2nd review of hess-2018-604
"Time-variability of the fraction of young water in a small headwater catchment"
by Michael P. Stockinger et al.

The authors responded to all of my comments in detail and addressed my major concerns, which were

- Provide uncertainties of the individual 1-year Fyw estimates
- Justify the choice of the threshold value
- Account to hydro-climatic variations in the data set
- Explain how month-specific Fyw values were extracted from 1-year Fyw values
- Be consistent in the inter-comparison of Fyw values
- Better relate the current study to a previous one where 1-year Fyw have already been calculated and compared

Unfortunately, the track-changed version of the manuscript did not show deletions, and thus it was quite difficult to reconstruct all changes made by the authors during this round of revisions. My comments below refer to the version of the manuscript that shows track-changes.

The authors added a more detailed analysis of the 2015 summer heat wave, which has contributed largely to increased uncertainties in 1-year Fyw estimates around the period. As a result, the authors decided for a part of their analysis to remove 4 moths of heat wave-affected isotope data from the 4.5-year time series. For this new 4.1-year isotope time series, the authors obtained generally more consistent Fyw values. In addition, the authors discussed the potential effects of winter precipitation isotope values in the 1-year data set, e.g., sampling parts of two winter seasons instead of only one likely result in different Fyw estimates.

Despite the additional analyses of the 2015 summer heat wave and winter precipitation effects on Fyw, most of my major concerns with this study remain:

The authors introduced a new threshold value of 0.04 to test their three hypotheses. Although this threshold value is two times larger than the previous one, the authors still reject all three hypotheses; thus, the overall outcome of the manuscript did not change. However, because the revised Figure 4a now shows the uncertainties (i.e., standard errors) of the individual 1-year Fyw values, I doubt that hypothesis 1 can be rejected so readily. The authors decided to not show the Fyw-uncertainties in Figure 6, which they use to illustrate that >10% of the individual 1-year Fyw values fall outside the 0.04-threshold. If the uncertainties of Fyw would be considered here, I suppose that >90% of all Fyw values would lie within the boundary conditions, i.e. hypothesis 1) can be accepted.

Similarly, hypothesis 2) needs to be re-evaluated considering the uncertainties of the individual 1-year Fyw values. In P15L27-28, the authors refer again to Figure 6 to point out that some Fyw-values that are 4 weeks apart differ by more than 0.04. I doubt that these differences are statistically significant given the large standard errors of the 1-year Fyw values around summer 2015 and at the end of the 4.5-year time series.

The analysis around hypothesis "3) Fyw estimates are similar for calculation years that are centered around a given calendar month (seasonal-invariance)" is still not satisfying to me since I am still puzzled about the question of what is actually tested here. For instance, in P8L5-9 the authors state: "If the hypothesis is accepted, it would indicate seasonal changes in the Fyw result as a function of the start date of a one-year sampling campaign." I do not follow this train of thought. If all 1-year Fyw values values centered around e.g. August would be similar within 0.04, how can we conclude that 1-year Fyw values varies seasonally? Please be more specific about the goal of this analysis and why it is important.

As far as I understood, the main objective of the study was to show that individual 1-year Fyw values might not be representative for the long-term Fyw (P16L2-3). As I have pointed out in my first review, this finding seems rather trivial since a catchment's hydro-climatic conditions and flow pathways can change substantially between seasons and years so that the age of streamwater is likely to change as well. A specific 1-year Fyw value might therefore estimate the average fraction of young water for this particular year, whereas a multi-year Fyw-value will be representative for the average fraction of young water for these multiple years. In case of hydro-climatic conditions and isotope values being highly variable during these multiple years, the uncertainty in the multi-year Fyw value would accordingly be large. As a consequence, comparing 1-year Fyw values between catchments is not per se a useless analysis as long as the same time periods are used and uncertainties are considered.

Overall, I find that the reviewed version of the manuscript is more challenging to read due to confusing wording and too general or contradicting statements. Below I provide some examples, however, my list is not complete and I suggest a thorough review of the language in a revised version of the manuscript.

In order to improve readability, I would suggest to be more specific about and more consistent in addressing the different Fyw values. In the manuscript, the distinction between the different Fyw values is often not clear and causes confusion, e.g. what is a "single Fyw" in P13L6-7: "If the isotope data and Fyw results of the period of low R^2_{adj} values was left out, the average Fyw of the 189 sine waves compared even better to the single Fyw (approximately 0.07 in both cases)." It might be easier to refer to the 4.5-year Fyw (often referred to as "single Fyw" in the manuscript) as $F_{yw,4.5}$, to the individual 1-year Fyw values as $F_{yw,i}$ (with *i* denoting the *i*-th 1-year time series), and to the average of all 189 $F_{yw,i}$ values as $\overline{F_{yw,186}}$.

The phrase "data-inherent uncertainty of the complete timeseries" is repeatedly used (e.g., P7L13: "In doing so, the time-variable Fyw results were tested against the data-inherent uncertainty of the complete timeseries"). The phrase "data-inherent uncertainty of the complete timeseries" is inaccurate as it is not clear what time series are referred to (temperature, streamflow, isotopes?) and what "data-inherent" means; I would suggest to be more specific and simply say "the standard error of $F_{yw,4.5}$".

P8L15-16: "Therefore, these results would represent a runoff with a fraction of young water that systematically varies with the start of the sampling campaign, from a catchment with stable environmental conditions and water transport properties, and low sampling uncertainties." Can you elaborate on this? It would be good to add a statement about whether this condition would be good/bad for estimating Fyw or whether these conditions would result in small/large uncertainties in Fyw.

P8L10-14: This paragraph is very confusing due to poor wording. E.g., "Despite it having a time-variant young water fraction, all three hypotheses are accepted." This statement is contradicting hypothesis 1) "Fyw estimates do not change over time (time-invariance)"! Please re-phrase. Also, the statement "On a long term basis, the young water fraction does not deviate significantly from its overall mean value (time-invariance)" is not easy to understand. Please use more specific wording to make clear what Fyw-values you are referring to.

What do you mean by "short-term changes in the start of a one-year sampling campaign" (P10L30)? How can a "start" exhibit short-term changes? What does "short-term" mean here? I suggest that you actually refer to the shift of the starting and end time of the time series by 1 week. Also, the following sentence (P10L30-31) "The hypothesis is accepted if during any consecutive four weeks Fyw did not differ more than 0.04", reads as if you have calculated Fyw based on 4-week isotope data sets. Please be more specific.

Your justification of why streamflow at Wüstebach is mainly comprised of groundwater is not convincing (P11L21-27). You infer from the similar $\delta^{18}O$ values in precipitation (-8.53‰) and groundwater (-8.43±0.17‰), that groundwater is mainly fed by rainwater (which is a somewhat trivial observation since precipitation is usually the main source of recharge); however, streamwater has a mean $\delta^{18}O$ signature (-8.40±??‰) that is very similar to that of precipitation, too. Instead, aren't the Fyw values (both, $F_{yw,4.5}$ and $F_{yw,i}$) actually more informative here as they suggest that roughly >90% of streamwater is older than 3 months?

Please be more specific when you talk about "peaks" and "amplitudes". E.g., (P12L12-15): "However, the relationship between precipitation and streamflow considerably changed due to the influence of the 2015 European heat wave: while the double-peak of precipitation in summer 2015 was not transferred to streamflow (Figure 3), the amplitudes of both lost their close relationship at the same time (supplementary Figure S2a)." The "double-peak of precipitation" is actually the "double-peak of the sine fits to the precipitation isotopes"; the "amplitudes of both" are the "seasonal cycle amplitudes of the isotopes in streamwater and precipitation".

The authors claim in P13L1-2 that "… this hydrological information about the Wüstebach catchment [that precipitation mixes with a quasi-constant $\delta^{18}O$ source] would have been impossible to detect with a single sine wave fit." I feel that the authors over-sell their results here. The sine fits to the entire 4.5-year streamwater and precipitation isotope data set would lead to the same conclusion (that is, the sine fit to streamwater isotopes resembles very similar patterns to the sine fit to precipitation isotopes).

Similarly, the authors conclude that the baseline Fyw value of 0.05 could only be found because of calculating Fyw for individual 1-year periods (P15L8-11). However, their estimate of Fyw based on the entire 4.5-year isotope record was 0.12±0.04 (or 0.11±0.04, see my other comment), so that a minimum Fyw value of 0.8 can be obtained. Thus, I disagree with the authors' statement that "Using a single sine wave would not have revealed this lower boundary" (P15L10), because it would still have revealed a very similar lower boundary (0.8).

Specific comments (referring to the version of the manuscript that shows track-changes):

P1L19: "For a given calendar month …" sounds like as if you have calculated Fyw for one month only. Similarly inaccurate expressions are used throughout the manuscript and should be corrected everywhere.

P1L23: Define "adjusted R^2", since the reader won't know where this value comes from without reading the rest of the manuscript.

P9L17: From your numbers of As and Ap, I obtain a 4.5-year Fyw value of 0.11... Can you please provide the standard errors of As and Ap?

P10L19: On P9, you stated that the average of all 189 1-year Fyw values was 0.08, not 0.09!

P10L6-7: You could actually calculate the effects of Ap on the standard error of Fyw through the Gauss error propagation approach. Through this, you might find that Ap has a much larger influence than As simply because Ap>As and standard error(Ap)> standard error(As).

P12L17-20: "Thus, considering the general hydrological observations obtained from the isotope data discussed above, we conclude that a certain percentage of precipitation became groundwater while another

percentage that might or might not be Fyw quickly generated runoff, conserving the precipitation d18O signal in streamflow and resulting in the similar shapes of the 189 sine wave pairs." This sentence is very general ("… a certain percentage of precipitation…") and does not tell us anything specific or interesting. Please rephrase.

P1416-30: What about catchments in Mediterranean climates that receive highly seasonal precipitation inputs? This case has already been discussed in Kirchner (2016), e.g. Figure 3.

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, 2016.

P16L15-16: The statement "Furthermore, snow blankets also change the isotopic signal potentially to a degree that obscures seasonal isotope patterns [Cooper, 2006]" is seems arbitrary. Why snow blankets and not snow cover? What do you mean by "obscures"? What seasonal isotope pattern is obscured? How much can the seasonal snow cover change the streamwater (?) isotope signal, and would this be significant in case of the Wüstebach catchment?

P16L27-26: "Only two Fyw were calculated in contrast to the 189 results of the present study (approximately 1%), making insights into the possible causes and a judgement if varying Fyw results are an isolated result or the rule impossible." Poor language, please rephrase.

P17L7-8: "(1) a potential strong influence of the 2015 European heat wave on Fyw estimates and uncertainties was discovered, which is a problem which could magnify in the future considering global warming;" Why is this a problem? It could very much be true that the 1-year Fyw values between June 2014 and October 2015 are representative for this particular period.

P17L9-10: "(2) precipitation and groundwater seemed to be the only end-members in streamflow which is information that isotope hydrograph separation studies can greatly benefit from;" As far as I can tell, these were the only two endmembers measured. So how can you be sure that there are not more endmembers, such as soil moisture or deep groundwater? What about isotopic fractionation effects due to evaporation?

P17L11-12: "Testing three hypotheses about the time-variability of Fyw we found that both in the long and short term Fyw is time-variable …" I do not understand how the long-term variability of Fyw was tested. Please clarify.

Figure 2, 4 and 5: Unit for Fyw is missing.
Figure 6: Please include the uncertainty bounds for the Fyw values, similar to Figure 4.

Figure 7: Tick marks are missing, unit for Fyw is missing.