Reply to Referee #1

The manuscript “Exploring the interplay of new and young water fractions with hillslope topography in a subtropical headwater catchment” measures the stable isotope composition ($d^{18}O$ and $d^{2}H$) of precipitation, seepage, and stream water. They use the stable isotope composition of these pools of water to determine new and young water fractions – a commonly used metric in hydrology to understand how water moves through landscapes. Determining the new and young water fractions of hillslope seepages is interesting, as many studies primarily focus on new and young water fractions in streamflow.

Reply: We appreciate Reviewer #1 for providing valuable comments on this manuscript. Reviewer #1 acknowledged the fundamental value of our analysis on the new and young water fraction of hillslope seepages. These comments will benefit our paper significantly.

There are areas of the manuscript that require major revision, namely:

- Contextualizing the study within the existing literature. The manuscript cites only 17 papers and lacks discussion of some of the key studies in this field.

Reply: We agree with Reviewer #1's comment. We will cite more papers on the young water fraction and new water fractions in streamflow. It is important to place our study within the existing literature. We will add a paragraph in the introduction providing background information about new and young water fractions in streamflow and highlight that our study sampled the hillslope seepages. Additionally, we will rephrase our discussion to compare our findings with previous studies.

- Analytical rigor: the analysis of the new and young water fractions lacks key details that are standard for the field. For example, there is no discussion of accounting for precipitation amount or stream discharge in the new/young water fraction calculations. The authors need to clarify if they have calculated volume-weighted new/young water fractions or not, and discuss the implications of however they have accounted for this in their approach. This extends to the cited ranges of Fnew/Fyw – are these volume weighted or not? It’s crucial to state this, so that rigorous comparisons can be made (e.g., one cannot expect to meaningfully compare unweighted Fnew/Fyw in one region to weighted Fnew/Fyw in another region).

Reply: We fully agree with the reviewer's comment. We will recalculate the volume-weighted F_{yw} and F_{new} to meaningfully compare them with other literature values. Additionally, we will state whether these literature values had been weighted or not. However, we only measured the river discharge at site S2 during 2012-2015, and we missed the discharge for 2016. Therefore, we will use a hydrological model (HBV) to estimate the discharge for 2016. For other seepage and streamflow sites, we will estimate the discharge based on the drainage area.

- Relationship between HAND and Fnew/Fyw. I'm unconvinced that the observed relationship between HAND and Fnew/Fyw is significant, especially given the small sample size of n =4 for seepage waters. There are no statistical analyses performed to show that this relationship is meaningful, and I think stating that there is a "threshold" behavior is over-interpreting the data.

Reply: Although there may be considerable uncertainty due to the small sample size, we conducted segmented regression analysis to identify the breakpoint of HAND for F_{yw}, F_{new}, and α. By testing different HAND values, we obtained coefficients of determination and p-values. The results suggest a statistically significant threshold of 10-15 m. We will include this finding in the revised manuscript. However, the small sample size remains a concern. We will rewrite the discussion to emphasize that this "threshold" is a hypothesis and will compare it with other literature on the mixing zones of shallow subsurface flow and deep groundwater.

- Discussion of how the sampling frequency and hydroclimatic conditions have the possibility to impact the results. It is likely that during storms or wetter vs. drier periods, the dynamic between hillslope and stream Fnew/Fyw could change. Some helpful citations related to this topic are:

Reply: We agree with the reviewer's comment. We will test the influence of sampling frequency and conduct a time-variant F_{yw} and F_{new} analysis to examine the influence of hydroclimatic conditions. These two tests will also deepen our study's focus on understanding the influence of topography. We thank the reviewer for providing the following literature.

  - Gallart, F., Valiente, M., Llorens, P., Cayuela, C., Sprenger, M., and Latron, J.: Investigating young water fractions in a small Mediterranean mountain

catchment: Both precipitation forcingand sampling frequency matter, Hydrol. Process., 34, 3618–3634, https://doi.org/10.1002/hyp.13806, 2020

- Gallart, F., von Freyberg, J., Valiente, M., Kirchner, J. W., Llorens, P., and Latron, J.: Technical note: An improved discharge sensitivity metric for young water fractions, Hydrol. Earth Syst. Sci., 24, 1101–1107, https://doi.org/10.5194/hess-24-1101-2020, 2020
- 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.
- 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, J. Hydrol., 541, 952–964,https://doi.org/10.1016/j.jhydrol.2016.08.007, 2016.

- The isotope dataset really should be made available with the paper to download as a supplement or posted concurrently to a data repository to achieve traceability of results.

Reply: We will make the isotope dataset available for download as a supplement to the paper to ensure traceability of the results.

I have provided many more comments for the authors to take into consideration directly on the manuscript. The paper needs considerable revision, but I believe that there is room for it to become a stronger and more rigorous study.

More comments in the pdf

L48-51: These aren't null hypotheses, just hypotheses. Also you need to revise number 3, as written it doesn't make sense/isn't a hypothesis.

Reply: Thank you for the reminder. The text now reads: "*This study posits that if the relationships observed at the catchment scale between topography and hydrology can be extrapolated to hillslope scale, specific hypotheses can be tested*" and "*(3) the relationship between HAND and both F_{yw} and F_{new} is a linear relationship.*"

L56: average slopes across the whole catchment?

Reply: Yes. For clarification, the sentence now reads: "*The average catchment slopes of R1, R2, and R3 are 27.0, 26.6, and 25.4 degrees, respectively.*"

L60: Did you measure actual precipitation or discharge during the study period?

Reply: We have measured actual precipitation and discharge (for R2) from July 2012 to May 2015. These hydrometric data partially overlap with our sampling period (June 2014 to May 2016). To calculate the weighted F_{yw} and F_{new}, we will run a simple hydrological model (HBV) to fill in the missing daily streamflow data for June 2015 to May 2016. We will include this information in the Materials and Methods section.

L62: delete "such".

Reply: revised as requested.

L62: seasonal cycle of what? temperature, precipitation amount, precipitation isotopes? please be more specific.

Reply: It refers to the precipitation amount. We have specified this sentence: "*The seasonal cycle of precipitation amount is distinctly divided into summer and winter periods, with the former extending from May to October and the latter from November to April.*"

L66: Biweekly can be a tricky word here, because some folks interpret it as twice a week, while others interpret it to mean twice a month. I'd suggest saying "twice-weekly" or "twice-montly"/"fornightly" to clarify your frequency of sampling.

Reply: Thanks for your comment. We have replaced "biweekly" with "twice-monthly" here.

L68: delete "meticulously".

Reply: revised as requested.

L76-77: How have you determined that d2H is more accurate for calculating Fnew and Fyw? Is this due to your measurements of d2H being more accurate than d18O, or due to something that you've determined from the literature. Please elaborate and provide a citation as to why you think d2H is more accurate.

Reply: We selected δ^2H for estimating F_{yw} and F_{new} due to the higher precision of δ^2H measurements and the similar estimations of F_{yw} and F_{new} from both $\delta^{18}O$ and δ^2H. The text now reads: "*Our study focused on δ^2H data to more accurately calculate the fractions of young and new water (F_{yw} and F_{new}) because (1) we have more precise measurements for δ^2H compared to $\delta^{18}O$ (as noted by Rodriguez et al., 2021), and (2) the strong correlation between $\delta^{18}O$ and δ^2H suggests that similar estimates for F_{yw} and F_{new} can be obtained from either isotope.*"

Rodriguez, N. B., Pfister, L., Zehe, E., & Klaus, J. (2021). A comparison of catchment travel times and storage deduced from deuterium and tritium tracers using StorAge Selection functions. Hydrology and Earth System Sciences, 25(1), 401-428.

L78: This is not the seminal citation for young water fraction calculation, should be Kirchner et al., 2016 a, b.

Reply: Thank you for the comment. The text has been revised as requested.

L114: try to be more specific with language here, you're refering specifically to the seasonal cycle amplitude. "variation" could mean anything. I'd swap out "variation" with "amplitude".

Reply: Thank you for the comment. The text has been revised as requested.

L114: I think you mean larger, not smaller? It looks like S2 and S4 have larger seasonal cycle amplitude that the streams, while S1 and S3 have smaller or the same amplitudes as streams.

Reply: Sorry for the mistake. We meant that seepage has a larger amplitude than the streams. The text has been revised as requested.

L115-116: Rather than stating larger and smaller fluctuations, tell the reader what the exact amplitudes are.

Reply: revised as requested.

L116-117: Awkward phrasing, you might say "discussed below" instead.

Reply: We revised this sentence. Now the sentence reads: "*Intriguingly, S4 and S3 are geographically proximate, a detail that will be discussed below.*"

Fig. 2: As these figures are currently drawn, precipitation, stream and seepage water are only differentiated by color. I recommend using different symbols for the three water types (i.e., square, triangle, circle). Also, you could improve the figure design by clearly differentiating between seepage on the left and streams on the right. Readers might not quickly be able to distinguish the meaning between "S" and "R". You could put label at the top of each column of figures saying "seepage" and "rivers" or something like that.

Reply: Thank you for providing detailed comments on Fig. 2. We will revise it as you requested.

L124-125: You only present the phase data in the tabel below, and it is in radians. Much better would be to convert the phase to day of year and calculate the lag between the stream/seepage phase and the precipitation phase.

Reply: We agree with the reviewer's comment. We will revise Table 1 and include the phase value in the sentence after performing the volume-weighted estimation.

L125-126: The range of offset between the seepage and the stream water is quite similar:
Seepage: -70.2 – -74.1
River: -72.1 – -74.2
Have you done any statistical analysis to justify being able to say that the offset term actually shows greater variation in the seepage water than the stream? I'm not convinced that these two ranges are significantly different, especially given the small sample size of n = 4 (seepage) and n = 3 (streams).

Reply: Sorry for the oversight. We conducted a t-test for the offset value between seepage and streamwater, and the result was statistically insignificant (p-value = 0.28). Therefore, we will revise this sentence based on the statistical analysis.

Table 1: Rather than reporting the RMSE of the sine wave fit results, it would be much better to report the following:
amplitude ± SE
Fyw ± SE
Fnew ± SE
See table 3 in von Freyberg, et al., 2018 as a good example of how to robustly report these data.

Reply: We will revise Table 1 following the format of von Freyberg et al. (2018).

Table 1: Too many significant figures on the amplitude. Should be one significant figure, given that the precision on the d2H is only 0.8 ‰

Reply: Thank you for your comment. The significant figure has been revised to one.

Table 1: Convert this into a more reader-friendly format, like days of the year, and state the units once you've converted it.

Reply: Thank you for your comment. We have revised the unit to days of the year and stated the unit accordingly.

Table 1: Too many significant figures on the offset. Should be one significant figure, given that the precision on the d2H is only 0.8 ‰

Reply: Thank you for your comment. The significant figure has been revised to one.

L137: I think you need to eliminate the phrase "different systems" here. Instead, specifically describe where you think the seepage vs. stream waters could be sourced from. For example, do you mean shallow soil vs. deeper groundwaters? Also, I think it may be a better idea to move this sentence and the one beneath it to the discussion and clearly elaborate on the details. As it reads now, these two sentences are vague and overly general.

Reply: We thank the reviewer for suggesting a more direct way to describe seepage versus stream water. We will revise our terminology to use "shallow system" vs "deeper system" accordingly. Additionally, we will move the two sentences to the Discussion section 4.2 and make them clearer.

Fig. 3: Again, consider different symbols for them rather than just color differences.

Reply: revised as requested.

L144: delete “s”

Reply: revised as requested.

Fig. 4: Why are the gray dots excluded from the regression analysis?

Reply: New water fraction refers to the fraction of water younger than a specified time, indicating the cumulative fraction of younger water. Our regression analysis aims to calculate the cumulative transit time distribution (TTD). In the context of TTD, the new water fraction should be greater than or equal to that of the previous period. However, since the gray dots are lower than the highest F_{new}, we suggest that the δ^2H cannot identify water age older than the period length at the highest F_{new}. Therefore, we excluded the gray dots from the regression analysis. We will add the above text to Results section 3.2.

L158: You describe this section as exploring the relationship between topography and Fnew/Fyw, yet you only provide a discussion of the HAND metric. Are there any other metrics you can include here? If not, change the title of this section to state that it only focuses on HAND.

Reply: Yes, we have other topographic metrics that can be included here. We will perform a correlation matrix analysis to explore additional relationships between topography and F_{yw} or F_{new}.

L162-163: I do think that describing this as a threshold is overinterpreting. There are only 4 data points, so you really need to be careful.

Reply: We fully understand the reviewer's concern regarding the four data points. Therefore, we will perform a segmented regression analysis to identify any potential threshold. Additionally, we will be very careful in describing this potential threshold and will cite related works about the mixing area of shallow water and deeper water to support the potential threshold.

L183: has

Reply: revised as requested.

L184-185: You need to discuss whether the Fyw you've calculated for your own data and reference here from the literature are flow-weighted volumes or not. This will greatly affect the ranges of Fyw.

Reply: Thank you. We will recalculate the flow-weighted F_{yw} for our data and identify whether the literature values are flow-weighted volumes or not.

L188: you should also cite, at a minimum:

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.
Lutz, S. R., Krieg, R., Müller, C., Zink, M., Knöller, K., Samaniego, L., & Merz, R. (2018). 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

Reply: Thank you. We will also cite other related works.

L188-191: A more apt citation here is:
Burt, E. I., Coayla Rimachi, D. H., Ccahuana Quispe, A. J., Atwood, A., and West, A. J.: Isotope-derived young water fractions in streamflow across the tropical Andes mountains and Amazon floodplain, Hydrol. Earth Syst. Sci., 27, 2883–2898, https://doi.org/10.5194/hess-27-2883-2023, 2023.
This paper discusses the range of Fyw in multiple streams across the the Andes mountains to Amazon floodplain transition, which is 5–52 % for flow-weighted Fyw

Reply: Thank you. We will also cite other related works.

L201: There are many other factors that could drive differences between your observed Fnew/Fyw and the Burt et al., dataset. Be more thorough.

Reply: We compared our environmental settings with those of Burt et al. and found similarities in altitude, soil type, and lithology. We suggest that differences in precipitation amount and regional storage could explain why our F_{yw} and F_{new} values are larger than those of Burt et al. Our annual precipitation is around 4,900 mm, compared to 4,100 mm in Burt et al.'s study. Higher precipitation might increase F_{new} and F_{yw}. Additionally, within a 5 km radius, Burt et al. have higher mountains (3,800 m) compared to our study area (3,400 m). Higher mountains indicate larger storage and older water age, while lower mountains correspond to smaller storage and younger water age. We will include this information in the discussion.

L202: Awkward phrasing.

Reply: We have revised this sentence. Now the sentence reads: “Revisiting the comparison between the ages of seepage and stream water, the

observation that seepage water generally has a higher F_{yw} than stream water suggests that seepage water follows shorter flow paths."

L202-203: Are you actually able to calculate flowpath length? You should be able to do that for the stream and for the seepages.

Reply: Yes, we have calculated the flow path length. We will include this information along with other topographic metrics in the results and discussion sections.

L227: again, careful with overinterpreting.

Reply: Thank you for the reminder.

L230-234: so basically what you're saying is that at high elevations, seep water is sustained by groundwater, then at more mid-elevations, seeps are sustained by younger water, then at the lowest elevations, there is a switch back to older water?

Reply: Yes, the reviewer basically understands our point. High elevation seepage is sustained by shallow groundwater, while seepage at the lowest elevation is sustained by a mix of young water and a large amount of deep groundwater. Mid-elevation seepage is sustained by young water due to faster underground flow paths through fractured lithology. We will revise this sentence and Fig. 6 for easier reading.

L234-235: this isn't actually a contrast! you're stating that at low elevations, you found low Fyw, while at mid/higher elevations, Fyw was higher. if these two studies found low Fyw at low elevations, and higher Fyw at higher elevations, then you cannot say your results contrast with theirs.

Reply: We agree with the reviewer's comment. The different patterns in Fyw between their study and ours suggest that their lysimeter water (depth <1 m) reflects very shallow soil water, while our seepage reflects shallow groundwater (depth >1 m). We will add this explanation to the discussion.

L235-237: how do you know that they sampled at lower hand levels? I think this sentence is problematic and needs to be removed, while the sentence prior needs to be reworked.

Reply: We cannot accurately calculate their HAND values, so we will remove this sentence.

L244: I wouldn't say that traditional perspectives in hydrology tend towards linearity. There are rarely linear relationships in hydrology.

Reply: We agree with the reviewer's comment and have removed this sentence.

L274: I'm concerned about the lack of citations here. You've only cited about 15 papers – well-written studies should include at least ~3 times as many citations.

Reply: We agree with the reviewer's comment. We will cite more papers for this study and appreciate you providing these classic works on young water fraction. In response to your major comment, we will add paragraphs to highlight the significance of our work compared to other literature.