Instytut Geografii i Przestrzennego Zagospodarowania
Polska Akademia Nauk
Zakład Badań Geośrodowiska

Institute of Geography and Spatial Organization
Polish Academy of Sciences
Department of Geoenvironmental Research

31-018 Kraków, ul. św. Jana 22; tel./fax +48 12 422-40-85; e-mail: zbg@zg.pan.krakow.pl

Kraków, 01.04.2026

Dear Editor and Reviewers,

Thank you for your feedback and the opportunity to revise the paper for publication in Biogeosciences. All the suggestions have been carefully considered and the manuscript has been revised accordingly. Please see our responses and detailed explanations of the subsequent modifications (marked in red) below:

Editor: Streamlining the paper, clarify focus, leave away side aspects. This may include a change in the presentation framework such as replacement of the inadequately treated hydro-geomorphic scheme by a simpler comparison of large/old vs. small/young ponds.

Answer: We appreciate this suggestion and agree that the presentation should be as clear and focused as possible. However, after re-evaluating our dataset, we decided not to replace the original framework with a simplified ‘large and old vs. small and young” classification. Age and type (in-channel vs. overflowing) should not be merged because they represent two different sources of variation: temporal development versus local geomorphic setting. While many of the studied ponds were indeed either old and large or young and small, we also identified old but small ponds and young but large ponds. This indicates that pond size/type is not a direct consequence of age, but largely reflects local geomorphological controls, such as valley width and channel morphology. We have added Fig. 1b, which shows the valley cross-sections of the studied streams at the beaver dam sites, to illustrate the valleys’ geomorphological characteristics. The main limitations of the suggested grouping include statistical constraints and sample size - further analysis should exclude sites that do not belong to these two main groups: “old and large” and “young and small”. Our study covered nine diverse sites in the Carpathians. The application of the combined division proposed by you would lead to the creation of subgroups with very low numbers, which would prevent the reliable execution of statistical tests (e.g. Mann–Whitney U, or Kruskal–Wallis tests). Separating these two factors allows us to statistically analyze the impact of age and type on water parameters independently, which is crucial given the limited number of objects.

Furthermore, our results indicate that the age of the pond and its type control different biogeochemical processes in the water. The age of the pond is more strongly associated with the accumulation of organic matter over time. We have shown that it is age (and not type) that is the key factor influencing the decrease in dissolved oxygen, pH, and sulfate concentration. Geomorphic type (overflowing vs. in-channel), on the other hand, determines parameters dependent on the water surface area or retention time, such as the concentration of nitrates, and water temperature. Combining these variables into one 'type + age' category could blur these significant differences in the mechanisms of pond impact.

Editor: Data presentation must be improved, this includes minor things such as using more clearly distinguishable color gradients and symbols as well as homogeneity in data presentation (e.g. sites are numbered, at other places they are named).

Answer: Thank you for this helpful comment. We have revised the data presentation throughout the manuscript to improve clarity and consistency. In particular, we unified site identification across all figures and tables by using the same numbering system throughout the paper, including Figure A1. We also improved the visual design of the figures by applying more clearly distinguishable color gradients and symbols, and we revised Figure 3 to enhance readability. These changes were made to ensure a more homogeneous and reader-friendly presentation of the dataset.

Editor: The current results presentation suggests confounding issues among type, age and landuse - these may be addressed up-front with a changed framework, or through improved results/discussion.

Answer: Thank you for this important remark. In the revised manuscript, we set out more precise and definitive exclusion criteria for results that could be affected by the correlation between the pond age and its type. This led to the rejection of the relation between pH and the pond type. This parameter was found to be strongly related to the pond age. At the same time its relation with pond type, despite p-value being lower than 0.05, did not exceed the corrected, more rigorous, threshold. We clearly marked this case in the revised Table 3 and Result section.

Editor: Adequate treatment of data quality issues. Noise has already been a concern in a previous round. Are there calibration standards below the detection limit? How do measurement errors propagate when computing ratios or percent change? When a careful assessment of noise shows its importance, this may be reflected in the wording of the discussion.

Answer: Thank you for this important comment. In the revised manuscript, we addressed data-quality and noise-related issues more explicitly. To ensure the reliability of our results, we have accounted for the measurement uncertainty resulting from the impact of the analytical errors (percent relative errors). For statistical tests, we decided to include those parameters that: 1) show clear changes in value due to the presence of beaver dams ($|Dc|>15\%$), and 2) whose changes are greater than the differences resulting from analytical errors. The analytical errors are significant in cases of large percentage differences (Dc %) between values of chemical parameters in the beaver ponds/streams below the beaver ponds and values above the ponds. The impact of the analytical error increases as Dc increases. Therefore, we calculated the lowest (Dc_{min}) and highest (Dc_{max}) possible Dc values of a given physico-chemical parameter for each sample, with maximum percent relative errors (+/-) assumed for each (Fig. A3a). To illustrate the method used more clearly, we have provided an example of SO_4 (Fig. A3a) showing the maximum and minimum possible Dc values, respectively Dc_{max} and Dc_{min}, for a given, measured Dc. For most of the selected chemical parameters, such as DO, Cl^-, K^+, SO_4^{2-}, and NO_3^- the majority of the maximum possible differences in Dc (Dc_{min}-Dc_{max}) were less than 10–15% (Fig. A3b). However, these differences were higher for H^+ and NH_4^+, though still many times smaller than the mean absolute difference $|Dc|$ values, which allowed for their inclusion in the further analysis. However, analytical errors in the case of eg. Ca^{2+} and HCO_3- could indeed have generated Dc differences large enough to exceed the measured concentration differences caused by the presence of beavers. Finally, taking into account the uncertainty associated with analytical errors, we increased the Dc threshold from 5% to 15% - this excluded Ca and HCO_3 from further statistical tests (Mann–Whitney U, or Kruskal–Wallis tests).

As regards the values below detection limit (DL), as Palarea-Albaladejo and Martín-Fernández (2013) write, whom we have cited: 'One of the most popular ad hoc solutions among practitioners is replacing less-thans by an arbitrary small value, usually half or 75% the DL" (in Section 1) or "That value can be freely chosen by the practitioner, including the common 0.5DL or 0.75DL values, (...)" (in Section 3.1). It does not matter much whether we set the threshold for values below the detection limit at 50% or 75% of the detection limit. $DL/\sqrt{2}$ is also often used, which is approximately 71% of the detection limit. In all these cases, the concentrations will be very low. For our work, we have decided to use 75% of the detection limit.

Editor: There are a lot of minor issues you may still want to take care of. The absolute differences used in the PCA are not clearly defined (is equ (2) showing a mean or the measure used in the PCA?). What happens if differences were negative as well as positive, but only their absolute value was used in the PCA?

Answer: Thank you for pointing out this ambiguity. In the revised manuscript, PCA was removed entirely from the analysis, so the issue of whether absolute differences were used in PCA is no longer applicable. All the PCA calculations and site-to-site comparisons were based on Dc (equ1). The purpose of introducing |Dc| (equ2) was to determine the parameters for which potential changes generally fall within the uncertainty range, in order to exclude them from further analyses from the very beginning. Since we removed PCA analyses and it is clearly stated in the Methods that seasonal and site-to-site variations were based on Dc, there was no need for further clarification in the text.

Editor: Detail how Bonferroni correction was used, Table 3 currently resembles a "fishing expedition" for P-values. Also, in Table 3 you report usage of data in an aggregated version - clarify what this means (a random factor? pooling? averaging?).

Answer: We have updated the verification process and supplemented Methods with its description:

'...Despite statistical significance being obtained, the results of site-to-site comparisons were further verified to exclude those deemed unreliable due to the potential influences of noise and the relationship between the grouping variables. When comparing the differences between groups based on pond age, the only parameters retained were those for which there was a clear upward or downward trend in the median difference between groups as age increased. Additionally, to mitigate the risk of family-wise error for multiple comparisons based on the same dataset, we applied a Bonferroni correction to the significance threshold ($p < 0.017$) (Armstrong, 2014). If the same parameter was found to be dependent on both pond characteristics (age and type), the result that failed to meet this stricter threshold was excluded, to account for any potential correlation between the characteristics…'

In the previous version of the manuscript, we treated the *p* threshold correction as additional information that was only interpreted textually. However, to reduce the risk of misinterpretation, we decided it is better to definitely exclude the result for which the family-wise error might have occurred (pH vs. type).

We recognised that some readers might view Table 3 as the primary outcome of the study and disregard parts of the textual descriptions. Therefore, we have supplemented it with information about the exclusion of results following the verification process. We marked the results deemed unreliable in the first version of the manuscript, which were previously indicated only in the text (K and NO_3 vs. age), as well as the results excluded based on an improved approach to the partial overlap of both categorisations (pH vs. type). We believe that this modification, and use of definite exclusion instead of descriptive interpretation, has made both the table and the text more scientifically sound and easier to interpret. We have also corrected the description of the seasonal comparison in the caption.

Editor: Fig 7 is really hard to read, Fig 8 is a potentially nice conceptual figure but currently hard to read - why are variables ordered along a pond sequence and seemingly associated with specific ponds in the sequence?

Answer: We made some visual modifications to Figure 7. For example, we highlighted the boundaries between Profile 1, Profile 2, and the outflows, to make this division clearer. Following the reviewer's suggestion, we added information about the age and type of the study sites to the figure caption to make it easier to interpret.

Thank you for noticing both the potential and shortcomings of Figure 8. To avoid confusion related to the order of the ponds, we present all variables within a single pond now.

Editor: Consider this a request for major revision. I support publication of this study, but the manuscript still needs substantial attention and major changes. Given the ecopolitical context of beavers, their removal or support, the study must be presented in a scientifically sound AND accessible way, thus making sure it actually achieves impact while preventing abuse.

Answer: Thank you for your feedback and for highlighting areas for improvement. We fully agree that, given the ecopolitical context of beaver management, the manuscript must be both scientifically rigorous and clearly accessible to readers. We believe that the revision has substantially improved the clarity of the paper. In particular, we replaced controversial parts that did not provide any obvious explanations (PCA) with additional materials on our main analyses. We have adopted a more rigorous approach to exclude potentially unreliable results. Such results were clearly marked in Table 3, beside the description of the process that led to their exclusion in the main body of the text. Consequently, they did not appear in the summary Figure 8. Furthermore, providing information on absolute data diversity corresponding to the studied percentage changes has been an important improvement in maintaining clarity of outcomes in the summary (see Figure A4). Overall, we are very grateful for the opportunity to improve the manuscript, and for your help in addressing important shortcomings.

Reviewer: 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.

Answer: Thank you for your valuable opinion and for recognizing our manuscript as an important contribution to the research on the impact of beavers on hydrochemistry. We appreciate all the suggestions that helped us to improve our work and its presentation. We have improved treatment of noise and overlap between both pond categorisations and prepared additional materials to clarify ambiguities. Moreover, we paid attention to avoiding overstatement, for example we have removed some remarks on 'water purification' processes for which we did not have enough evidence.

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

Answer: We made some changes to the abstract based on the comments that followed.

Specifically, we replaced 'geomorphic type' and 'geomorphic categories' with 'flooding extent' and 'flooding extent categories', to not confuse the factor depending on a dam-floodplain interaction with typical geomorphic channel properties.

We also rewrote the main paragraph of the abstract, to increase clarity: '...The results showed that beaver dams have a strong impact on stream water chemistry during warm periods, when high temperatures accelerate biogeochemical activity. The age of the beaver ponds was a more statistically significant factor in altering stream water chemistry than the flooding extent of the ponds. There was a greater decrease in dissolved oxygen, pH and SO_4^{2-} concentrations in older ponds compared to younger ones. These changes might be regulated by the decomposition and aerobic/anaerobic oxidation of organic matter, which is more abundant in older ponds. Moreover, greater decreases in NO_3^- concentrations were observed in overflowing ponds than within in-channel ponds, which suggests a more intense denitrification process in the former. The denitrification process requires anoxic conditions, which were favoured in overflowing ponds due to their higher temperatures and longer water residence time compared to in-channel ponds. The decrease in NO_3^- concentrations may also be due to the increased uptake of mineral nitrogen by vegetation and the accumulation of nitrogen-rich organic sediment at the bottom of overflowing ponds. Most of the effects observed within the ponds were attenuated in downstream river sections. However, further research based on a larger dataset and a more frequent sampling strategy is needed to identify the biogeochemical processes that occur in beaver ponds under specific hydro-meteorological conditions.'.

Reviewer: 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.

Answer: Thank you for this remark. We used such terms as 'geomorphic type' for several reasons. Firstly, we considered beaver dams to be geomorphic features per se. Consequently, their extent appeared to be a geomorphic characteristic, particularly in terms of the relationship between the dam and the channel or the dam and the floodplain. Furthermore, we expected the extent of the flooding to be largely controlled by the geomorphology of the channels and valleys. However, after thorough consideration, we agree that such a broad term as geomorphology might suggest the use of river channel or valley parameters as explanatory variables, which could confuse the reader. Therefore, we opted for a more straightforward phrasing and replaced the 'geomorphic type' with 'type' or 'flooding extent' throughout the revised text.

Reviewer: 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.

Answer: We have added a map of general geological units to the revised Figure 1, as well as their brief description in the text. However, geology was previously described with respect to local composition, based on 1:50 000 maps, which are referenced in the text.

Reviewer: 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.

Answer: Thank you for this suggestion. Although our focus was on the characteristics of beaver dams and ponds rather than the properties of river valleys, we found the presentation of valley configurations very valuable since they could influence the latter. The cross-sections of the study sites added to Figure 1 may help readers to situate those streams within a broader geomorphic context, particularly given the mountainous character of the study area. Moreover, the lack of information about the channel substrate was indeed a significant oversight. We have now added that all the study sites were gravel-bed streams.

Reviewer: 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?

Answer: Thank you for this important comment. We agree that a consistent hydro-geomorphic framework would require additional explanatory variables, such as discharge, water residence time, catchment size, and more detailed characterization of valley and channel morphology, which were beyond the scope of the present study. We therefore revised the manuscript to avoid framing hydro-geomorphic context as the primary driver directly tested here. Following your suggestion, we now interpret the observed differences more explicitly in terms of pond age and flooding extent, which are the variables actually represented in our dataset. Accordingly, we revised the terminology and the overall narrative so that the interpretation is better aligned with the available data.

Reviewer: 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.

Answer: We appreciate the suggestion to simplify the classification into 'large and old' vs. 'small and young' groups; however, after a thorough analysis of our data and considering the specificity of the studied objects, we would like to maintain our original division into age and geomorphic type (in-channel vs. overflowing) as two separate variables. One of the main reasons for our decision is the lack of a clear cause-and-effect relationship between the age of a beaver dam and the type of beaver pond. The type of beaver pond is actually related to the geomorphological characteristics of the valley (e.g., the width of the valley) and riverbed (e.g., the depth of a riverbed). Therefore, although most of the beaver ponds we studied are indeed “old and large” and “young and small”, there were also “old and small” and “young and large” ponds. The main limitations of the suggested grouping include statistical constraints and sample size - further analysis should exclude sites that do not belong to these two main groups: “old and large” and “young and small”. Our study covered nine diverse sites in the Carpathians. The application of the combined division proposed by you would lead to the creation of subgroups with very low numbers, which would prevent the reliable execution of statistical tests (e.g. Mann–Whitney U, or Kruskal–Wallis tests). Separating these two factors allows us to statistically analyze the impact of age and type on water parameters independently, which is crucial given the limited number of objects.

Furthermore, our results indicate that the age of the pond and its geomorphic type control different biogeochemical processes in the water. The age of the pond is more strongly associated with the accumulation of organic matter over time. We have shown that it is age (and not type) that is the key factor influencing, among others, the decrease in dissolved oxygen, pH, and sulfate concentration. Geomorphic type (overflowing vs. in-channel), on the other hand, determines parameters dependent on the water surface area or retention time, such as the concentration of nitrates, and water temperature. Combining these variables into one 'type + age' category could blur these significant differences in the mechanisms of pond impact.

Reviewer: 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?

Answer: Thank you for this important comment. We agree that the original presentation did not make the absolute upstream and downstream differences sufficiently clear, which may also have made it harder to distinguish signal from noise. Although Figure 3 compared upstream and downstream values directly on the original measurement scales (mg/L, pH and uS/cm), these contrasts were not presented in a sufficiently explicit way. In response, we added Figure A4, which shows the absolute differences between upstream sections and beaver ponds/downstream sections for the discussed physico-chemical parameters. We believe this addition makes the magnitude of change much easier to assess and improves the transparency of the interpretation.

Reviewer: I tried to download the original dataset from Zenodo, but didn't find it under the cited DOI.

Answer: We changed the citation format from 'doi: 10.5281/zenodo.17937225' to the direct link https://doi.org/10.5281/zenodo.17937225 to make access more straightforward.

Reviewer: 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?

Answer: Thank you for this detailed and important comment. We think that removing values below the detection limit (DL) would be inappropriate, as this would result in the loss of information regarding the lowest concentrations of ions. Such information is crucial for comparing concentrations above and below beaver dams. For example, if the concentration of a certain ion is very low (just below the detection limit) at a sampling site above the beaver pond and high in the pond, this information would be lost if values below the detection limit were removed. Removing values below the detection limit would also artificially alter the distribution of the variables, increasing the mean, median and quartile values, as noted by Hites (2019 - Environ. Sci. Technol. 2019, 53, 11059−11060). Furthermore, in our study, concentrations below the limit of detection were observed only for NO_3 and NH_4. Their share in all measurements did not exceed 20%, which, according to Hites (2019), justifies their inclusion in the analyses.

As Palarea-Albaladejo and Martín-Fernández (2013) write, whom we have cited: 'One of the most popular ad hoc solutions among practitioners is replacing less-thans by an arbitrary small value, usually half or 75% the DL" (in Section 1) or "That value can be freely chosen by the practitioner, including the common 0.5DL or 0.75DL values, (...)" (in Section 3.1). It does not matter much whether we set the threshold for values below the detection limit at 50% or 75% of the detection limit. DL/√2 is also often used, which is approximately 71% of the detection limit. In all these cases, the concentrations will be very low. For our work, we have decided to use 75% of the detection limit.

We agree with your remark that most values on Figure 3 cluster around the 1:1 line. The purpose of this figure was to illustrate the overall image of the absolute values constituting our database. Therefore we also showed those parameters for which we did not identify any trends (such as SEC, Mg, Na, HCO_3). Following your comment, however, we have decided to highlight the parameters on which we focused in the Results and Discussion. These parameters are now emphasised in the revised figure, so as to reduce the distracting effect of the others. The typo in the legend was also fixed. X axis presents values measured above beaver ponds, and Y axis presents values measured in beaver ponds (squares) and below beaver ponds (Xs). Which means that the vertical distance between the point and 1:1 line represents the difference between inlet value and pond/downstream value. Also, the vertical distance between square and its respective X represents the difference between values obtained in the pond and downstream of the pond. To increase the efficiency of this figure, an additional parameter is illustrated by the colour of the signature: pond age. We could not find a way to illustrate two parameters (together with the pond type) at once without blurring the image with an excessive number of signatures, so we decided to present the more prominent one. In case of pH, DO and SO_4, differences between the blue (old), yellow (moderate) and red (young) squares (ponds) can be observed. Some squares

(young ponds represented in red) indeed cluster close to the 1:1 line, while for others the distances are greater, which is consistent with our point raised from the further analyses.

Reviewer: 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.

Answer: We agree with the reviewer's comment. To ensure the reliability of our results, we have accounted for the measurement uncertainty resulting from the impact of the analytical errors (percent relative errors). For statistical tests, we decided to include those parameters that: 1) show clear changes in value due to the presence of beaver dams (|*Dc*|>15%), and 2) whose changes are greater than the differences resulting from analytical errors. The analytical errors are significant in cases of large percentage differences (*Dc* %) between values of chemical parameters in the beaver ponds/streams below the beaver ponds and values above the ponds. The impact of the analytical error increases as Dc increases. Therefore, we calculated the lowest (Dc_{min}) and highest (Dc_{max}) possible Dc values of a given physico-chemical parameter for each sample, with maximum percent relative errors (+/-) assumed for each (Fig. A3a). To illustrate the method used more clearly, we have provided an example of SO_4 (Fig. A3a) showing the maximum and minimum possible Dc values, respectively Dc_{max} and Dc_{min}, for a given, measured Dc. For most of the selected chemical parameters, such as DO, Cl^-, K^+, SO_4^{2-}, and NO_3^- the majority of the maximum possible differences in *Dc* (Dc_{min}-Dc_{max}) were less than 10–15% (Fig. A3b). However, these differences were higher for H^+ and NH_4^+, though still many times smaller than the mean absolute difference |*Dc*| values, which allowed for their inclusion in the further analysis. We thank the reviewer for this comment. Analytical errors in the case of Ca^{2+} and HCO_3− could indeed have generated Dc differences large enough to exceed the measured concentration differences caused by the presence of beavers. Taking into account the uncertainty associated with analytical errors, we increased the Dc threshold from 5% to 15%. This excluded Ca and HCO_3 from further statistical tests.

As the reviewer suggested, we have included a graph showing the differences in the absolute physicochemical parameter values identified in the Mann–Whitney U, or Kruskal–Wallis tests (see Fig. A4).

Reviewer: 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?

Answer: We agree with this comment - we decided not to include the PCA as it didn't add enough to the work. We also considered regression-based approaches, but decided not to present a multiple regression model because our dataset is relatively small, heterogeneous, and includes group-based comparisons rather than a continuous explanatory framework. In this context, a regression model would be difficult to interpret robustly and would be sensitive to outliers. Since our primary aim was to test whether water-chemistry differences varied between categories of pond age and flooding extent/type, we considered non-parametric group-comparison tests more appropriate. Therefore, in the revised manuscript, we rely on the Mann–Whitney U test for two-group comparisons and the Kruskal–Wallis test for comparisons among multiple groups, as these methods are better suited to the structure of our data and are less affected by non-normality and outliers. However, following this logic, we updated our statistical approach. In the previous version of the manuscript, we used a parametric test (ANOVA) for the temperature and SO_4, which met the assumption of normal distribution based on the Shapiro–Wilk test. However, after considering the potential limitations of parametric tests on small samples, we decided to avoid this approach in our analysis due to the small number of objects studied. Thus, non-parametric tests were used for all the parameters in the revised manuscript. This resulted in small changes in the results obtained for the changes in temperature and SO_4 concentrations. Although the test yielded slightly less favourable p-values for our hypothesis, the differences were still significant.

Reviewer: 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.

Answer: Thank you for this helpful suggestion. We revised the captions of both Figures 6 and 7 to make their purpose and relevance clearer. In particular, we now explicitly indicate the age/type characteristics of the sites shown in these figures and explain why these examples are important for illustrating the downstream changes discussed in the manuscript.

Reviewer: 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.

Answer: We agree with the reviewer's comment. This paragraph was a remnant from a previous version of the manuscript, which included a dedicated chapter on the sources of variability in upstream values. While that section was removed to streamline the paper and focus on core findings, this specific paragraph was inadvertently retained. As the results to which it refers are no longer part of the study, we have now removed this passage to ensure consistency and clarity.

Reviewer: 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.

Answer: It should indeed have stated 'most' instead of 'all' to better reflect the pattern visible in the data. We have corrected this.

Reviewer: I believe you don't refer to TOC, but rather DOC or POC? Why is this data not shown?

Answer: We planned to focus on the parameters that were measured in all of the three approaches (seasonal, longitudinal, and vertical sampling). Unfortunately TOC was measured only during the last field campaign, which means that we have this kind of data only for the vertical profile approach. That is why we decided to use it only as an additional background information in the discussion. However, as it has been noticed, it might be reasonable to present it as well, to avoid using 'not shown' data as reference.

Finally, we presented TOC on revised Figure 6, added a respective sentence to the Methods in the paragraph regarding vertical sampling: '...In addition to the previously mentioned parameters, this scheme also involved determining the total organic carbon (TOC) concentration ($mg \cdot L^{-1}$)...' and transferred the text that mentioned this data into the Result section, e.g.: '...The TOC concentration in the beaver pond at site 3 was found to be more than 13 $mg \cdot L^{-1}$, which was almost three times higher than in the stream flowing into the pond….'.

Reviewer: 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).

Answer: Thank you for this important comment. In Figure 3, most NO_3 values are below the 1:1 line, which means that NO_3 concentrations are lower in the ponds/streams below the ponds than in streams above the ponds. For all signatures to be visible, they must be small. We have changed the color of the signatures to make them easier to identify. We have also added Figure A4, which clearly shows that NO_3 concentrations [mg/L] are lower in ponds/streams below the ponds than in streams above the ponds during the warm seasons.

Reviewer: 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?

Answer: Following recurring suggestions, we decided not to include PCA in the manuscript. Instead, we focused on the main analyses and improved the exclusion criteria for potentially unreliable results. The related Discussion fragments were then removed.

Reviewer: Why does your summary figure (Figure 8) has numbers without standard errors?

Answer: Thank you. We agree that it is reasonable to include more information on original data and its variability. Therefore, we added Figure A4, which shows relations between seasons/age/type and parameters affected by

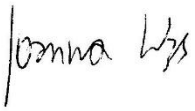
those variables, presented in their respective measurement units. The numbers in parentheses in Figure 8 referred to medians. Standard errors, on the other hand, refer directly to standard deviations and averages. In our opinion, it would be incorrect to present standard errors as classic measures alongside medians as positional measures. We have considered other unparametric measures that could describe the data variability. However most of them, such as quartile range or quartile deviation can be easily derived from Figure A4 that has been added to the manuscript. We decided to remove medians from Figure 8 and refer to Figure A4 in the caption instead to lead readers to this information.

Reviewer: 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”.

Answer: Thank you for this comment. We revised the indicated sentence for clarity. It now reads: ‘The closure points for all characterised catchments were located at the sampling sites downstream of the beaver ponds.’ In addition, the revised manuscript was carefully proofread and linguistically corrected by Cambridge Proofreading LLC.

On behalf of all authors of the submitted manuscript, sincerely

Joanna Wąs



corresponding author: joanna.was@zg.pan.krakow.pl