Dear Editor,

Here we provide a revised version of our manuscript. The revised paper contains the necessary changes to address the reviewers' and editors' comments on our previous submission. Below we provide a detailed point-by-point response (in bold) to those comments (*in italics*).

Sincerely,

Marius Floriancic, Michael Stockinger, James W. Kirchner & Christine Stumpp

---

Reply to RC 1:

Comments on 'New water fractions and their relationships to climate and catchment properties across Alpine rivers' by Floriancic et al., 2023.

General comments ：

Understanding runoff generation processes in high mountain catchments is important as it is water tower for providing water for drinking, agriculture, and hydropower production. The main findings in this manuscript reveal that Alpine rivers tend to have larger new water fractions at low elevations, in flatter terrain and while in smaller catchments with large forest cover. The findings are interesting to scientific community, however, is also controversial in different perspective. As claimed by the authors, there seems less studies linking new water fractions to hydroclimatic drivers and physical catchment properties across small to very large basins. We know that the application of chemistry tracing methods is usually underlain by many basic assumptions, for instance, applicants must account for the heterogeneity in rainfall isotope referring to Pinder and Jones (1969). Hence, the methods adopted as well as the conclusions in this manuscript should be carefully discussed to justify the rationality of the methods and the reliability of this conclusions.

Thank you. In the manuscript we use two methods to assess the fraction of more recent precipitation in streamflow (calculation of "young water fractions" and "new water fractions"). For the assessment of young water fractions, we fit a sinusoidal curve to the typical long-term seasonal precipitation cycle; for the assessment of new water fractions we use the recently developed ensemble hydrograph separation method. Our results reveal the typical long-term averages of young and new waters in stream and do not reveal any information on short term variations as introduced by heterogeneities outside the bounds of the typical seasonal precipitation cycle. Unfortunately, measurements on isotope ratios in precipitation are rarely

available at high spatial resolution. Therefore, we used the available gridded precipitation dataset provided by Nelson et al (2021), where factors like altitude effects are considered. In the revised version, we added a discussion section on “Limitations in data availability” where we describe the limitations in more detail and compare different data products of the available precipitation isotope data (shown now in the supplement – Figures S5 & S6).

Importantly, both young water fractions and new water fractions do not depend on the absolute values of either precipitation or streamflow isotopes, but only on their variations over time (their seasonal cycles for young water fractions, and their month-to-month fluctuations for new water fractions). Thus any offsets introduced by terrain effects will have little or no impact on calculated young water fractions and new water fractions.

Moreover, as there are cross-correlations between many catchment properties, and hydroclimates potential drivers, the explanations that which may be a first-order control on new water fractions should be further verified with more cautions as indicated by the authors. For example, the authors found that high fractions of new water (Fnew) were more likely in small catchments, at low elevations, with small total relief and larger forest cover, and following months with high precipitation. However, they also found that Fnew tended to decrease downstream, from smaller headwaters to larger river basins, in which it can be inferred that altitude, relief, slope gradient, and even precipitation (see in Ménégoz et al., 2020) all decrease with altitude from smaller headwaters to larger river basins in Alps. A reasonable explanation to the issue is that it is the impact of water storage in lakes and reservoirs, as well as the potential effects of anthropogenic flow regulation when moving from the headwaters downstream. But it seems a little bit contradiction in the context. So, the authors should provide more deepen and persuasive discussions for this key conclusion in the manuscript. I wonder whether water stored as snow or glaciers in high altitude basins have led to relatively lower Fnew in headwaters, which is quite inconsistent with our intuition about runoff generations. Cause in alpine basins more areas with naked rocks or thin soils could be observed in high altitude regions as erosion rates increase with local relief (Heimsath et al., 2012), thus, more rainfall directly drains to drainage networks as new fractions in high altitude areas?

Thank you. In the revised version of the manuscript, we now provide a more detailed discussion of the dependencies between different catchment descriptors. Of course, we acknowledge that many catchment characteristics are correlated; this was the reason why we included a cross-correlation matrix in the manuscript (see Figure 9). Regarding the comment about the importance

of snow in higher-elevation catchments of the Alps we'd like to point out that this issue is not yet fully resolved; we reported this in the original version of the manuscript and now extended the discussion on this issue. It has not been fully understood yet to what extent the lower fractions of more recent precipitation found in Alpine streams is due to snow processes or due to the large elevation gradient (and thus large subsurface storage volumes). With our sample of catchments (as in many other studies before) we cannot test these hypotheses independently (catchments with large topographic gradient are also dominated by snow across the European Alps). Still, we adopted the discussion to point out that the smaller fractions of more recent precipitation in the higher Alpine catchments have multiple potential explanations but need to be considered in another study based on a targeted dataset.

In addition, Fyw as defined by the authors is the ratio of seasonal amplitudes of sinusoidal fits to precipitation and streamflow isotope time series. However, as shown in Figure 2, the seasonal fluctuations in streamflow isotopes among all the basins are quite similar. Therefore, Is Fyw only related to the corresponding fluctuations in precipitation isotopes? We have known that interannual variations of precipitation isotopes have a significant impact on the young water fraction (Gou et al., 2023; Dai et al., 2022), which raises doubts about the results. Could it be consistent in the results obtained by using isotopic data with different time series lengths? Furthermore, the sampling frequency of precipitation isotopes also has a significant impact on the young water fraction (Gallart et al., 2020; Stockinger et al., 2016). So, is it reliable to estimate Fyw through using monthly-scale data?

The young water fractions can only be calculated as the amplitude ration between multiple years of precipitation data AND streamflow data; splitting of these data on an interannual basis should never be done as it will lead to flawed results (see the descriptions in Kirchner 2016). This was confirmed by the case studies of Gallart et al. (2020) and Stockinger and Stumpp (2024), as their studies found high uncertainty associated with Fyw derived from one-year precipitation and runoff data, and Stockinger and Stumpp (2024) suggest a minimum time series length of six years for a robust, long-term Fyw result. Also, it is important to point out that the sampling frequency does not impact the results, as long as the samples are bulk precipitation samples from the entire month and the isotopic signals are volume-weighed by the precipitation amount (which they are in the case of our study). The Stockinger et al. 2016 and Gallart et al. 2020 results do not indicate sensitivity on the sampling frequency of precipitation but rather of streamflow, and in any case

their results are primarily an artifact of the under-representation of high streamflow in the low-frequency sampling procedures that they used. Furthermore, since the precipitation and streamflow timeseries of all catchments are in monthly intervals, streamflow is biased towards lower streamflow values in all catchments. Thus Fnew at high flows can not be reliably estimated from monthly data, however we can infer the relationships between Fnew and hydroclimatic and physical catchment descriptors. It is important to point out that several published studies examine Fyw but do not use the method outlined by Kirchner 2016 as it was intended (and benchmark tested); thus, conclusions drawn from these studies should not be interpreted as reflecting on the original Fyw method.

In general, the manuscript needs more deepen discussions, and the authors should provide more persuasive evidences to clarify the first order control of runoff generation across headwaters to down streams for credible conclusions.

Thank you. We present all dependencies of hydroclimatic variables and physical catchment properties in Figure 9, and we also adapted parts of the discussion to stress the importance of the potential interdependencies of certain catchment characteristics i.e., we report the potential dependencies in the respective parts of the discussion and added a notice of caution for interpretation in the conclusions.

Specific comments: (L means lines)

L121-131: added how precipitation varies with altitude.

We present the correlation of mean, max and min elevation and annual, summer and winter precipitation in Figure 9, whereas correlations between elevation and precipitation are weak overall.

L140-150: the accuracy of the monthly gridded precipitation isotope reanalysis database Piso.AI should be evaluated before adopted in the study regions.

While we agree that the isotope reanalysis product Piso.AI is not the perfect, unfortunately it is the only gridded dataset that can be used for such large catchments. The use of point measurements is not reliable for such large basins, and other reanalysis methods (i.e., Seeger & Weiler 2016) yield similar results to Piso.AI. The gridded data and point measurements within our study area have been evaluated in great detail by Nelson et al. 2021. In addition, in the supplement of the revised manuscript we now show a comparison of data derived from the approach of Seeger & Weiler 2016 and the data from Nelson et al. 2021 for selected catchments (Figures S5 & S6), confirming that both approaches yield similar results.

L258: The catchment areas range from 29 km2 to 103'946 km2. Such huge difference in area would lead to quite different patterns of rainfall and discharge distribution in different catchments which dramatically impact isotope fractions in downstream rivers, and weaken the basic assumption for isotope estimations, and there is especially large spatial heterogeneity of rainfall and discharge concentration across a catchment with larger areas. So how do you calculate monthly precipitation isotope particularly in catchments with larger area (e.g., >1000 km2)?

As outlined in the manuscript, we use the gridded dataset of Nelson et al. 2021 that is to date the most reliable available product that allows the estimation of precipitation isotope ratios for larger basins. We emphasize this now in the revised manuscript and discuss potential uncertainties in the additional discussion section 4.1.

Technical corrections:

(1) In figure 1 Gauge names should be provided and drainage areas should also be added and listed in table 1.

We added gauge names in Figure 1 of the revised manuscript. Drainage areas are already available in Table 3.

References:

Dai, J., Zhang, X., Wang, L., Luo, Z., Wang, R., Liu, Z., ... & Guan, H. (2022). Seasonal isotopic cycles used to identify transit times and the young water fraction within the critical zone in a subtropical catchment in China. Journal of Hydrology, 612, 128138.

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

Gou, J., Qu, S., Guan, H., Shi, P., Zhang, Z., Yang, H., ... & Han, X. (2023). Seasonal variation of transit time distribution and associated hydrological processes in a Moso bamboo watershed under the East Asian monsoon climate. Journal of Hydrology, 617, 128912.

Heimsath, A., DiBiase, R. & Whipple, K. Soil production limits and the transition to bedrock-dominated landscapes. Nature Geosci 5, 210–214 (2012). https://doi.org/10.1038/ngeo1380.

Ménégoz, M., Valla, E., Jourdain, N. C., Blanchet, J., Beaumet, J., Wilhelm, B., Gallée, H., Fettweis, X., Morin, S., and Anquetin, S.: Contrasting seasonal changes in total and intense precipitation in the European Alps from 1903 to 2010, Hydrol. Earth Syst. Sci., 24, 5355–5377, https://doi.org/10.5194/hess-24-5355-2020, 2020.

Pinder, G., Jones, J., 1969. Determination of the ground-water component of peak discharge from the chemistry of total runoff. Water Resour. Res. 5 (2), 438–445.

Stockinger, M. P., Bogena, H. R., Lücke, A., Diekkrüger, B., Cornelissen, T., & Vereecken, H. (2016). Tracer sampling frequency influences estimates of young water fraction and streamwater transit time distribution. Journal of hydrology, 541, 952-964.

Reply to RC 2:

The work of Floriancic et al. (2023) addresses the estimation of new water fractions with the ensemble hydrograph separation in 32 catchments to investigate their relationships to climate and catchment properties across Alpine rivers. The work is novel since, according to my knowledge, there are no previous studies that have related "new" water fractions with hydroclimatic and physical *properties. Nevertheless, there are some major issues that should be taken into consideration before considering the manuscript for publication in HESS.*

Thank you for the positive feedback.

Main comments:

1) You have used monthly streamflow isotope for 12 Austrian sites and 8 Swiss stations. Moreover, you have used monthly gridded precipitation isotopes from Nelson et al. 2021. It is well known that the sampling resolution affect the young water fraction estimates (Gallart et al., 2020; Stockinger et al., 2016). When you compare your results with those of previous studies you should include the effect of sampling resolution since some discrepancies between your results and previous results can be also due to the low sampling resolution of isotope data used in your study. Additionally, by considering this sampling resolution you consider as "new" some water that is younger than 1 month. This threshold age is of the same magnitude of the typical threshold age for young water (2-3 months), as also reflected by similar Fyw and Fnew for some catchments very close to the 1:1 line in Figure 3. Accordingly, I would underline this when you state that similar relationship to catchment properties is obtained with young water fractions. This happens since they have a similar threshold age. I would add in the conclusion that in future studies is necessary to investigate the relationship between new water fraction and climate/catchment properties by using high-resolution (e.g., daily) isotope data (and accordingly much lower threshold ages). Please, specify in the manuscript title the age of new water: as soon as I read the title I thought that you have investigated event water with a threshold age of the order of few days.

Thank you. We changed the title to "Monthly new water fractions and their relationships to climate and catchment properties across Alpine rivers" to avoid this misconception and added that the ensemble hydrograph separation method (calculation of F_{new}) in the presented dataset allowed us to calculate the amount of streamflow that is younger than one month. Thus, in the manuscript we consider all water that is "new" to be younger than one month. However, it is important to mention that the sampling frequency of precipitation isotopes does not impact the *<u>young</u>* water fraction (F_{yw}) results, as long as the samples are bulk precipitation samples from the entire month and the isotopic signals are volume-weighed by the precipitation amount (which they are in the case of our study). In the Gallert and Stockinger studies, the main difference between the high-frequency and low-frequency sampling was that the low-frequency sampling

under-represented high-streamflow conditions. Thus, given the well-known discharge sensitivity of Fyw (which itself is a key characteristic of catchment transport behavior), when high-streamflow conditions were sampled more often, Fyw understandably increased. Thus, the issue is not a bias in the Fyw method, but instead under-representation of high-streamflow conditions in the low-frequency sampling used by Gallert and Stockinger.

2) About your conceptual scheme reported in Figure 10. I think that it is biased from the use of very large catchments. Indeed, some catchments you have used in your dataset, due to their large extension, cover the entire landscape (reported in Fig. 10) from high to low elevations (in some catchments the elevation difference is higher than 4000 m) including both steeper and gentler terrain, largely heterogenous land cover etc. Thus, the fraction of new water you estimate is the result of hydrological processes occurring across the entire catchment area. Indeed, Fnew could depend more on processes occurring at higher elevations or more on processes occurring in plain areas (e.g., with influence of natural and/or artificial lakes). In this regard I would suggest looking at your results by not considering very large catchments and to better detail your conceptual scheme by subdividing it in two/three different elevation ranges (e.g., a possible elevation threshold could be 1500 m a.s.l. since previous studies (e.g., Ceperley et al. 2020) have found a change in hydrological regime above this threshold).

Thank you. It was one main aim of this study to compare catchments of different scales to understand to which extent fractions of more recent precipitation are dominated by catchment characteristics in a wide range of different-sized catchments. Certainly, larger catchments integrate over many processes, and we do not normally know which features predominantly control Fyw and Fnew. Still, our specific analysis on how Fnew changes along large streams indicates the dampening effects with increasing catchment size. While differences between catchments exist (also in studies that only consider small catchments) we think it is interesting to see that most of the established relationships between young & new water and catchment descriptors hold both for small and for large catchments. Therefore, the figure appropriately combines our findings across smaller and larger catchments.

3) The database for geology is very coarse, but I guess nothing is available. Is it possible to better discuss its implications?

Unfortunately, this is true for many studies, including ours. More systematic databases for geological information would be a big help, and we support any efforts in this direction. Even the CAMELS-CH dataset, only covering Switzerland, does not have better-resolved (and homogenized) geology information.

4) In the paper you cited in relation to quaternary deposits, Gentile et al. (2023) in this same journal, it has been discussed the relation between baseflow and Fyw. Can you compare their results with yours, although the different size of the catchments?

Thank you for this remark. We show the relation of F_{yw} and q_{95} in the supplementary material and found similar negative correlations as Gentile et al. 2023. We added the refence to Gentile et al (2023) to the discussion.

5) Less important (the paper structure is a personal choice) : you have chosen to separate results from discussion. Nevertheless, I think that in many parts of the "Results" section you have also discussed the results. Moreover, in the "Discussion" section there are many parts that are redundant since you recall the results or methods sections. Accordingly, I would suggest changing the paper structure in "Results and Discussion". This will shorten the manuscript length and improve its readability.

Thank you, we discussed this internally prior to the first submission and decided to go for a separate results and discussion section for the manuscript. We now reconsidered this choice but decided to stick with the separated results and discussion format, to be able to discuss all potential drivers in a step-wise manner. We carefully edited the manuscript to minimize the redundancies between results and discussion.

Specific comments:

Figure 1: please, add the catchment areas on the Map by indicating the Site code.

We added the site code in Figure 1 in the revised version of the manuscript.

Lines 216-221: Please, explain better how you have split your dataset. It is not fully clear to me.

We updated the explanation of splitting the dataset accordingly.

Lines 320-321: Have you considered a delayed input or a direct input (von Freyberg et al. 2018) for estimating new water in snow-dominated catchments? I think that should be discussed more about the role of snowpack on the estimation of the new water and what are the consequences of choosing a direct or delayed input on new water fraction (similarly to what von Freyberg et al. 2018 did for young water fraction).

Thank you for this remark. In the revised version of the manuscript, we updated the discussion on the impact of delayed snowmelt on young and new waters.

Lines 505-508: Are there previous studies that support this result? If not, please provide an explanation to this counterintuitive result.

We now explain the potential reasoning (i.e., cross-correlation with elevation and steeper slopes) for this in more detail in the revised version of the manuscript.

Table 3: Add the fields "min elevation" and "max elevation"

We added "min" elevation (the location of the gauge) in Table 3, "max. elevation" can be inferred from the already available column "elevation difference".

Figure 5b: Please add the site code in panel b. If possible, think about a possible alternative representation of Figure 5 since it is really intricate.

Thank you, we understand that the representation is not ideal, however upon discussion we did not find a more suitable way to present this. However, we believe that the important overall

message of the figure is that only in roughly half of the catchments F_{new} substantially increases with increasing precipitation, thus we moved panel (b) of the Figure to the supplementary material (now in the Supplementary Material in Figure S2).

Figure 7 & Figure 8: I would suggest to represent rainfall-dominated, hybrid and snow-dominated catchments by using three different colors. This could lead to additional insights to improve the discussion about these results.

Thank you for this suggestion, we updated the Figures accordingly and now show the different precipitation regimes in different colors.