

Hydrol. Earth Syst. Sci. Discuss., author comment AC3
<https://doi.org/10.5194/hess-2021-355-AC3>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC1

Ravindra Dwivedi et al.

Author comment on "Tandem use of transit time distribution and fraction of young water reveals the dynamic flow paths supporting streamflow at a mountain headwater catchment" by Ravindra Dwivedi et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2021-355-AC3>, 2021

RC1: 'Comment on hess-2021-355', Anonymous Referee #1, 14 Oct 2021

Recommendation to Editor:

1.1 I recommend this paper be rejected for publication in Hydrology and Earth System Science (HESS) in its current form. I recommend the authors resubmit after major revision. The topic is certainly of great interest in scientific hydrology. The combination of data sets might be better leveraged to make clearer inferences about the true range of water residence times in small headwater catchments. I provide three general criticisms here that are reiterated in a numbered list of specific comments about the manuscript and graphics.

Dear reviewer, Thank you for your in-depth review of our paper. We plan on addressing each of your comments related to novelty and organization of our paper in the revised version of the paper documents. Please see also our responses to the comments # 1.2 and #1.4 below.

1.2 One major criticism is that the work relies heavily on antiquated methodologies. Major portions of the results are based on application of the so-called lumped-parameter-transport model based on time-invariant transit-time distributions (TTDs; equation 1). The TTDs have time-invariant parameters, which assumes that the distribution of flow pathways within the landscape and associated water velocities are constant in time. This assumption defies intuition, but was applied out of convenience for decades [e.g., most works reviewed by McGuire and McDonnell, 2006]. A theoretical basis for analysis based on time-variable TTDs was presented as early as Lewis and Nir [1978]. The theory has been advanced by many recent works [Botter et al., 2010; 2011; Harman, 2015; Rinaldo et al., 2011; van der Velde et al., 2012]. That TTDs should be time-invariant was shown to be theoretically implausible for low-order watersheds with dynamic flow [Botter et al., 2011]—a result that has been supported by empirical results from manipulative tracer experiments [e.g., Kim et al., 2016b]. I don't think our premier disciplinary journals should continue publishing results based on this antiquated approach.

Our response: We appreciate this concern but respectfully disagree. The papers cited in the above comment mostly focus on the highly dynamic and not the relatively stable part of a catchment's flow system (e.g., baseflow through fractured bedrock aquifers). The

papers cited above or those listed at the end of the reviewer's comments or similar papers related to dynamic transit time distributions (TTDs) have used either stable water isotopes (e.g., *Heidbüchel et al.* [2013]; *Heidbüchel et al.* [2012])) or similar tracers (e.g., chloride tracer in *Harman* [2015]; *Hrachowitz et al.* [2016]), which are applicable on shorter time scales [*Suckow, 2014*]. In contrast, our study uses tritium sampled under low flow or baseflow conditions, and this tracer is applicable under a longer time scale than stable water isotope tracers [*Suckow, 2014*]. Most importantly, the use of any time-variable transit time method (e.g., rSAS functions of *Harman* [2015], master equation method [*Botter et al., 2011; Heidbüchel et al., 2012*] or wavelet analysis methods [*Dwivedi et al., 2021*]) requires high-frequency data on hydrologic fluxes and tracer concentrations in inflow and outflow. These high frequency datasets of hydrologic fluxes and tritium concentrations are generally not available on the time scale for which tritium-based TTDs are estimated; this includes our study site, Marshall Gulch catchment, as well as many global sites as acknowledged in *Gleeson et al.* [2015]. Therefore, several studies, when modeling TTD using tritium sampled under baseflow conditions or in groundwater, use steady state TTDs (e.g., *Stewart et al.* [2017]; a paper published recently in HESS). Even with a coarser resolution, tritium observations are able to shed light on long-period dynamics in the transit time distributions for deep fractured bedrock aquifer groundwaters. In the revised version of our paper, we plan to clearly highlight the aforementioned point as a rationale for using the steady state version of TTDs.

1.3 A second major criticism is that there are significant shortcomings in the data, especially the measurements of Tritium in surface waters. There appear to be only 6 data points representing Tritium abundance in stream water. That's not many, but the authors rely on that data to calibrate and compare a range of models. Also, the time interval over which precipitation is sampled is coarse. Much of the temporal dynamics of tracer concentration in that inflow will be lost. For these reasons, most of the parameters for the various TTD models are not uniquely identifiable. The different models generate markedly different estimates of different water age metrics. There is inadequate guidance, or rationale, for the reader to understand which, if any, should be considered correct.

Our response: The purpose of this study was not to explicitly simulate high temporal resolution dynamics of tracer concentrations in stream water. In contrast, by using the stable water isotope tracers, our study seeks to estimate the fraction of young water metric in a time averaged sense to compare and contrast the value and information contained in that metric at our field site to literature-reported data from other sites, in order to better understand dynamic flow path behavior. The temporal resolution of the stable water isotope data in precipitation and stream water was sufficient to meet these study objectives, based on estimated Fyw values that were within the range reported in the literature using the TTD-based method. Please also note that the stable water isotope data (collected during water year 2008 through 2012) in our study were collected by the Santa Catalina Mountains and Jemez River Basin Critical zone observatory prior to this study. In our study, the TTD model performance was not only assessed by the value of the Kling-Gupta efficiency or KGE', but also by: (i) evaluating the reliability of the estimated optimal model parameters by running the same model three times with separate initial model parameter guesses and (ii) determining whether the estimated model parameters are within the permissible parameter space. Thus, while KGE' model performance may be lower than the KGE' obtained from a Gamma TTD, the response surface for an ADE-1x TTD type (Figure 4C) suggests that the estimated model parameters were not reliable and sometimes at the edge of the permissible parameter space. In contrast, the estimated model parameters with a Gamma TTD were unique (Figure 4B vs. 4C). A similar explanation also applies when comparing a Gamma to an exponential TTD type (Figure 4A vs. 4B) where the response surface of an ADE-nx TTD type is similar to the response surface for the exponential TTD in the sense that the estimated model parameters are at the edge of their permissible parameter spaces. Therefore, ADE-1x, ADE-nx and Exponential TTD types do not meet our set criteria for

selecting an appropriate TTD type and its parameters. For the sake of brevity, we only discussed piston flow and gamma TTD types in section 4.1.1. In section S3 in supporting information we provide more in depth information about the performance of each TTD type for three separate model runs. **To address this comment, we plan on including some more details about the other TTD types in the revised version of this section.** Please see also our response to comment # 1.2 above.

1.4 The third major criticism is that the article does not clearly convey what outstanding question/problem in scientific hydrology is likely to be resolved through the elaborate set of methodologies employed here. The discussion section does not convey any new insights about flow processes in headwater catchments. Rather, that section seems to emphasize intricacies of the various technical approaches that lead to order-of-magnitude differences in water age metrics such as the fraction of young water and mean transit time.

Our response: We used multiple metrics including the state-of-the-art fraction of young water (Fyw) metric and the mean transit time metric in conjunction with both young and old groundwater residence time tracers to better understand the dynamic nature of hydrologic flowpaths at a sub-humid mountain catchment in Arizona, USA. Please note that our work builds upon previous efforts [Kirchner, 2016a; b; Stewart et al., 2017] that show that spatial and temporal aggregation errors are lower when using the fraction of young water metric compared to the mean transit time metric. However, as acknowledged by the reviewer, these efforts are made using a virtual experimental setup. Therefore, *the principal contribution of our study is co-application of these two metrics, i.e., fraction of young water and mean transit time, using multiple tracer types for a real world catchment.* Additionally, our study makes the following specific contributions:

1. Use of multiple methods to estimate Fyw that demonstrate the variability associated with sampling frequency, hydroclimate, the method used in its estimation, and the processes that dictate streamflow generation.
2. As most of the existing literature on fraction of young water metric is focused on Fyw for annual or seasonal tracer cycles, in our work we estimated Fyw for not only annual cycles but also for several periods ranging from 2 days to 5 years when using stable water isotope tracers.
3. Delineation of a consistent mathematical framework to estimate Fyw using both young- and old- groundwater age tracers. Our proposed framework is flexible and can be easily employed at sites with long-term observations of young and old groundwater age tracers in inflow and outflow
4. Characterization and discussion of the limitations of tritium-based Fyw estimates due to a common lack of long-term data, coarse and/or sparse tritium concentration time series observations, and/or lack of measurement precision
5. Description of an alternative approach to more reliably estimate deep subsurface storage
6. Characterization of dynamic flowpaths that reorganize and restructure with catchment storage through the use of multiple metrics
7. Identification of a threshold short-term storage that once reached, increases the propensity for precipitation to infiltrate and activate deeper flow paths

In light of these contributions, we believe our study will be broadly useful to hydrological researchers and practitioners that rely on either Fyw or mean transit time metrics to

understand the subsurface residence time of water, or that aim to constrain the links between water quantity and quality as water moves along subsurface flow paths.

However, based on this reviewer's comment, we now recognize that we did not do a sufficient job of communicating these contributions to hydrologic scientists, and will improve the discussion in this regard in the revision.

1.5 My first recommendation for revision would include omission of methods and results that are based on lumped-parameter transport modeling using time-invariant TTDs. My second recommendation for revision would be a deep consideration of what is the specific gap in knowledge that the research would address, and a deeper discussion about what all these (somewhat abstract) age metrics tell us about flow processes, or the linkage between water age and catchment structure, that we didn't already know. Emphasis should be placed on new knowledge that may be generalizable across watersheds. This is especially important since the study focuses on a single watershed where quite a lot of tracer-aided flow and transport studies have already been conducted [Heidbuchel et al., 2013; Heidbuchel et al., 2012; Lyon et al., 2008; 2009], including multiple previous works by the lead author.

Our response: Thank you. We will revise accordingly. Please see our response to comments 1.1 and 1.2 above.

Specific comments on content in the text:

- Line 23: The phrase "single age tracer" is unclear and not conventional. Please omit or rephrase.

Our response: Thank you. The rephrased sentence between lines 21 and 23 now reads: "Current understanding of the dynamic flow paths and subsurface water storages that support streamflow in mountain catchments is inhibited by the lack of long-term hydrologic data and the frequent use of short residence time tracers that are not applicable to older groundwater reservoirs."

- Lines 42-51: The rationale provided here is not very strong. To say that the processes being studied are "still incompletely understood" is not a very effective way to communicate (1) what aspects of the processes are well understood (because certainly a lot of prior knowledge does exist), (2) what is/are the explicit knowledge gap/s, and (3) how this research is designed to specifically address that knowledge gap.

Our response: In the revised version of the main document, we plan on revising text between lines 42 and 51 to clearly state existing knowledge gaps and study objectives. Please see also our response to comment # 1.4 above.

- Lines 59-61: Doesn't really make sense to say that an underestimated quantitative metric affects actual substrate weathering in Earth's crust. The conjunctive phrase "As a result" in the following sentence seems out of place. Consider rephrasing.

Our response: Our intention was to point out that an underestimated residence time can lead to inaccurate understanding or estimate of subsurface mineral weathering rate (as shown by *Frisbee et al.* [2013]). In the revised version of the document, we plan to revise text between lines 59 to 62 to address this comment.

- Lines 70-71: I think you should temper the language here. Robust? That implies the result should be representative across a range of systems. Yet the papers you cite apparently rely on simulated experiments in synthetic landscapes. Any evidence from the real world that you could cite?

Our response: In the revised document, we used “more accurate” instead of the word “robust”. Please note that the use of “robust” in the original document refers to only one site and not to any range of systems. Our use of the word is also motivated by the results reported by previous efforts [Kirchner, 2016a; b; Stewart et al., 2017] that show that the spatial and temporal aggregation errors are much less when using the fraction of young water metric, in contrast to mean transit time metric. However, as acknowledged by the reviewer, these efforts are made using a virtual experimental setup. Therefore, co-application of these two metrics, i.e., fraction of young water and mean transit time, using multiple tracer types for a real catchment, is the novel contribution of our study.

- Lines 74-76: To help emphasize the knowledge gap, I think you need to clarify what exactly is meant by “only one period”.

Our response: Here, our intention was to highlight that most of the literature on fraction of young water metric is focused on Fyw for annual or seasonal tracer cycles. In our work, we estimated Fyw for not only annual time frames but also for shorter periods ranging from 2 days to 5 years (Figure 6A and B).

- Line 90: The question “what is the appropriate TTD type” is somewhat unclear. Precipitation is episodic. There is not a continuous inflow of water volumes with different ages entering any watershed. Therefore, the distribution of transit times (i.e., exit time – entry time) must also be discontinuous. This fact is illustrated by real TTDs observed from active tracer introductions [e.g., Kim et al., 2016a]. They are quite messy and not continuous distributions. Any continuous function that is chosen as a TTD for application in lumped-parameter-transport modeling is therefore just an approximation of reality. If that is accepted as true, then it seems your question could be restated as “what mathematical distribution yields simulation results that best fit the data from this particular watershed?”. That is not a question of great relevance for scientific hydrology in general, in my opinion.

Our response: Note that our complete research question # 1 is “what is the appropriate TTD type and mTT for the deep groundwater system that supports streamflow?” Thus, our focus is estimating/finding an appropriate transit time distribution and distribution for deep groundwater. Please see also our response to comment # 1.2 above.

- Lines 91-93: I have a very hard time interpreting this sentence. Please rephrase. Again, I would suggest carefully explaining, or omitting, the phrase “age tracers”. Are you meaning to distinguish stable isotopes from radioactive isotopes?

Our response: As most of the existing literature on the fraction of young water metric is based on the stable water isotope or chloride tracers, the aim here was to extend this literature by including tritium tracer-based fraction of young water estimates. The rephrased question 2 between lines 91 and 93 reads “What do the Fyw and storage estimates vary between shallow and deep groundwaters for a high elevation mountainous catchment?” Please see also our response to comment # 1.2 above.

- Lines 93-94: Suggest deleting “...as determined by stable water isotope tracers”. It implies to the reader that the answer to your more general question (i.e., “what is the discharge sensitivity of F_{yw} ”) is somehow conditional on this particular data type? Is that in fact what you think? If so, it raises some concern about the generality of the results.

Our response: Thank you. We plan on revising the third research question in the revised version of the main document.

- Lines 95-100: Suggest deleting all of this. A prelude to the methods elaborated on the

following pages is unnecessary. The concluding paragraph of the introduction should highlight the identified knowledge gap then state the objectives of this study and how they address that gap. The final sentence raises some concern that the current work is partially redundant.

Our response: In the revised version of the main document, we plan on revising the text between lines 94 and 100 to better highlight the knowledge gap and to more concisely state the study objectives and how these study objectives address the identified knowledge gaps.

- Lines 112-113: So it was a notably drier than average 9 years, or the PRISM results are biased high here?

Our response: It was notably drier.

- Lines 130-131: This is a very coarse sampling resolution for the intended application of the data. Undoubtedly there are tremendous temporal dynamics in the stable-isotope composition of precipitation within and among individual storms that occur during 5-7 day intervals. The range of stable-isotope abundances in precipitation observed during individual storms may be comparable or greater to the range observed among monthly-aggregated samples collected across years [e.g., Rozanski et al., 1993]. The true temporal dynamics of tracer concentration in precipitation are lost in a lumped sample that aggregates over 5-7 days. Any quantitative model that uses those tracer concentrations as input will be very limited in its ability to accurately simulate the temporal dynamics of the same tracer in the stream. That limitation seems very germane to the stated objectives of this study. Passive, sequential sampling devices are easy to make and deploy. Analysis of stable isotope abundances by laser spectrometry for large sample numbers is relatively inexpensive. This data limitation is hard to excuse.

Our response: The purpose of this study was not to explicitly simulate high temporal resolution dynamics of tracer concentrations in stream water. In contrast, by using the stable water isotope tracers, our study seeks to estimate the fraction of young water metric in a time averaged sense to compare and contrast the value and information contained in that metric at our field site to literature-reported data from other sites, in order to better understand dynamic flow path behavior. It is important to note that:

- Whatever sampling interval is chosen, there will be shorter periods of data variation that are not sampled, a characteristic Nyquist frequency, and a range of frequency responses that cannot be addressed. We have limited our interpretations to responses that can be addressed, using the data available.
- The physics of our system will act to filter out very high frequency variations in isotopes in precipitation. Soils are wet by rain and remain wet as more rainwater is added – mixing is inevitable. Runoff flowing in the main stem of a stream is a mixture of flow from small tributaries of different length, water held on wet leaves and water held in leaf litter or very shallow soil. Given that most of our summer storms are < 1 hour in length, mixing at shorter time scales is likely.
- High frequency variations (which cannot indeed be addressed in our study because of the 5-7 day sampling) are of less interest than lower-frequency, i.e., longer-period, phenomena, as we seek estimates of mean transit time and fraction of young water metrics in a time-averaged sense.
- Lines 149-151: I can't quite understand what this means. Please consider rephrasing.

Our response: In the revised version of the main document, we plan on revising the sentence between 148 and 151 for a better readability.

- Line 164-166: Simplify the headings and sub-headings. Here and elsewhere there are sub-headings with no content underneath. Suggest deleting.

Our response: In the revised version of the document, we plan on providing an improved structure of each section. We further plan to simplify section headings.

- Line 171: When you say "thereafter", do you mean over longer time increments than 1 month? Please rephrase to clarify.

Our response: On line 171, when we say "thereafter", we meant for longer periods. We plan to rephrase this line for clarity in the revised version of the main document.

- Line 176: Some formatting inconsistencies with citations here and throughout the manuscript. Uneven use of open and closed parentheses and lack of spacing between cited papers within in-line citations. Proofread carefully. Suggest using "(" instead of duplicate parentheses. Also, I cannot find the Dwivedi 2019b entry in your bibliography. Is it missing? Put spaces between entries in the bibliography. It is terribly difficult to read through single spaced.

Our response: Thank you. In the revised version of our paper, we plan to address this comment by: (i) using a consistent representation for multiple citations, (ii) providing the complete reference to *Dwivedi et al.* [2019] citation, and (iii) using dual-spacing for the whole main document for a better readability.

- Line 177: "expand on these results" again seems to suggest this is somewhat redundant with the previous works from the same catchment.

Our response: Our statement regarding "expanding on these results" when referring to previous work of *Ajami et al.* [2011] or *Dwivedi et al.* [2019], is meant to say that our present work aims to further improve our understanding of deep groundwater flow paths by characterizing their transit time distributions and evaluating if the state-of-the-art fraction of young water metric is appropriate for deep groundwater. This has not been reported in the literature, and thus our work makes a novel contribution of assessing the appropriateness of Fyw metric for deep groundwater flow paths. Please also note that *Ajami et al.* [2011] have not considered either transit time distribution or Fyw for deep groundwater and *Dwivedi et al.* [2019] have not considered various TTD types for deep groundwater.

- Lines 185-190: Pretty sure $h(\tau)$ is the specified functional form of the TTD, but that is not stated in the paragraph.

Our response: In our work, a TTD is referred to as $h(\tau)$ where τ is the transit time (in years). To address this comment, in the revised version of the paper $h(\tau)$ will be used between lines 185-190.

- Line 193: I am not familiar with the DownHill Simplex method. It is described in a single sentence, yet it is apparently the method for evaluating how appropriate is one versus the other TTD model. Could you please elaborate just a little bit on what this is for the unaffiliated reader? The KGE is used as the "model performance criteria" but you say that the Downhill Simplex was used to evaluate "the performance of each TTD". This is confusing to me. Equation 1 is the model, but the variable performance of the model is due only to the selection of different functional forms of the TTD. So the performance of the model is a direct reflection of the performance of the function selected as TTD, no? Please clarify.

Our response: We agree with the reviewer that the performance of the model, i.e.,

Equation 1 in our paper, is based on the performance of a selected TTD type. As far as the Downhill Simplex method is concerned, it a way to "search" for the model parameters (e.g., mean age and the shape parameters for a gamma TTD) such that the model performance is optimal, which in our work is assessed by using the modified Kling Gupta efficiency. In the revised version of our paper, we plan to explain this method in some more details for an unfamiliar reader.

- Line 228, equation 5: Use "C" with "Q" and "P" subscripted to indicate concentration in streamflow versus precipitation. You already adopted this notation in equation 1. Be consistent here and in subsequent equations.

Our response: Please note that Equation (5) on line 228 in the original version of the paper is for input tracer flux, which is a product of tracer concentration (C) and precipitation flux (P). Similarly, Equation (6) is for tracer flux in stream water, which is the product of tracer concentration C and streamflow (Q).

- Lines 348-357: What about all the other models? You only discuss PF and Gamma. The KGE of the 1d-ADE falls exactly between the values for the Gamma and PF models, yet the mTT estimated by the 1d-ADE is factors of 8-9 less than the mTT from those models, respectively. Why do you ignore the other models and what do you conclude from this order-of-magnitude difference? If I understand Figure 4 correctly, then only the "ADE-nx" and Exponential function as TTDs seem to generate uniquely identifiable parameters. Is that correct? Neither model is discussed at all here.

Our response: Please see our response to comment # 1.3 above.

- Lines 390-392: The data are also far too sparse to reliably fit the parameters for TTDs used in equation 1. Isn't this confirmed by (1) the lack of unique solutions illustrated in most cases shown in Figure 4 and (2) the generally poor accuracy of all model simulations shown in Figure 5? I would argue yes.

Our response: Between lines 390-392 in the main text, our emphasis is fitting sinusoidal curves to the tritium concentration data for tritium-based Fyw estimation. We noted that sinusoidal curve fitting was not appropriate to tritium tracer data due the coarse resolution of our dataset, which has also led to a higher standard error in the estimated tracer cycle amplitude. However, we were able to identify unique solutions for gamma distribution type TTD parameters when using tritium tracer under low-flow conditions (Figure 4B in the main document). Please note that Figure 4 shows the response surface for various TTD types. Therefore, neither applicability nor poor fit of a particular TTD type should be considered as an indication of data limitation alone. For example, assumptions implied during derivation of the equation model of a particular TTD type may also contribute to a poor data fit. A case to cite is the multiple-paths advection and dispersion model type TTD (ADE-nx) of *Kirchner et al.* [2001]. When deriving model for this TTD type, it is assumed that the recharge to an aquifer is spatially distributed. When tracer concentration prediction from this TTD is tested against the observations for our study site, the results show a very poor model fit (Figures 4D and 5D). Therefore, a poor fit for the ADE-nx model can be hypothesized as indicating that "recharge to the fractured bedrock aquifer is not spatially distributed at our field site". This is an important finding, because such recharge pathways can lead to replenishment of deep subsurface storages that support mountain block recharge to valley fill aquifers.

- Lines 400-414: The text makes no allusion at all to Figure 7D, which has unusual qualitative axes and cannot be easily interpreted by the reader. The figure caption only provides a citation to a previous work to explain the graphic. More explanation is needed in this section of the Results, or Figure 7B should be deleted.

Our response: We are confused by this comment because Figure 7 in our work only has A and B panels: Figure 7A is cited in section 3.3 (line # 302) and in section 4.4 (line # 400), Figure 7B is cited on line 549 and 550 in section 5.4.

- Lines 437-440: I am unclear what is the importance or relevance of this concept of "short-term storage". What is it and why does it matter? In any case, you present estimates of this metric based on three competing approaches that vary by a factor of approximately 125 (0.08, 0.22, and 10.7)! Which, if any, should we believe is correct, and why?

Our response: In contrast to the traditional approach for estimating subsurface storage that requires values for a subsurface property (e.g., porosity), the short-term storage can be estimated without requiring information on porosity by simply using the Fyw metric in conjunction with streamflow and threshold age for young water [Jasechko *et al.*, 2016]. By quantifying this metric for our study site, we noted (between lines 562 and 564 of the original paper document) "Thus, after a threshold of 0.05 m short-term near-surface storage at MGC, the current study supports that infiltration may activate deeper groundwater flowpaths [Dwivedi *et al.*, 2019]." Before our reported value of this storage, no such estimates are available/reported in literature at our study site. Please see also our response to comment # 1.1.

- Lines 472-477: The results are highly dependent on the temporal resolution of the input time series. As I noted in a comment above, if a temporal dynamic in the tracer concentration in precipitation is hidden within a sample that accumulated over 5-7 days, then the model can't possibly simulate the effects of that dynamic in streamflow. The results are entirely dependent on the resolution of sampling the tracer concentration in inflow, and the resolution used in this study is quite coarse.

Our response: We agree with the reviewer that we cannot ask a model to reveal a higher resolution pattern than the resolution of the data used in the model. Further, we also agree with the reviewer that the Fyw results are dependent on the temporal resolution of the data used in computing Fyw, which is the point we tried making in the lines cited by the reviewer. However, we respectfully disagree with the reviewer that "the resolution used in this study is quite coarse". Please note that when using stable water isotopes for stable water isotopes-based Fyw, our data have a high resolution, i.e., daily for streamflow and approximately weekly for precipitation as it does not rain every day at our study site [Heidbüchel *et al.*, 2012]. When using the tritium data for tritium-based Fyw, while our data have coarse resolution as we sampled low-flow conditions, they nonetheless facilitate a better understanding of the longer period component which will otherwise be hidden from sampling dynamic flow conditions from a catchment.

- Lines 499-501: You make a sweeping assumption here that the bedrock at several research sites is "water tight". That seems quite speculative. What evidence supports this assumption? Most rocks are fractured and jointed to some extent. Even exposed, granitic plutons commonly have sufficient fracturing and water storage capacity to host woody-stemmed plant communities and support inter-storm flow from emergent springs. More support for this assumption is needed here, perhaps through more extensive synthesis of the geology of the sites used in these cited studies.

Our response: Thank you for your comment. We concur with you completely. However, our use of the term "water tight" was a quotation from Gallart *et al.* [2020] who were describing their field site. We have made no assumptions about the tightness of the bedrock aquifer at other sites. Between lines 499 to 501 in the main document we stated the following "however, the fractured bedrock at MGC is functionally distinct than the "watertight" bedrock characterized by Gallart *et al.* [2020] and the majority of humid sites in Jasechko *et al.* [2016] that are comparable in size to MGC."

- Lines 538-539 and 544-546: So, you're saying the fraction of young water estimates are invalid when based on the use of Tritium as a tracer?

Our response: To avoid ambiguity regarding the F_{yw} metric when using tritium for low flow conditions, we stated "A negligible F_{yw} at MGC calls into question of the suitability of the ^3H -based F_{yw} approach for deeper groundwater." Thus, our intent is to suggest (based on our study results) that this metric is unsuitable when applied to the baseflow or deeper groundwater component of a catchment's flow system due to its longer residence time and significant mixing-amplitude dampening in the subsurface.

- Lines 555-564: Here there are a series of sentences elaborating some intricate details of methodology which seem misplaced in the discussion. They lead to the ultimate conclusion at the end of the paragraph that "infiltration may activate deeper groundwater flowpaths". That is not a novel conclusion in scientific hydrology, and it is not even stated definitively here (i.e., may..). This paper uses a wide ensemble of methodological approaches, which, from my view, has only created ambiguity in how the markedly contrasting results can be interpreted. I find no new insight into hydrological processes resulting from all this computational effort.

Our response: Thank you! The three cases mentioned in these lines will be placed in the methods sections in the revised main document. Please see also our response to comment # 1.1 above.

Comments on Figures and Tables:

- Figure 3: Does the inset have a linear scale? If not, please make it linear. If so, please add more tick marks to the vertical axis so we can approximate the numeric values of the data points. Are these six data points all you have to calibrate the parameters of the TTD models used with equation 1? If so, that seems inadequate.

Our response: The inset plot has a logarithmic scale as the main plot. This will be clearly stated in the revised version of this figure. Please see also our response to the comment # 1.2 above.

- Figure 4: Is the vertical axis the KGE? If so, please label it that way. I am unclear what the variable "response surface" on the vertical axis indicates.

Our response: A revised version of this document that addresses your comment will be provided with the revised main document.

- Table 1: Words and numbers should not be split between rows. Use emboldened lines, or no lines, to better delineate the content. This is not acceptable for a journal article. Please make it more presentable.

Our response: A revised version of Table 1 that addresses your comment will be provided with the revised version of the main document.

- Figure 5: Y-axis labels should have "3" as a superscript preceding "H" to conform to established conventions of symbolizing isotopes. Would suggest compressing this into a single graph with a legend indicating the results from different TTD models. The gray dots are the same across all 5 subplots.

Our response: In the revised version of Figure 5, we plan on: (i) including all TTDs into a single plot and (ii) properly label the y-axis.

- Figure 7: Here and elsewhere the font size is illegible. Please enlarge font on axes and

in legends.

Our response: Thank you. This figure will be revised in the next version of the main document.

Works Cited:

Botter, G., E. Bertuzzo, and A. Rinaldo (2010), Transport in the hydrologic response: Travel time distributions, soil moisture dynamics, and the old water paradox, *Water Resources Research*, 46, doi:W0351410.1029/2009wr008371.

Botter, G., E. Bertuzzo, and A. Rinaldo (2011), Catchment residence and travel time distributions: The master equation, *Geophysical Research Letters*, 38, doi:L1140310.1029/2011gl047666.

Harman, C. J. (2015), Time-variable transit time distributions and transport: Theory and application to storage-dependent transport of chloride in a watershed, *Water Resources Research*, 51(1), 1-30, doi:10.1002/2014wr015707.

Heidbuchel, I., P. A. Troch, and S. W. Lyon (2013), Separating physical and meteorological controls of variable transit times in zero-order catchments, *Water Resources Research*, 49(11), 7644-7657, doi:10.1002/2012wr013149.

Heidbuchel, I., P. A. Troch, S. W. Lyon, and M. Weiler (2012), The master transit time distribution of variable flow systems, *Water Resources Research*, 48, doi:W0652010.1029/2011wr011293.

Kim, M., L. A. Pangle, C. Cardoso, M. Lora, T. H. M. Volkmann, Y. Wang, C. J. Harman, and P. A. Troch (2016), Transit time distributions and StorAge Selection functions in a sloping soil lysimeter with time-varying flow paths: Direct observation of internal and external transport variability, *Water Resources Research*, 52(9), 7105-7129, doi:10.1002/2016WR018620.

Lewis, S., and A. Nir (1978), On tracer theory in geophysical systems in the steady and non-steady state. Part II. Non-steady state - theoretical introduction, *Tellus*, 30, 260-271.

Lyon, S. W., S. L. E. Desilets, and P. A. Troch (2008), Characterizing the response of a catchment to an extreme rainfall event using hydrometric and isotopic data, *Water Resources Research*, 44(6), doi:W0641310.1029/2007wr006259.

Lyon, S. W., S. L. E. Desilets, and P. A. Troch (2009), A tale of two isotopes: differences in hydrograph separation for a runoff event when using delta D versus delta O-18, *Hydrol. Process.*, 23(14), 2095-2101, doi:10.1002/hyp.7326.

McGuire, K. J., and J. J. McDonnell (2006), A review and evaluation of catchment transit time modeling, *J. Hydrol.*, 330(3-4), 543-563, doi:10.1016/j.jhydrol.2006.04.020.

Rinaldo, A., K. J. Beven, E. Bertuzzo, L. Nicotina, J. Davies, A. Fiori, D. Russo, and G. Botter (2011), Catchment travel time distributions and water flow in soils, *Water Resources Research*, 47, doi:10.1029/2011wr010478.

Rozanski, K., L. Araguas-Araguas, and R. Gonfiantini (1993), Isotopic patterns in modern global precipitation, *American Geophysical Union Monographs*, 78, 1-36.

van der Velde, Y., P. J. J. F. Torfs, S. E. A. T. M. van der Zee, and R. Uijlenhoet (2012), Quantifying catchment-scale mixing and its effect on time-varying travel time distributions, *Water Resour. Res.*, 48(6), W06536, doi:10.1029/2011wr011310.

References

Ajami, H., P. A. Troch, T. Maddock, T. Meixner, and C. Eastoe (2011), Quantifying mountain block recharge by means of catchment-scale storage-discharge relationships, *Water Resources Research*, 47(4), 1-14.

Botter, G., E. Bertuzzo, and A. Rinaldo (2011), Catchment residence and travel time distributions: The master equation, *Geophysical Research Letters*, 38(11), 1-6.

Dwivedi, R., T. Meixner, J. McIntosh, P. A. T. Ferré, C. J. Eastoe, G.-Y. Niu, R. L. Minor, G. Barron-Gafford, and J. Chorover (2019), Hydrologic functioning of the deep Critical Zone and contributions to streamflow in a high elevation catchment: testing of multiple conceptual models, *Hydrological Processes*, 33, 476-494, doi: 10.1002/hyp.13363.

Dwivedi, R., C. Eastoe, J. F. Knowles, L. Hamann, T. Meixner, P. A. T. Ferre, C. Castro, W. E. Wright, G.-Y. Niu, R. Minor, G. A. Barron-Gafford, N. Abramson, B. Mitra, S. A. Papuga, M. Stanley, and J. Chorover (2021), An improved practical approach for estimating catchment-scale response functions through wavelet analysis, *Hydrological Processes*, 35(3), 1-20.

Frisbee, M. D., J. L. Wilson, J. D. Gomez-Velez, F. M. Phillips, and A. R. Campbell (2013), Are we missing the tail (and the tale) of residence time distributions in watersheds?, *Geophysical Research Letters*, 4633-4637.

Gallart, F., M. Valiente, P. Llorens, C. Cayuela, M. Sprenger, and J. Latron (2020), Investigating young water fractions in a small Mediterranean mountain catchment: Both precipitation forcing and sampling frequency matter, *Hydrological Processes*, 34(17), 3618-3634.

Gleeson, T., K. M. Befus, S. Jasechko, E. Luijendijk, and M. B. Cardenas (2015), The global volume and distribution of modern groundwater, *Nature Geoscience*, doi: 10.1038/ngeo2590.

Harman, C. J. (2015), Time-variable transit time distributions and transport: Theory and application to storage-dependent transport of chloride in a watershed, *Water Resources Research*, 51, 1-30.

Heidbüchel, I., P. A. Troch, and S. W. Lyon (2013), Separating physical and meteorological controls of variable transit times in zero-order catchments, *Water Resources Research*, 49(11), 7644-7657.

Heidbüchel, I., P. A. Troch, S. W. Lyon, and M. Weiler (2012), The master transit time distribution of variable flow systems, *Water Resources Research*, 48(6), 1-19.

Hrachowitz, M., P. Benettin, B. M. van Breukelen, O. Fovet, N. J. K. Howden, L. Ruiz, Y. van der Velde, and A. J. Wade (2016), Transit times-the link between hydrology and water quality at the catchment scale, *Wiley Interdisciplinary Reviews: Water*, doi: 10.1002/wat2.1155.

Jasechko, S., J. W. Kirchner, J. M. Welker, and J. J. McDonnell (2016), Substantial proportion of global streamflow less than three months old, *Nature Geoscience*, 9(2), 126-129.

Kirchner, J. W. (2016a), Aggregation in environmental systems- Part 1: Seasonal tracer cycles quantify young water fractions, but not mean transit times, in spatially heterogeneous catchments, *Hydrology and Earth System Sciences*, 20(1), 279-297.

Kirchner, J. W. (2016b), Aggregation in environmental systems -Part 2: Catchment mean transit times and young water fractions under hydrologic nonstationarity, *Hydrology and Earth System Sciences*, 20(1), 299-328.

Kirchner, J. W., X. Feng, and C. Neal (2001), Catchment-scale advection and dispersion as a mechanism for fractal scaling in stream tracer concentrations, *Journal of Hydrology*, 254, 82-101.

Stewart, M. K., U. Morgenstern, M. A. Gusyev, and P. Maloszewski (2017), Aggregation effects on tritium-based mean transit times and young water fractions in spatially heterogeneous catchments and groundwater systems, and implications for past and future applications of tritium, *Hydrology and Earth System Sciences*, 21, 4615-4627.

Suckow, A. (2014), The age of groundwater – Definitions, models and why we do not need this term, *Applied Geochemistry*, 50, 222-230.