This review was prepared as part of graduate program Earth & Environment (course Integrated Topics in Earth & Environment) at Wageningen University, and has been produced under supervision of dr Ryan Teuling. The review has been posted because of its potential usefulness to the authors and editor. Although it has the format of a regular review as was requested by the course, this review was not solicited by the journal, and should be seen as a regular comment. We leave it up to the authors and editor which points will be addressed.

Dear Peter Jansson and Ryan Teuling,

we would like to thank you for having selected our paper for this class. The input is highly appreciated and will help to improve our paper. For simplicity, we cut the first part, the overall impression section, of the comment in this response. But we thank the authors for the summary and the appreciation of our work.

We would like to point out here that we find the review extremely helpful but overly harsh at instances. The wording implies that we made mistakes. While this wording style is omnipresent (unfortunately!), we take the opportunity to call for a change in wording style in scientific reviews. A simple example is the following: “the authors should further stress the relevance of this study” could be reformulated to “the study would gain from a more concise presentation of its relevance”.

Please find below a point-by-point response to both your general and specific comments.

With kind regards,

The Authors

Overall impression

(..)

Another useful finding of this paper is the introduction of a new, formal classification of alpine catchments into 3 hydro-climatic regimes. Though similar classification systems have existed, the system proposed in this paper used formal and objective criteria, making it suitable for cross-catchment datasets even beyond the alps. This classification system also includes a new proxy for estimating snow coverage using only discharge data – Q_{june}/Q_{DJF}. *This can be a very useful tool for future research looking to estimate snow cover requiring the necessary satellite data to estimate* F_{SCA} *directly.*

Thank you for this comment. As detailed in our answer to reviewer 1 and reviewer 2 (J. von Freyberg), we might consider modifying the classification scheme to make it more straightforward to link to previous classification. But it will remain transferable to other regions.

At first glance, the results and conclusion appear impressive. Interpretations of the results were synthesised to address a relevant knowledge gap regarding the understanding of alpine catchment processes. However, the perceptual model proposed by the authors do not appear to be sufficiently backed up; the explanatory variables used may not accurately represent the catchment processes as the authors intended. In my view, the authors should adapt their methodology to limit the model shortcomings, while also investigating other explanatory variables which can give more credibility to their conclusions. Overall, I believe the paper requires significant revisions before it can be considered for publication.

We would like to thank you for your critical comment. At this stage, this critical comment remains vague for us, and we do not provide an answer here, but to the detailed comments below.

General Comments

We number the comments below for cross-referencing.

Use of Fbf to represent groundwater storage processes

[1] Arguably a larger weakness of the paper compared to the one above, F_{bf} *is by definition related to* F_{yw}*, so the correlation found may not be strong enough evidence to convincingly corroborate the perceptual model.*

We agree that in snow-free systems, F^*_{yw} **('*' indicates a flow-weighted variable) is by definition related to** F_{bf}**: baseflow is composed of groundwater and groundwater is the dominant source of old water in snow-free systems (in absence of large lakes). However, in snow-influenced systems, part of the old water is temporarily stored in the snowpack. This is at a first glance not measured by** F_{bf}**. We nevertheless show here that** F_{bf} **obtained from streamflow alone (with the selected baseflow filter) leads to an approximately complementary relationship to** F^*_{yw} **(**$F^*_{yw} + F_{bf} \simeq 1$**), which is an important result for catchments where we do not have isotope measurements. Why is this so? Simply because the snowmelt in snow-dominated systems largely transits through the groundwater store and leads to high baseflow. This will be made much clearer in the revised version.**

[2] It could already be reasonably assumed that streamflow with higher mean residence times would indicate a slower flow path. Additionally, F_{bf} *does not directly infer the type of catchment storage; catchments with high snow storage and low groundwater storage may also yield higher* F_{bf}*.*

Yes, we agree that if snow is being stored for a long time on the surface, it will also lead to high shares of old water. However, if potential groundwater storage is low, the snowmelt cannot transit through the groundwater storage, i.e., it flows off quickly and cannot lead to a high baseflow. As mentioned in our response to J. von Freyberg, high baseflow in summer (and thus high F_{bf} **in high elevation catchments) means that snowmelt is transiting via groundwater. We will make this very clear in the paper.**

[3] Baseflow could also account for a myriad of other catchment characteristics, such as the mean slope and shape coefficient. Hence, the assumption made by the authors that F_{bf} *could be used as a proxy for groundwater storage processes is questionable.*

We do not understand this comment. Baseflow is "the slowest responding and longest lasting component of streamflow" and "it has been described as flow from groundwater storage" (Duncan, 2019 cum bibl.) In absence of big lakes, it is simply the groundwater released to the stream.

[4] As a result, the perceptual model and conclusions may have been supported with the wrong evidence. In line 549 the authors state that the inverse correlation between F_{bf} *and* F_{yw} *"clearly indicates that the observed patterns of* F_{yw} *are related to water partitioning between the surface and*

subsurface". However, F_{bf} may be inferring to other processes, inferring the need for cross-validation.

Thanks for pointing this out, we agree that we took a shortcut here not explaining the link between F_{bf} and groundwater storage explicitly. However, we do not think that F_{bf} might be inferring any other processes besides groundwater (in absence of big lakes).

[5] It is hence highly recommended that the authors run additional correlations to identify how F_{bf} links to groundwater-related processes. Figure 8 a) shows such an example, where F_{qd} is shown to be positively correlated with WFI. This helps justify the inclusion of F_{qd} and WFI as explanatory variables that act as proxies for groundwater flow. A correlation between WFI and F_{bf} would hence illustrate the strength of F_{qd} as a proxy for groundwater flow and storage. If WFI and F_{bf} are indeed positively correlated, the use of F_{qd} would be further justified. A correlation between F_{SCA} and F_{bf} would show the extent of snow-related processes being incorporated in F_{bf}.

Thanks for this comment. We will pay attention to better explain the link between all presented variables.

Trends with precipitation not analysed

[6] Precipitation can have a large effect on F_{yw} variations: catchments with higher rainfall may have faster flow paths and hence higher F_{yw}. Figure 12a illustrates differences in precipitation levels, and seems to indicate an increasing trend between F_{yw} and precipitation. That would indicate that precipitation should be included as an explanatory variable. Neglecting to include this variable further weakens the strength of the discussion and conclusion; the authors intended to investigate potential drivers of F_{yw} variations with elevation, without including a variable that is known to affect catchment residence times and to vary with elevation. I therefore suggest that the authors include correlations between F_{yw} and precipitation. Doing so would provide a more holistic view of the alpine catchment processes, and further enrich the perceptual model, conclusions, and the scope for further research in this new and exciting field.

Thanks for this comment, which is similar to a comment by J. von Freyberg. We copy here the answer from our response to J. von Freyberg:

"We decided to use variables that were not previously considered for explaining young water fractions variations; Jasechko et al. (2016) wrote: "Although topographic gradient provides the strongest correlation with young streamflow fractions in our data set, the fraction of unexplained variance is large, suggesting that other variables also play a significant role. <u>We observe no significant correlations between the young streamflow fraction and</u> catchment size, <u>annual precipitation</u>, bedrock porosity, population density, or the fraction of catchment area comprised of pasture land or open water".
We explain in the manuscript that, below roughly 1500 m, the increase of F^*_{yw} with elevation also depends on the increase of precipitation with elevation. In fact, in such cases, annual precipitation can be considered as a proxy of catchment wetness since we mainly observe a liquid water input.
Above 1500 m, using mean annual precipitation as a proxy for catchment wetness is misleading because the seasonal snowpack leads to a very dry period of the year despite high (solid) water input. In other words, the total amount of precipitation is not the variable of interest, rather the temporal concentration of water input is the relevant variable. It is possible to observe the saturation of the system (i.e., high wetness conditions) also when annual precipitation is low if a large volume of water (stored in the snowpack) is released in a relatively concentrated time interval. After the long winter period, we expect high infiltration that recharges the groundwater storage.

This process can bring the system to saturation (high wetness) so that ultimately rain or snowmelt can more rapidly reach the stream as overland flow.
Additionally, Lutz et al. (2018), estimating the young water fraction for 24 catchments in Germany, have found exactly the opposite of what is generally thought: the young water fraction decreases with increasing mean annual precipitation. They stated that this result reflects "the impact of various factors relevant in the mountainous region, resulting in the decrease of young water fractions for higher-elevation catchments" (Lutz et al. 2018). Thus, what we did in our work is searching for the "various factors" that lead to low F^*_{yw} in mountainous catchments not considering the mean annual precipitation as an explanatory variable."

Specific Comments

[7] From the methodology of the paper, a new hydro-climatic regime classification came about, including a new proxy variable that can be used to estimate F_{SCA} with high accuracy ($R^2=0.99$). Though useful in strengthening the methodology, the authors only dedicate a small section of the discussion and a couple of sentences in the conclusion on this classification scheme. The development of the hydro-climatic regime classification is given a lot of attention to despite not being part of the research objectives. If the authors believe that this new method has potential in future studies, it is recommended that these methods could be mentioned more explicitly: either being incorporated in the abstract and treated as a research objective, or moving elaborations to the appendix or to a separate publication.

Thanks for this suggestion. We will decide on the final classification scheme during the work on the revised version and consider moving technical detail in the supplementary material.

[8] Isotope data timespan and resolution not stated. This uncertainty propagates to form some of the uncertainty in F_{yw}. However, by including the temporal data span/resolution, and/or the goodness of fit of the sine curves, readers could get some idea whether the uncertainty stems from the lack of data or from the dynamic nature of the catchment. This can be done in a separate table in the methodology, or in the supplementary material.

Thank you for this comment. We will add info about isotope timespan and resolution.

[9] Implications of problem statement/results could be elaborated. This paper provides a perceptual model that addresses a clear knowledge gap in existing literature. However, the paper fails to link the implications of the results in the wider context of this field. Sure, the knowledge of high-altitude catchment processes is improved, but what does this mean for the environment? Existing literature (i.e. Hayashi 2019) point out that if high-altitude catchments have a large capacity for groundwater storage, the flow dynamics would respond differently to climate change than previously thought, with consequences to downstream settlements. Reflecting their results in this larger scope may give more importance to the study by putting their results in context.

In answer also to J. von Freyberg comment, we will make the research objective and the relevance of the outcome much clearer in the revised version. We will also pick up the above suggestion to broaden the discussion, without becoming speculative. In fact, it is important to point out that the storage in the high alpine environments studied here is relatively large but this is a local effect. We cannot make statements about the role of this headwater storage at a more regional scale.

[10] Significance of results compared to existing knowledge. While the perceptual model manages to synthesise the results in an effective way, a lot of the results gained have already been known (or

could be inferred) when looking at existing literature. For instance, the strong relationship between F_{bf} and F_{yw} could be inferred by definition, (..)

As discussed in our response to comment [1] above, we respectfully disagree that this result was known before. Specifically, to the best of our knowledge, a complementarity between F_{bf} (using a baseflow filter without parameters calibration) and F^*_{yw} was never presented before. Moreover, in answer also to J. von Freyberg comment: *"The potentially dominant water flow and storage processes driving young water fractions are of course discussed in previous work (Jasechko et al. 2016, von Freyberg et al. 2018, Lutz et al. 2018); to the best of our knowledge, this is the first attempt to present a perceptual model that harmonizes these known processes with the surprising low F^*_{yw} values of high elevation catchments. This will be explicitly stated in the introduction."*

[11] … while the fact that a significant portion of precipitation is stored in the seasonal snowpack at high-elevated catchments is also previously known.

We do not think that we presented this as a new result.

[12] Hence, the authors should further stress the relevance of this study (e.g. using a new tracer-based empirical method to investigate catchment processes to explain "counterintuitive" evidence found in previous studies) in the aim and conclusions of the paper.

See also our response to comment [10], we will make this clearer.

[13] Limited number and variety of catchments in dataset. Though I appreciate the practical issues when it comes to gathering this much data, I feel that stronger, more significant conclusions could have been drawn with a larger dataset. As stated in activity 1, only 2 catchments have limestone-dominated bedrock, despite a large portion of the alps having such geology. Only 1 catchment had a significant coverage of a glacier (NBPV), despite them covering a large number of high-altitude alpine catchments. Indeed, NBPV was seen as an outlier in many of the correlations – would this have still been the case if more glacier-dominated catchments were included?

We would of course be extremely happy to include more case studies but this will only be possible once more data becomes available. Year-around isotope sampling in very high elevation environments is rare.

[14] Use of F_{qd} to represent groundwater storage processes. The paper uses F_{qd} and F_{bf} to represent the groundwater storage- related processes that may explain the variation of F_{yw} with elevation.

We would argue that we use them as proxies but not to "represent processes". We will carefully revise the manuscript to avoid misuse of terminology.

[15] However, I find that the use of these variables have inherent shortcomings that limit the strength of the conclusions and the validity of the perceptual model. F_{qd} represents only a limited part of the catchment geology responsible for groundwater flow, as suggested by the weak correlation between F_{qd} and F_{yw}. This results clashes with those obtained by Arnoux et al., (2021) who found that quaternary deposits played an important role in groundwater storage in alpine catchments, though that study used a modelling approach instead of an empirical one. Additionally, Arnoux et al., (2020) investigated 13 catchments, 4 of which are included in the dataset of Gentile et al., suggesting that not all alpine catchments have quaternary deposits as a major store of groundwater. Hayashi (2020) hypothesised that groundwater storage in alpine catchments could be controlled by the amount of quaternary deposits, the type of underlying bedrock or the presence of cracks and fissures in the

underlying bedrock. This leads me to believe that the quaternary deposit coverage can only act as a first-order measure of geological groundwater storage in alpine catchments, since there can be large differences in the geological structures between various catchments, and how they function to store groundwater.

Yes, you are exactly right about F_{qd}: it is a first-order measure, but the best that we can do for the catchments at hand. In answer also to reviewer 1: "*<u>The missing information</u> about the portion of fractured bedrocks, the thickness of quaternary deposits and the bedrock topography <u>will demand future attention for a complete picture of the role of geology</u> (potentially resulting in a statistically significant correlation with F_{yw}).*"

[16] As a result, the perceptual model lacks information on groundwater storage processes.

The proposed model is supposed to summarize what we can know at the moment. Especially, we want to link well-known hydrological processes with information related to the water age in order to harmonize empirical evidence of F^*_{yw} values with the processes that lead to these values. In other words, many of the processes are known, but how these processes are linked with F^*_{yw} is poorly addressed in the past literature: this last point is what we focus on in the manuscript. We will be pleased if future studies can add additional information to refine our perceptual model.

[17] Though the shortcomings of representing groundwater storage using F_{qd} has been adequately explained in the discussion and conclusions, it is recommended that the authors explore other geological information as potential explanatory variables. For instance, bedrock type could have been added as an additional explanatory variable for a geological form of groundwater storage. Additionally, the depth of the deposits are not considered using this methodology, though that data may be difficult to obtain.

Yes, the above information cannot be easily obtained for all catchments. The bedrock type does not seem to add relevant information (Fig. 1 below) and part of this information is already included in F_{qd}.

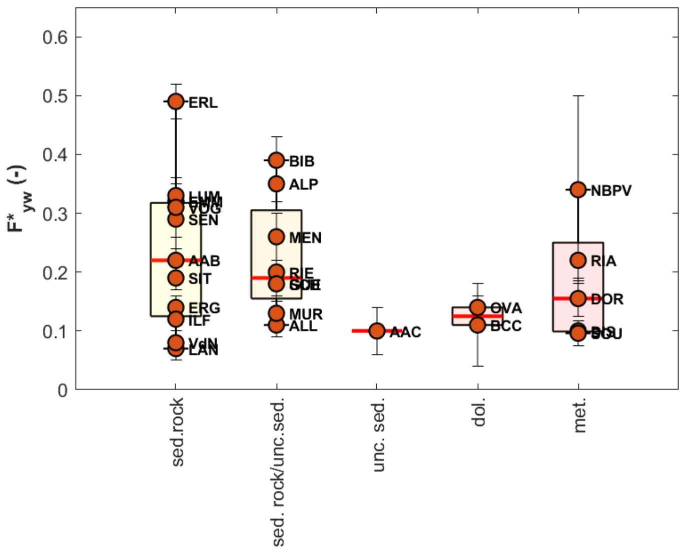

Fig.1 F^*_{yw} with varying bedrock type. Sed.rock: sedimentary rock, unc.sed.: unconsolidated sediments, dol: dolomite, met: metamorphic.

Minor issues: line-by-line

Thanks for these detailed suggestions, highly appreciated. We will implement them in the revised version.