from the slow flow reservoir, which is consistent with recent geophysical evidence (see Chen et al, 2018). Since W3 and W4 are located at high permeability areas, water at W3 and W4 was contributed mostly by fast flows (mixing with the young water), particularly during rainfall events (e.g. 9/7, and 20/7 in Fig 6b). So the high negative values of δD at W3 and W4 below the uncertainty bands were reasonable. We will revise the descriptions accordingly in the new manuscript.

Line 368: KE is also quite insensitive. Can you also show the entire 500 ensamble (or confidence limits)?

Reply: The simulations from the entire 500 ensemble are shown in Fig r5. Since the parameter *Ke* that determines the exchange amount between the fast and slow flow reservoirs is insensitive, the simulated exchange flux is highly uncertainty, though much smaller, compared to the water fluxes from the rainfall recharge and hillslope flow.

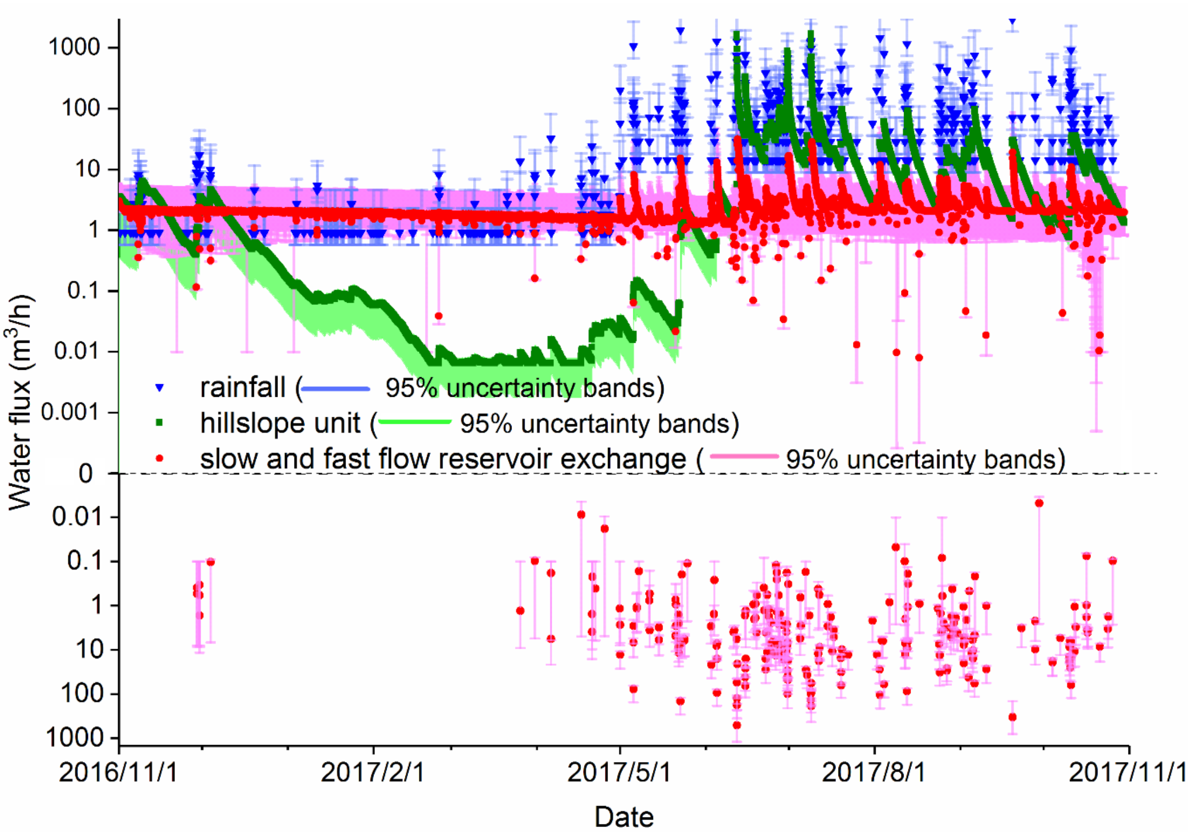


Figure r5 Source contributions to the underground stream flow (fast reservoir) at the catchment outlet. The red dots above and under the dotted line represent transient reverse water fluxes from the slow reservoir to fast reservoir and fast reservoir to slow reservoir, respectively.

Line 387: Please double-check this with literature values. Fast flow components in karst systems provide water with ages mostly between days or weeks (including temporal storage in the epikarst). Mostly, the ages found here are too large, even in the wet period.

Reply: We believe that the estimated ages are reasonable. Most models do not include mixing processes with stored water so tend to under-estimate water ages.

Here we listed δD values at the sampling points in this catchment for the two largest rainfall events in 2017 (the details refer to Chen et al., 2018, https://doi.org/10.1002/hyp.13232).

Date	rainfall amount	rain water	outlet water	hillslope spring
12/6	86.6 mm	-85	-48 ~ -70	-62~-67
9/7	83.4mm	-80	-62~ -73	-59~ -70

It shows that δD values at outlet and hillslope spring are much less negative than rainwater. So there was strong mixing of the "new" rainwater with "old" stored water during and after the rainfall although the response of discharge to rainfall is fast.
Also, our estimated ages in the manuscript refer to the mean of the ages over a long period of time. For short-term (event based) responses to the rainfall, the ages of water from hillslope flow and fast reservoirs can be shortest as 4 and 2 days, respectively. There were 8 and 23 events for the fast flow with the ages of water less than 5 and 10 days, respectively (see the lowest values in Fig 9). So, the results are not inconsistent with previous work, rather capture the time-variance of water ages. We will add these explanations in the revised manuscript.

Line 398: See comment above
Reply: The same response as for Line 387.

Line 409: Some recent example how this can be done with water quality data in karst:
Hartmann, A., Barberá, J. A., & Andreo, B. (2017). On the value of water quality data and informative flow states in karst modelling. Hydrology and Earth System Sciences, 21, 5971–5985. https://doi.org/10.5194/hess-2017-230
Reply: We will add this relevant literature.

Line 441: Large fractions of the fast reservoir have ages larger than several months, which appears a bit slow. (see also comments above)
Reply: See response to Line 387.