Final Peer Review – Gentile et al.

By Peter Jansson

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.

Overall Impression

Gentile et al. investigate possible mechanisms that may cause variations in young water fraction (F_{yw}) in streamflow to vary with altitude in elevated alpine catchments. Previous studies have found that F_{yw} increases in catchments with higher elevations, however catchments higher than 1500 m a.s.l. show an inverse relationship (F_{yw} decreases with elevation), the causes of which are poorly understood.

The authors of this paper correlated F_{yw} in 27 alpine catchments with measures for snow storage (F_{SCA} – fractional snow cover area) and groundwater storage (F_{bf} – fractional baseflow, F_{qd} – fractional quaternary deposits). The results led to the authors developing a perceptual model showing that catchments with higher F_{SCA} (prominent in elevated catchments with persistent snowpacks) promote more infiltration and gradual discharge in winter, leading to lower F_{yw}. This model was corroborated with the strongly inverse relationship between F_{yw} and F_{bf}.

This paper presents a reasoning for 'confounding' results found in existing research that was not previously explained. The inverse relationship between elevation and F_{yw} in elevated catchments can be seen in the work of von Freyenberg et al., (2018) and Ceperley et al., (2020). Gentile et al. consequently investigate whether these factors indeed drive F_{yw} variations. The results obtained show a strong relationship between F_{yw} and F_{bf}, and a puzzling "bell-shaped" relationship between F_{yw} and F_{SCA}. These results on their own right would provide a limited explanation to the 'confounding results'. However, by tying together the different hypothesised processes indicated by the results in the perceptual model, the authors manage to formulate a coherent and thorough concept of catchment processes that would explain the results. Figure 13 illustrates the different processes in the perceptual model very well, allowing for further discussion and research on the topic.

I feel that the reasons for undertaking this research have been justified. Though it was not linked to a pressing environmental concern, this paper continues the use of a relatively new method (F_{yw} variations) to understand catchment dynamics. The paper directly addresses a knowledge gap found in existing research and sought to explain what appeared to be confounding results.

The method used for determining water age at the catchment scale – F_{yw} – has been developed fairly recently (Kirchner, 2016a), yet appears to be robust, and has been used in earlier papers (i.e. Ceperley et al., 2020; Stockinger et al., 2019; von Freyenberg et al., 2018). Though it appears logical to use the same type of data as the earlier research this paper is based on, the authors should be credited for using such a contemporary empirical method to describe catchment behaviour.

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.

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.**

General Comments

Use of F_{bf} to represent groundwater storage processes

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. 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}. 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.

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. **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}.

Trends with precipitation not analysed

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.

Specific Comments

Minor findings may deserve more attention. 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.

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.

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.

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

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?

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

Minor issues: line-by-line

- Title – "young water fraction" instead of "F_{yw}" to make the paper accessible to readers unacquainted with the topic
- Line 22 – briefly justify the necessity of the proposed formal classification scheme
- Line 36 – "low F_{yw} is found" instead of "this results in low F_{yw}"
- Line 70-71 – were these catchments above the F_{yw}-reset as well? If so, please specify to stress how the findings are "in line" with the previous one
- Line 82 – what is meant by "efficient groundwater recharge"? It may mean high rates of recharge, or that most precipitation is recharged with minimal runoff, making it ambiguous to readers
- Line 94 – here, "dynamic storage" is mentioned as a knowledge gap, but this concept is not referred to again in the rest of the paper, especially regarding whether this concept connects to how the explanatory variables were chosen. Perhaps mention dynamic storage in the discussion when discussing F_{qd} results
- Line 104-105 – Briefly justify the intention to introduce the new classification system
- Line 129-130 – Why were these catchments included to the dataset? Presumably to better represent high-altitude catchments; justifying this would clarify readers about this section
- Line 154 – Rhaetian Alps (Rätische Alpen/Alpi Retiche)
- Line 155 – would avoid using "a good range" – instead perhaps "allows us to explore F_{yw} variations in large range of geological and climatic conditions…"
- Figure 1 – the shading of the catchments are unclear. Perhaps use a cleaner basemap (black/white) so that each catchment colour stands out better (BCC was almost invisible).
- Table 1 – Column 3 shows average elevation, perhaps explain how the averaging was conducted (i.e. area-weighted)
- Line 171 – Would this sub-section not be better suited under section 2, since data on each study site was already being discussed?
- Line 205 – unsure what "according to volume of precipitation" implies; either elaborate, or state "accounting for volume of precipitation…"
- Line 222-223 – this may be better suited in the introduction to help justify the new classification system
- Line 233 – "ambiguities" instead of "tricky-points"
- Figure 2 – the source of the data (Staudinger et al., 2007) was already included in the caption, no need to repeat it in the legend
- Line 248 – Enjoyed reading the overall description in this section!
- Line 267 – Nice scrutiny of methods
- Line 311 – "allows *for* estimation"
- Line 325 – Section 4.1 appears to contribute little to the discussion/conclusions, perhaps it can be removed?

- Line 345 – “evenly” instead of “equally”
- Line 354-355 – Nice exploration of outliers!
- Line 367 – Perhaps further elaborate in the introduction/methods that slope will also be considered as an explanatory variable
- Line 370-371 – Good preliminary result, well-written to explain the subsequent steps in the data analysis
- Figure 5 – the dotted line appears to be some sort of average, clarify whether this is a mean or median
- Line 419 – nice finding (strong correlation between F_{qd} and WFI), would be good to mention this in the discussion to support the choice of including F_{qd} in the methodology
- Line 472-473 – nice wording to introduce perceptual model
- Line 482 – “…recharge. This…”
- Line 487 – “reduce*s*”
- Line 494 – “However, this effect…”
- Line 500 – “depending on elevation.”
- Line 519 – Jansson et al. (2003) was not included in the reference list. Also, please review the paper to confirm whether the fact stated was indeed referred to in the paper
- Line 596 – “few streamflow events *are* characterised….”