

This paper presents a numerical experiment to estimate the relative influence of the different sampling periods in the estimate of the fraction of young water (fyw). The authors used 1-year long subsets of precipitation and stream tracer data sampled sequentially over a 4.5-year long record. This resulted in 189 different to estimate fyw based on sine function fits. I find the paper interesting as this approach to estimate the event water fraction is becoming popular among hydrologist. However I dough this paper provides information useful outside the catchment where the data was collected. The authors made no case on how these findings would be relevant to other locations. As such, it reads like a case study. Therefore, I suggest this paper not be consider for publication in HESS in its present form. In addition, I found the study lacks proper justification for the used of 2% difference in fyw as indicative of a significant difference.

We thank reviewer #1 for the helpful comments.

Usefulness to other catchments

The manuscript was adapted to account for the following points and supplementary material was added:

The main aim of this study is to present a *generic method* to analyze the time-variance of the fraction of young water. The reviewer already mentioned that Fyw is becoming more popular. Still, we lack information on when to best use this method and how sensitive it is to different datasets (e.g. frequency of sampling and length of observation time). Thus, investigating its use, limits and pitfalls is very important before we apply it to any catchment and particularly when comparing results from different catchments.

Many catchment studies showed that the transit time of water strongly varies (e.g. Harman, 2015; Heidbüchel et al., 2013), and it is thus very likely that Fyw also varies in other locations than ours. While previous studies focused on hydroclimatic and methodological influences on Fyw, this study is the first to focus on the influence of the sampling period and length. This is a first step, and it is highly recommended that this is repeated in other catchments to assess if this is a general situation or only a few catchments have time-varying Fyw (which we doubt because of strongly varying transit times in general).

Ultimately, catchment comparison studies that rely on Fyw should be based on comparable Fyw results. For example, Stockinger et al. 2016 already showed that only changing the sampling frequency of isotopes data led to drastically changed Fyw results. The present study goes further and shows that also the sampling period can influence Fyw. Based on reviewer comments, we now also present the associated uncertainties that are also varying in time.

Information on the various influences on Fyw results is critical for catchment comparison studies. We encourage hydrologists to use our generic method to test the existence of strong time-variances of Fyw in other catchments. However, the application of this method to a large set of catchments is beyond the scope of this study.

Additionally, we found strong evidence of the 2015 European heat wave significantly increasing Fyw uncertainty. With this knowledge we were able to reduce uncertainty. Additionally, we found indications that snow also potentially influenced Fyw uncertainty. Thus, we now present suggestions on how to reduce uncertainty in estimating Fyw.

Using 2% as difference

We now applied Gauß error propagation to estimate uncertainty in Fyw and used it to derive a data-driven threshold. This threshold was based on the uncertainty when estimating Fyw with a single sine wave to the complete data set. The threshold is now 4%. Figure R1 shows the uncertainty bands of the initial 189 Fyw results (influenced by the 2015 European heat wave) and was taken from the supplementary material.

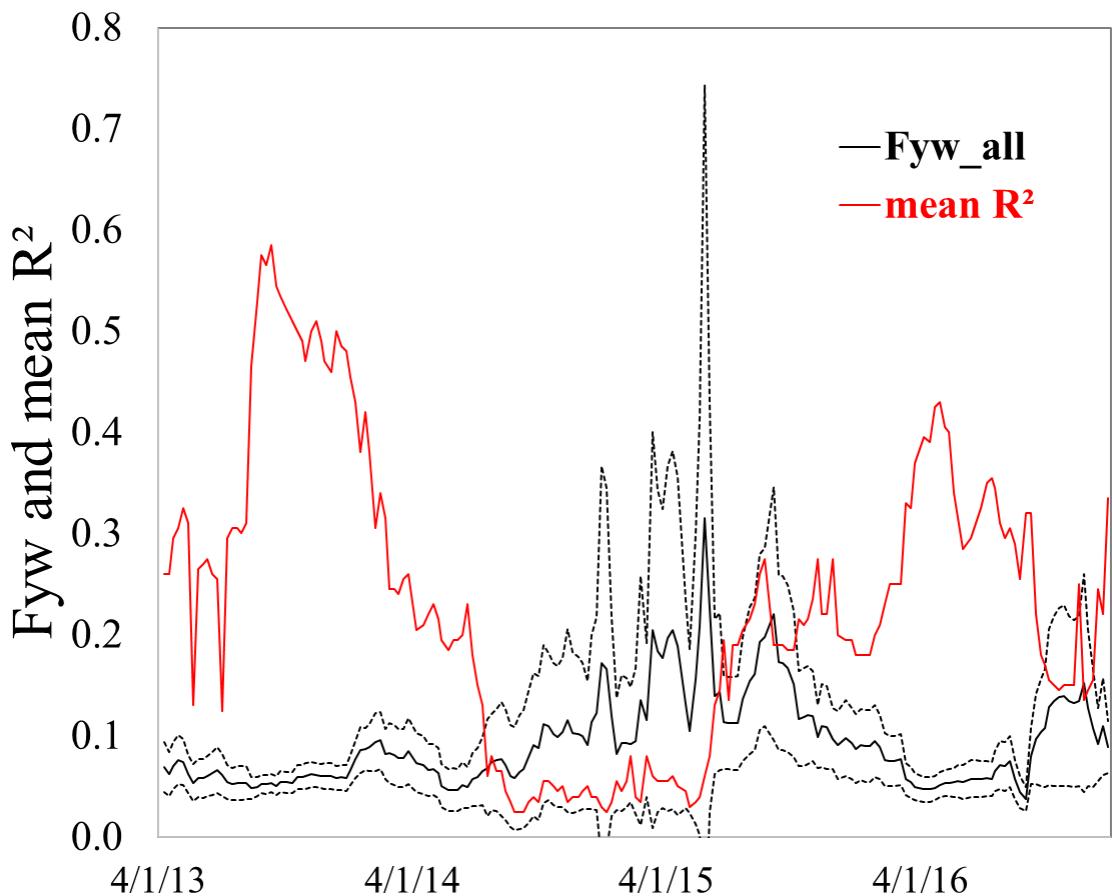


Figure R1. Fyw and its uncertainty using all data (black solid and dashed lines) compared with the average of streamflow and precipitation adjusted R^2 values of the respective sine wave fits (mean R^2).

The following can be said from this result:

- a) with a drop in R^2 below approx. 0.2 the uncertainty increases drastically. This, together with the strongly fluctuating Fyw results, indicates that in the Wüstebach an R^2 of at least 0.2 should be reached. We highly recommend conducting similar studies in different catchments to test whether different R^2 threshold values exist in other catchments.
- b) Fyw using a single sine wave had an uncertainty of appr. $\pm 4\%$. This is the new threshold that was used for re-evaluating our hypotheses.
- c) the Fyw results become highly uncertain during 2014/2015. This was partly due to the 2015 European heat wave and we managed to correct for its influence.

The statistical approach is also somehow vague. For example, it would be important to know how does the r^2 fits of the input compare to the r^2 fits of the output. This would allow understanding what is driving to low mean r^2 values that were observed for some of the results.

A comparison of input and output R^2 is shown in Figure 4: the mean R^2 , R^2 of throughfall (input to the catchment) and R^2 of streamflow (output from the catchment). Both the throughfall and streamflow R^2 values became low during this period too. In a further analysis we found that the main driving force for low R^2 (and thus high Fyw uncertainty) was the magnitude of the throughfall amplitude (Figure 4b). This is featured in the manuscript now.

Other specific comments:

Line 21 (P1): Sentence in poorly worded.

Based on the new uncertainty analysis, we suggest:

“Based on an increase of Fyw uncertainty when the mean adjusted R^2 was below 0.2 we recommend further investigations into the dependence of Fyw and its uncertainty to goodness-of-fit measures.”

Line 23 (P1): The abstract indicates that they recommend an r^2 threshold for future studies. However nowhere in the text, the authors offer any justification for the limit.

The added uncertainty of Fyw supports a certain threshold of quality for the fit. Very low goodness-of-fit values increased the uncertainty (as expected). Otherwise, R^2 values close to 0 would also be accepted and the respective Fyw results accepted as they are.

We changed the text (page 13, line 6):

“[...], we assumed that in our case the Fyw calculation method reached its limit below an average $R^2_{adj} = 0.2$. Fyw became highly sensitive to a small change in input data and in consequence highly uncertain. We recommend further investigations of the sensitivity of Fyw to the goodness-of-fit (not necessarily only measured with R^2_{adj}) for future studies. It remains to be seen if a value of 0.2 for R^2_{adj} is a general critical threshold for Fyw or if different catchments show varying results.”

Line 6 (P2), Line 15 (P13) and elsewhere: Better to refer "water stable isotopes" rather than "stable isotopes of water"

We changed it.

Line 16 (P3): Indicates that the hypotheses were tested against rules of acceptance that were based on whether differences in Fyw exceeded a threshold value of $\pm 2\%$. A more comprehensive justification for the 2% threshold should be included.

This was changed to the data-driven 4%.

Line 17-18 (P4): Please explain how this precision was estimated. Did you collect duplicate samples? This is the long-term precision derived from the uncertainty of 10,000s of measurements of various water samples conducted during the last years. Each unique sample is measured 6 times.

Line 21 (P4): Did you consider using deuterium instead of ^{18}O ?

D and ^{18}O are strongly correlated ($R^2 = 0.97$ throughfall, 0.87 streamflow) so we did not consider using it. We added this information to the manuscript.

Line 14-16 (P5): It would be interesting to see the distributions or R^2 of both fits independently.

R^2 is shown in Figure 4a for both throughfall and streamflow (orange lines labeled TF R^2 and Q R^2).

Line 1-5 (P6): Since the 2% threshold is mentioned in the introduction this explanation belongs there. We moved the suggested sentences and adapted them to the new threshold value.

Line 22 (P7): These values are very low. An $r^2 = -.08$ would indicate that a sine wave function is weak to describe the variability of the data.

The low R^2 mentioned are values for the single sine wave fit to the full 4.5-year time series and are 0.09 for TF and 0.23 for Q. We fully agree that those sine wave functions are gross simplification of the inter-annual variability of isotopes; which is the point of this study: a single sine wave fit oversimplifies naturally occurring, annual variations. However, even if a sine wave is weak to describe the data, the Fyw calculation method is based on using the amplitudes of sine wave functions (Kirchner, 2016). We added this discussion in the revised version.

Line 24-25 (P7): Can you provide information about the range of the r^2 of these fits

We added:

“The 189 fitted sine waves had a wide range of R^2_{adj} values: precipitation ranged from -0.02 to 0.63 with a mean of 0.22 and streamflow ranged from 0.00 to 0.55 with a mean of 0.25.”

Line 4-9 (P8): Would this indicate that the sine fit method is not appropriate for much of 2014? How confident can one be of the Fyw estimates when the r2 are below 0.2?

We added the uncertainty estimate and Fyw indeed becomes highly unreliable in this period. Hydrometeorological evidence pointed to a strong influence of the 2015 European heat wave, and part of the uncertainty could be corrected. Using only a single sine wave would not have revealed this. In the discussion we encourage studies of the reliability of Fyw based on goodness-of-fit measures of the sine waves.

Line 12-14 (P8): It is not clear what is the significance of this clustering of points.

We removed the sentence, as this was just an observation on our part.

Line 19 (P8): Considering how skew the data is would it be better to use the median? Also I suggesting some standard deviation or standard error.

We ultimately decided against using the median versus the mean value in hypotheses testing as we assume that a single sine wave rather averages the data (mean result) instead of finding the median. Furthermore, the mean Fyw (0.09) did not deviate much from the median Fyw (0.08) and both lie within the uncertainty bounds of the Fyw derived by the single sine wave approach (0.12 ± 0.04).

Line 20 (P8) Please consider some measure of error or uncertainty in the fyw estimates.

We added Fyw uncertainty.

Line 23-345 (P8): Please elaborate, that is indicate how many of the 189 were between this ranges.

The manuscript now reads:

"Out of the 189 Fyw results 159, i.e. 84%, were within those boundaries (Figure 6a). It could be possible that the period between July 2014 and October 2015 with low R²adj values and erratic Fyw behavior significantly influenced the rejection of the hypothesis. Therefore, in a second step we excluded this period, calculated the mean for those values and evaluated Fyw results again (Figure 6b). The new mean Fyw was 0.07 with 93% of results found between 0.03 to 0.11."

Line 29 (P8): Please provide some statistical information about the strength of the correlation.

We added:

"[...]the runoff coefficient (Q/P) was negatively correlated with R² = 0.25 and p-value = 1.7E-11 (Figure 7d). Leaving out again the period from July 2014 to October 2015 reduced the correlation to R² = 0.08 and p-value = 9.8E-4."

Line 10 (P9): This is confusing about figure 9. Are these the 189 fits? That is, are these fits over a one-year duration time series?

We agree that the sentence was confusing. We changed it to:

"As mentioned in the methods, the Fyw results were put in the middle of the one-year calculation period (calculating from February 2016 to February 2017, the result would be displayed as a data point in August 2016). We grouped together all Fyw results that were assigned to a specific calendar month and used a box plot to detect possible seasonality (Figure 8)."

Line17 (P9): Where is the value of d18O for ground water coming from?

The information about groundwater sampling was added to the methods:

"Isotope data was complemented by $\delta^{18}\text{O}$ values of groundwater sampled in four different locations in weekly intervals since 2009."

Line 19-20 (P9): please elaborate some more in the parallel to the Weigand et al. [2017] study

We extended the sentence:

"The study by Weigand et al. [2017] came to the same conclusion for the Wüstebach catchment using wavelet analysis of nitrate and DOC data collected at mainstream and tributary locations. While lower altitude locations of the catchment near the outlet were dominated by groundwater, higher altitude areas were less affected. This finding was additionally supported by field observations of shallow groundwater."

Line 18-19 P13): How do we know this conclusion is relevant to other catchments?

We expect that changes in flow paths over time alter the Fyw result of a catchment. Because of this, a one-year long time series of tracer data might lead to very different Fyw results depending on when the tracer data was sampled. Therefore, the results of comparison study could vary greatly simply by shifting the one-year sampling window by a couple of weeks.

Please see also our answer to the major comment regarding usefulness of our results for other catchments.

Line 24-25 (P 13): This sentence is vague. Please explain.

We adapted the sentence to make it clearer:

"If feasible, we recommend investigating a multi-year time series of tracer data with the method suggested in this study to enhance our knowledge of the sensitivity of Fyw to the chosen time frame in different catchment situations and the behavior of its uncertainty."

Line 26-28 (P13) It would be important to understand if the variability observed here would be relevant to understand difference across catchments.

As discussed above, different sampling periods yield (occasionally very) different Fyw result that would influence the interpretation of catchment comparison studies. Please also refer to the discussion points above regarding the usefulness to other catchments.

Figure 1: The markers for the precipitation and runoff gauges are too similar. Add latitude and longitude grids to the map. The contour should be in the legend indicating the units. In addition, the font next to the contours is difficult to read.

We added it.

Figure 4: In the legend, please clarify that "mean" refers to

We changed it to "Mean R²", "TF R²" and "Q R²".

Figure 7: Please add a legend.

Done.

Figure 8: The caption should include an explanation of what is hypothesis 2

Explanations were added to all figures.

References

Harman, C. J. (2015), Time-variable transit time distributions and transport: Theory and application to storage-dependent transport of chloride in a watershed, *Water Resour. Res.*, 51, doi:10.1002/2014WR015707.

Heidbüchel, I., P. A. Troch, and S. W. Lyon (2013), Separating physical and meteorological controls of variable transit times in zero-order catchments, *Water Resour. Res.*, 49, 7644–7657, doi:10.1002/2012WR013149.

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 Sc*, 20(1), 279-297, 2016.

Lutz, S. R., Krieg, R., Müller, C., Zink, M., Knöller, K., Samaniego, L., and Merz, R.: Spatial patterns of water age: Using young water fractions to improve the characterization of transit times in contrasting catchments. *Water Resources Research*, 54, 4767–4784. <https://doi.org/10.1029/2017WR022216>, 2018.

Stockinger, M.P., H.R. Bogena, A. Lücke, B. Diekkrüger, T. Cornelissen and H. Vereecken (2016): Tracer sampling frequency influences estimates of young water fraction and streamwater transit time distribution. *J. Hydrol.* 541: 952-964, doi:10.1016/j.jhydrol.2016.08.007.

Weigand, S., R. Bol, B. Reichert, A. Graf, I. Wiekenkamp, M. Stockinger, A. Lücke, W. Tappe, H. Bogena, T. Pütz, W. Amelung and H. Vereecken (2017): Spatiotemporal dependency of dissolved organic carbon to nitrate in stream- and groundwater of a humid forested catchment – a wavelet transform coherence analysis. *Vadose Zone J.* 16(3), doi:10.2136/vzj2016.09.0077.

This manuscript investigates the temporal variability of the young water fractions (Fyw) based on a 4.5-year time series of $\delta^{18}\text{O}$ in precipitation and streamwater in the German Weiherbach catchment. For this, the authors fit sine curves to the entire 4.5-year data set to estimate the long-term average Fyw. Then, they cut out 189 individual 1-year $\delta^{18}\text{O}$ time series from the 4.5-year data set (i.e., shifting each 1-year period by 1 week) and fit individual sine curves to these 1-year periods to estimate 189 Fyw-values. The goodness-of-fit of the sine curves to the $\delta^{18}\text{O}$ data was quantified through adjusted R² values. Three hypotheses were tested: "(1) Fyw estimates do not change over time (time-invariance) (2) Short-term changes in the start of a tracer sampling campaign do not influence the Fyw estimate (sampling-invariance) (3) Fyw estimates are similar for a given calendar month of different years (seasonal-invariance)" (P3L13-15). By applying a Fyw-threshold value of 2%, the authors reject all hypotheses and conclude that Fyw-values based on 1-year isotope data sets can be highly variable over time. This time-variability of Fyw hampers catchment-comparison studies that utilize tracer data of different time series lengths or time periods. I find the critical evaluation of new metrics (young water fraction Fyw) interesting and useful as this allows to better plan sampling campaigns or to use existing data sets more efficiently for a robust estimation of Fyw. In that regard, I consider the testing of hypothesis 2 the most useful scientific contribution of this study (Sect. 3.4) because it quantifies how much a 4-weeks delay can change the Fyw-values that are estimated from 1-year data sets. Unfortunately, these changes in Fyw do not correlate with any of the tested hydro-meteorological variables (Sect. 4.4) so that the authors cannot provide any suggestions about the optimal sampling strategy.

We thank reviewer #2 for evaluating our study and the helpful comments.

We now conducted an extended analysis of hydro-meteorological data and found that the 2015 European heat wave significantly influenced Fyw estimates and uncertainty, as well as indications of snow influencing it. From this, we derived the recommendation to sample only one year's winter in a one-year long sampling campaign.

Besides the testing of hypothesis 2, I find it difficult to identify a clear motivation and the novel scientific contribution of this study. The fact that Fyw responds to changes in precipitation, discharge and/or catchment wetness has already been established (e.g., Kirchner, 2016b; Lutz et al., 2018; von Freyberg et al., 2018; Wilusz et al., 2017), and thus it can be expected that Fyw changes over time in a catchment with a variable hydro-climatic regime such as the Weiherbach. Thus, the temporal changes in Fyw that have been identified in the present study are likely related to the hydro-climatic conditions at the site, however, the scientific analysis of these relationships often remains too superficial (which is surprising, given that the Weiherbach catchment is an intensively studied research site). As such, this study does not teach us something new about the catchment but rather shows that different tracer time series provide different young water fractions. Although it is interesting to quantify these temporal differences in Fyw, it remains to be tested how these findings for the 0.385 km² Weiherbach catchment are transferable to other landscapes and climates.

First, to avoid any misunderstandings we would like to make clear that our study area is the Wüstebach catchment and not the Weiherbach catchment as indicated by the reviewer. The primary focus of this study was not to identify catchment influences on Fyw but to investigate the extent and significance of temporal Fyw changes using variable 1-year time series of isotope tracer data. We primarily aim to improve the robustness of the Fyw method and not to analyze specific Wüstebach influences on Fyw. This is essential information for planning future sampling strategies in other catchments and for the layout of catchment comparison studies.

As mentioned above, we found evidence of the 2015 European heat wave negatively influencing Fyw uncertainty and accounted for it. Furthermore, a sampling recommendation could be derived.

Even before that, a recommendation was given in the original submission: estimating Fyw with data from a single year is not enough (page 12, 25f); the time-variant Fyw for a catchment should be calculated to understand the behavior and uncertainty for a given location (page 13, lines 2ff).

The reviewer already mentioned that previous studies found Fyw reacts to changes in precipitation, discharge, and other factors. Thus it is safe to assume that other catchments also have a time-varying Fyw and applying our method would yield more information about the Fyw behavior and uncertainty in each catchment. We highly suggest conducting the same study for other catchments in other landscape or climatic units to be able to generalize the findings.

These discussion points were added to the manuscript and the new supplementary material.

Major comments:

We first answer major comment #6 as it is of great importance:

6. At the very end of the Discussion section the authors state that a previous analysis has been carried out that used a 3-year isotope times series from the Weiherbach catchment. This previous study already showed that the Fyw values differed substantially between three 1-year periods (Stockinger et al., 2017, data in the supplement). In the present study, the authors simply repeat this analysis with a 4.5-year isotope data set from the same site knowing that their hypothesis "(1) Fyw does not deviate more than $\pm 2\%$ from the mean of all Fyw results indicating long-term invariance [: : :]" will likely be rejected. I was surprised to read about a very similar previous study at the very end of the current manuscript and wondered why is it necessary to repeat the analysis when the result (rejection of hypothesis 1) is already known?

The focus of Stockinger et al., 2017 was to correct canopy-induced isotope changes in throughfall for the complete time series. 3 individual years were cut out as a test without going into further detail. We extended this test into its own study to focus on investigating the extent and dynamics of the time-variability of Fyw, which was not the focus of Stockinger et al., 2017 and would have been out of scope for the previous study.

We had no possibility to know the results beforehand since for hypothesis 1 more than 90% of Fyw results must be within $\pm 4\%$. It is not possible to estimate from 3 out of 189 results if 19 results (10%) would be outside the $\pm 4\%$.

1. One of my largest concern is that the presented analysis did not provide any information on the uncertainties of the individual Fyw estimates. Only the adjusted R² values of the sine fits are presented. Previous studies that calculated young water fractions for several other catchment reported uncertainties in Fyw between 1% and 41% (e.g., Jasechko et al., 2016; Stockinger et al., 2016; von Freyberg et al., 2018). Thus, it should be tested whether the individual young water fractions that were calculated from the 1-year time series are indeed statistically significantly different from each other when their uncertainties are considered. Looking at the low adjusted R² values for the July 2014–October 2015 period (e.g. Figure 4), I would expect the uncertainties of the 1-year Fyw values to be rather large. However, instead of analyzing the uncertainties in Fyw, the authors mainly focus on the time-variability of the individual 1-year Fyw values and conclude rather boldly (P12L22) "The obtained Fyw could be a potential outlier, a larger value or part of the Fyw baseline". I would argue that the uncertainty in Fyw (e.g., expressed as standard error) would allow us to objectively judge whether we can believe our Fyw estimates or not. Such an analysis is, however, missing here. In fact, knowing the uncertainties of the individual 1-year Fyw values would allow a more informative analysis of how the Fyw-uncertainty (not Fyw itself) is controlled by hydroclimatic conditions. Such an analysis might provide concrete guidelines for planning targeted sampling campaigns to robustly estimate Fyw.

We thank the reviewer for pointing out this very important issue. It is the aim of this study to present a generic method to analyze Fyw for time-variance and thus improve the robustness of the Fyw method. We added Fyw uncertainty estimates using Gauß error propagation (Figure R1, taken from the supplementary material):

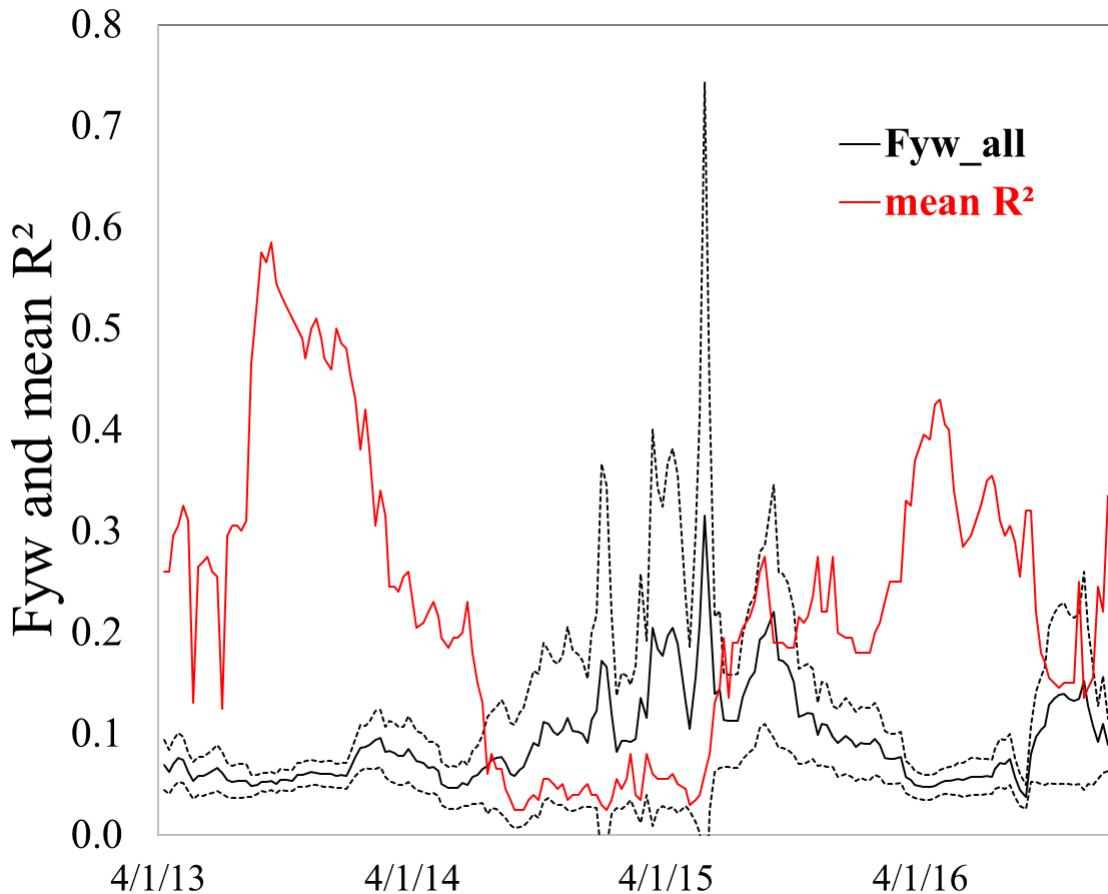


Figure R1. Fyw and its uncertainty using all data (black solid and dashed lines) compared with the average of streamflow and precipitation adjusted R^2 values of the respective sine wave fits (mean R^2).

The following can be said from this result:

- a) with a drop in R^2 below approx. 0.2 the uncertainty increases drastically. This, together with the strongly fluctuating Fyw results, indicates that in the Wüstebach an R^2 of at least 0.2 should be reached. We highly recommend conducting similar studies in different catchments to test whether different R^2 threshold values exist in other catchments.
- b) Fyw using a single sine wave had an uncertainty of appr. $\pm 4\%$. We used this new data-driven value for evaluating our hypotheses.
- c) the Fyw results become highly uncertain during 2014/2015. This was partly due to the 2015 European heat wave and we managed to correct for its influence.

Furthermore, we found indications of snow influence on Fyw uncertainty and now recommend sampling only one year's winter.

2. Furthermore, given that the uncertainties in Fyw values can potentially be much larger than 2% (as it was shown in the previous studies cited above), to me the 2% threshold seems too low and the authors' justification for that 2% threshold is not convincing.

We now used data-driven 4% based on the uncertainty estimation of using all data.

3. The hydro-meteorological conditions in the Weiherbach catchment were highly variable during the 4.5-year study period. For instance, only the winter 2013/2014 was snow-free in contrast to the other winters when a snowpack built up (Sect. 3.1). In addition, 21% "...of the forest were clear-cut in August/September 2013: : :" (P3L27), which significantly altered the streamflow regime of the Weiherbach creek (Wiekenkamp et al., 2016). In Fig. 8d (P24) we find that the runoff coefficients for the Weiherbach catchment ranged between roughly 0.8 and 1.25, suggesting that hydro-climatic conditions at the site varied considerably over time. The authors do not, however, provide any data or figures that present the hydro-climatic conditions during the study period except for the scatter plots in Figure 8, which contrasts 1-year averages of four hydroclimatic metrics with the respective Fyw-values. Despite the highly variable streamflow regime of the catchment and the authors citing another study where flow weighting of the streamwater isotope values resulted in "...significant changes in Fyw..." (P2L4), the authors should more thoroughly investigate how catchment wetness might control Fyw. Why was streamflow-weighting not done here? Why was there no further analysis of potential factors that may control the large variability in 1-year Fyw values, particularly in the period July 2014–October 2015? It seems likely, that individual storm events may have had strong effects on the discharge of young water, so it may be useful to investigate extreme events rather than average behavior.

Precipitation and streamflow weighing were both done for Fyw calculation, but unfortunately not clearly mentioned in the submission. We clarified this now

"Precipitation isotope values were weighed using collected precipitation volumes, while streamflow was weighed using runoff volumes."

As mentioned above, we found influences of the 2015 European heat wave and snow. This is now extensively discussed in the new manuscript version and in the supplementary material.

4. Sect. 3.5 and Figure 9: It is not clear to me how the Fyw values for testing hypothesis 3 (seasonal invariance) were determined. As far as I understood, Fyw-values were calculated for 189 1-year periods (Sect. 2.3). How were month-specific Fyw-values extracted from these annual Fyw-values? Wouldn't each 1-year Fyw-value be affected by the isotope values of all 12 months that comprise this 1-year period? If so, I doubt that the analysis presented in Sect. 3.5 and Figure 9 provides useful information.

No monthly Fyw were extracted from the data. Each of the 189 Fyw results was assigned to the date that lies in the middle of the calculation period. For example, Fyw was calculated from 1 January to 31 December and the corresponding result was assigned to 1 July.

We then grouped all Fyw results according to the month they were assigned to. All results in the boxplot (now Figure 8) are still 1-year calculation results. Should a seasonal trend be observable, one could argue that e.g., a 1-year sampling campaign centered around July would lead to higher/lower Fyw estimates compared to when it is centered on March. Ultimately, we found indications of snow potentially increasing Fyw uncertainty if the winters of two different calendar years are featured in a one-year sampling campaign. This is now discussed in the revised version of the manuscript.

5. Part of the analysis presented in "4.2 Fraction of young water" is not valid. First, the authors calculated Fyw from the entire 4.5-year data set ($Fyw_{4.5}=10.8\%$) and compared this to the average of the 186 1-year Fyw values (9.3%), concluding that both values are similar with regard to their 2% threshold. A second comparison was carried out with $Fyw_{4.5}$ and the average of a much smaller number of 1-year Fyw values that neglects the Fyw values from the period July 2014–October 2015 (7.5%). This second comparison should, however, use another Fyw value as a reference based on the same isotope data set (i.e., 4.5 years minus the period July 2014–October 2015)– otherwise the authors compare apples with oranges.

We thank the reviewer for pointing this out. We accounted for this now.

"If we use the averages of the 189 sine wave amplitudes to calculate Fyw, the result would be 0.08 instead of 0.12 of the single sine wave. This is less than the 0.04 difference in Fyw defined by this study as the data-driven threshold value for significant differences. Leaving out the period of low R²adj values the single sine wave and the average of 189 amplitudes would both yield approximately 0.07."

Minor comments:

P3L8-9: "However, it remains to be tested how sensitive the Fyw method is towards the timing and the length of the available data." Why does this need to be tested? Can you provide an example of where the length and the timing of the isotope data resulted in different Fyw values? Otherwise, a clear motivation for your analysis is missing.

We rephrased the introduction

"The mentioned studies highlight the current research interest in the new measure of Fyw. For this reason, it is necessary to investigate the sensitivity of Fyw and its uncertainty to different datasets. This is especially important for catchment comparison studies where the conceptualization of calculating Fyw might vary between catchments or datasets of different catchments may vary in quality. The question to answer is how much of the difference between individual Fyw estimates stems from actual, catchment-borne differences in flow path distributions and which part is merely based on e.g., different data quality or quantity."

P4L25: "Because of this on average 43 isotope values were available for precipitation compared to 53 values for streamflow." Does this average refer to a 1-year period? Please clarify. It would also be nice to provide the total number of streamwater and precipitation samples of the entire 4.5-year period.

Yes, this refers to the one-year calculation periods. The total number of P and Q samples (156 and 195, respectively) is now mentioned on page 5, line 8.

P5L22-23: I would suggest to move these two sentences to the beginning of the chapter to make clear where the number "189" comes from.

We agree.

P5L13: 24*365.25 is 8766 not 1/8766

We changed it

"(i.e., if CP(t) and CS(t) are calculated in hourly time steps then the frequency f is 1/8766; once per 24 x 365.25 hours)"

P5L31-32: What do you mean with "the timing of peaks and the individual amplitudes"? Do you refer to the isotope time series or to the fitted sine functions?

We referred to the fitted sine functions and changed it:

"Apart from the 189 Fyw results we also calculated Fyw for the whole time series with one sine wave as was the standard of previous studies. We compared its peak timing and amplitude to the timing of peaks and amplitudes of the 189 sine waves."

P6L3: Here you switch units of Fyw (0.02 and 2%). Also, in the text you express Fyw in percent, whereas in the figures you use the scale from zero to one. Please be consistent throughout the manuscript.

We now consistently use e.g., 0.02 instead of 2%.

P7L13-14: Please be more specific about what water isotopes you are talking about, e.g. add δ 18O.

The sentence is now:

“Precipitation isotope ratios ranged from -3.04 to -17.80‰, spanning a range of 14.76‰ in δ 18O values.”

P8L28-30: Please provide some metrics for the strength of these correlations (e.g., Pearson correlation coefficients).

We added statistical information to the text:

““[...]the runoff coefficient (Q/P) was negatively correlated with $R^2 = 0.25$ and p-value = 1.7E-11 (Figure 7d). Leaving out again the period from July 2014 to October 2015 reduced the correlation to $R^2 = 0.08$ and p-value = 9.8E-4.”

P8L29: Was the runoff coefficient calculated with catchment-average precipitation or throughfall? I would suggest to add the runoff coefficients to Fig. 3 since the relationship between Q/P and the sine wave fits to the isotope data are discussed in Sect. 4.1.

The runoff coefficient was calculated with throughfall, but we changed to open precipitation to enable comparability to other studies. It is now prominently featured in the supplementary material.

P9L16-20: You suddenly present groundwater isotope data without providing information about the source (location, sampling procedure, number of samples etc.) of these data. Please include this information into Sect. 2.2.

This was an oversight on our part. We added:

“Isotope data was complemented by δ 18O values of groundwater sampled in four different locations in weekly intervals since 2009.”

P10L13: “The double-peak in precipitation of autumn 2015 was not found in streamflow (Figure 3).” Do you refer to the δ 18O in precipitation and streamflow or to the sine fits to the isotope data?

To the sine fits, we adapted the sentence to also account for the 2015 European heat wave:

“However, the relationship between precipitation and streamflow considerably changed due to the influence of the 2015 European heat wave: while the sine wave 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).”

P11L33: “Thus, during the 4.5-years Fyw never fell below the baseline of 5% [...]” This statement is incorrect. Figures 6 and 7 clearly show that Fyw fell below 5% on several occasions, such as around June 2014 and September 2016.

We changed it

“Thus, during the 4.5-years Fyw seldom fell below the baseline of 0.05 and [...]”

P12L5: “The variability in Fyw of this study could not be explained by most meteorological or hydrometric variables”. Could a lack of correlation be explained by the large distance (3km) of the meteorological station to the study site? What about median values of the hydro-climatic variables or metrics that describe extreme events?

Correlations of precipitation amounts ($R^2 = 0.95$), temperature ($R^2 = 0.99$) and relative humidity ($R^2 = 0.94$) of the 3 km distant climate station with the respective climate data from the clear-cut area of the Wüstebach catchment showed good R^2 . We did not use the on-site climate station for our study as its data does not cover the full study period.

While most hydrometeorological data still did not have a strong correlation with the time-variable Fyw, we found evidence of the 2015 European heat wave increasing Fyw uncertainty.

P12L9: "...the different sampling periods of all mentioned studies...". This contradicts a previous statement: "...Lutz et al. [2018] used the same sampling period for precipitation and streamflow for all 24 investigated catchments." (P11L25).

We changed it

"Such contradictions could be explained by the different sampling periods of our study and the mentioned studies but also by differing catchment characteristics."

P12L23: "As the violation of hypothesis 2 did not correlate with any meteorological or hydrometric data :: :". How can a violation correlate with anything? Please clarify.

We referred here to the timing of the violation of hypothesis 2: if the timing could be connected to hydrometric or meteorological data.

We adapted:

"As the timing of the violation of hypothesis 2 did not[...]"

Figures: The date formats in all figures are confusing. Does 4/10/13 mean 4th October 2013 or 10 April 2013? Also, I would suggest to have each tick mark at the first of the month and to have consistent date axes in all figures.

4/10/13 refers to April 10th, 2013. We will adapt the figures to uniformly start on 4/1/13 with the exception of Figure 2 (just a theoretical example) and Figure 3 (showing the input data and starting on a different date than the Fyw result figures; see also explanation below).

Figure 4: This figure misses a proper legend (e.g., What does "Mean" stand for?). The unit and numbers of Fyw on the right vertical axes don't match. Do panels a and b share the same legend? Why are the shown time series much shorter than 4.5 years?

We will change the legend entries to "Mean R^2 ", "TF R^2 " and "Q R^2 ".

We changed part b of the figure drastically to avoid any confusion of units.

The time series are shorter than 4.5 years since each Fyw result was placed in the middle of the year it was calculated for. The time series starts on 10/10/12, thus the first Fyw result is placed on 4/10/13. Doing this cuts off the first half year and the last half year of the complete time series, explaining the shortening.

Stockinger et al. presents a study to evaluate the temporal variability of young water fraction (Fyw) based on 189 sine curve fits of 1-year subsets of a 4.5-year rainwater and streamwater ^{18}O isotope dataset. The Fyw, developed by Kirchner 2016, has become a powerful descriptor of streamwater flow path as the substitute for mean transit time. It is important to test the how Fyw change with different timing and sampling time coverage of water isotopes. The results showed “high” temporal variability of Fyw but no seasonality of Fyw based on the criterion defined by the author. The variability due to sampling time chosen is very useful for the isotope hydrology community. This study sheds new light on the development and application of Fyw, which is interesting and suitable for HESS. I find this paper is generally well-written but not strong enough. One of my concern is that how and why the 2% difference was defined as significant for the three hypotheses? This threshold value is introduced in the paper but not clearly explained. The discussion section lacks the discussion of the importance of the results. It would be a stronger paper if the author can explain the cause of the Fyw timevariability, which is ambiguous in the current form. Other specific comments on this paper are listed below.

We thank reviewer #3 for the helpful comments.

We now estimated Fyw uncertainty by Gauß error propagation (Figure R1 taken from the supplementary material). Fyw of the single sine wave fit had an uncertainty of $\pm 4\%$ (not shown in Figure R1). We used this new data-driven value instead of the $\pm 2\%$ for evaluating our hypotheses.

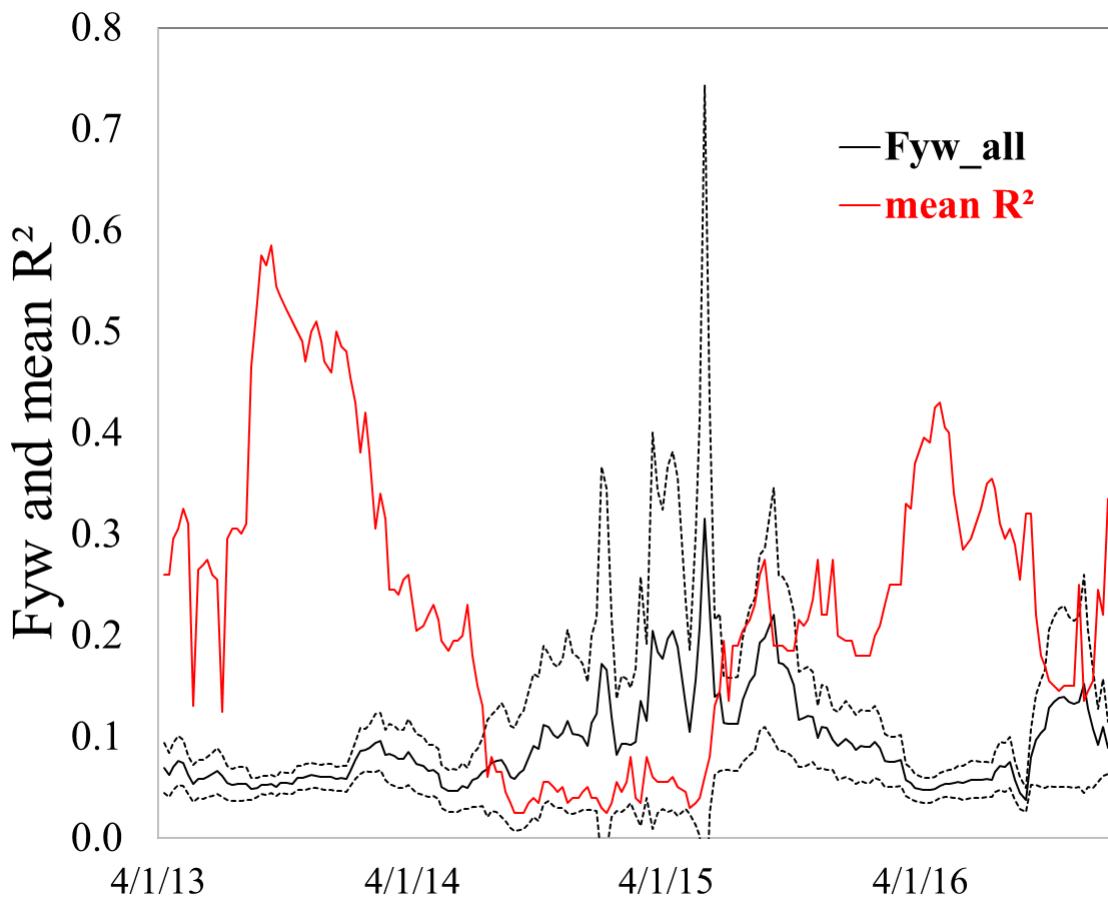


Figure R1. Fyw and its uncertainty using all data (black solid and dashed lines) compared with the average of streamflow and precipitation adjusted R^2 values of the respective sine wave fits (mean R^2).

The following can be said from this result:

- a) with a drop in R^2 below approx. 0.2 the uncertainty increases drastically. This, together with the strongly fluctuating Fyw results, indicates that in the Wüstebach an R^2 of at least 0.2 should be reached.

We highly recommend conducting similar studies in different catchments to test whether different R^2 threshold values exist in other catchments.

b) Fyw using a single sine wave had an uncertainty of appr. $\pm 4\%$. We used this new data-driven value for evaluating our hypotheses.

c) the Fyw results become highly uncertain during 2014/2015. This was partly due to the 2015 European heat wave and we managed to correct for its influence.

Furthermore, we found indications of snow influence on Fyw uncertainty and now recommend sampling only one year's winter.

The importance of our results is now emphasized in the discussion:

We will first add the new Fyw uncertainties and discuss these. Additionally, an extended analysis of hydro-meteorological data was added in the supplementary material. The discussion was expanded by the 2015 European heat wave and possible influence of snow.

Previous studies (e.g., Lutz et al., 2018; von Freyberg et al., 2018) showed that Fyw reacts to changes in e.g., precipitation and discharge. Thus, it is safe to assume that catchments other than the Wüstebach also have a time-varying Fyw. Applying our method would yield information about the Fyw behavior and its uncertainty which is important before applying the method to a catchment and especially when comparing results of different catchments. We emphasize this now more in the manuscript.

P3-L27: change "8" to Eight

Done.

P5-L14: Add "reciprocal of" or similar phrase before "24 hours..." since frequency (f) should be $1/T$.

We rephrased this "(i.e., if $CP(t)$ and $CS(t)$ are calculated in hourly time steps then the frequency f is $1/8766$; once per 24×365.25 hours)"

Figure 1: It would be nice to add latitude and longitude to the map. An alternative way is giving the latitude and longitude of the sampling location in the text. Square brackets with "-" can be removed, it may be misread as minus.

Latitude and longitude were added. [-] were removed in all Figures.

Figure 4a: Which line represent the R^2 ?

The red and orange lines are R^2 (orange = R^2 of TF and Q, red = mean). We changed the legend entries to "Mean R^2 ", "TF R^2 " and "Q R^2 " to clarify.

Figure 7, 8, and 9: The hypotheses should be explained in the captions.

We added explanations.

References

Lutz, S. R., Krieg, R., Müller, C., Zink, M., Knöller, K., Samaniego, L., and Merz, R.: Spatial patterns of water age: Using young water fractions to improve the characterization of transit times in contrasting catchments. *Water Resources Research*, 54, 4767–4784. <https://doi.org/10.1029/2017WR022216>, 2018.

von Freyberg, J., Allen, S. T., Seeger, S., Weiler, M., and Kirchner, J. W.: Sensitivity of young water fractions to hydro-climatic forcing and landscape properties across 22 Swiss catchments. *Hydrol. Earth Syst. Sci.*, 22, 3841–3861, <https://doi.org/10.5194/hess-22-3841-2018>, 2018.

Time-variability and uncertainty of the fraction of young water in a small headwater catchment

Michael Paul Stockinger^{1,2}, Heye Reemt Bogena¹, Andreas Lücke¹, Christine Stumpp², Harry Vereecken¹

¹Forschungszentrum Jülich GmbH, Agrosphere Institute (IBG-3), Wilhelm-Johnen-Straße, 52425 Jülich, Germany

²University of Natural Resources and Life Sciences Vienna, Institute of Hydraulics and Rural Water Management, Muthgasse 18, 1190 Vienna, Austria.

Correspondence to: Michael Paul Stockinger (michael_stockinger@boku.ac.at)

Abstract. The time precipitation needs to travel through a catchment to its outlet is an important descriptor of a catchment's susceptibility to pollutant contamination, nutrient loss and hydrological functioning. The fast component of total water flow

10 can be estimated by the fraction of young water (Fyw) which is the percentage of streamflow younger than three months. Fyw is calculated by comparing the amplitudes of sine waves fitted to seasonal precipitation and streamflow tracer signals. This is usually done for the complete tracer time series available neglecting annual differences in the amplitudes of longer time series. Considering inter-annual amplitude differences, we employed a moving time window of one-year length in weekly time steps over a 4.5-years $\delta^{18}\text{O}$ tracer time series to calculate 189 Fyw estimates and their uncertainty. They were then tested against the

15 following null hypotheses, defining a difference of 0.04 in Fyw (4% young water) as significant based on data-inherent uncertainty: (1) At least 90% of Fyw results do not deviate more than ± 0.04 from the mean of all Fyw results indicating long-term invariance. Larger deviations would indicate changes in the relative contribution of different flow paths; (2) for any four-week window Fyw does not change more than ± 0.04 indicating short-term invariance. Larger deviations would indicate a high sensitivity of Fyw to a 1-4 weeks shift in the start of a one-year sampling campaign; (3) for a given calendar month Fyw does

20 not change more than ± 0.04 indicating seasonal invariance of Fyw. In our study, all three null hypotheses were rejected. Thus, the Fyw results were time-variable, showed variability in the chosen sampling time and had no pronounced seasonality. We furthermore found evidence that the 2015 European heat wave and including two winters into a one-year sampling campaign increased the uncertainty of Fyw calculation. Based on an increase of Fyw uncertainty when the mean adjusted R^2 was below 0.2 we recommend further investigations into the dependence of Fyw and its uncertainty to goodness-of-fit measures.

25 Furthermore, while investigated individual meteorological factors did not sufficiently explain variations of Fyw, the runoff coefficient showed a moderate negative correlation of $r = -0.50$ with Fyw. The results of this study suggest that care must be taken when comparing Fyw of catchments that were based on different calculation periods and that the influence of extreme events and snow must be considered.

1 Introduction

Precipitation water uses slow and fast flow paths on its way through a catchment to the outlet where it becomes streamwater [Tsuboyama *et al.*, 1994]. Slow flow paths are for example the saturated and unsaturated flow through the soil matrix [Gannon *et al.*, 2017] while fast flow paths include preferential flow [Wiekenkamp *et al.*, 2016a] and overland flow [Miyata *et al.*, 2009].

5 The distribution of slow and fast flow paths varies in time and depends on a catchment's spatiotemporal characteristics [Harman, 2015; Heidbüchel *et al.*, 2013; Stockinger *et al.*, 2014; Tetzlaff *et al.*, 2009a; Tetzlaff *et al.*, 2009b]. Knowledge of this distribution helps in assessing the risk of streamflow contamination with pollutants or nutrient loss since nutrients and pollutants are transported through the soil by hydrological pathways [Bourgault *et al.*, 2017; Gottselig *et al.*, 2014].

10 The water stable isotopes ($\delta^{18}\text{O}$ and $\delta^2\text{H}$) are widely applied in the study of flow paths and transit times of precipitation through a catchment [McGuire and McDonnell, 2006]. One method that utilizes the water stable isotopes for investigating fast flow paths is the fraction of young water (Fyw). Developed by Kirchner [2016a], Fyw estimates the streamflow fraction that is younger than three months since entering the catchment as meteoric water. It does so by comparing the amplitudes of sine waves fitted to the seasonally-varying isotope tracer signals of precipitation and streamflow. The seasonally-varying isotope 15 signal in precipitation is caused by different evaporation/condensation temperatures, vapor source areas and evaporation amounts of falling rain droplets during warmer and colder seasons, leading on average to higher $\delta^{18}\text{O}$ values in summer and lower ones in winter [Dansgaard, 1964]. As rainfall passes through a catchment to reach the outlet, this signal is attenuated and shifted in time, leading to a much smoother but still seasonally-varying isotope signal in streamflow. The ratio of the fitted streamflow sine wave's amplitude A_S divided by the fitted precipitation sine wave's amplitude A_P equals the percentage of 20 water in streamflow younger than three months. Kirchner [2016a,b] showed the robustness of Fyw against spatial catchment heterogeneities (aggregation bias error) where previous methods of transit time estimation by sine wave fitting produced highly uncertain results.

Catchment influences on Fyw were, e.g., investigated globally by Jasechko *et al.* [2016]. They calculated Fyw for 254

25 catchments and concluded that one third of global streamflow consists of water younger than three months with catchments in steeper terrains having smaller contributions of young water to their runoff. Wilusz *et al.* [2017] coupled a rainfall generator with rainfall-runoff and time-varying transit time models to determine the young water fraction. They found an increase of annual rainfall amounts of 1 mm/d led to an increase of 0.03-0.04 in the modeled Fyw (percentage point increase of 3-4%, from here on written as 0.03-0.04, where the value 1 would mean that 100% of streamflow is younger than three months). 30 Similar to this, von Freyberg *et al.* [2018] found a positive correlation between Fyw and high-intensity precipitation events. This dependence of Fyw on precipitation characteristics could lead to long-term changes in Fyw due to global warming. Global warming was found to increase precipitation intensity and the frequency of droughts [Pendergrass and Hartmann, 2014; Trenberth, 2011]. For Europe, the chance of extreme heat waves and thus dry conditions has substantially increased since 2003

[Christidis *et al.*, 2015]. Previous studies highlighted that the distribution of fast and slow flow paths is time-variable [Harman, 2015; Heidbüchel *et al.*, 2013]. Since Fyw focuses on fast flow paths we expect it to be variable in time as well. However, so far previous studies focused on comparing Fyw between different catchments to derive relationships between catchment characteristics and Fyw, but no study investigated the temporal variability of Fyw for a given catchment yet.

5

Besides catchment characteristics, the conditions and conceptualizations of the Fyw calculation also influenced results in past studies. The effect of varying sampling frequencies of tracer data was investigated by Stockinger *et al.* [2016]. A higher sampling frequency led to higher Fyw highlighting the sensitivity of Fyw to the temporal resolution of the available tracer data. Lutz *et al.* [2018] investigated 24 catchments in Germany and used 10,000 Monte Carlo simulations with random errors 10 in the isotope data of precipitation and streamflow to derive the 95% confidence intervals of Fyw. Their confidence intervals indicated a robustness of Fyw against random errors in input data. The study of von Freyberg *et al.* [2018] focused on three influences on Fyw: (a) spatially interpolating precipitation isotopes, (b) including snow pack and (c) weighing streamflow in fitting sine waves. They found that weighing streamflow led to significant changes in Fyw while the other factors had a negligible effect.

15

The mentioned studies highlight the current research interest in the new measure of Fyw. For this reason, it is necessary to investigate the sensitivity of Fyw and its uncertainty to different datasets. This is especially important for catchment comparison studies where the conceptualization of calculating Fyw might vary between catchments or datasets of different catchments may vary in quality. The question to answer is how much of the difference between individual Fyw estimates stems 20 from actual, catchment-borne differences in flow path distributions and which part is merely based on e.g., different data quality or quantity.

The present study aims at answering one aspect of this open research question by focusing on the time-variance of Fyw and its associated uncertainty. Past studies fitted one sine wave to the complete time series available, varying from less than a year 25 to several decades [Ogrinc *et al.*, 2008; Song *et al.*, 2017; von Freyberg *et al.*, 2018]. To our knowledge, only the study of Stockinger *et al.* [2017] calculated Fyw for two different 1-year periods of a multi-year time series but did not test the temporal variability of Fyw nor influencing factors on it or its uncertainty. Thus, the sensitivity of the Fyw method towards the timing and the length of the available data remains to be tested in detail. The present study investigated the temporal variability of Fyw when different calculation periods of a multi-year isotope data set are used. We used a one-year time window which was 30 shifted in 7-days steps to calculate 189 Fyw estimates over a 4.5-year time series of isotope data. The 189 Fyw results were tested against the following null hypotheses:

- (1) Fyw estimates do not change over time (time-invariance)
- (2) Short-term changes in the start of a tracer sampling campaign do not influence the Fyw estimate (sampling-invariance)

(3) Fyw estimates are similar for calculation years that are centered around a given calendar month (seasonal-invariance)

The three hypotheses were tested against whether Fyw differences exceeded a threshold value of ± 0.04 which is the Fyw uncertainty when fitting a single sine wave to the 4.5-years time series (data-inherent uncertainty derived by Gauß error propagation, see results). We used hydrometeorological and isotopic data to investigate possible influences on time-variable Fyw results and their associated uncertainties and, where applicable, to reduce uncertainty. In conclusion of this study we recommend a tracer sampling design that reduces Fyw uncertainty.

2 Material and methods

2.1 Study site

10 The Wüstebach headwater catchment (38.5 ha) is located in the Eifel National Park (Germany, Figure 1). It is also part of the Lower Rhine/Eifel Observatory of the Terrestrial Environmental Observatories (TERENO) network [Bogena *et al.*, 2018]. The mean annual precipitation amounts to 1107 mm (1961 – 1990) with a mean annual temperature of 7°C [Zacharias *et al.*, 2011]. Soils are up to 2 m deep with an average depth of 1.6 m [Graf *et al.*, 2014]. Soil types of cambisol and planosol/cambisol are found on hillslopes, whereas gleysols, histosols and planosols are found in the riparian zone. The catchment is mostly covered
15 with Norway spruce (*Picea abies*) and Sitka spruce (*Picea sitchensis*) [Etmann, 2009]. Eight ha (~21%) of the forest were clear-cut in August/September 2013 [Wiekenkamp *et al.*, 2016b]. A severe heat wave occurred in the Wüstebach during summer 2015 [Duchez *et al.*, 2016].

2.2 Data preparation

We used hourly hydrometric and weekly $\delta^{18}\text{O}$ isotope data of precipitation (composite sample) and streamflow (grab sample)
20 from October 2012 to June 2017. We did not use $\delta^2\text{H}$ due to the strong correlation of $\delta^{18}\text{O}$ and $\delta^2\text{H}$ ($R^2 = 0.97$ for throughfall and 0.87 for streamflow) and therefor redundancy of information content. Precipitation depths were measured hourly in 0.1 mm increments for rainfall and daily in 1 cm increments for snowfall at the meteorological station Monschau-Kalterherberg of the German Weather Service (Deutscher Wetterdienst DWD station 3339, 535 m asl), located 9 km northwest of the catchment. Runoff was measured at the outlet by a V-notch weir for lower and a Parshall flume for higher runoff depths in 10-
25 minute intervals. We collected throughfall samples for isotopic analysis as the Wüstebach catchment is forested and canopy-passage of precipitation influences Fyw [Stockinger *et al.*, 2017]. The samples were collected with six RS200 samplers (UMS GmbH, Germany) with a distance of 2 m to each other and to trees. The samplers consisted of a 50 cm long, 20 cm diameter plastic pipe which was buried in the ground. On top of it a 100 cm long plastic pipe with the same diameter was installed. An HDPE sample bottle (max. volume of 5000 ml) was placed inside the buried pipe and connected with plastic tubing to a funnel
30 on top of the 100 cm long pipe. The funnel had a collecting area of 314 cm² and was protected by a wire mesh against foliage

and a table tennis ball in the funnel served as an additional evaporation barrier. Tests of the system showed the reliability in protecting the collected water from evaporation and in consequence isotopic fractionation for several weeks [Stockinger *et al.*, 2015]. Two samplers of the same design were placed in a clearing of the Wüstebach catchment to sample open precipitation, i.e., precipitation that has not passed through the spruce canopy. Streamflow samples for isotopic analysis were collected 5 weekly as grab samples in HDPE bottles at the outlet of the catchment.

Isotopic analysis was carried out using laser-based cavity ringdown spectrometers (models L2120-i and L2130-i, Picarro Inc., USA). Internal standards calibrated against VSMOW, Standard Light Antarctic Precipitation (SLAP2) and Greenland Ice Sheet Precipitation (GISP) were used for calibration and to ensure long-term stability of analyses [Brand *et al.*, 2014]. The 10 long-term precision of the analytical system was $\leq 0.1\text{\textperthousand}$ for $\delta^{18}\text{O}$.

We calculated weekly volume-weighed means of $\delta^{18}\text{O}$ for throughfall and open precipitation, which were further weighed according to the respective land-use percentage of spruce forest (79%) and clear-cut (21%) areas to generate a time series of precipitation $\delta^{18}\text{O}$ for the whole catchment. The derived precipitation isotope time series was then used together with the 15 weekly streamwater grab samples to calculate Fyw. While streamflow never ceased and thus a time series of weekly isotope values was available for the whole time series, there were weeks of no precipitation and thus gaps in the time series. Because of this for a 1-year calculation window on average 43 precipitation isotope values compared to 53 streamflow values were available. The total number of isotope values amounted to 156 for precipitation and 195 for streamflow. We could not always sample precipitation in weekly intervals, leading to bulk samples of 2-3 weeks on occasion. In this case, we assigned the 20 measured bulk isotope value to each week, while the measured bulk precipitation depth was proportionally assigned to each week according to the distribution of hourly precipitation measured at the meteorological station Kalterherberg.

For further hydro-meteorological and isotopic analyses several additional data were collected: we measured air temperature and relative humidity in 10-minute intervals at the TERENO meteorological station Schleiden-Schönesieffen (Meteodata 25 station, 572 m asl), located 3 km northeast of the catchment. We also calculated the runoff coefficient from runoff (Q) and open precipitation (P) as Q/P and used it for further analysis. Isotope data was complemented by $\delta^{18}\text{O}$ values of groundwater sampled in four different locations in weekly intervals since 2009. Groundwater was sampled by pumping first to avoid sampling stagnant water. Lastly, we calculated the d-excess of the precipitation samples using the slope and intercept of the global meteoric water line (d-excess = $\delta^2\text{H} - 8 * \delta^{18}\text{O}$) [Craig, 1961; Merlivat and Jouzel, 1979].

30 2.3 Fraction of young water

We used a one-year time window which was moved in 7-days steps to calculate 189 Fyw estimates over the 4.5-year time series. A minimum time window length of one year was chosen to fully capture the annual isotope signal. Fyw is calculated

by fitting sine waves to both the seasonally-varying precipitation and streamflow isotope signals, respectively. We used the multiple regression algorithm IRLS (iteratively reweighted least squares, available in the software R) to minimize the influence of outliers:

5 $C_P(t) = a_P \cos(2\pi ft) + b_P \sin(2\pi ft) + k_P,$
6 $C_S(t) = a_S \cos(2\pi ft) + b_S \sin(2\pi ft) + k_S$ (1)

with $C_P(t)$ and $C_S(t)$ the simulated precipitation and streamflow isotope values of time t , a and b regression coefficients, and k and f the vertical shift and frequency of the sine wave. The difference of $C_P(t)$ and $C_S(t)$ to the measured isotope time series in

10 precipitation and streamflow is minimized to fit the parameters a , b and k , while the frequency f of the sine wave is known due to its annual character (i.e., if $C_P(t)$ and $C_S(t)$ are calculated in hourly time steps then the frequency f is 1/8766; once per 24 x 365.25 hours). Precipitation isotope values were weighed using collected precipitation volumes, while streamflow was weighed using runoff volumes. The goodness-of-fit of the sine waves are expressed as the adjusted coefficient of determination R^2 (R^2_{adj}). If not otherwise stated we will use the mean of the streamflow and precipitation R^2_{adj} , as both sine waves are needed
15 to estimate the fraction of young water. After fitting the multiple regression equations, the amplitudes A_P and A_S and Fyw can be calculated:

$$A_P = \sqrt{a_P^2 + b_P^2}, \quad A_S = \sqrt{a_S^2 + b_S^2},$$
$$Fyw = \frac{A_S}{A_P} \quad (2)$$

20 Shifting the calculation window in 7-days steps resulted in a time series of varying Fyw estimates. Of course, the Fyw estimates cannot be considered independent from each other precluding the use of regression analysis to derive predictor variables (e.g., temperature, relative humidity) for the independent variable (Fyw). However, we used regression analysis to describe the average meteorological conditions during each Fyw time window. The thus derived “predictor” variables may have influenced
25 Fyw and could be investigated in future studies that use independent Fyw estimates.

Fyw calculation was done in a two-step process as the initial 189 Fyw results had large uncertainties that originated from a strong influence of the 2015 European heat wave (see results and supplementary material). Thus, in a second step we considered its influence and recalculated results while omitting precipitation isotope data of summer 2015. This greatly reduced uncertainty. Apart from the 189 Fyw results we also calculated Fyw for the whole time series with one sine wave as was the standard of previous studies. We compared its peak timing and amplitude to the timing of peaks and amplitudes of the 189 sine waves.

2.4 Hypotheses testing

For clarity we want to highlight that each Fyw result was placed in the midpoint of the year it represents. That is, a data point located at any date represents the value for the six months before and six months after this date. For example, a Fyw result of 0.2 on 6th August 2013 means that between 5th February 2013 to 4th February 2014 on average 20% of runoff consisted of

5 water younger than three months. The same logic applies to R^2_{adj} values, amplitudes, phase shifts and hydrometeorological data if not explicitly stated otherwise. The hydrometeorological data was calculated as mean values for the 189 individual calculation years to facilitate comparison to the Fyw results that are averages valid for the respective calculation time frame.

Prior studies in the Wüstebach catchment identified changes of Fyw between 0.02-0.04 as significant [Stockinger *et al.*, 2016;

10 Stockinger *et al.*, 2017]. Here, we employed Gauß error propagation on the sine wave fit parameters to carry their respective standard errors through to the Fyw results. Doing this resulted in the uncertainty of the 189 Fyw results as well as the uncertainty of Fyw calculated with the complete time series. We used the latter as the threshold value for testing the null hypothesis. In doing so, the time-variable Fyw results were tested against the data-inherent uncertainty of the complete time series. In our study we found that a threshold value of 0.04.

15

Based on this definition of a significant change in Fyw, three hypotheses were tested according to the following rules of acceptance:

1) Fyw estimates do not change over time (time-invariance)

20

This hypothesis is accepted if more than 90% of Fyw values are within ± 0.04 of the mean value of all Fyw results. We chose a minimum percentage of 90% to ensure that the long-term time-invariance is captured. Larger changes of Fyw over time would indicate either flow path changes or a change in the relative contribution of different flow paths.

25 2) Short-term changes in the start of a tracer sampling campaign do not influence Fyw estimate (sampling-invariance)

30

This hypothesis is accepted if four consecutive Fyw results (i.e., four weekly shifts of the one-year time window) do not differ more than ± 0.04 . We thus investigated 186 four-week time windows of the in total 189 Fyw estimates. The short time span of four weeks ensures that the influence of possible long-term changes in catchment flow paths are not captured and only the influence of the start and end time of sampling one year of isotope data is investigated. In the case that Fyw shows stronger variations, the sampling time will likely have influenced Fyw results. Patterns to help identify such situations beforehand are then searched by analyzing the time of occurrence of these situations.

35 3) Fyw estimates are similar for calculation years that are centered around a given calendar month (seasonal-invariance)

This hypothesis is accepted if the Fyw results centered around a specific month do not differ more than ± 0.04 within this month. To clarify, we did not calculate Fyw on a monthly basis but simply sorted the 189 Fyw results by the month they were assigned to (midpoint of the calculation year, see also explanation above). 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. This would allow the pre-planning of sampling campaigns to establish comparable Fyw results. However, it is also possible that the hypothesis is accepted if Fyw is constant for all 189 results, as only the intra-month variance matters with this hypothesis. Contrary to the acceptance of the hypothesis, rejecting it for most months would indicate that there are no distinct seasonal patterns imprinted on Fyw.

5

10 An example of a theoretical Fyw time series is given in Figure 2. Despite it having a time-variant young water fraction, all three hypotheses are accepted. On a long term basis, the young water fraction does not deviate significantly from its overall mean value (time-invariance), choosing to start a one-year long sampling camping on a specific date or e.g., two weeks later would not significantly alter the result (sampling-invariance) and results show a seasonal behavior that is stable over longer time frames (seasonal-invariance). 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.

15

3 Results

3.1 Isotopic and hydrometric data

Precipitation isotope ratios ranged from -3.04 to -17.80‰, spanning a range of 14.76‰ in $\delta^{18}\text{O}$ values. In comparison, 20 streamflow values ranged from -7.78 to -8.74‰ with a range of 0.96‰ or only 1/15th of precipitation values. The volume-weighed groundwater isotope value was $-8.43 \pm 0.17\text{‰}$. The maximum and minimum air temperatures were 27.0 and -7.4 °C, respectively, with a mean value of 7.6 °C. Relative humidity ranged from 96.8 to 32.3% with a mean of 82.2%. All the sampling years except winter season 2013/14 experienced a build-up of snow pack with a mean height of 15 cm. The absence of snow in 2013/14 correlated with on average higher temperatures (3.5 times the average temperature of the other years) and lower 25 relative humidity (5% lower average relative humidity compared to the other years). The hydrometeorological and isotope data are presented in more detail in section 3.3.

3.2 Climatological influence on preliminary data set analysis

Before presenting final Fyw estimates we briefly introduce the detection and subsequent remedy of a climatological influence on the initial Fyw results and their uncertainty: the initial 189 Fyw estimates and their uncertainty significantly increased from July 2014 to December 2015 (supplementary Figure S1). The uncertainty of Fyw reached peak values of ± 0.43 . Concurrent 30 with this, R^2_{adj} values dropped close to 0 while being above 0.2 for most other Fyw results. The low goodness-of-fit and the

consequential large uncertainty could have been caused by outlier values or extraordinary catchment conditions in the Wüstebach. The hydrometeorological and isotopic data pointed to an influence of the 2015 European heat wave (see supplementary material). The heat wave was detectable in the Wüstebach catchment by the lowest relative air humidity, second lowest rainfall amounts, lowest runoff coefficient, high temperatures, and the complete disconnection of precipitation and streamflow amplitudes (supplementary Figure S2). In addition, the 2015 European heat wave coincided with the lowest surface water temperatures of the North Atlantic since 1948 [Duchez *et al.*, 2016] which were visible by the loss of the seasonal d-excess signal. This created a situation where several months of precipitation isotope signal did not reach streamflow in the Wüstebach. The Fyw methods depends on comparable signals in precipitation and streamflow. Consequently, this disconnection of precipitation and streamflow added uncertainty to Fyw estimation. Therefore, we decided to omit the precipitation isotope values between April to July 2015 (11 out of 156 precipitation isotope data; 7% of the measurements; Figure 3a) resulting in less Fyw uncertainty (average: 0.08, maximum: 0.31). We did not omit streamflow data during the same period as it contained Fyw information of the previous three months of precipitation and streamflow sine wave fitting had no impact on Fyw uncertainty (see results of Figure 4b below).

3.3 Isotopic and hydrometric data

After omitting summer 2015 precipitation data the sine waves for the whole study period had an R^2_{adj} of 0.09 for precipitation and 0.23 for streamflow, respectively (Figure 3). The precipitation amplitude $A_P = 0.72\text{‰}$ and the streamflow amplitude $A_S = 0.08\text{‰}$ resulted in a Fyw of 0.12 ± 0.04 . Thus, the threshold value for hypothesis testing was chosen as the absolute value 0.04. The 189 fitted sine waves had a wide range of R^2_{adj} values: precipitation ranged from -0.02 to 0.63 with a mean of 0.22 and streamflow ranged from 0.00 to 0.55 with a mean of 0.25. The mean R^2_{adj} (arithmetic average of precipitation R^2_{adj} and streamflow R^2_{adj}) for each calculation year ranged from 0.03 to 0.59 with a mean of 0.24. The sine waves showed strong variations in terms of amplitudes and phase shifts leading to distinct deviations from the sine wave fitted to the whole time series (Figure 3). Precipitation amplitudes ranged between 0.35 to 2.60‰ with a mean value of 1.26‰ while streamflow amplitudes ranged between 0.03 to 0.19‰ with a mean value of 0.10‰. The mean of all streamflow amplitudes was closer to the single sine wave amplitude (0.10‰ vs. 0.08‰) than those for precipitation (1.26‰ vs. 0.72‰). If we use the averages of the 189 sine wave amplitudes to calculate Fyw, the result would be 0.08 instead of 0.12 of the single sine wave. This is less than the 0.04 difference in Fyw defined by this study as the data-driven threshold value for significant differences. The overall pattern of the individual peaks was similar to the single sine wave peaks, except for the period of the 2015 European heat wave when between June to October 2015 a distinct double-peak in precipitation was visible. The individual sine waves followed the general pattern of enriched isotopic values during summer months and depleted values in winter.

The mean R^2_{adj} showed a marked decrease during July 2014 to October 2015 with values falling well below 0.2 (Figure 4a). Approximately at the same time the Fyw results varied strongly (mean and maximum change of Fyw between consecutive one-year windows: 0.02 and 0.12) and the uncertainty was large (mean uncertainty: ± 0.11). Contrary to this, during periods of

larger R^2_{adj} the change in Fyw was more modest (mean and maximum change of Fyw between consecutive one-year windows: 0.01 and 0.05) with lower uncertainty (mean uncertainty: ± 0.04). To find possible modeling influences on the Fyw uncertainty we first compared the mean R^2_{adj} with it and found that they were correlated (Figure 4b inset, $R^2 = 0.65$). Following this we further investigated relationships between Fyw uncertainty and the amplitudes, phase shifts and vertical shifts of the 189 sine waves but only show results for throughfall amplitudes, as the other parameters had no correlation (Figure 4b). The throughfall amplitudes were correlated with an $R^2 = 0.79$ while contrary to this streamflow amplitudes had an $R^2 = 0.04$. Thus, the Fyw uncertainty was strongly controlled by the amplitudes of the precipitation sine waves while the streamflow sine waves barely influenced it.

The baseline for Fyw was around 0.05 (Figure 4). Before the low R^2_{adj} period Fyw was around 0.05, increased to about 0.1 for a short time and then fell back to 0.05. After the low R^2_{adj} period Fyw also fell to about 0.05, before rising in the end. Thus, during the 4.5-years Fyw seldom fell below the baseline of 0.05 and we assumed that during any one-year period the Wüstebach catchment will have at least 0.05 Fyw. Overall, the 189 Fyw results were positively skewed (Figure 5). Around 30% of results indicated a Fyw of 0.06, followed by 55% of results indicating a Fyw up to 0.08. Few Fyw values are higher than 0.16 with possible outliers between 0.26 to 0.28. Leaving out the period of low R^2_{adj} values does not change the skewness of the histogram. However, values of Fyw larger than 0.16 disappeared in favor of 0.06 that shifted from 30% to 40% relative frequency.

3.4 Hypothesis 1: Time-invariance

The mean value of all Fyw results was 0.09. Consequently, 90% of all Fyw results must lie within 0.05 to 0.13 to accept hypothesis 1. Out of the 189 Fyw results 159, i.e. 84%, were within those boundaries (Figure 6a). It could be possible that the period between July 2014 and October 2015 with low R^2_{adj} values and erratic Fyw behavior significantly influenced the rejection of the hypothesis. Therefore, in a second step we excluded this period, calculated the mean for those values and evaluated Fyw results again (Figure 6b). The new mean Fyw was 0.07 with 93% of results found between 0.03 to 0.11. Thus, contrary to using all data the hypothesis could be accepted if the period of large uncertainty was left out. We then compared the time-variable Fyw to hydrometeorological measurements (Figure 7) and found that neither temperature nor relative humidity were correlated with Fyw (not shown). While throughfall volume, runoff volume and snow height were also not correlated (Figure 7a-c) the runoff coefficient (Q/P) was negatively correlated with $R^2 = 0.25$ and p-value = 1.7E-11 (Figure 7d). Leaving out again the period from July 2014 to October 2015 reduced the correlation to $R^2 = 0.08$ and p-value = 9.8E-4.

3.5 Hypothesis 2: Sampling-invariance

Here we tested if short-term changes in the start of a one-year sampling campaign could significantly influence Fyw. The hypothesis is accepted if during any consecutive four weeks Fyw did not differ more than 0.04. On multiple occasions this rule was violated for the full data set, as well as for the reduced one (discounting the low R^2_{adj} period), so we rejected hypothesis 2

(Figure 6). Thus, the start time of a one-year long sampling campaign could significantly influence Fyw. The periods when hypothesis 2 was violated were neither equally spaced in time (Figure 6) nor did they show significant correlations to hydrometric (Figure 7) or meteorological (not shown) variables. The only observation made was that hypothesis 2 seems to have preferentially failed around the 2015 European heat wave.

5 3.6 Hypothesis 3: Seasonal-invariance

As mentioned in the methods, the Fyw results were put in the middle of the one-year calculation period (calculating from February 2016 to February 2017, the result would be displayed as a data point in August 2016). We grouped together all Fyw results that were assigned to a specific calendar month and used a box plot to detect possible seasonality (Figure 8). Only in January and February was the difference in Fyw below 0.04. When leaving out the period with low R^2_{adj} , January to August stayed within ± 0.04 . Thus, we also rejected hypothesis 3 based on all data as our results did not indicate pronounced seasonality. Nonetheless, a trend of declining Fyw from January to June was visible that reversed from July onwards. Additionally, the standard deviation of Fyw, the interquartile range of the boxplots and the number of outliers increased starting with June until October/November. We compared this behavior qualitatively to the start and end time of snow influence in the Wüstebach, which usually started in December and the last melt event happened in February. Since the influence of this delayed signal transmission from precipitation to streamflow does not immediately end with the final snowmelt in February, we assumed that snowmelt still influenced streamflow for the following two months, i.e., until April. This comparison showed that calculation years that included one year's winter had lower interquartile ranges, a lower number of outliers and smaller standard deviations. On the other hand, calculation years that included winters of two different years (e.g., a calculation year starting and ending in December) matched the boxplot results with increased uncertainty (Table 1).

20 4 Discussion

Judging by the isotope data, we generally expect that groundwater was recharged locally from precipitation as the long-term, volume-weighed $\delta^{18}\text{O}$ of precipitation with $-8.53\text{\textperthousand}$ was close to the quasi-constant $\delta^{18}\text{O}$ of groundwater with a 5-year mean of $-8.43 \pm 0.17\text{\textperthousand}$. Streamflow was substantially comprised of groundwater as its volume-weighed $\delta^{18}\text{O}$ was $-8.40\text{\textperthousand}$. The study by Weigand *et al.* [2017] came to the same conclusion for the Wüstebach catchment using wavelet analysis of nitrate and DOC data collected at mainstream and tributary locations. While lower altitude locations of the catchment near the outlet were dominated by groundwater, higher altitude areas were less affected. This finding was additionally supported by field observations of shallow groundwater.

4.1 Sine wave fits

The single sine wave fits to all data had low R^2_{adj} values (0.09 for throughfall and 0.23 for streamflow). Compared to this, the 189 individual sine waves reached a maximum R^2_{adj} of 0.63 and were often larger than 0.2. This indicated that the single wave

fit to multi-year data is an oversimplification of the inter-annual variability in meteoric and streamflow isotope data and annual sine waves better capture the variability. One might argue that sine waves are a non-adequate function to describe the data variability if their R^2 is low. However, Fyw estimation is based on comparing sine wave amplitudes [Kirchner, 2016a] and no similar method exists to calculate it with different functions.

5

Completely undetectable by a single sine wave fit, the 189 sine waves highlighted a hydrologic change in the Wüstebach catchment caused by the 2015 European heat wave: the disconnection of precipitation and runoff. First, the general shapes of the 189 precipitation and 189 streamflow sine waves were similar (Figure 3), which can be seen, e.g., in the positive and negative peaks occurring around September 2014 and 2016 and February 2013 and 2014, respectively. Additionally, throughfall and streamflow amplitudes generally matched each other (supplementary Figure S2a). This indicated that throughout the 4.5-year time series the characteristic of the precipitation $\delta^{18}\text{O}$ signal was for the most part consistently and quickly transferred to the streamflow $\delta^{18}\text{O}$ signal within a year. 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). After the heat wave the general shape of precipitation and streamflow sine waves matched each other again while their respective amplitudes regained their former relationship, albeit weakened: the large amplitude peak in throughfall in April 2016 again led to increasing streamflow peaks. 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 $\delta^{18}\text{O}$ signal in streamflow and resulting in the similar shapes of the 189 sine wave pairs. The 2015 European heat wave greatly disturbed the usually occurring runoff-generation process in the Wüstebach, leading to a disconnection of precipitation and streamflow signal.

A fast transmission of precipitation to streamflow was also found by *Jasechko et al.* [2016], and the fact that a part of precipitation quickly becomes streamflow is already inherent in Fyw. The new insight of the present study is the unexpected close resemblance of the 189 sine waves for precipitation and streamflow although the groundwater influence seems to have dominated in the Wüstebach. The simultaneous strong attenuation of the $\delta^{18}\text{O}$ streamflow signal while at the same time retaining much of the precipitation $\delta^{18}\text{O}$ signal characteristics can be explained by mixing with a quasi-constant $\delta^{18}\text{O}$ source, e.g., with groundwater. This would not alter the pattern but only attenuate the signal. Thus, the 189 sine waves gave a strong indication that streamflow in the Wüstebach consisted of precipitation and groundwater with no additional, unaccounted sources of runoff such as subsurface flows from outside the catchment boundaries. This supports a previous study that closed the water-balance for the Wüstebach catchment using only precipitation, evapotranspiration and runoff data [*Graf et al.*, 2014] and is essential information for e.g., endmember-mixing analysis [*Barthold et al.*, 2011; *Katsuyama et al.*, 2001] or isotope

hydrograph separation [Klaus and McDonnell, 2013]. Similar to before, this hydrological information about the Wüstebach catchment would have been impossible to detect with a single sine wave fit.

4.2 Fraction of young water

The fact that Fyw calculated with the average amplitudes of 189 precipitation and streamflow sine waves was within the ± 0.04 boundary to Fyw calculated with a single sine wave (0.08 vs. 0.12) indicated that the single sine wave generally averaged the behavior of the 189 ones.

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). Thus, if a study is interested in the overall behavior of a multi-year time series, a single sine wave fit would seem sufficient. Nevertheless, hypothesis 1 was rejected for both cases as Fyw varied significantly within this multi-year time series (Figure 6). Using a moving time window

10 to calculate a host of Fyw values ensures that the entire range of possible Fyw estimates is considered with an average estimate and most importantly its uncertainty.

Most of the isotope data between 7-day calculation window shifts were the same. Still, during the low R^2_{adj} period Fyw occasionally fluctuated in the order of 0.12 between one-week shifts. From a hydrological standpoint it is difficult to imagine

15 a short-term change in flow paths of this magnitude for annual averages. Given that the Fyw calculation is based on comparing the amplitudes of precipitation and streamflow and a low R^2_{adj} indicates a weak fit to a sine wave shape, we assumed that in our case the Fyw calculation method reached its limit below an average $R^2_{adj} = 0.2$. Fyw became highly sensitive to a small change in input data and in consequence highly uncertain. We recommend further investigations of the sensitivity of Fyw to the goodness-of-fit (not necessarily only measured with R^2_{adj}) for future studies. It remains to be seen if a value of 0.2 for R^2_{adj}

20 is a general critical threshold for Fyw or if different catchments show varying results. Such studies should consider that the Fyw uncertainty was correlated with throughfall amplitudes (Figure 4b), raising the question if a curve fit with $R^2_{adj} = 0.6$ is objectively better than a fit with $R^2_{adj} = 0.3$ when the underlying isotope data have completely different amplitudes. A decrease in the goodness-of-fit of the sine wave when amplitudes are low was also found by Lutz *et al.* [2018].

25 A difference of ± 0.04 Fyw was defined as the data-driven threshold value for significant differences in Fyw by this study. The acceptance or rejection of our null hypotheses will thus inform if the time-variability of Fyw is large in comparison to the averaged Fyw value and its uncertainty. We recommend using different thresholds that are suited to the purpose of calculating a Fyw estimate. Purposes can range from any application of the method to answer questions about the quantity and quality of water resources for various industrial, touristic or infrastructural uses. First, a critical difference in Fyw should be defined by

30 each application that reflects e.g., the vulnerability of aquatic ecosystems to certain pollutant loads. If an increase or decrease by less than this value does not impact the results of an, e.g., risk assessment, then these Fyw changes are non-significant for the practical purpose at hand. The present study did not aim to answer any specific question related to Fyw that would justify setting a threshold value a priori but investigated the time-variability of Fyw and used the data-inherent uncertainty as its

threshold value. Thus, while our hypotheses are accepted or rejected, the results of the hypothesis tests might change completely if we would answer practical questions about the Wüstebach such as the vulnerability to pollutant loads of a certain chemical substance.

5 The 2015 European heat wave was among the top ten heat waves of the past 65 years and was accompanied by the lowest surface water temperatures of the North Atlantic in the period of 1948 to 2015 [Duchez et al., 2016]. The North Atlantic influences the European summer climate [Ghosh et al., 2017] and is an important vapor source for precipitation over Europe [Hurrell, 1995; Trigo et al., 2004]. The combined effects of low ocean water temperatures and high air temperatures in Europe were visible in the d-excess that lost its clear seasonal signal in summer 2015 (supplementary Figure S2d). The d-excess of

10 precipitation samples is strongly controlled by the relative humidity of the moisture source [Pfahl and Sodenmann, 2014; Steen-Larsen et al., 2014] which in turn would change with changing surface water temperatures and thus changing evaporation rates. Additionally, the increased European air temperatures during the heat wave would increase secondary evaporation of falling raindrops, further altering the d-excess of precipitation samples. The North Atlantic and European temperature anomalies of 2015 explain the behavior of the d-excess as well as the unusual double-peak of the 189 sine waves that was

15 observed for summer 2015 in the Wüstebach.

Apart from affecting the isotopic input signal into the Wüstebach catchment, the temperature anomalies of 2015 also changed the hydrological behavior of the Wüstebach: precipitation was largely disconnected from streamflow and the isotopic signal was not transferred (supplementary Figure S2a-c). This directly increased Fyw uncertainty during this period. Future studies

20 must be careful in comparing Fyw estimates of different time periods, especially if a heat wave occurred during those periods. We assume that mostly small headwater catchment with shallow soils are strongly affected by this effect but do not exclude the possibility of other catchments being affected in varying degrees too. It is highly advisable to investigate further in this direction, as the probability of heat waves in the period from 2021 to 2040 is poised to increase [Russo et al., 2015]. This, in extension, means that the probability of getting highly uncertain Fyw results will increase too. We argue that heat waves are

25 actively disturbing the estimation of Fyw by potentially decoupling the input from the output isotope signal. This can be more clearly illustrated by the theoretical worst-case scenario: the decoupling of precipitation and streamflow signal for a full year and streamflow being solely fed by another source, e.g., groundwater. Why, in this case, would we trust the Fyw result, no matter the magnitude of the uncertainty and goodness of sine wave fit? Thus, it is reasonable to assume that any amount of decoupling will add uncertainty to Fyw, as demonstrated by our data and results. Only by comparison to other time frames

30 where the uncertainty was smaller was it possible for us to detect that the uncertainties for summer 2015 were unusually large.

4.3 Hypothesis 1 – Fyw is time-variant

Hypothesis 1 was rejected because the Fyw varied in the long-term. For example, in December 2013 Fyw was 0.06 while two months later it increased to 0.1, almost doubling. From summer 2016 to the end of the time series Fyw even tripled from 0.06

to [0.15](#). These differences in Fyw results complicate catchment comparisons as the result does not only depend on catchment characteristics but also on when isotope data was collected. As far as we can tell, the recent Fyw catchment comparison study of *Lutz et al.* [2018] used the same sampling period for precipitation and streamflow for all 24 investigated catchments. In contrast, the studies of *Jasechko et al.* [2016] and *von Freyberg et al.* [2018] had isotope sampling periods varying in start date 5 and overall length for the 254 and 22 investigated catchments, respectively, potentially influencing the uncertainty for the inter-catchment comparison according to the results of our study.

In the Wüstebach catchment the baseline for Fyw was around [0.05](#). This lower boundary is useful in assessing pollutant risk and nutrient loss in the catchment as it defines a minimum expected load that will quickly appear in the stream if combined 10 with precipitation volumes and chemical substance concentrations. [Using a single sine wave would not have revealed this lower boundary.](#)

The variability in Fyw of this study could not be explained by meteorological or hydrometric variables. *Lutz et al.* [2018] found a negative correlation between annual precipitation and Fyw. The study of 22 Swiss catchments by *von Freyberg et al.* [2018] 15 found significant positive correlations between Fyw and mean monthly discharge and precipitation volumes. Fyw of this study neither correlated with precipitation nor with runoff (Figure [7a](#) and Figure [7b](#)). Such contradictions could be explained by the different sampling periods of [our study and the](#) mentioned studies but also by [differing](#) catchment characteristics. Additionally, the present study investigated the same catchment temporally while the other studies investigated spatially different catchments. Furthermore, *Lutz et al.* [2018] found complex interactions between several catchment characteristics and Fyw, 20 possibly resulting in nonsignificant linear regressions between Fyw and individual catchment characteristics. However, the runoff coefficient Q/P was negatively correlated with Fyw (Figure [7d](#)). Physically, this could be explained by the fact that if annual runoff [volumes increase per](#) annual precipitation [volume](#) then the additional runoff volumes were provided by catchment storage. This increased the percentage of old water in streamflow and [relatively](#) decreased the Fyw since catchment storage consists of old water [*Gabrielli et al.*, 2018].

25 **4.4 Hypothesis 2 & 3 – Fyw is sensitive to sampling and has no [clear](#) seasonal pattern**

While hypothesis 1 concentrated on long-term changes, hypotheses 2 focused on short-term changes where choosing to [start](#) a one-year sampling campaign by one to four weeks [later](#) could lead to significantly different results. On several occasion Fyw differed more than [±0.04](#) within four weeks (Figure [6](#)). This means that the choice of the sampling period has a large potential 30 for uncertainty in the Fyw estimates for studies that can monitor the [water](#) stable isotopes in precipitation and streamflow for only one year. The obtained Fyw could be a potential outlier, a larger value or part of the Fyw baseline [around 0.05 in the present study. As the timing of](#) the violation of hypothesis 2 did not correlate with any meteorological or hydrometric data it was not possible to determine the conditions under which the sampling period led to higher Fyw uncertainty. [A relationship with the 2015 European heat wave is possible, albeit not fully evident. Nonetheless, as discussed above, the choice of another](#)

threshold value beside the data-inherent ± 0.04 may lead to an increase in the number of significant short-term Fyw changes. The results of this study indicate that estimating Fyw with data of a single year might not be enough for fully understanding catchment behavior. Quoting Kirchner et al. [2004]: “If we want to understand the full symphony of catchment hydrochemical behavior, then we need to be able to hear every note.”. A single Fyw result is one note in the symphony of potential Fyw results slumbering in multi-year data sets.

Fyw did not have a clear seasonal pattern in that not all the months had Fyw differences of less than ± 0.04 (Figure 8). A pattern was visible with larger Fyw with less uncertainty when the sampling campaign was centered around winter months compared to lower Fyw with larger uncertainties when the campaign was centered around summer months. The behavior of Fyw uncertainty can potentially be explained by the influence of snow and is similar to the proposed problem that the 2015 European heat wave introduced: a tracer signal in precipitation/streamflow that does not have any instantaneous connection with its counterpart streamflow/precipitation. This disconnection by snow could be explained by the longer delay in signal transmission of snowfall compared to rainfall due to snow blanket build-up. Consider a winter at the start of a sampling campaign: it is likely that streamflow will feature the snowmelt isotope signal originating from snowfall of e.g., several weeks ago that is not featured in the precipitation isotope data of this calculation year. Furthermore, snow blankets also change the isotopic signal potentially to a degree that obscures seasonal isotope patterns [Cooper, 2006]. This in turn would affect the Fyw estimate and its associated uncertainty. Currently, we recommend that if studies can only sample one year of data in snow-influenced catchments to not sample winters of two different calendar years and to design the sampling such that only one year’s winter is in the time series. Future studies should provide more evidence if Fyw calculated by one year of isotope data shows a seasonal behavior or not and how snow influences the uncertainty. We highly recommend calculating a time series of Fyw, e.g., with the method of this study, to understand the temporal behavior of Fyw for the investigated catchment and to be able to evaluate possible uncertainties for Fyw estimation.

A difference in Fyw when only one year of isotope data is available was also observed by Stockinger et al. [2017] for the same catchment using only two calculation years without any further investigations in this direction, as it was not the main objective of their study to investigate Fyw time-variability and uncertainty. 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. Fyw for these years were 0.06 and 0.13, respectively. The authors assumed that using the complete time series averages sub-sets of the time series as the Fyw for the whole time series was approximately 0.13, so in between 0.06 and 0.13. However, this happened by coincidence. The present study shows that the two Fyw could have been very different, e.g., both near 0.05. Then, Fyw of the whole time series would not have averaged the results of the two individual years. Thus, only the complete picture of all 189 individual Fyw results allowed a better judgment of Fyw time-variability and uncertainty. With knowledge from the current study, we would even consider one of the hydrological calculation years of Stockinger et al. [2017] as highly uncertain and possibly influenced by the 2015 European heat wave.

5 Conclusions

The fraction of young water (Fyw) is a promising new measure to estimate the fast transport of precipitation through a catchment to the stream. To calculate Fyw, sine waves are fitted to the water stable isotopes in precipitation and streamflow and their respective amplitudes compared. This is usually done for the complete time series available, ranging from less than

5 a year to multiple years. This study used a moving one-year window to investigate the temporal variance of Fyw and its uncertainty for a 4.5-year long time series. Using 189 Fyw results instead of a single multi-year one, we were able to increase our hydrometeorological knowledge about the study catchment: (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; (2) precipitation and groundwater seemed to be the only end-members in streamflow which is information

10 that isotope hydrograph separation studies can greatly benefit from; (3) a lower boundary of 0.05 Fyw was found, aiding e.g. pollutant risk studies in calculating minimum expected loads. Testing three hypotheses about the time-variability of Fyw we found that both in the long and short term Fyw is time-variable as defined by this study by the data-inherent ± 0.04 threshold, while showing no clear seasonal pattern. The long-term variability has implications for catchment comparison studies when different time periods are investigated. Short-term variability indicated a potentially high sensitivity to the sampling period,

15 15 where a shift of 1-4 weeks in the start of a one-year long sampling campaign significantly influenced Fyw. No pronounced seasonality of Fyw could be derived. However, a possible influence of snow pack led to the recommendation of sampling one year's winter and avoiding sampling the winters of two different years. If feasible, we recommend investigating a multi-year time series of tracer data with the method suggested in this study to enhance our knowledge of the sensitivity of Fyw to the chosen time frame in different catchment situations and the behavior of its uncertainty. That is, to use a one-year moving time

20 20 window and estimate an ensemble of Fyw results and its uncertainty. Based on the goodness-of-fit for all 189 calculated sine waves and the corresponding Fyw behavior, we recommend considering that Fyw based on R^2_{adj} below 0.2 might be highly uncertain. This must be verified by other dedicated studies of different catchments and would allow for a better comparability of Fyw results with various goodness-of-fits. The present study shows the importance of considering inter-annual fluctuations in the amplitudes of isotope tracer data and consequently of derived Fyw estimates in further learning about the uncertainty of

25 Fyw and in aiding in catchment comparison studies.

Acknowledgements

We gratefully acknowledge the support by the SFB-TR32 “Patterns in Soil-Vegetation-Atmosphere Systems: Monitoring, Modelling, and Data Assimilation” funded by the Deutsche Forschungsgemeinschaft (DFG) and TERENO (Terrestrial Environmental Observatories) funded by the Helmholtz-Gemeinschaft. Holger Wissel, Werner Küpper, Rainer Harms, 30 Ferdinand Engels, Leander Fürst, Sebastian Linke and Isabelle Fischer are thanked for supporting the isotope analysis, sample collection and the ongoing maintenance of the experimental setup. We appreciate the helpful comments of three anonymous

reviewers that greatly improved the present study and the work of editor Patricia Saco. The data used in this study can be acquired from the corresponding author.

References

Barthold, F. K., Tyralla, C., Schneider, K., Vaché, K.B., Frede, H.G. and Breuer, L.: How many tracers do we need for end member mixing analysis (EMMA)? A sensitivity analysis, *Water Resour. Res.*, 47, W08519, doi:10.1029/2011WR010604, 2011.

Bogena, H.R., Montzka, C., Huisman, J.A., Graf, A., Schmidt, M., Stockinger, M., von Hebel, C., Hendricks-Franssen, H.J., van der Kruk, J., Tappe, W., Lücke, A., Baatz, R., Bol, R., Groh, J., Pütz, T., Jakobi, J., Kunkel, R., Sorg, J., and Vereecken, H.: The TERENO-Rur Hydrological Observatory: A Multiscale Multi-compartment Research Platform for the Advancement of Hydrological Science. *Vadose Zone J.* 17(1): 1-22, doi:10.2136/vzj2018.03.0055, 2018.

Bourgault, R.R., Ross, D.S., Bailey, S.W., Bullen, T.D., McGuire, K.J., and Gannon, J.P.: Redistribution of soil metals and organic carbon via lateral flowpaths at the catchment scale in a glaciated upland setting. *Geoderma* 307:238–252, 2017.

Brand, W. A., Coplen, T.B., Vogl, J., Rosner, M., and Prohaska, T.: Assessment of international reference materials for isotope-ratio analysis (IUPAC Technical Report), *Pure Appl Chem*, 86(3), 425-467, 2014.

Christidis, N., Jones, G. S., and Stott, P. A.: Dramatically increasing chance of extremely hot summers since the 2003 European heatwave. *Nat. Clim. Change*, 5, 46–50, doi:10.1038/nclimate2468, 2015.

Cooper, L.W.: Isotopic Fractionation in Snow Cover. In: Kendall, C., McDonnell, J.J. (Eds.), *Isotope Tracers in Catchment Hydrology*. Elsevier Science B.V., Amsterdam, pp. 119–136, 2006.

Craig, H.: Isotopic variations in meteoric waters. *Science*, 133, pp. 1702-03, 1961.

Dansgaard, W.: Stable isotopes in precipitation, *Tellus A* 16:436–468. <https://doi.org/10.1111/j.2153-3490.1964.tb00181.x>, 1964.

Etmann, M.: Dendrologische Aufnahmen im Wassereinzugsgebiet Oberer Wüstebach anhand verschiedener Mess- und Schätzverfahren, M.S. thesis, Institut für Landschaftsökologie, University of Münster, Münster, Germany, 2009.

Gabrielli, C. P., Morgenstern, U., Stewart, M. K., and McDonnell, J. J.: Contrasting groundwater and streamflow ages at the Maimai watershed. *Water Resources Research*, 54, 3937–3957. <https://doi.org/10.1029/2017WR021825>, 2018.

Gannon, J.P., McGuire, K.J., Bailey, S.W., Bourgault, R.R., and Ross, D.S.: Lateral water flux in the unsaturated zone: a mechanism for the formation of spatial soil heterogeneity in a headwater catchment. *Hydrol. Process.* <http://dx.doi.org/10.1002/hyp.11279>, 2017.

Ghosh, R., Müller, W.A., Baehr, J., and Bader, J.: Impact of observed North Atlantic multidecadal variations to European summer climate: a linear baroclinic response to surface heating. *Clim Dyn* 48: 3547. <https://doi.org/10.1007/s00382-016-3283-4>, 2017.

Gottselig, N., Bol, R., Nischwitz, V., Vereecken, H., Amelung, W., and Klumpp, E.: Distribution of phosphorus-containing fine colloids and nanoparticles in stream water of a forest catchment. *Vad. Zone* J.13, 1–11.doi: 10.2136/vzj2014.01.0005, 2014.

5 Graf, A., Bogena, H.R., Drüe, C., Hardelauf, H., Pütz, T., Heinemann, G., and Vereecken, H.: Spatiotemporal relations between water budget components and soil water content in a forested tributary catchment, *Water Resour. Res.*, 50, 4837–4857, doi:10.1002/2013WR014516, 2014.

Harman, C. J.: Time-variable transit time distributions and transport: Theory and application to storage-dependent transport of chloride in a watershed. *Water Resour. Res.*, 51, doi:10.1002/2014WR015707, 2015.

10 Heidbüchel, I., P. A. Troch, and S. W. Lyon: Separating physical and meteorological controls of variable transit times in zero-order catchments. *Water Resour. Res.*, 49, 7644–7657, doi:10.1002/2012WR013149, 2013.

Hurrell, J.W.: Decadal trends in the North Atlantic oscillation: Regional temperatures and precipitation. *Science* 269(5224):676–679, 1995.

Jasechko, S., Kirchner, J. W., Welker, J. M., and McDonnell, J.J.: Substantial proportion of global streamflow less than three months old, *Nat. Geosci.*, 9, 126-129, 2016.

15 Katsuyama, M., Ohte, N., and Kobashi, S.: A three-component end-member analysis of streamwater hydrochemistry in a small Japanese forested headwater catchment. *Hydrological Processes*, 15, 249– 260, 2001.

Kirchner, J.W., Feng, X., Neal, C., and Robson, A.J.: The fine structure of water-quality dynamics: the (high-frequency) wave of the future. *Hydrol. Process.* 18, 1353–1359. <https://doi.org/10.1002/hyp.5537>, 2004.

20 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 Sc*, 20(1), 279-297, 2016a.

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

Klaus, J., and McDonnell, J.J.: Hydrograph separation using stable isotopes: review and evaluation, *J. Hydrol.*, 505, pp. 47-64, 2013.

25 Lutz, S. R., Krieg, R., Müller, C., Zink, M., Knöller, K., Samaniego, L., and Merz, R.: Spatial patterns of water age: Using young water fractions to improve the characterization of transit times in contrasting catchments. *Water Resources Research*, 54, 4767–4784.<https://doi.org/10.1029/2017WR022216>, 2018.

McGuire, K. J., and McDonnell, J.J.: A review and evaluation of catchment transit time modeling, *J. Hydrol.*, 330(3–4), 543–563, 2006.

30 Merlivat, L. and Jouzel, J.: Global climatic interpretation of the deuterium-oxygen 18 relationship for precipitation, *J. Geophys.Res.*, 84, 5029–5033, 1979.

Miyata, S., Kosugi, K., Gomi, T., and Mizuyama, T.: Effects of forest floor coverage on overland flow and soil erosion on hillslopes in Japanese cypress plantation forests, *Water Resour. Res.*, 45, W06402, doi:10.1029/2008WR007270, 2009.

Ogrinc, N., Kanduč, T., Stichler, W., and Vreča, P.: Spatial and seasonal variations in $\delta^{18}\text{O}$ and δD values in the River Sava in Slovenia. *Journal of Hydrology*, 359(3–4), 303–312. <https://doi.org/10.1016/j.jhydrol.2008.07.010>, 2008.

Pendergrass, A. G., and Hartmann, D. L.: Changes in the distribution of rain frequency and intensity in response to global warming. *Journal of Climate*, 27(22), 8372–8383. <https://doi.org/10.1175/JCLI-D-14-00183.1>, 2014.

5 Pfahl, S. and Sodemann, H.: What controls deuterium excess in global precipitation?, *Clim. Past*, 10, 771–781, doi:10.5194/cp-10-771-2014, 2014.

Russo, S., Sillmann, J., and Fischer, E.M.: Top ten European heatwaves since 1950 and their occurrence in the coming decades. *Environ. Res. Lett.* 10, 124003, 2015.

Song, C., Wang, G., Liu, G., Mao, T., Sun, X., and Chen, X.: Stable isotope variations of precipitation and streamflow reveal 10 the young water fraction of a permafrost watershed. *Hydrological Processes*, 31, 935–947. <https://doi.org/10.1002/hyp.11077>, 2017.

15 Steen-Larsen, H. C., Sveinbjörnsdóttir, A. E., Peters, A. J., Masson-Delmotte, V., Guishard, M. P., Hsiao, G., Jouzel, J., Noone, D., Warren, J. K., and White, J. W. C.: Climatic controls on water vapor deuterium excess in the marine boundary layer of the North Atlantic based on 500 days of in situ, continuous measurements, *Atmos. Chem. Phys.*, 14, 7741–7756, <https://doi.org/10.5194/acp-14-7741-2014>, 2014.

Stockinger, M.P., Lücke, A., Vereecken, H., and Bogena, H.R.: Accounting for seasonal isotopic patterns of forest canopy 20 intercepted precipitation in streamflow modeling. *J. Hydrol.* 555:31–40. doi:10.1016/j.jhydrol.2017.10.003, 2017.

Stockinger, M. P., Bogena, H. R., Lücke, A., Diekkrüger, B., Cornelissen, T., and Vereecken, H.: Tracer sampling frequency influences estimates of young water fraction and streamwater transit time distribution, *Journal of Hydrology*, 541(Part B), 952–964. <https://doi.org/10.1016/j.jhydrol.2016.08.007>, 2016.

Stockinger, M. P., Lücke, A., McDonnell, J.J., Diekkrüger, B., Vereecken, H., and Bogena, H.R.: Interception effects on stable 25 isotope driven streamwater transit time estimates, *Geophys Res Lett*, 42(13), 5299–5308, 2015.

Stockinger, M. P., Bogena, H.R., Lücke, A., Diekkrüger, B., Weiler, M., and Vereecken, H.: Seasonal soil moisture patterns: Controlling transit time distributions in a forested headwater catchment, *Water Resour. Res.*, 50, 5270–5289, doi:10.1002/2013WR014815, 2014.

Tetzlaff, D., Seibert, J., and Soulsby, C.: Inter-catchment comparison to assess the influence of topography and soils on 30 catchment transit times in a geomorphic province; the Cairngorm mountains, Scotland, *Hydrol. Processes*, 23(13), 1874–1886, 2009a.

Tetzlaff, D., Seibert, J., McGuire, K.J., Laudon, H., Burn, D.A., Dunn, S.M., and Soulsby, C.: How does landscape structure influence catchment transit time across different geomorphic provinces?, *Hydrol. Processes*, 23(6), 945–953, 2009b.

Trenberth, K.E.: Changes in precipitation with climate change. *Clim Res* 47:123-138. <https://doi.org/10.3354/cr00953>, 2011.

Trigo, R. M., Pozo-Vazquez, D., Osborn, T.J., Castro-Diez, Y., Gamiz-Fortis, S. and Esteban-Parra, M.J.: North Atlantic oscillation influence on precipitation, river flow and water resources in the Iberian Peninsula. *International Journal of Climatology* 24: 925–944, 2004.

Tsuboyama, Y., Sidle, R. C., Noguchi, S., and Hosoda, I.: Flow and solute transport through the soil matrix and macropores of a hillslope segment. *Water Resources Research*, 30(4), 879–890, 1994.

von Freyberg, J., Allen, S. T., Seeger, S., Weiler, M., and Kirchner, J. W.: Sensitivity of young water fractions to hydro-climatic forcing and landscape properties across 22 Swiss catchments. *Hydrol. Earth Syst. Sci.*, 22, 3841–3861, 5 <https://doi.org/10.5194/hess-22-3841-2018>, 2018.

Weigand, S., Bol, R., Reichert, B., Graf, A., Wiekenkamp, I., Stockinger, M., Luecke, A., Tappe, W., Bogena, H., Puetz, P., Amelung, W., Vereecken, H.: Spatiotemporal analysis of dissolved organic carbon and nitrate in waters of a forested catchment using wavelet analysis. *Vadose Zone Journal*. DOI: 10.2136/vzj2016.09.0077, 2017.

Wiekenkamp, I., Huisman, J. A., Bogena, H. R., Lin, H. S., and Vereecken, H.: Spatial and temporal occurrence of preferential 10 flow in a forested headwater catchment. *Journal of Hydrology*, 534, 139–149. <https://doi.org/10.1016/j.jhydrol.2015.12.050>, 2016a.

Wiekenkamp, I., Huisman, J.A., Bogena, H., Graf, A., Lin, H., Drüe, C., and Vereecken, H.: Changes in Spatiotemporal Patterns of Hydrological Response after Partial Deforestation. *J. Hydrol.* 542: 648-661, doi:10.1016/j.jhydrol.2016.09.037, 2016b.

15 Wilusz, D. C., Harman, C. J., and Ball, W. P.: Sensitivity of catchment transit times to rainfall variability under present and future climates. *Water Resour. Res.*, 53, 10231–10256, <https://doi.org/10.1002/2017WR020894>, 2017.

Zacharias, S., Bogena, H.R., Samaniego, L., Mauder, M., Fuß, R., Pütz, T., Frenzel, M., Schwank, M., Baessler, C., Butterbach-Bahl, K., Bens, O., Borg, E., Brauer, A., Dietrich, P., Hajnsek, I., Helle, G., Kiese, R., Kunstmann, H., Klotz, S., Munch, J.C., Papen, H., Priesack, E., Schmid, H.P., Steinbrecher, R., Rosenbaum, U., Teutsch, G., and Vereecken, H.: A Network of 20 Terrestrial Environmental Observatories in Germany, *Vadose Zone J*, 10(3), 955-973, 2011.

Months from January (01) to December (12)

07	08	09	10	11	12	01	02	03	04	05	06	07
08	09	10	11	12	01	02	03	04	05	06	07	08
09	10	11	12	01	02	03	04	05	06	07	08	09
10	11	12	01	02	03	04	05	06	07	08	09	10
11	12	01	02	03	04	05	06	07	08	09	10	11
12	01	02	03	04	05	06	07	08	09	10	11	12
01	02	03	04	05	06	07	08	09	10	11	12	01
02	03	04	05	06	07	08	09	10	11	12	01	02
03	04	05	06	07	08	09	10	11	12	01	02	03
04	05	06	07	08	09	10	11	12	01	02	03	04
05	06	07	08	09	10	11	12	01	02	03	04	05
06	07	08	09	10	11	12	01	02	03	04	05	06

Table 1. The calculation years used for the boxplots of Figure 8. For example, the first row shows a calculation year starting in July and ending in July, where the Fyw result was assigned to January. Grey shaded areas are the usual beginning of snowfall and the final snowmelt (Dec to Feb, dark shaded) with an assumed prolonged influence of snowmelt on streamflow until April (light-shaded).

5 **Green coloured calculation years highlight snow influence of only one winter within this year, while red coloured calculation years highlight influence of two different winters.**

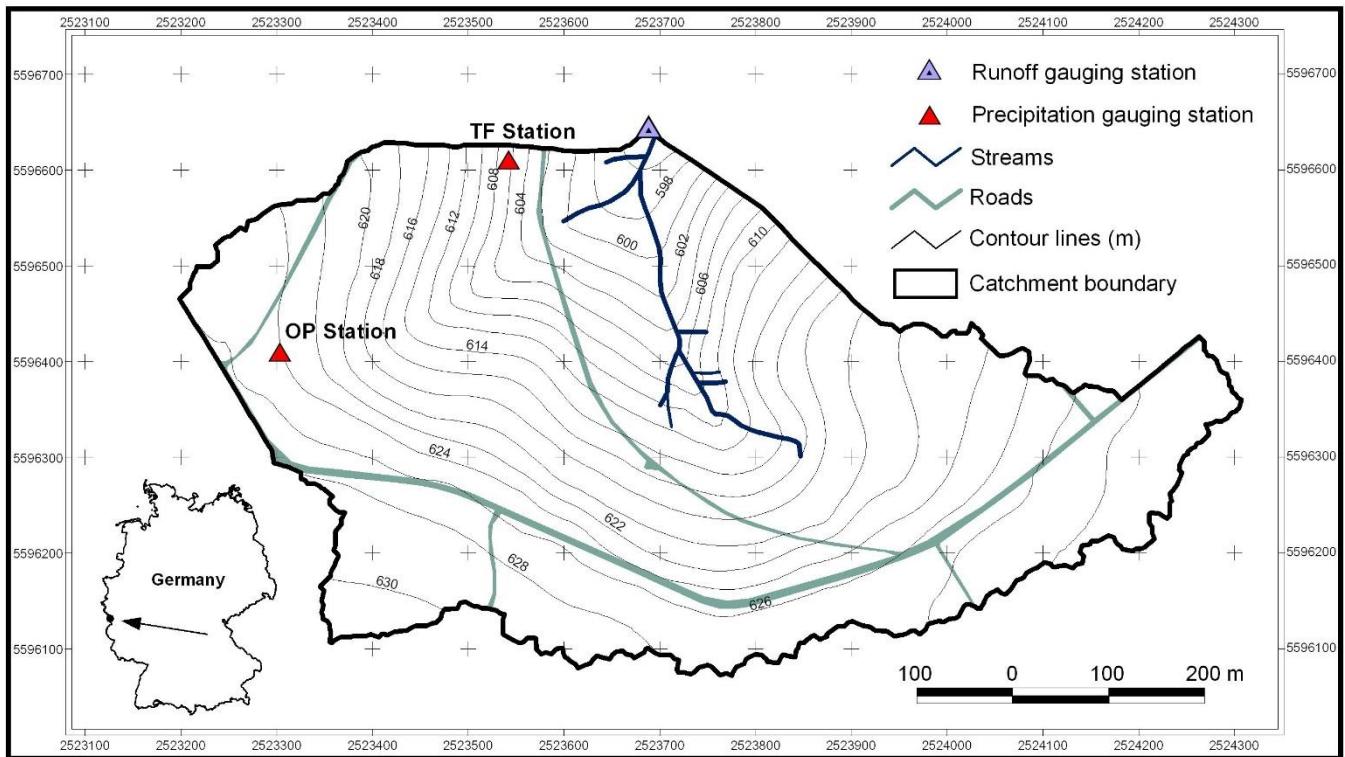


Figure 1. Map showing the Wüstebach catchment and the used monitoring stations. OP Station is the open precipitation collection site, while TF Station is the throughfall station.

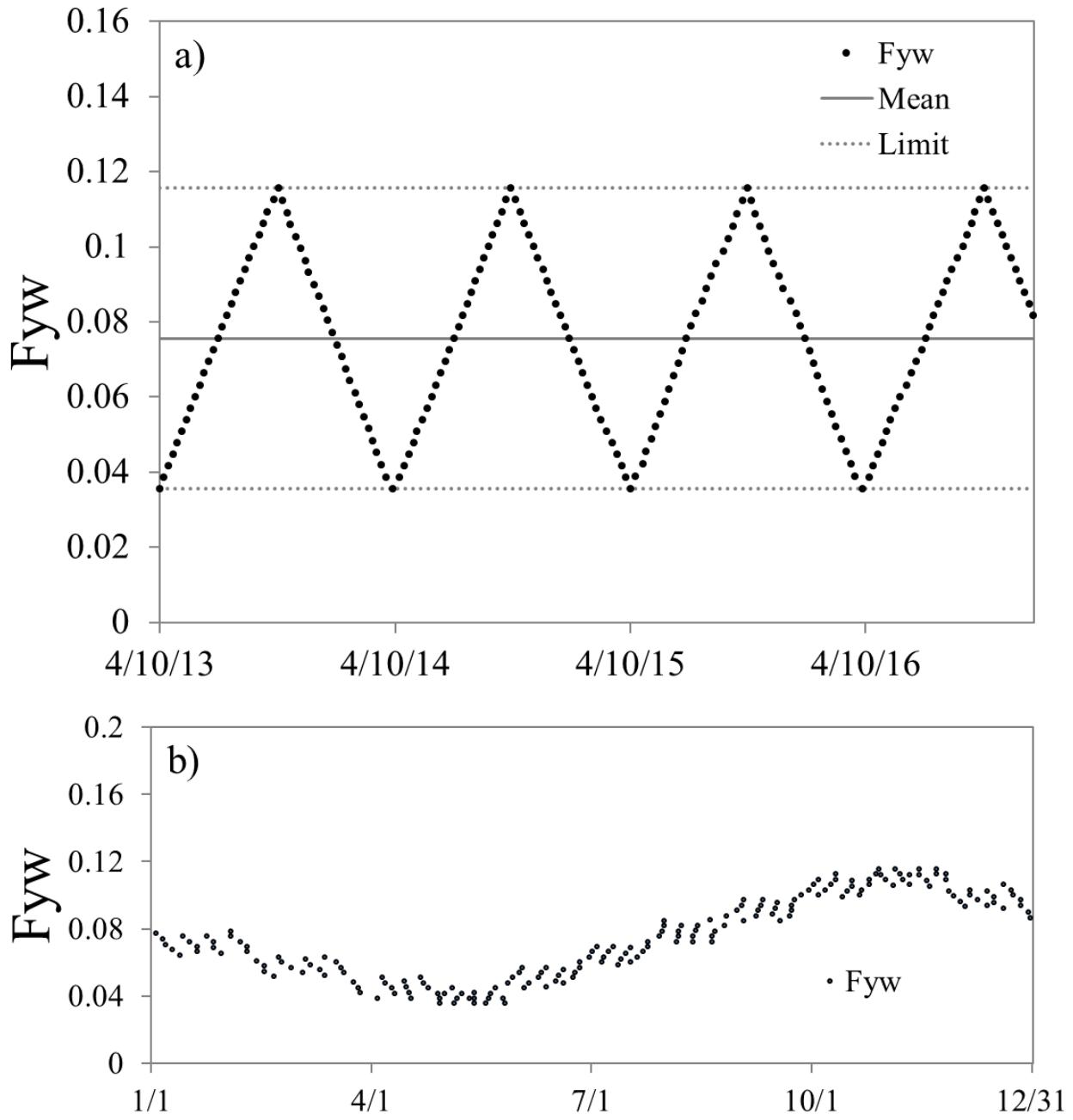


Figure 2. Panel (a): Example of a theoretical Fyw time series where despite the time-variance all three null hypotheses are accepted: (1) more than 90% of Fyw values lie within ± 0.04 of the mean of all values; (2) Fyw does not change more than ± 0.04 over the course of four weeks; (3) Fyw for each month does not change more than ± 0.04 within a month (panel (b)).

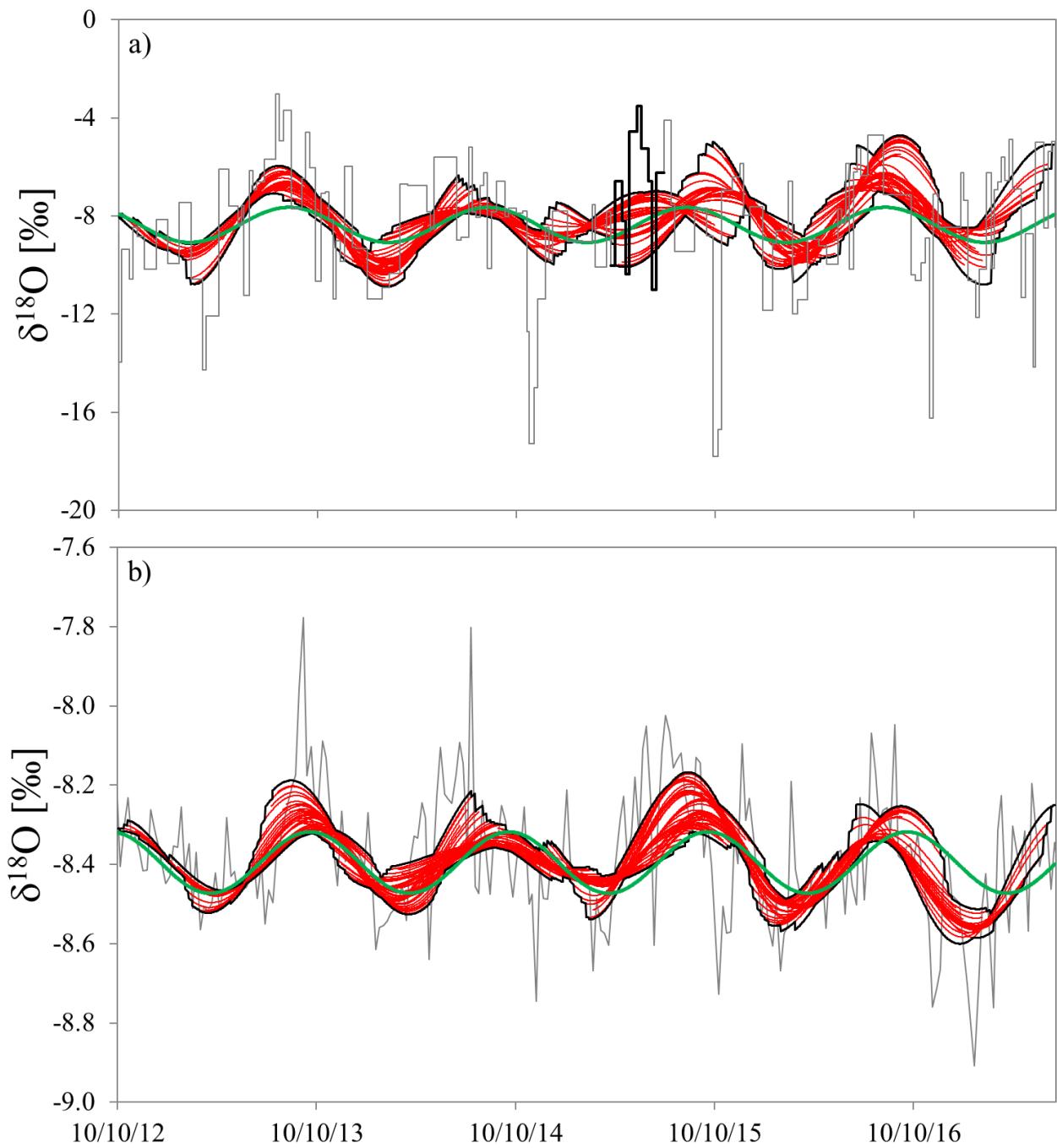


Figure 3. Sine waves (red lines) were fitted to (a) throughfall and (b) streamflow stable isotope data (grey line) with maximum and minimum values at each point in time (black enveloping curve). In comparison a single sine wave was fitted to the complete data set for both throughfall and streamflow (green lines). The omitted precipitation isotope values of the 2015 European summer heat wave are shown in panel (a) with bold black lines.

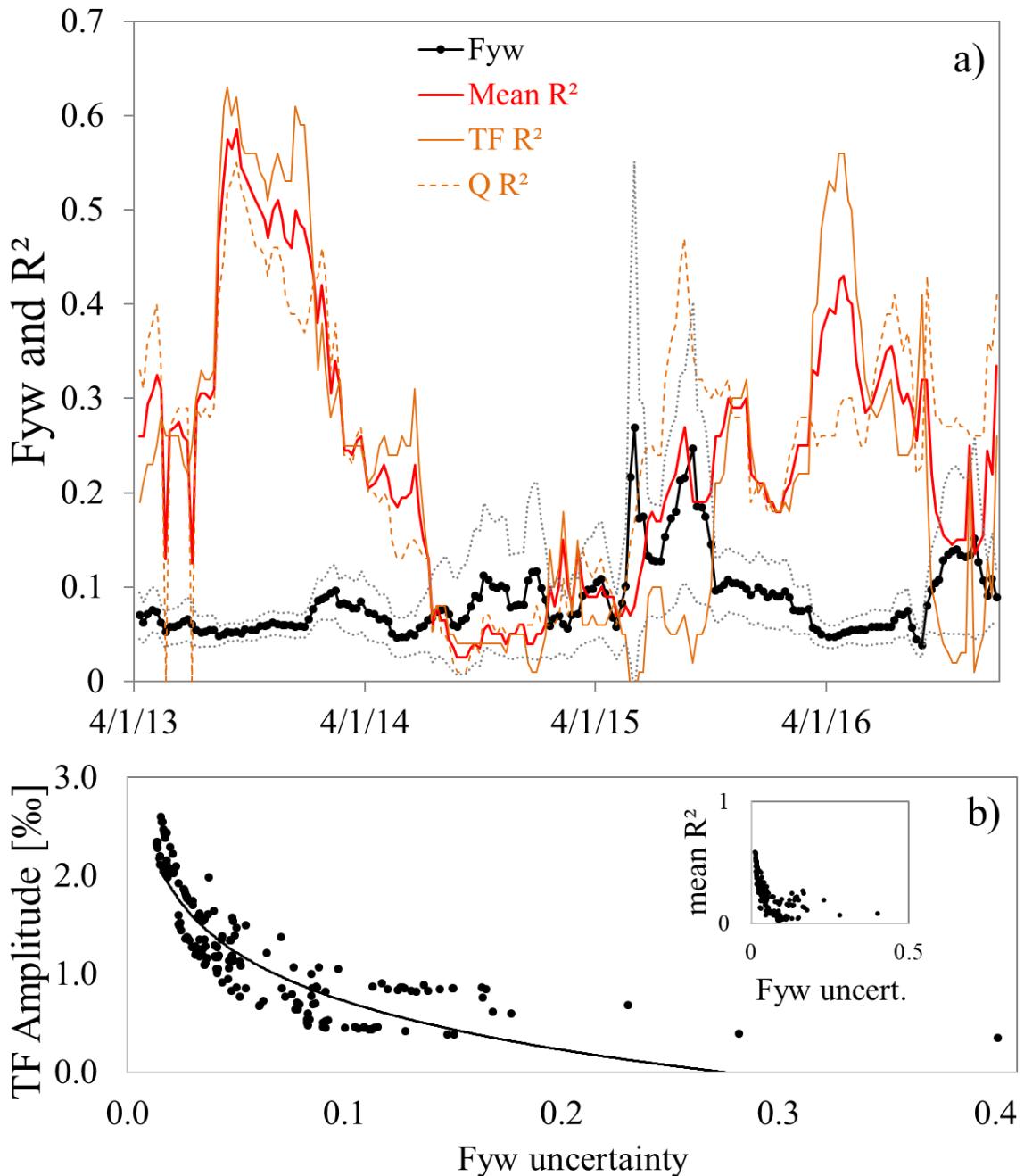


Figure 4. (a) Fyw results and their uncertainty (black and grey lines) plotted against R^2_{adj} for throughfall (TF R^2 , solid orange line) and runoff (Q R^2 , dashed orange line) sine wave fits and their average (Mean R^2 , red line). All values are shown at the midpoint of the respective year they are valid for. Panel (b) shows throughfall amplitudes (TF Amplitude) versus the Fyw uncertainty. The regression equation is $TF\ Amplitude = -0.716\ ln(Fyw\ uncertainty) - 0.9236$ with an R^2 of 0.79. A similar comparison between runoff amplitudes and Fyw uncertainty showed no relationship (R^2 of 0.04, not shown). The inset shows the Fyw uncertainty against mean R^2_{adj} values of streamflow and precipitation.

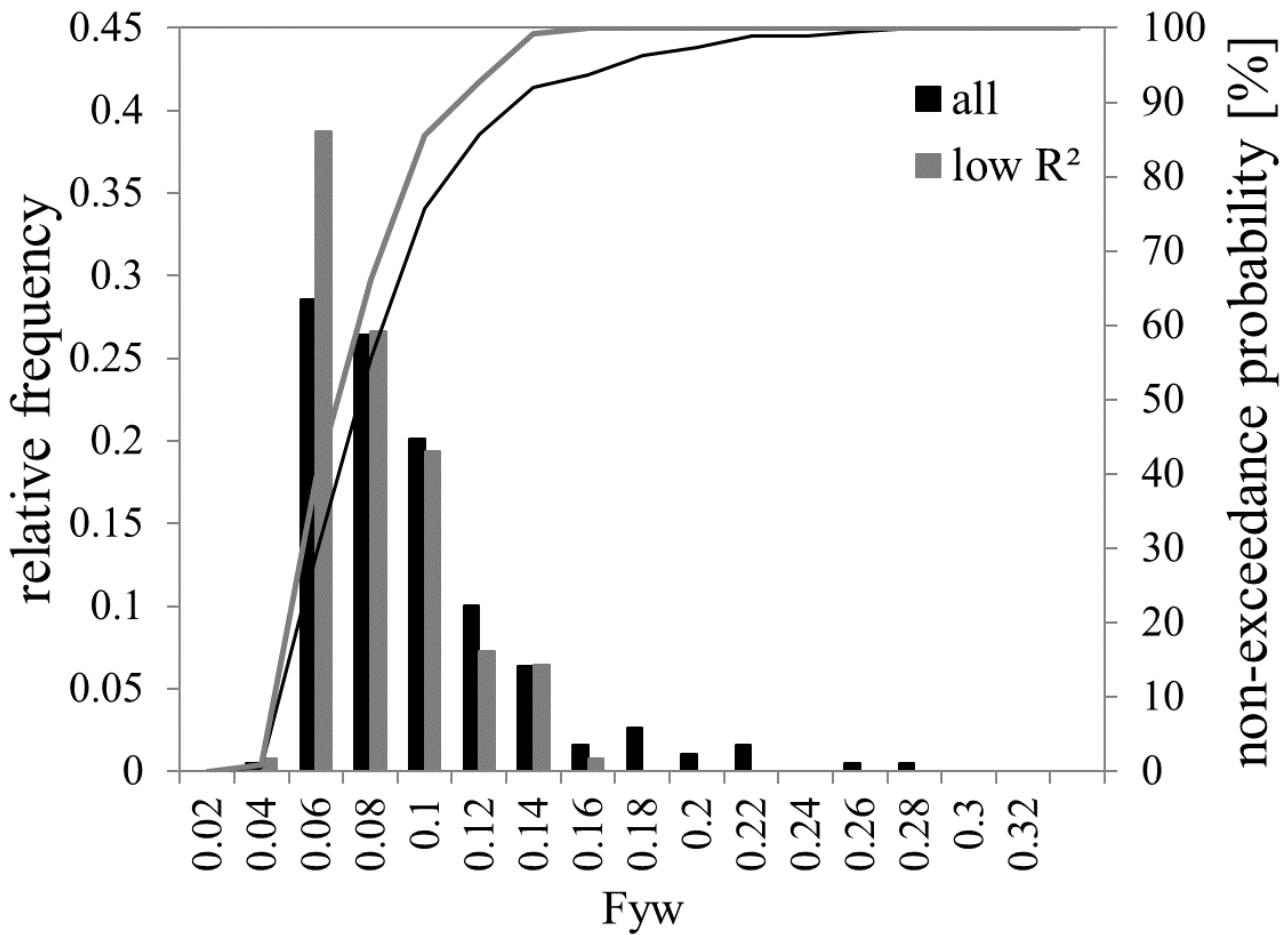


Figure 5. Histograms and cumulative distribution functions of all F_{yw} results (black) and of the results when the low R^2_{adj} period is left out (low R^2 , grey).

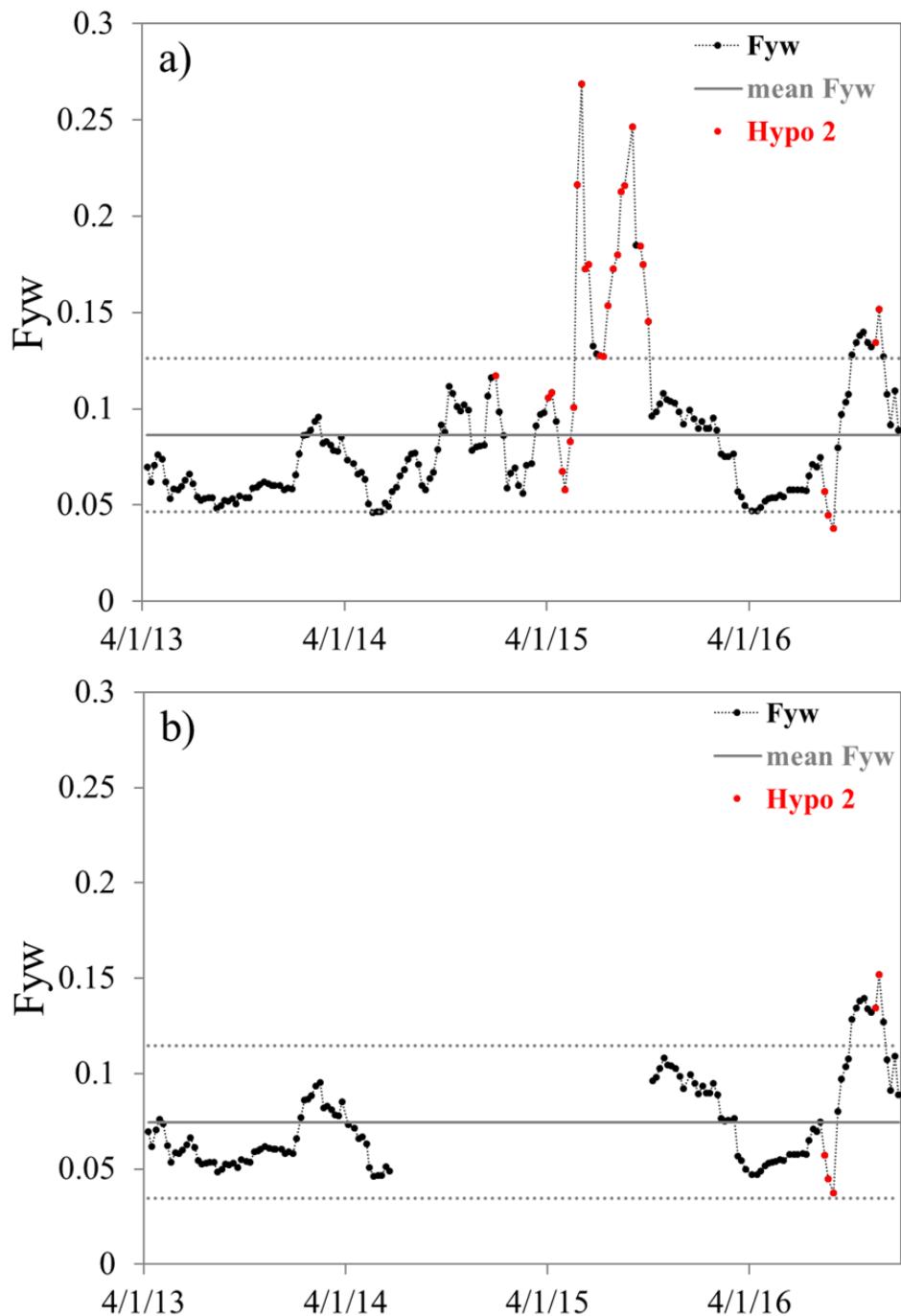


Figure 6. Fyw compared to the mean Fyw (solid grey line) and a ± 0.04 margin around it (dotted grey lines) to test hypothesis 1 (90% of all Fyw results are within the mean Fyw ± 0.04). Red data points are periods where within four weeks Fyw differed more than 0.04 (testing hypothesis 2). Once all data was used (panel a) and subsequently data of the low R^2_{adj} period between July 2014 to October 2015 was left out (panel b).

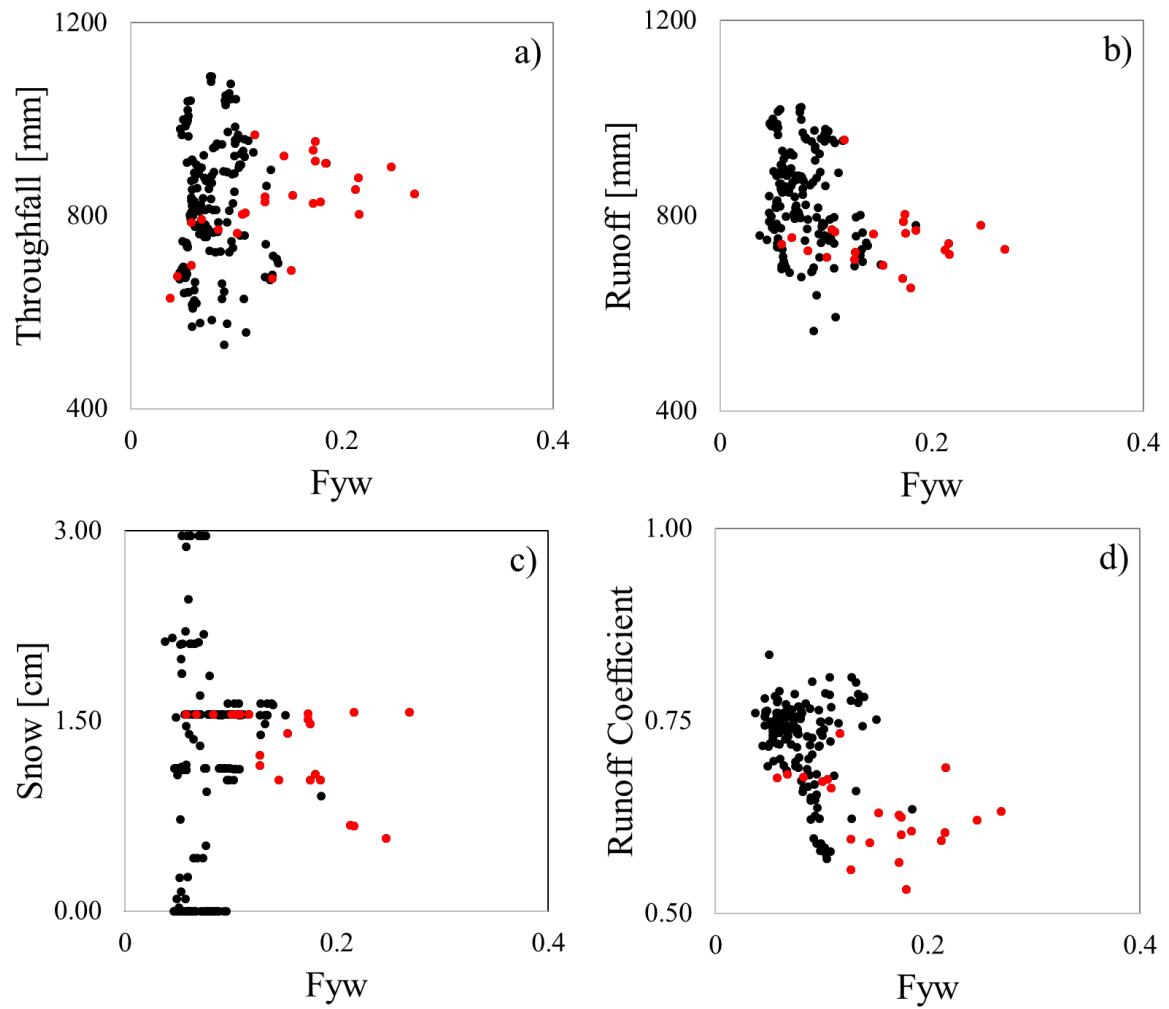


Figure 7. Fyw plotted against hydrometric data (red and black dots): a) throughfall volumes, b) runoff volumes, c) snow height, d) the runoff coefficient. Red dots are data points where hypothesis 2 was rejected ([Fyw does not differ more than \$\pm 0.04\$ within four consecutive weeks](#)).

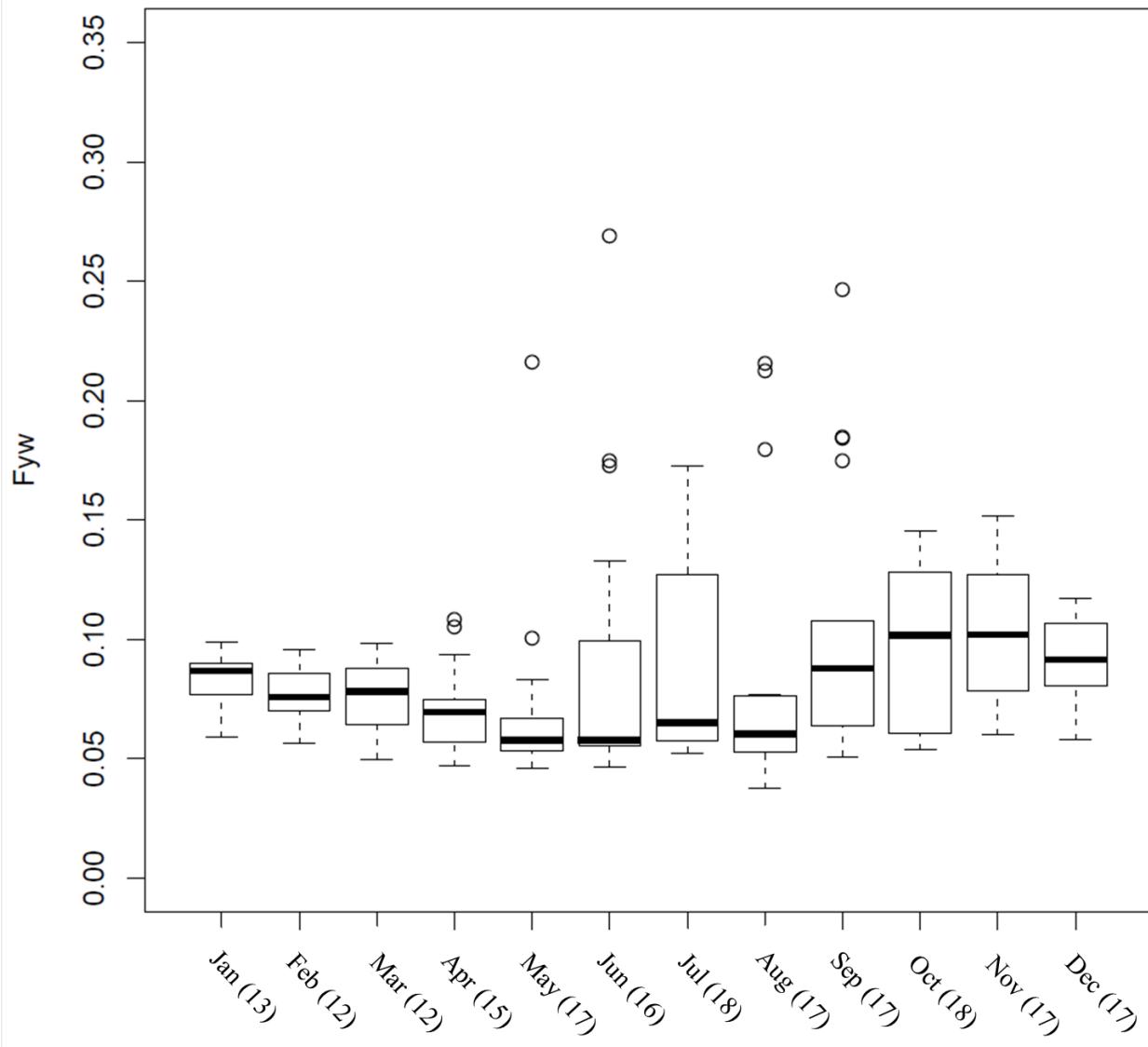


Figure 8. Testing hypothesis 3 (Fyw centred around a specific month does not differ more than ± 0.04 within this month): Boxplot of all Fyw results of a specific month. Whiskers are the upper and lower 1.5 interquartile range and circles are outlier values. The number of data points for each month is given in the brackets on the horizontal axis.