2nd Review of:

The effect of beaver ponds dams on water physico-chemical composition properties of stream water in the Carpathians (Poland and Slovakia) by Joanna Wąs et al.

First, I would very much like to see this manuscript published, as it represents an important contribution and a clear stepping stone for further work on the controls of water chemistry in beaver-influenced reaches. At the same time, a degree of caution is warranted. We have seen cases in which NGOs/governments have taken marginal or context-dependent positive evidence of beaver impacts, combined with the occasional overstatement often typical of scientific papers, and used it out of context to inform policy. I therefore wish to emphasize a note of caution, and ask that my review be read in this light.

I believe the main argument of this manuscript is that reach geomorphology and pond age controls water chemistry, but the abstract is confusing.

Firstly, regarding geomorphology, the current distinctions appear largely limited to overbank flooding versus in-channel conditions, which I would not characterize as geomorphology sensu stricto. You further suggest that water chemistry is related to substrate, yet there is no table or figure that clearly situates the study sites with respect to geological units. If geomorphology, or perhaps more precisely landform, is a key explanatory factor, I would expect at least a basic characterization of the valley bottom, for example valley width, channel and hillslope configuration, longitudinal river profile, and substrate. What you appear to be arguing for is a hydro-geomorphic framework, which would also need to include parameters such as estimated discharge, water residence time, and catchment size. If this line of argument is pursued consistently, it would imply that the primary driver of change, namely the hydro-geomorphic context, has not been adequately investigated in the present study.

As a practical way forward, you could consider quantifying the flooded area and framing the observed differences between upstream and downstream primarily in terms of flooding extent and age, which may account for the variation you observe? If it makes sense, you might consider a simplified classification of beaver-influenced reaches into “large and old” versus “small and young.” This would allow you to streamline the argument by discussing area and age jointly as key explanatory variables. The main exception appears to be site 5, which is described as old but is very small and apparently in-channel; this case would require specific discussion.

Further, I am concerned that the analysis may primarily be capturing noise rather than signal. I find it frustrating that there is no straightforward comparison of upstream and downstream values. Figure 3 appears intended to address this, but it doesn’t give me the absolute changes between up- and downstream – did I miss this somewhere? I tried to download the original dataset from Zenodo, but didn’t find it under the cited DOI. In

Figure 3, most values cluster around the 1:1 line, suggesting that they may largely reflect natural variability, with the exception of DO, ammonium, and pH. Even for these variables, however, a substantial number of data points exhibit trends opposite to those being argued for. This concern is fueled by the way the 'less-than values' (data below detection limits) are treated. The cited reference (Palarea-Albaladejo and Martín-Fernández, 2013) in fact argues against the approach adopted here for handling this data. In my view, values below the detection limit should be excluded from the analysis, but not sure they are? The authors state that NO_3^- and NH_4^+ concentrations were below detection in 17 and 20 out of 102 samples, respectively, corresponding to nearly one fifth of the dataset, which is substantial. Figure 3 legend has a typo (stream). Here, I also don't understand the different symbols – each value is plotting upstream versus downstream?

My concern about this manuscript interpreting mostly noise is also caused by the fact that measurement uncertainty is not propagated in the percent-change calculations, but as far as I know, ratios amplify and propagate measurement error, resulting in increased variability and poorer statistical behaviour compared to analyses based on absolute values or log-transformed changes, especially when original concentrations are low. I still don't understand why the authors have not chosen a multiple regression but pca, it seems an odd choice, as in my opinion is a clustering exercise and very helpful in developing an argument, but should not be the final result?

I like Figure 6 (downstream changes) which has absolute numbers, and I assume contrasts an old and a young site? If this is true, then please annotate and describe in the captions, the same is true for Figure 7. Please write in the captions why those are relevant.

The discussion starts with variability of water properties depending on catchment geology, but I don't see a figure showing this - there is also no reference to one in the text. The next point of the discussion is about DO, and states that there is a coherent change at all sites, but Figure 3 clearly shows values on both sites of the 1:1 line. This section does not cite a table or figure. I believe you don't refer to TOC, but rather DOC or POC? Why is this data not shown? After that, you report a marked difference in NO3 concentrations, but again, I cannot really see this in Figure 3. Overall, the argument makes sense, but I specifically don't see the decrease in NO3 in the dataset (Figure 3).

The discussion would benefit from substantial streamlining and a clearer focus. As I understand your central argument, you propose that decomposition of organic matter is the primary driver of the observed changes in water chemistry, with higher organic matter availability in older and larger ponds leading to the reported chemical patterns. If your evidence supports this claim, then this should be stated clearly and directly, without complicating the argument with secondary explanations, I am still doubtful if

the PCA is the best method, as the axes cannot be clearly attributed? Why does your summary figure (Figure 8) has numbers without standard errors?

Please check on the English as well, there are a few unclear sentences, e.g. “The catchment characteristics refer to the catchments closed at the sampling sites below the beaver pond sequences”.