This review was prepared as part of graduate program course work at Wageningen University and has been produced under supervision of Ryan Teuling. The review has been posted because of its good quality and likely usefulness to the authors and editor. This review was not solicited by the journal.

Overview

Floriancic et al. analysed how spatial and temporal patterns of water age across 32 Alpine rivers are related to hydroclimatic and physical catchment properties by calculating young and new water fractions from already existing isotope timeseries of ^{18}O and ^{2}H. Lower young and new water fractions were observed in catchments with higher mean elevation, steeper slopes, larger catchment area, less antecedent precipitation as well as in catchments dominated by snow or affected by large water reservoirs such as lakes and dams. It is one of the first studies to use new water fractions on a regional scale, which has promising potential in improving the understanding of runoff generation processes. However, many of the observed relationships do not significantly advance on previous research and the study does not convincingly identify the dominant process controls, which even leads to some contradicting claims. Interesting results are found in exploring subsets of data, which is made possible by the novel methodology and would deserve a publication, although the paper overall requires a more detailed discussion and justification of certain analyses.

Main comments

1) Cross-correlations leading to spurious conclusions

Much of the paper is dedicated to correlating new water fractions to a range of hydroclimatic variables and physical catchment properties, which are highly cross-correlated and thus lead to issues in identifying the driving controls. This issue has already been long acknowledged and several similar relationships of young water fractions to catchment properties have been published on this topic (e.g. Jasechko et al., 2016 or von Freyberg et al., 2018). It is unclear from the text why new water fractions should have different relationships to catchment properties than young water fractions and how repeating such a broad analysis contributes to the process understanding. Recent research has instead focused on identifying individual factors within this variability such as the effect of snow cover (Gentille et al., 2023). The significant influence of the hydroclimatic regime has been confirmed even in this study, showing that new water fractions are much lower in snow-dominated catchments where snowpack provides longer water storage. Yet this effect is neglected when exploring the downstream propagation of new water fractions (which is presented as one of the focal points of this study) because sub-catchments of the Rhine and Danube are sorted by area without considering their location and climate. Due to sampling bias, many of the smaller catchments of the Rhine are rain-dominated, while snow-dominated catchments make up four of the six catchments with largest areas. This leads to a spurious and contradicting conclusion about the effect of catchment area because catchments typically shift from being snow-dominated to hybrid and rain-dominated along a single river. Mixing these effects could be avoided if only sampling locations along a single river (and not its tributaries) were considered.

The Danube case is slightly more convincing, although a statistically significant correlation is obtained only after the removal of two snow-dominated headwaters and with the acknowledgment that the three most downstream sampling locations are largely affected by dams. Thus, the main conclusion is a confirmation that dams indeed lower new water fractions without being able to say much about the effect of catchment area which could be transferrable to other rivers. I would suggest that the authors consider the perceptual model of downstream propagation of water age outlined by Gentille et al. (2023) based on the shifts in hydroclimatic regimes with low young water fractions expected in high-elevation snow-dominated headwaters, higher fractions in lower hybrid catchments and low fractions in downstream rainfall-dominated reaches. To observe such a relationship, a more sophisticated statistical method than a single Spearman correlation covering the whole dataset would be needed. Ideally, the relationship to catchment area would be derived from a completely rain-dominated catchment to eliminate the influence of hydroclimatic regimes altogether.

2) Superficial analysis of data subsets

I see the most significant contribution of this study in section 3.4 comparing different subsets of data, which takes advantage of ensemble hydrograph separation as opposed to other approaches and deserves a more detailed analysis. The study finds that new water fractions strongly increase after a precipitation threshold of 70 to 225 mm in most catchments. It would be interesting to see a discussion of what might be causing these differences in precipitation thresholds, possibly by relating it back to the catchment characteristics. It might also be worth focusing on how the precipitation threshold of 150-200 mm/month for most catchments relates to the 5 mm/day threshold observed by Knapp et al. (2019) because it appears as almost linear extrapolation of the daily precipitation intensity to the monthly scale at first glance, although the distribution of the precipitation during the month might influence the relationship. Additionally, ensemble hydrograph separation could be exploited further by distinguishing between winter and summer precipitation, which could yield considerably different thresholds in hybrid (and possibly snow) catchments according to Gentille et al. (2023). Such an effect is also suggested by observing higher new water fractions for the summer half of the year than for the wettest half of the year, indicating that precipitation in winter behaves differently (probably due to snowfall and snowpack storage in some catchments), which could be analysed further. Furthermore, monthly isotope timeseries would be appropriate to analyse more detailed patterns of seasonality or wetness conditions than just splitting the dataset into two halves (winter and summer, wet and dry) which averages over too much of the dataset to distinguish the control processes.

Concerning the interpretation of the catchments which do not experience increases in new water fractions with higher precipitation, it appears much of it can be explained by hydroclimatic regimes and catchment specificities instead of just stating relationship to elevation and slope. Unsurprisingly, all the catchments dampened by upstream lakes (INE, AAT and RHW) had low new water fractions no matter the precipitation magnitude. And the other catchments without precipitation thresholds are mostly snow-dominated catchments, which can be related to the effects of snow on water age as discussed by Gentille et al. (2023).

It should also be noted that the results of this section are confusingly interpreted in the discussion and conclusion where only 8 instead of 18 catchments are reported to have the precipitation threshold effect, overturning the whole message.

3) Potential issues of data compatibility

As the study relies on secondary data, compatibility of the individual datasets should be elaborated on more to strengthen the conclusions. Notably, the CH-IRP dataset was obtained at a fortnightly resolution (Staudinger et al., 2020), whereas WISA and ISOT provide streamflow isotope sampling data only at a monthly resolution. The methodology does not explain how this discrepancy was dealt with, even though higher sampling frequency leads to higher young water fractions (Stockinger et al., 2016) and lower new water fractions (Knapp et al., 2019) and thus potentially spurious comparisons. Information on the data sampling period is also missing (in Table 1) so it is unclear what years does the analysis cover and if it is the same for all locations (only number of samples per location is provided).

It is good to see use of a recently developed precipitation isotope reanalysis dataset which seems to have promising accuracy (Nelson et al., 2021), however, the study could be strengthened by providing a validation of the dataset against the sampling locations within the catchment boundaries, if possible, as this could give indication about the uncertainty arising from this factor. Furthermore, averaging the precipitation isotopes over the catchment area could lead to inaccuracies in large catchments with high precipitation gradients. It could be beneficial to weigh the averaging by mean precipitation to get a better representation of the relative contribution of precipitation to streamflow in different parts of the catchment or at least provide a sensitivity analysis of the weighting.

Minor comments

Abstract

The abstract covers slightly different results than the conclusion. Notably, no results about the precipitation thresholds are presented here despite being an important conclusion.

Line 21: Highest new water fractions (9.6%) were found in hybrid (not rainfall-dominated) catchments according to the results. Means across all catchments for young and water fractions were never presented in the text in results.

Line 26: Missing word – the fraction of slopes steeper *'than'* 40°.

Line 28: Replace elevation gradients by elevation difference or relief.

Introduction

The introduction provides a good background about previous studies of water age in rivers, although new water fractions (and their benefit over other measures) should also be introduced in this section already since they are part of the research questions. Hypothesis about the downstream propagation of new water fractions would also be welcome.

Line 107: The reference should be *(Kirchner and Knapp, 2020b).*

Line 109 (Research Question 1): The absolute value of new and young water fractions across Alpine rivers does not provide much information if it is known that the value sampled at a monthly resolution might be different from those sampled at higher frequencies, the observed values are also never related to other studies in the discussion. The question rather appears to be answered primarily by comparing young and new water fractions across different hydroclimatic regimes, hence it would make more sense to formulate the question according to that.

Line 113 (Research Question 3): Is monthly precipitation total really a measure of precipitation intensity? Consider rephrasing or going deeper in the analysis of precipitation distribution in the month (could potentially be based on daily streamflow or rain gauge records).

Methods and Available Data

Section 2.1 and 2.2: Missing information on the time period of obtained datasets (provided only for the precipitation dataset).

Figure 1: It would be more helpful to provide site codes in the map rather than stating their coordinates in Table 1.

Section 2.2 and 2.3: How were hydroclimatic regimes classified? Does it follow Weingartner and Aschwanden (1992) in line with similar publications? Motivation of studying relationships to all catchment properties (e.g. PET or elevation difference) should be mentioned in the methodology or introduction.

Lines 223-229 belong to Physical catchment properties. Missing justification of using fractions of slopes below 10° and above 40° (arbitrary value). Only six topographic properties are mentioned, although seven is written on line 224.

Line 236-238: Explanation of boxplots is redundant here.

Results

Table 2: The column $q\ P^{-1}$ is presented as fractions but with the units of %.

Line 258: Not consistent in use of ' in large numbers in different parts of the text and in figures.

Line 306: It should be specified more clearly that Pearson correlation was used for the relationship between new and young water fractions if that is the case or use r_s to signify Spearman rank correlation.

Figure 3a, 3b and Figure 6d: Two catchments are labelled as AAB, although one of them corresponds to AAR instead.

Figure 4: Rainfall and hybrid catchments cannot be easily distinguished; different colour choice would make the figure clearer.

Line 334: Catchment DOW appears in both 175 and 200 mm threshold groups, should be replaced by ALP in the 200mm group.

Line 394: Excess word *'or'*.

Lines 411-416 repeat methodology and can be omitted. This section mentions ratio of elevation difference to catchment area which has not been used anywhere else. While it might be a better indicator than elevation difference itself (which is inherently related to catchment area), how does this measure differ from mean slope?

Figure 7 and 8: It would be interesting to see the effect of hydroclimatic regimes here the by plotting them in different colours such as in Figure 4.

Discussion

In general, results are well related to previous studies and expected physical drivers, however, the processes possibly causing the discrepancies of the results are mostly not explained in enough detail. It would also be beneficial to provide links to more studies using different measures of water age than young and new water fractions.

Line 446: How do you explain the opposite correlation compared to von Freyberg et al. (2018) despite using part of the same dataset? Do you get the same correlation for the 12 overlapping catchments as was found in their study?

Line 448: Rain-dominated and hybrid catchments appear to have similarly high values (hybrid catchments even have higher new water fractions than rain-dominated), only snow-dominated catchments have significantly lower values.

Line 452: Provide process explanation of why catchment size should matter in explaining the discrepancy with Gentille et al. (2023).

Line 469: Since correlations tend to improve when catchments with large lakes are removed (Jasechko et al., 2017), would removing these catchments have a significant impact on the derived correlations?

Line 505: What could be driving the discrepancy between the expected effect of evapotranspiration and the results? Can you imply that the effect of PET is not as significant as the other variables?

Line 509: Excess words *'were found'*.

Line 515-519: These results were not presented in the results section.

Lines 517 and 550: Correlation should be signed r_s instead of R to be consistent.

Line 545: Von Freyberg (2018) should start with a capital V at the beginning of the sentence.

Figure 10: Use elevation difference or relief instead of elevation gradient.

References

Data sources are not properly referenced (e.g. Umweltbundesamt, 2022 or FOEN, 2022).

Line 695: Incomplete reference