Response to the interactive comment of Reviewer #2 on

"Sensitivity of young water fractions to hydro-climatic forcing and landscape properties across 22 Swiss catchments" by Jana von Freyberg et al.

This manuscript analysis stable water isotope signals in a range of contrasting catchments in the Swiss Alps to better understand what controls catchment storage and release dynamics. Based on a recently developed metric, the young water fraction (Fyw), the analysis provides a highly interesting and new perspective on the topic: the sensitivity of Fyw to stream flow. From my point of view, this topic alone would already merit publication. In fact, I would even argue that much of the additional analysis provided in the manuscript, specifically the comparison of the interpolation methods and the snow storage considerations, do not really add much value and actually somewhat dilute the really interesting story. I thus think these parts could easily be removed or at least be considerably shortened, but I leave this decision open to the authors. Notwithstanding the well-designed experiments and in-depth analysis, the manuscript would benefit from some restructuring and, in places, from more precise and detailed explanations (see detailed comments below).

My only major comment is the rather superficial discussion of the relationships between young water fractions and catchment characteristics (section 5). There were quite a lot of studies over the last 10-15 years (e.g. that looked into the relationships of the very same variables, e.g. soil types, L/G, drainage densities, area, TWI, precipitation intensity, etc., with mean transit times (e.g. McGlynn et al., 2003, HP; McGuire et al., 2005, WRR; Laudon et al., 2007, JoH; Broxton et al., 2009, WRR; Tetzlaff et al., 2009, HP; Hrachowitz et al., 2010, WRR, 2010, HP; Soulsby et al., 2010, HP; Speed et al., 2010, HP; Asano and Uchida, 2012, WRR; Hale and McDonnell, 2016, WRR; and many others). Although the Fyw is an arguably more stable and thus reliable metric, it would be interesting to see and understand how the results and interpretations of the analysis presented here compares to these earlier studies. Can similar conclusion be drawn for Fyw than previously for MTTs? If yes, what does that mean? If no, why? Such a more detailed discussion would lend an additional, interesting edge to the manuscript. In any case, I would be glad to see this work eventually published and I hope that the authors find my comments helpful.

We thank Dr. Hrachowitz for his thoughtful comments, which we have addressed in detail below.

Comments of the reviewer are shown in italics. Responses from the authors are presented in regular font below each comment. Citations from the manuscript are in Times New Roman, changes of the manuscript text are underlined.

Detailed comments:

(1)P.2, l.8-9: "usually" is a quite unfortunate term here. Clearly, while there are quite some studies using "lumped-parameter" models (I suppose the authors referred to convolution integral approaches), there many(!) other studies that go far beyond that with many different types of models ranging from fully coupled 3D models to more conceptual models based on suites of storage tanks and the associated mixing coefficients/SAS functions. Please rephrase.

We will change that: "Transit time distributions are <u>often</u> inferred from concentrations of conservative tracers, such as stable water isotopes in precipitation and streamwater, using lumped-parameter models…"

(2)P.2, l.10: "catchment storage" is inaccurate. It rather expresses some (essentially unknown) storage that is significantly affected by exchange processes. For many systems, there may well be significant additional storage below that, which remains essentially undetectable with stable isotope

data due to potentially very long time scales of these exchange processes at depth (mostly molecular diffusion?). Please rephrase.

We will clarify this by writing: "Because the mean transit time expresses the ratio between <u>mobile catchment storage</u> and the average flow rate, it is widely used in catchment inter-comparison studies…"

(3)P.2, l.12-13: is this generally true or is it not mostly due to the assumption of time-invariance? Again, please note that most model approaches, except lumped parameter convolution integral approaches, do *not* rely on time-invariance of TTDs.

The cited references substantiate the problem of bias and unreliability in estimates of mean transit times. The conjecture that this problem disappears when TTD's are assumed to be time-variant is interesting but, as far as we know, not (yet) demonstrated – and it side-steps the very difficult problem of estimating what time-varying transit time distributions actually are (in the real world, based on real-world data, not in models).

(4)P.2,l.16: perhaps better to use "estimated" than "obtained"

We will change that.

(5)P.2,l.13: to be precise, it should read as:". . .from the differences in the amplitudes. . ."

Strictly speaking, it is the ratio of the seasonal cycle amplitudes that defines the young water fraction. We will clarify this in the revised version of the manuscript.

(6)P.3,l.3-21: this is quite lengthy and written in an unnecessarily complicated way. The bottom line is, in my opinion, if only liquid water input to/storage in the system is considered or the total water input/storage.

We find it important to properly explain both cases (catchment storage including/excluding snow storage) to the reader so that the relevance of this distinction regarding the interpretation of young water fractions can be grasped. Since the young water fraction framework is/will be applied to catchments in very different climatic regimes, we would like to emphasize these conceptual descriptions of catchment storage early on in the manuscript.

(7)P.3,l.26: please clarify what is meant by "coefficients" of the seasonal cycles.

We will write "seasonal cycle amplitudes", instead.

(8)P.4,l.5-17: some of the above references, analysing the relationships of catchment characteristics with MTTs would fit in nicely here and would place your manuscript into a somewhat wider context.

We will include some of the references in the Introduction: "Because the young water fraction can be estimated from sparse and irregular tracer data, it has been suggested as a useful metric for catchment inter-comparison studies (Kirchner, 2016). <u>To date, however, most catchment inter-comparison studies have investigated controls on mean transit times instead. Mean transit times have been variably found to be correlated with (for example) flow path lengths and gradients (McGuire et al., 2005), drainage density (Soulsby et al., 2010), the areal fraction of hydrologically responsive soils (Tetzlaff et al., 2009), bedrock permeability (Hale and McDonnell, 2016), or combinations of multiple factors (Hrachowitz et al., 2009; Seeger and Weiler, 2014).</u>

(9)P.4,l.32ff: also here, sine-wave fitting has been used already quite long time ago to understand transit times. Please add some references (e.g. DeWalle et al., 1997, HP; Soulsby et al., 2006, JoH)

We will include these references.

(10)P.5,l.13: see comment (3)

Please, refer to our reply to comment (3).

(11)P.5,l.17ff, eqs.(3) and (4): redundant with eqs.(1) and (2). Instead of amplitude and phase eqs.(3) and (4) give the same information only expressed in sine and cosine components. I think eqs. (1) and (2) can be removed.

We would like to keep Eqs.(1) and (2) as they introduce the general idea of fitting sine curves with coefficients A and k to streamwater or precipitation isotope time series. By only presenting Eqs. (3) and (4) instead, the meaning of A and k might not be clear without explanation. We thus consider Eqs. (1) and (2) the most efficient way to do this.

(12)P.6,l.26: what does "i. Br." mean?

It means "im Breisgau". We have removed this expression in the main text but kept it in the affiliations of the authors.

(13)P.6,l.28: "accuracy" or "precision"?

Accuracy, which is consistent with the information given in Seeger et al. (2014).

(14)P.7, section 3.3: also here, some references to earlier papers that used similar and partly the same predictor variables would be good

We will add more references here.

(15)P.7,l.32: do flow path length and gradient refer to subsurface or total length and gradient to the outlet? Please be more specific.

We used the same indices as in Seeger et al. (2014), which were derived with the SAGA module "Overland Flow Distance to Channel network". The flow path length L refers to the total (surface) length of the stream network, while the gradient G was calculated from the ratio of the horizontal and vertical components of L (L_h and L_v). We will clarify this in the revised manuscript.

(16)P.8,l.20-22: was the use of multiple linear regressions considered to better identify potentially spurious correlations? If not, why?

We tried this. Multiple linear regression analysis on such a small sample (22 sites), with such a large number of candidate explanatory variables, leads to results that are strongly dependent on the specific model selection criteria that are used. Thus, we feel that a simple table of rank correlations is a more realistic, if more modest, representation of our results.

(17)P.8,l.25ff, section 4.1: does this section actually add value to the manuscript? I think, the section can at least be considerably shortened if not condensed altogether.

(18)P.8,l.25-P.9,l.14: this would fit much better into the methods section

In contrast to Reviewer #2, Reviewer #1 asked to expand more on the description of methods 1 and 2, and therefore we will move this part into Chapter 3 and keep the short discussion of the results in Section 4.1. With this, Chapter 4 becomes considerably shorter, while both interpolation methods are still described in a short manner in the manuscript (new: Sect. 3.4 Precipitation isotope data). A detailed description of method 2 will still be available in the Supplement.

(19)P.9,l.28-29: although this term is widely used in our community, I do not think that in any environmental system application we can actually "validate" a model in the actual sense of the word. The best we can do is to rigorously test our models.

We will change that.

(20)P.10,l.1ff, section 4.2: see comment (17). If you decide to keep the section, more detailed descriptions of the model used for the snow dynamics (including parameters, calibration procedure, uncertainties involved, etc.) is needed and can be placed in the supplementary material. In addition, I may have missed it, but it is unclear what PREVAH stands for.

PREVAH stands for PREcipitation-Runoff-EVApotranspiration HRU model. It was used here to interpolate hydro-climatic variables at the study sites (See Sect. 3.1 Hydro-climatic data).

(21)P.11,l.1: not clear what is meant by ". . .shifts the seasonal isotope pattern toward later in the season." Does this refer to the amplitudes? If yes, please say so.

We will change that: "As can be seen in Figure 4a, the delayed meltwater input shifts <u>the phase of the seasonal isotope pattern </u>toward later in the season."

(22)P.11,l.2-3,fig.4: it would be easier for the reader to appreciate the information content of figure 4, if the phase would be given in days (or months) rather than in radians.

We will change Fig. 4 to show the phase and phase shifts in fractions of 1 year instead of radians.

(23)P.11,l.19-23: "...young water fractions. . .that are larger. . .because high flows generally contain more young water. . .". This seems a bit of circular reasoning to me.

We argue that the statement "…because high flows generally contain more young water..." puts the increase in young water fractions after flow-weighting into a process-based context. We therefore won't change the sentence.

(24)P.12,l.3-4: repetition of what was said earlier. Can be omitted.

We will remove this sentence.

(25)P.12,l.1ff,section 5: again please see comment (8)

We agree with the reviewer, that numerous earlier studies have looked into the relationship between MTT's and catchment characteristics. However, for the reasons outlined in the Introduction of the manuscript (an in much more detail by Kirchner (2016a,b)), that is that MTT's are prone to severe aggregation bias and are thus likely to be uncertain, we don't think a direct comparison of our results with those earlier studies is useful.

(26)P.12,l.31-34: sure, a few studies could identify area as potential control on MTTs, but others clearly could not (see in the given references above). Thus please rephrase this statement.

We did not claim that this is a universal relationship, but rather indicate that some studies did find a significant correlation. We will change the sentence, to make this more clear: "Some studies have identified catchment area as a major control on mean transit times (e.g., DeWalle et al., 1997; Soulsby et al., 2000), however, the negative correlation of F_{yw}^* and F_{yw} with catchment area only becomes significant (ρ=-0.49, p<0.05) when the five high-elevation, snow-dominated sites are omitted from the analysis (Fig. 6)"

(27)P.13,l.28: this interpretation is of course possible, but it surprisingly seems to not consider the potentially important influence of fast, lateral preferential flow pathways (e.g. macropores), which can be abundant in particular at (steep) forest sites. It may be worth reflecting on this a bit more.

We will add this alternative explanation to the revised version of the manuscript: "One would normally expect tree roots to increase soil permeability, resulting in greater infiltration and groundwater recharge (Brantley et al., 2017). However, at steep, forested slopes, abundant lateral preferential flow pathways (e.g. macropores) may facilitate rapid transport of water (Whipkey, 1965)."

(28)P.14,l.4: what is meant by "bigger" cycles?

With bigger we mean that the seasonal cycles of streamwater isotopes are less dampened, which is redundant with larger values of Fyw. We will remove this part of the sentence.

(29)P.14,l.17: the description of how this was in detail done remains quite vague. Please provide a more detailed description in the methods section. Were samples from time periods outside the individual quartiles simply removed and the sine wave refitted on the remaining samples? How many samples on average were the individual fits then based on? The information content of the 4th quartile and the top 20% is very similar. One can be removed.

The separation of the flow regimes was carried out in dependence of the flow at the time of sampling, so that roughly similar numbers of data points were available for each flow regime. For instance, at the Erlenbach site, the total number of streamwater isotope samples was 140, and thus each quartile of *Q* comprised 35 samples, while the upper 20 % and 10 %, of daily discharges comprised 28 and 14 samples, respectively. At other sites with much smaller numbers of streamwater samples, this separation procedure would not yield enough isotope samples to reliably estimate F_{yw} for each flow regime. Therefore, we used the alternative approaches presented in section 6.2. We will include a more detailed description of this approach in the revised version of the manuscript.

(30)P.15,l.1-12: it is not entirely clear in how this is different to what was done in 6.1. Please also here, provide a more detailed description in the methods section of what was done and how.

We will add an explanatory sentence: "As a first-order estimate of the sensitivity of F_{yw} to discharge across all 22 study catchments, we calculated the linear slope of the relationship between Q and F_{yw}, using a method that does not require breaking the streamwater isotope time series into separate flow regimes (and thus has more modest data requirements than plots like **Error! Reference source not found.**). Thus, instead of fitting a linear slope to the few data points shown in **Error! Reference source not found.**, we estimated the linear slope of the Q-F_{yw} relationship directly from the tracer time series $c_S(t)$ and $c_P(t)$. For each site, we assume that […]"

(31)P.17,l.23-24: which, in turn, would imply (to maintain the fraction of young water in spite of increasingly more young water in the system) an increasingly preferential sampling of older water as the system gets wetter.

This scenario is possible (besides the scenario that the age distribution remains the same with increasing streamflow), however, we can only speculate about this.

(32)P.17,l.26ff: this is a very interesting analysis, but it remains unclear, which parts of it are actually supported by the available data/results and which are mere speculation. Please try to make it clearer, which evidence supports these interpretations.

The analysis in Sect. 6.3 is based on the findings presented in Sects. 6.1 and 6.2. We will include these references to make this more clear to the reader. Besides that, we link the interpretations to the results they are based on by referencing specific figures of some study sites for which the three cases are distinguishable.

References:

Brantley, S. L., Eissenstat, D. M., Marshall, J. A., Godsey, S. E., Balogh-Brunstad, Z., Karwan, D. L., Papuga, S. A., Roering, J., Dawson, T. E., Evaristo, J., Chadwick, O., McDonnell, J. J., and Weathers, K. C.: Reviews and syntheses: on the roles trees play in building and plumbing the critical zone, Biogeosciences, 14, 5115-5142, 2017.

Broxton, P. D., Troch, P. A., and Lyon, S. W.: On the role of aspect to quantify water transit times in small mountainous catchments, Water Resour. Res., 45, 10.1029/2008WR007438, 2009.

DeWalle, D. R., Edwards, P. J., Swistock, B. R., Aravena, R., and Drimmie, R. J.: Seasonal isotope hydrology of three Appalachian forest catchments, Hydrol. Process., 11, 1895-1906, 10.1002/(Sici)1099-1085(199712)11:15<1895::Aid-Hyp538>3.0.Co;2-#, 1997.

Kirchner, J. W.: Aggregation in environmental systems-Part 2: Catchment mean transit times and young water fractions under hydrologic nonstationarity, Hydrol. Earth Syst. Sci., 20, 299-328, 10.5194/hess-20-299-2016, 2016b.

Soulsby, C., Malcolm, R., Helliwell, R., Ferrier, R. C., and Jenkins, A.: Isotope hydrology of the Allt a' Mharcaidh catchment, Cairngorms, Scotland: implications for hydrological pathways and residence times, Hydrol. Process., 14, 747-762, 2000.

Whipkey, R. Z.: Subsurface stormflow from forested slopes, International Association of Scientific Hydrology, 10, 74-85, 1965.

Hale, V. C., and McDonnell, J. J.: Effect of bedrock permeability on stream base flow mean transit time scaling relations: 1. A multiscale catchment intercomparison, Water Resour. Res., 52, 1358-1374, 10.1002/2014WR016124, 2016.

Hrachowitz, M., Soulsby, C., Tetzlaff, D., Dawson, J. J. C., and Malcolm, I. A.: Regionalization of transit time estimates in montane catchments by integrating landscape controls, Water Resour. Res., 45, 2009.

Kirchner, J. W.: Aggregation in environmental systems-Part 1: Seasonal tracer cycles quantify young water fractions, but not mean transit times, in spatially heterogeneous catchments, Hydrol. Earth Syst. Sci., 20, 279-297, 10.5194/hess-20-279-2016, 2016.

McGuire, K. J., McDonnell, J. J., Weiler, M., Kendall, C., McGlynn, B. L., Welker, J. M., and Seibert, J.: The role of topography on catchment-scale water residence time, Water Resour. Res., 41, 2005.

Seeger, S., and Weiler, M.: Reevaluation of transit time distributions, mean transit times and their relation to catchment topography, Hydrol. Earth Syst. Sci., 18, 4751-4771, 2014.

Soulsby, C., Tetzlaff, D., and Hrachowitz, M.: Are transit times useful process-based tools for flow prediction and classification in ungauged basins in montane regions?, Hydrol. Process., 24, 1685-1696, 10.1002/hyp.7578, 2010.

Tetzlaff, D., Seibert, J., and Soulsby, C.: Inter-catchment comparison to assess the influence of topography and soils on catchment transit times in a geomorphic province; the Cairngorm mountains, Scotland, Hydrol. Process., 23, 1874-1886, 2009.