We thank the reviewer for spending their time reviewing our manuscript and for providing constructive comments. Please find below our point-to-point response to the reviewer's comments (text in black) and proposed modifications (text in blue) of the original manuscript.

In this study, the authors studied the uncertainty in transit time estimation. Two sources of uncertainty were considered: the assumed shape of the StorAge Selection (SAS function (and the uncertainty in the associated parameters) and the interpolation scheme for the precipitation tracer data. The reported uncertainty is large, resulting in a 90% confidence interval between 286 – 895 days for the median transit time. The uncertainty was greater in dry conditions than in wet conditions. The uncertainty depended more on the SAS function parameterization and the temporal interpolation of the precipitation tracer data than the spatial interpolation of the precipitation tracer data. Importantly, the authors argued that it could be useful to utilize the young water fraction, Fyw, in estimating the SAS function parameters, as it could constrain the SAS function and reduce uncertainty.

Thank you for this summary. The main intent of this paper is to explore the uncertainty in the and simulated isotopes and median transit times arising from the differences in model inputs, structure, and parameters. In our case, we want to study the uncertainty when different interpolation techniques are used to construct the high-frequency behavior of tracer data, for the application of SAS-based modeling framework. The use of Fyw is a subordinate objective as an attempt to suggest a further metric that might be helpful in constraining the model simulations of an already calibrated SAS model.

Though understanding the uncertainty in the SAS function and transit time is important, it is unclear what readers could learn from this manuscript in its present form other than the summarized results above for the specific catchment. As the results were not discussed enough in detail, it is not easy to think about their implications (see major comment 1). The suggestion of using the young water fraction in the SAS function estimation is interesting, but the authors' argument regarding using it must be more convincing (see major comment 2). In addition, some additional potential sources of uncertainty should be considered or mentioned explicitly (see my major comment 3). Recent advances in estimating transit should be mentioned (see my major comment 4). Thus, I think a significant revision is required before this manuscript can be considered for publication in HESS again.

Thank you for raising these issues. We agree with the reviewer that these issues were not addressed properly in the original manuscript; therefore, we will revise the manuscript accordingly. Please see below our detailed responses to these comments.

- Discussion of the results

I think it is necessary to discuss the results further to make the implications of this study clearer. The current manuscript focuses more on describing the results for the specific catchment and dataset than discussing the results, so it is not easy to think about those implications. For example, why is the uncertainty in the estimated TTD (or the median transit time) large? Why is the uncertainty greater under drier conditions? Why does the spatial interpolation method not substantially affect the water age simulation? Without discussing that type of question for each finding, it is not easy to truly understand the described results. uncertainty.

Thanks for this comment. We would like to point out that the original manuscript has already an entire subsection (5.4) that discusses the implications of uncertainty in water transit times for water quantity and quality studies. However, we agree with the reviewer that further discussion, including the points mentioned above, should be included in the revised manuscript, and they are outlined in the following.

Previous studies used only a specific SAS function and/or specific tracer data interpolation technique. Our work shows that there could be a wide range of different results (in terms of water age and instream isotopes, as well as model performances), and parameter uncertainties due to distinct model setups regarding SAS parameterization and tracer data interpolation technique, at least for our study area. With this, we want to encourage similar studies in other catchments to explore these uncertainties and examine whether or not they can influence their conclusions/implications for water quality and quantity management.

We would like to specifically address the questions posed in the comment in the revised manuscript as follows:

1) Firstly, the uncertainty analysis is done among all 12 tested setups corresponding to different combinations of spatial/temporal data interpolation techniques and SAS functions. We found that the uncertainty in the median transit time is large (Fig. 4), which is mainly due to the temporal interpolation of isotopes, which resulted in very different reconstructed input data depending on whether the step function or sine interpolation is used (Fig. S2). This explains why the simulated water transit times are different between the two interpolations or, in other words, why the uncertainty in the water transit time is large. Then, the choice of which SAS function to use also leads to differences in the simulated water transit times (Fig. 4), thus large uncertainty associated with the model parameterization. For example, choosing a time-invariant function created a time series with more moderate fluctuations (Fig. 4a, c, d, f, g, i, j and l), while the choice of a time-variant function led to more marked fluctuations (Fig. 4b, e, h, k), specifically between wet and dry conditions.

Secondly, the uncertainty analysis is done among the behavioral solutions within each single model setup. Here, we also found a large uncertainty which might be due to the poor identifiability of some model parameters (Table S1), such as the evapotranspiration parameter and the storage parameter - the latter being a key factor that deserves further attention in the application of the SAS framework for modeling outflow isotope signals.

2) We found greater uncertainty under drier conditions especially visible when using time-variant SAS functions (Fig. 4b, e, h, k), because the uncertainty increases along with the median transit time (Fig. 4), which is actually longer during drier periods. When the catchment is wetter nearly all flow paths are active and contribute to the streamflow. Thus, flows are generally "smoother"

and water ages are easier to constrain. Conversely, under dry conditions only selected flow paths are active, usually those longer flow paths carrying older water to the stream partly through a drier soil zone, where the flow is more erratic as the conductivity is controlled by the soil moisture. Hence, the flows in the soil matrix are less uniform, which could make it more difficult to constrain these older water ages. We will further elaborate this part in the revised manuscript.

3) The spatial interpolation method did not substantially affect the simulations (at least in our particular case) because there is no big difference between kriging and raw isotopes as it is possible to see in Figs. S1 and S2.

Reference:

Benettin, P., Soulsby, C., Birkel, C., Tetzlaff, D., Botter, G., and Rinaldo, A. (2017), Using SAS functions and high-resolution isotope data to unravel travel time distributions in headwater catchments, Water Resour. Res., 53, 1864– 1878, doi:10.1002/2016WR020117.

Overall, it needs to be clarified if the use of Fyw constrained the SAS function parameters in the right way.

Thank you for this comment. We agree with the reviewer that, in general, the robustness in Fyw from the sine-wave fitting approach to constrain SAS function parameter should be checked, for example, by looking at the simulated isotopes and model efficiencies from the SAS framework after constraining with Fyw. This can allow us to know if Fyw reduces the model outputs towards the 'right' or 'wrong' values. However, developing this in a more elaborate approach, which also considers the uncertainties associated with Fyw and the young water threshold (τyw) would go beyond the scope of this paper. So, we do not intend to present the use of Fyw in the revised manuscript, but we plan to develop a more elaborate way to account for uncertainties in Fyw and τyw in another study. Please also refer to the more elaborate response in the other reviewer's comments document.

The authors somehow decided to state that the estimated young water fraction indicates the fraction of water younger than $\tau yw = 75$ days (in L171). However, as the authors mentioned, for example in L171 and L256, the method of Kirchner (2016) does not provide a single value for τyw that can be utilized universally. Rather, it varies with the shape of TTD. While the authors argued that the arbitrary decision is okay since they considered the uncertainty in the estimated Fyw (in L 171-172), that argument was made without clear reasonings that support it.

Thank you for this observation. We acknowledge that the uncertainty in τyw was not properly addressed in the original manuscript. We agree with the reviewer that, in general, there is the need to consider the uncertainty in both Fyw (which we accounted for in the original manuscript) and τyw (which we did not) when fitting the sine function to the tracer data in inflow and outflow. However, as we have already said, there is need to develop this in a more elaborate way that would go beyond the scope of this paper, so we do not intend to present this in the revised manuscript. Please also refer to the more elaborate response in the other reviewer's comments document.

Also, it needs to be clarified if the estimated Fyw based on the method of Kirchner is a good estimate that can be used to constrain the SAS function. The method of Kirchner is based on a set of assumptions. It seems like the authors want to argue that it's okay to use the estimate regardless of all assumptions because the estimated uncertainty of Fyw is low, but I think it would be better if the authors could provide more concrete arguments to convince readers why it's okay. Why is the uncertainty low? And how does the low uncertainty support that the estimate is a good estimate regardless of all the assumptions? In addition, the method approximates the precipitation and outflow tracer signal using sinusoidal functions, which was shown to be not a good approximation for the precipitation tracer data by the authors in this manuscript (e.g., in L305-308).

Thank you for this comment. As explained above, in the revised manuscript we do no longer present this approach. We agree with the reviewer that, in general, the assumptions of the sine-wave fitting approach do not apply to SAS modelling. For example, the isotope signal in inflow and outflow might not perfectly follow the sinusoidal and the marginal transit time distribution might not perfectly follow a gamma distribution. Therefore, there is uncertainty in the use of Fyw from the sine-wave fitting approach. We partly accounted for it in the original manuscript, but there is need of a more elaborate method that would go beyond the scope of this paper. Please also refer to the more elaborated response in the comments document of the other reviewer.

It also seems that the uncertainty in Fyw could depend on the temporal resolution of data (if the finer resolution data shows more deviation from the sinusoidal signal) and other properties that the authors mentioned in L348-350. Overall, based on the limitation discussed by the authors (in L348-350), I feel that the authors are unsure whether the utilization of Fyw will be useful for other datasets.

Thank you for raising this point. It is true that collecting temporally refined tracer data could potentially help infer a more robust Fyw from the sine-wave fitting approach. However, it has been demonstrated that Fyw is the only approach that can robustly estimate some age statistics for discontinuous and low-frequency tracer time series for at least 2-3 years (Benettin et al., 2022). Low-frequency measurements are more readily available than high-frequency measurements, especially over a wide spatial domain. For this reason, Fyw has the advantage to be much more largely applicable, and its use from less temporally refined tracer data should be acknowledged and further explored. There is certainly a need to account for its uncertainties but, as we will no longer include Fyw, there is no need to make this clearer in the revised manuscript.

Reference:

Benettin, P., Rodriguez, N. B., Sprenger, M., Kim, M., Klaus, J., Harman, C. J., et al. (2022). Transit time estimation in catchments: Recent developments and future directions. Water Resour. Res., 58, e2022WR033096, https://doi.org/10.1029/2022WR033096.

- A minor comment related to Fyw

L163-164: The method used to estimate Fyw was described too briefly. For example, L163-165 is not enough for readers to understand the method.

Thank you for this observation. As we will remove the Fyw part, so there is no need to make this point clearer in the revised manuscript.

- Other sources of uncertainty

I believe that the uncertainty in precipitation, discharge, and evapotranspiration rates could propagate into the uncertainty in the estimated SAS function. The list of potential sources of uncertainty provided by the authors (L39-43) needs to include them. It would be helpful for readers if the authors provided a more concrete list of potential sources of uncertainty. Also, it would be essential to provide why this manuscript, where the authors consider only a few sources of uncertainty, is still useful.

Thank you for mentioning this. We agree with the reviewer that the uncertainty in precipitation, discharge, and evapotranspiration rates could propagate into the uncertainty in the estimated SAS functions. However, in the current study, we do not consider these sources of uncertainty as we use the hydrologic simulations from the mesoscale Hydrologic Model (mHM, Samaniego et al. (2010); Kumar et al. (2013); Zink et al. (2017)), which is an established model to simulate daily discharge and evapotranspiration time series. However, we agree with the reviewer's suggestion to include in the revised manuscript this type of uncertainty in the list of potential sources of uncertainty affecting transit time-based models.

We have decided to focus on only a few sources of uncertainty that we think are the most significant and critical for SAS modelling, and also the most linked directly to the questions of data interpolation and SAS parameterization. Firstly, since there is no general agreement on which SAS function to use, we explored the uncertainty generated by the use of different SAS functions (i.e. model parameterization and parameters). Secondly, as SAS modelling requires continuous time series of input tracer data, we tested how different gap-filling techniques (i.e. temporal interpolations) affect the model results. Finally, as the SAS models depend on the type of input data, we tested what happens when using regionalized or non-regionalized isotopic datasets (i.e. spatial interpolations). In the revised manuscript, we will make this part clearer in the Introduction and Discussion by emphasizing the reasons why we decided to explore these specific sources of uncertainty.

Reference:

Samaniego, L., Kumar, R., and Attinger, S. (2010) Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resour. Res., 46, W05 523, 10.1029/2008WR007327, 2010.

Kumar, R., Samaniego, L., and Attinger, S. (2013) Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, Water Resour. Res., 49, 360–379, 10.1029/2012WR012195.

Zink, M., Kumar, R., Cuntz, M., and Samaniego, L. (2017) A high-resolution dataset of water fluxes and states for Germany accounting for parametric uncertainty, Hydrol. Earth Syst. Sci., 21, 1769–1790, 10.5194/hess-21-1769-2017.

- Missing new methods of estimating TTDs

There have been some recent advances in the estimation of TTDs that are not discussed in this manuscript. The method of Kirchner (2019) and the method of Kim and Troch (2020) can estimate time-variable (or state-dependent) TTDs without assuming their form a priori. The estimated TTDs can be converted to the SAS functions. Thus, some descriptions of the motivation of this study, such as what is in L55-59, need to be revised. useful.

Thank you for noting this. In the revised manuscript we will integrate the motivation of our study with the suggested literature.

- Minor comments

fSAS/rSAS: I think the authors should make it clear that they are discussing only the fSAS (fractional SAS) function. Another form of the SAS function, the rank SAS (rSAS) function, may have different uncertainty characteristics, especially because of the difference in how the storage is considered

In our study SAS functions are expressed as function of the normalized age-ranked storage (i.e., fractional SAS functions). We will clarify this in both Methods and Discussions.

Naming of the "BETA" case: Better to name the case more clearly. While the beta distribution is used without any state dependency or time-variability in this study, several studies utilized state-dependent beta distribution that can consider the time-variable flow pathways (e.g., Van der Velde et al., 2015). Thus, it could confuse readers when the authors state something like "BETA could be appropriate where the catchment release scheme is expected to be relatively constant" (in L290-291). In this manuscript, the time variability is mentioned explicitly in the case names only for the power law cases (PLTI and PLTV).

We will say explicitly that we tested this study for a time-invariant beta distribution, and will re name the BETA parameterization as BETATI (i.e. time-invariant beta).

L11-12: Make it clear that this confidence interval is for the median transit time.

We will clarify this.

L113: V(t) (mm) "is"

We will correct it.

L185: When using GLUE, the authors determined the behavioral parameters using an arbitrary criterion. The top 5% of the parameters, in terms of KGE, were selected as the behavioral parameters. However, there is no statement to support the decision regarding the criterion. It may be better for readers if the authors could explain why the criterion was chosen and whether this seemingly arbitrary choice is okay.

Thank you for pointing this out. We acknowledge that we have not supported the reasons why we chose to define a behavioral solution based on the top 5% of the parameters in terms of KGE. As also said in the response to the other reviewer, in this work we explore the impact of different model setups by checking the uncertainties in the simulated outputs. The goal is not to identify the best model efficiency or setup. Given this objective, and considering the large number of model setups used, we find it appropriate to define a behavioral solution based on a fixed sample size identified with the lower bound dependent on the maximum KGE for each model setup, in our case 5%. In the revised manuscript we will make it clear the reasons regarding the chosen criterion by providing the supporting motivation described above. However, in the revised manuscript we will apply the SUFI-2 approach rather than the GLUE approach, as in the former the uncertainty of parameters is described by a uniform distribution and not by a formal likelihood (see answer to other reviewer for further explanations).

L198: “more positive”: I am not sure if “more positive” is the right expression here. The isotope ratios are always negative in the data.

We will replace ‘’more positive’’ with ‘’less negative’’.

L188-190 I believe that the authors in general talked about the “backward” transit time distribution throughout this manuscript. However, the sentence in L188-190 is not clearly written if they talked about “forward” TTDs or the “backward” TTDs. Please revise.

Here, we consider the backward formulation of the transit time distribution. We will make it clear in the revised manuscript.

Figure 2: I think it would be better to enumerate the subfigures using numbers (based on Table 1) instead of alphabets.

As we understand it, the HESS guidelines for language and typesetting for submission state that subfigures labels should be enclosed in parentheses around lowercase letters (e.g. (a), (b), etc.). Therefore, we would like to keep the enumeration of subfigures using letters. To be consistent between subfigures and model setups we will enumerate the setups displayed in Table 1 using letters instead of numbers. Figure 3 will be changed accordingly.

L290-291: Not clear if this trivial sentence is necessary.

We would like to keep this sentence as it is relevant to this study in order to inform when specific SAS parameterization should be used. For clarity, we mean that the catchment preference to discharge water of a certain age is time-invariant with time-invariant SAS functions (PLTI and BETA) as these functions do not consider time variability as a function of the catchment wetness (such as PLTV does). Therefore, their use might be more appropriate for catchments experiencing relatively constant hydrologic conditions without highly pronounced seasonality.

L298-303: This paragraph is also trivial and not necessary for the manuscript.

We will delete this paragraph.

L305: I think the figure illustrating the interpolation results is more important for this manuscript than some unnecessary paragraphs mentioned above.

We will put more focus on this part by addressing the temporal interpolation of stable water isotopes in precipitation.

L345: Typo: “ETET”

We will change this to “ET”.

L360: Typo: “te modern”

We will change this to “the”.

L395: “smooth changes” are unclear.

We will replace it with ‘’relatively constant hydrologic conditions’’.