Dear Editor,

Thank you very much for providing us with the opportunity to respond to the reviewers' comments. We sincerely appreciate the constructive feedback and valuable suggestions from the two anonymous reviewers. However, given the substantial revisions required in the Introduction, Results, and Discussion sections, it may be challenging for us to fully complete all requested changes within one month. Therefore, our current response primarily outlines our plan for revising the manuscript in line with the reviewers' recommendations.

We would like to express our deepest appreciation for your consideration and for the valuable input from the reviewers.

Reviewer 1

General Comments

Overall, the manuscript lacks critical context, particularly regarding management actions that have been taken since the 1970s to reduce tributary phosphorus inputs to Lake Erie. Specifically, under the 1978 Great Lakes Water Quality Agreement (GLWQA), Canada and the United States adopted an 11,000 metric ton/year phosphorus target for Lake Erie and, at least initially, the lake responded quite favorably; by the early 1990s Lake Erie was considered to be restored. However, by the early 2000s algal blooms began to reoccur, resulting in a 2012 update to the GLWQA, recommendations for new targets in 2015 (Annex 4 Objectives and Targets Task Team Final Report to the Nutrients Annex Subcommittee 2015) and in 2016 these new, lower phosphorus targets were approved. Management actions to achieve the lower targets have been underway since that time, and their effectiveness in reducing phosphorus inputs is currently of great interest. However, the authors do not include any of this important information in their introduction or the interpretation of their results. This absence of context leads to inaccuracies and missed opportunities, and conveys a sense that the authors may be only peripherally familiar with the system they are reporting on. It also contributes to making the future projections of phosphorus inputs analysis in this manuscript extremely dubious.

The tributary input analysis using seasonal trend decomposition is similar to one presented in a previous paper (Stow et al 2015); the quantitative approach is the same, as are many of the results, though this manuscript includes additional tributaries and more recent data. However, while Stow et al is cited to support a few technical details, there is no attribution indicating that they used this approach on much of the same data, or that they previously reported many of these results. Further, some of these results have been reported in other papers as well, and thus are not particularly novel. Because the analysis presented by the

authors includes more recent data (Stow et al covered the period through 2013, while this analysis extends through 2021) the results could have been presented as an update with an emphasis on more recent patterns. This approach would have been particularly insightful as an updated analysis could have evaluated how the tributaries were responding to the new 2016 targets. However, lacking the context of the updated targets, the authors apparently did not recognize this opportunity and fail to provide what could have been some relevant insight.

Response: We agree. We have included the relevant research progress from Stow et al. (2015) in the Introduction section. We have removed the seasonal trend decomposition analysis for the Maumee River covering 1971–2013 and will instead focus our discussion exclusively on the 2014–2021 period.

The attempt to project future loads using estimated load:flow regressions also has significant limitations. First, how the various methods, bagging and the extreme value analysis, were used together to project future river flow is unclear. There is a lot that goes on between the lines here that requires additional explanation. Second, and more importantly, load is equal to concentration times flow, and generally flow variability is greater than concentration variability. Thus, a relationship between load and flow is hardly surprising. Consequently though, regressing load on flow is almost like regressing flow on flow, which constitutes a subtle violation of the regression assumption that the explanatory variables are distributed independently of the error (often expressed as the predictor variables are fixed in repeated sampling). The main concern here though, is that the relationship between concentration and flow in a given tributary changes over time (figure 2 in the Choquette et al reference provides a straightforward depiction of this), thus using equations such as those in table 5, which were estimated from past data, to make future projections of load based on projected flows under climate change is flawed. That substantial changes in the concentration:flow relationship have occurred during the period of record can be seen in the results presented by Choquette et al (they examined some of the same tributaries) as well the results presented by the authors in Figure 8 and Table 5. Most of these results are quite noisy, which would be unlikely if the concentration:flow relationship was relatively stable with time. Further, if the updated phosphorus load targets for Lake Erie are effective, this relationship would be expected to experience additional change into the future as the phosphorus concentration at a given flow is reduced in response to watershed management actions. This is where an absence of appropriate context particularly limits the validity of the future load projections made by the authors; because we are currently making a deliberate effort to modify the concentration:flow relationships in the tributaries – it is unclear how relationships estimated from past data under different (and changing) concentration:flow relationships can be used for future inference.

Response: We agree. We acknowledge that the relationship between concentration and flow in tributaries significantly varied over time, being influenced by multiple factors including watershed management actions and climate-driven changes. we have deleted the load:flow regression analysis, and we will focus our analysis and discussion on the 2014–2021 period.

Specific Comments

Line 28 – There is a single mention of a one-time drinking water problem in Stow et al (2015); this reference should not be used to support a broad claim of "compromised drinking water supplies".

Response: We agree, and we have deleted this citation, and changed to Sayers et al. (2019)

Reference:

Sayers, M. J., Grimm, A. G., Shuchman, R. A., Bosse, K. R., Fahnenstiel, G. L., Ruberg, S. A., & Leshkevich, G. A. (2019). Satellite monitoring of harmful algal blooms in the Western Basin of Lake Erie: A 20-year time-series. *Journal of Great Lakes Research*, *45*(3), 508-521. https://doi.org/10.1016/j.jglr.2019.01.005

Line 31 – The authors need to define "4R nutrient strategies".

Response: We agree, and we have defined "4R nutrient strategies" as right source, right rate, right time, and right place.

Lines 33-34 – Note there are papers that do not support this perspective. See, for example, Oliver et al (2017), Wilkinson et al (2022), Sillen et al (2024), and Bosse et al (2024).

Response: We agree, and we have deleted this statement and added these papers.

Lines 36-38 – This sentence is lacking context and very misleading. Following the adoption of an 11,000 metric ton/year target phosphorus load that was introduced in the 1978 Great Lakes Water Quality Agreement, Lake Erie showed considerable improvement (DePinto et al 1986), in fact a special feature in the Journal Great Lakes Water Quality Agreement detailed the restoration of Lake Erie (Makarewicz and Bertram 1991). It was not until the early 2000s that hints of eutrophication began to reappear.

Response: We agree and have deleted this statement.

Beginning line 42: the authors repeatedly refer to P loss in Lake Erie. This language is confusing, I believe the authors are referring to P loss from the watershed into Lake Erie, not a P loss that is occurring in the lake. The authors should choose more precise language.

Response: We agree, and we have revised this statement to "P loss from the watershed into Lake Erie".

Lines 44-48: The sentence beginning: "An analysis by…" is a long non-sequitur; it is not clear if the 4.8 year return period is attributed to Baker et al or Zhang et al.

Response: We agree, and we have revised this statement, the 4.8 years return period is attributed to Zhang et al. (2020).

Lines 48-55: The authors incorrectly state that Choquette et al (2019) did a "static analysis". They did, in fact, examine long-term trends. The reason for the focus on the Maumee in many reports is that it is the single largest tributary phosphorus input, not only in Lake Erie, but all the Great Lakes, and is believed to be the main cause of the annual western Lake Erie algal bloom; this is important context that the authors fail to mention. Also, see Rowland et al (2021) which includes long-term analyses of the Raisin and Cuyahoga Rivers.

Response: We agree, and we have added these statements in our Introduction.

Lines 58-59: This is a very old report that is based on information that preceded the Clean Water Act, the Great Lakes Water Quality Agreement and many in-lake developments, such as the dreissenid mussel invasion, which substantially changed nutrient inputs and cycling in the lake.

Response: We agree and have deleted these statements in our Introduction.

Line 71-72: Rowland et al (2020) did not average daily P concentrations as the authors report. They examined entire distributions of daily flow-weighted mean concentrations.

Response: We agree and have deleted these statements in our Introduction.

Lines 74-75: The is a grammatical error or poor wording in this sentence: “they are difficult” doesn’t fit into the rest of the sentence. Also, I think the word “always” in line 75 should probably be “often”.

Response: We agree and have revised these statements.

Line 82: See Qian et al (2000) for an earlier use of STL for this purpose.

Response: We agree and have added citation of this paper in our Introduction.

Lines 90-91: Including “1974-2021” twice in this sentence is repetitive. Also, none of the periods of record reported in table 3 go back to 1974.

Response: We agree and have revised these statements. For Table 3 data, we added datasets back to 1974.

Lines 101 – 102: Tables 1 and 2 (and 3 in line 222) are never referenced in the text to provide context; they seem to be ancillary information for the reader to review if they are interested. Exhibits such as tables and figures should be included to support points in the results and conclusions, not just as general references.

Response: We agree and have added citations of Tables 1 and 2 in our Material and Methods part.

Lines 141-143: The sentence in lines 141-142 is repeated in the following sentence.

Response: We agree and have deleted these statements.

Lines 145-149: The selected window widths should be chosen based on the goal of the analysis which, generally, is to extract the important patterns from the data, leaving no pattern in the residuals; the goal is not to minimize the residuals. The authors justify their selected window widths based on precedent, but as neither of the cited references used these widths, the basis of their selection requires further justification.

Response: We agree and have revised these widths and reanalyze the seasonal trend.

Lines 208 -247: These two sections are vague; it is unclear how the "bagging model" results were used with the future predictions to estimate tributary flow. Also note the final sentence (lines 246-247) seems like a poor assumption, given the management actions currently underway to meet the updated phosphorus load targets.

Response: We agree. We have added more explanation regarding the use of bagging to simulate tributary flow in our Materials and Methods part. Briefly, we use daily precipitation, maximum and minimum air temperatures, relative humidity, wind speed, solar radiation, and reference evapotranspiration, as model input variables to predict daily river flow rates.

Lines 255, 259, 270-271, 367: I encourage the authors to review statements by the American Statistical Association regarding the use of p-values (Wasserstein and Lazar 2016, Wasserstein et al 2019).

Response: We agree and have deleted these statements.

Figure 1: the panels depicting trends are too small for readers to interpret. The axis labels are barely discernible, the periods of record are difficult to distinguish and without horizontal and vertical lines to guide the eye the resolution is insufficient to infer specific values. The caption also fails to explain a few important details, including the line depicting the Canada-US border. Note that insufficient axis labeling is a problem in most of the figures.

Response: We agree and have revised Fig.1 according to your suggestion.

Line 262 and other places: "Erie" is repeatedly misspelled as "Eire" in figure captions 1, 2, 4, 5, 6, and 8.

Response: We agree and have eliminated this mistake.

Lines 264-266: The authors need to explain how "a lack of consistent observations" justifies the use of linear regression. Too few observations would seem to make any quantitative analysis questionable. See also, lines 276-277.

Response: We agree and have added discussion about this research limitation.

Lines 296-297: The word “excessive” implies a value-judgement; in most contexts it means “too much”. In this context it is a comparison to the long-term trend; implying that it is “too much”, by some measure, lacks context. See also line 320.

Response: We agree and have deleted this statement.

Figures 4, 5, and 6: The vertical axis labels indicate negative loads or ratios. Either the captions should be edited to explain that these are deviations from the long-term trend, or the trends should be added back in so that the units are sensible.

Response: We agree and have revised these figures.

Lines 331-333: The logic here is strained. Peak snowfall and agricultural runoff are not simultaneous, peak runoff is generally from ~April-June, after peak snowfall. How does the observation that peak P concentration and snowfall are approximately simultaneous implicate runoff during snowmelt?

Response: Sometimes, high concentrations occur when discharge is low. This may be because the peak snowfall period coincided with a low discharge period. However, we will check the collected climate, discharge, and P load datasets and analyze it further.

Lines 341-343: The authors need to provide references to support the assertion that conservation practices increased after 1990.

Response: We agree, and we have added references Cousino et al. (2015), Jarvie et al. (2017), and USDA-NRCS (2016).

References:

Cousino, L. K., Becker, R. H., & Zmijewski, K. A. (2015). Modeling the effects of climate change on water, sediment, and nutrient yields from the Maumee River watershed. *Journal of Hydrology: Regional Studies*, *4*, 762-775. http://dx.doi.org/10.1016/j.ejrh.2015.06.017

Jarvie, H. P., Johnson, L. T., Sharpley, A. N., Smith, D. R., Baker, D. B., Bruulsema, T. W., & Confesor, R. (2017). Increased soluble phosphorus loads to Lake Erie: Unintended consequences of conservation practices? *Journal of Environmental Quality*, *46*(1), pp.123-132. https://doi.org/10.2134/jeq2016.07.0248

USDA–NRCS. (2016). Effects of conservation practice adoption on cultivated cropland acres in Western Lake Erie Basin, 2003–06 and 2012. *Conservation Effects Assessment Project–Cropland, Special Study Report*.

Lines 339-360, Figure 8 and Table 5: In many instances the relationships between load and flow are not very good, either visually or as reflected by relatively low r-squared values. This is surprising given that load is equal to flow times concentration, and strongly suggests that the relationship between concentration and flow has changed substantially during the period of record. Also, the pronounced heteroskedasticity exhibited in the load vs flow plots indicates a violation of the assumption that the error term has a constant variance, thus the estimated confidence intervals and equation parameter standard errors are not very representative. This kind of non-constant variance often occurs when the dependent variable has a hard lower boundary (load cannot go below zero) and can sometimes be accommodated by log transforming the dependent variable. A log transformation might also make the relationships more linear, eliminating the need to include polynomials in the regression equations.

Response: We agree, and we will reanalyze the relationships between load and flow.

Lines 387-424: See prior comment regarding lines 208 -247: It is not clear how the authors generated the information in figures 11, 12, and 13.

Response: We agree, and we will add more explanation about how we adopt bagging model to predict P loss.

Lines 407-8: The authors need to explain what characteristics of the diagnostics in this table 6 led them to the conclusion that the applied techniques “effectively captured flow rate dynamics”; a casual reference to the table is insufficient.

Response: We agree, and we have added numeric value of evaluation indices in this paragraph.

Lines 411-412: The authors need to offer context to explain what “the environmentally-friendly target P loads” is referring to. I suspect it refers to what I have included in the references as: Annex 4 Objectives and Targets Task Team Final Report to the Nutrients Annex Subcommittee 2015. I am not aware of any document authored by “Team et al”. Note this

does not even match the reference listed in line 775 with a sole author listed as “Team, T. (2015).

Response: We agree and have revised this citation.

Line 422: The index of agreement is not a familiar metric; what does it convey that differs from r-squared? Is Wang et al the principal reference?

Response: The r-squared (R^2) and index of agreement (d) values range from 0 to 1. If the evaluated model perfectly predicted the observed datasets, R^2 and d should be equal to 1. Index of agreement is more often used in terms of discharge simulation.

Lines 427-428: See prior comment regarding lines 411-412.

Response: We agree and have deleted this statement.

Lines 462-464: The authors attribute declining TN:TP ratios to reduced atmospheric N deposition and support this assertion by referencing Stow et al (2015). Note that Stow et al were quite circumspect regarding N deposition; the authors should not offer such an assertive statement (...can be partially attributed..., line 463) and support it with this reference.

Response: We agree and have deleted this statement.

Lines 464-466: Hellweger et al (2022) do not, in fact, indicate that “Lake Erie’s eutrophication problem may be shifting from P-limited to N-limited”; rather, they use a model-based approach to infer that toxin production could increase if P decreases more than N as a result of management actions.

Response: We agree and have deleted this statement.

Lines 470-471: Internal P release usually occurs under low oxygen conditions. Lake Erie’s western basin, where the algal blooms principally occur, is relatively shallow and rarely stratifies for duration sufficient to cause hypolimnetic anoxia and sediment P release.

Response: We agree and have added this explanation in our discussion part.

Line 533 - 534: The opening statement in the conclusions lacks context and is misleading. The authors claim a knowledge gap of Lake Erie P dynamics, but seem unaware of an extensive body of work that has been published regarding Lake Erie tributary P-loading. See, for example, the recent assessment (and references cited therein) prepared by the Canadian and US governments evaluating how the tributaries (and the lake) have responded since the phosphorus load targets were updated in 2016. This report is available on Binational.net (https://binational.net/2024/09/23/5-year-binational-adaptive-management-evaluation-for-lake-erie-2017-2021-measuring-the-ecosystem-response-to-nutrients/) . Also, within the 47 year record there are numerous signals, which appear at finer time-scales, reflecting management actions and changes in watershed activities over this period. These signals differ among the tributaries, depending on the watershed-specific characteristics. The Cuyahoga does, in fact, show an overall P decrease in Figure 1, and the Maumee shows an early P load decrease, which then becomes obscured by an increase in flow (noting that load = concentration times flow).

Response: We agree and have added this explanation in our discussion part.

Reviewer 2

General Comments

The study provides a broad-scale analysis of long-term phosphorus (P) trends in eight tributaries to Lake Erie, utilizing a large dataset of publicly available water quality data and applying novel statistical techniques. While the study offers valuable insights, it currently lacks clear research questions or objective. The primary focus appears to be on statistical techniques, but the hydrological interpretation of the results and their potential implications for water management are vague and insufficiently supported. The study spans a long time period (60 years) but does not frame its findings around specific research questions, limiting its impact. A stronger hydrological and environmental context is needed to improve its significance.

The discussion section is particularly broad and relies heavily on extrapolations and generalized explanations. Despite its potential to effectively summarize existing water quality data, substantial revisions are needed for the study to reach publication level and contribute meaningfully to the hydrological community. Providing clear research question, refining statistical interpretations, incorporating concentration analyses, and integrating a stronger environmental context will greatly improve the manuscript.

Key Areas for Improvement:

1. Context and Study Site Description:

o The study lacks a detailed description of the study site. In addition to a conventional study site description, a thorough characterization of each sub-watershed land cover and land use (both historical and current) is necessary. This should also include past and present management practices.

Response: We agree. We will add more information in Table 2.

o A map of the watershed, showing the location of the tributaries and data collection stations, should be included in the methods section.

Response: We agree, and we will revise Fig. 1 and add the location of the tributaries and data collection stations.

2. Research Questions and Justification:

o What are the authors' expectations for the observed trends?

Response: Thank you. And we expect the observed trends may decline according to previous efforts, while our 2014-2021 period analyses did not show a clear decline trend in most of tributaries.

o Why was the time span of 1974–2001 selected? There are mentions of changes in trends after 1990, but no explanation is provided. What significant changes occurred around 1990 that might explain these trends?

Response: We agree. The evident trends observed after 1990 are primarily attributable to increased adoption of conservation practices. We will further elaborate on this point in the revised manuscript to provide clearer explanations.

o Why were these specific rivers chosen? What is their significance to the lake's nutrient budget? What proportion of the total watershed do they drain?

Response: We chose these eight tributaries because they are the highest P-loading tributaries to Lake Erie (Han et al., 2012)

Reference

Han, H., Allan, J. D., & Bosch, N. S. (2012). Historical pattern of phosphorus loading to Lake Erie watersheds. *Journal of Great Lakes Research*, *38*(2), pp.289-298. 656 https://doi.org/10.1016/j.jglr.2012.03.004

3. Analysis of Concentrations vs. Loads:

o The study focuses on river flow, TP, SRP, TSS, and TN loads but does not analyze concentrations. Concentration data is essential for understanding loading dynamics both in the present and future. Without it, the interpretation of trends is incomplete.

Response: Agreed and we will add concentration analysis in our Result section.

Specific Comments

Line 90: The main objective of the study should be explicitly stated, followed by specific objectives. The methods section should then describe how each objective was addressed.

Response: We agree, and we will add more explanation to state our objectives more clear.

Line 100: The methods section should begin with a study site description subsection.

Response: We agree, and we have added more description about the study sites in this section.

Line 104: The study presents results in terms of loads but does not explain how these loads were calculated. Were data aggregated? How were missing values handled?

Response: We agree. In our study, we use average daily data for each month.

Line 106: Database citations should follow standard citation formats (e.g., NWIS, year), with access dates listed in the references. Please check this throughout the whole manuscript as it happens multiple times.

Response: We agree, and we have double checked our citations to make them follow the same citation format.

Line 120: Table 1: Are there two grand Rivers? You named them differently throughout the text. Please be consistent.

Response: We agree, and we have added brackets to differ these two Grand Rivers : (Ohio) and (Ontario)

Line 128: Statistical Tools: Clearly link each statistical method to the specific research question or objective it addresses. The mathematical details are well-presented, but the hydrological meaning of each method should be clarified. For instance, what is the difference between a long term trend and a long term monotonic trend (methods 1 vs 2 .)

Response: We agree, and we will add more explanation for each statistical tool to clearly explain our objectives of using each method.

Line 132: The use of the word seasonal here is confusing. In other parts of the text is used for climatological seasons but here the meaning is different. Can you replace it for a different term, so it is not confusing for the reader?

Response: We agree and will revise this statement.

Line 185: Were the GEV models used to investigate flow pattern or P loading, as expressed at the beginning of the sub-section? Please clarify.

Response: We agree and will add more explanation for GEV model objective.

Line 205: The statement about extreme events and return levels should be clarified, as they are typically defined using hydroclimatological data rather than calendar dates.

Response: We agree, and we will revise this statement.

Line 207: The study uses machine learning to predict river flow, yet presents predictions for P loads. It assumes concentrations remain stable, which is a significant assumption requiring more justification. The methodology for this transition should be explained in greater detail.

Response: We agree, and we will add more explanation for bagging model-based P predictions.

Line 250: An analysis of nutrient concentrations is necessary to complement the discussion on nutrient loads. Especially as down the line you venture on the impact of management practices, and climate change.

Response: We agree, and we will add concentration analysis in our Result section.

Line 255: “suggesting improvements in mitigating soil erosion”. This belongs to the discussion section.

Response: We agree, and we have deleted this statement in our Result section.

Line 262: Figure 1: The study site map should be placed in the methods section, while a separate figure presenting trends in a readable format should be included in the results.

Response: We agree, and we will revise Fig. 1, add the location of the tributaries and data collection stations, and put it into Material and Methods section.

Line 276: The management of missing data should be addressed in methods. In addition, the details for the linear regression should also be explained in methods. Please check this as it happens multiple times throughout the text.

Response: We agree, and we will add more explanation regarding missing data and details for the linear regression.

Line 292 and 304: Clarify why some figures show negative loads and ensure figure captions explicitly state differences in methodologies. In addition all these changes and exceptions in methodology should be explained in methods.

Response: the negative values of nutrient loading mean the gap between average trend value and observed value in that day. We have added more explanation in our Materials and Methods section.

Line 320: What do you mean by excessive? Greater than what?

Response: We agree and have deleted this statement.

Line 328: This is the first time you talk about concentration. It needs to be mentioned in methods. The analysis should be expanded as was suggested before. This analysis is not enough.

Response: We agree and will add concentration analyses.

Line 336: Figure 7: is this a trend or the monthly average over a time period?

Response: Fig 7 represents the monthly average, and we have revised statement to make it clearer.

Line 339: Subsection 3.2: A correlation between flows and concentrations would be more appropriate than between flows and loads, as flow is already a component of load.

Response: We agree and will delete this part analysis.

Line 340: None of these analyses was explained in the methods section.

Response: We agree and will add more explanations in the method section.

Line 341: Conservation practices should be described in the methods section, including historical changes and their expected effects on nutrient loads.

Response: We agree and will add more explanations in the method section.

Line 380: Figure 10: The caption needs to be re-arranged. First explain the figure and the methods and details for all the rivers. Then, explain Rouge River that is different. In addition to clarification in figures captions, you need to explain what you did for Rouge River due to lack of observations in the methods section.

Response: We agree. We will re-arrange Fig. 10 caption. For Rouge River, we will add explanation in Discussion part discussing the limitations of our work on the lack of observations.

Line 410 to 412: This sentence belongs to methods.

Response: We agree and will add this sentence to the method section.

Line 412: Clarify whether predictions refer to TP, SRP, or both. The assumption that increased flow directly translates to increased loads is overly simplistic. Biogeochemical interactions, seasonality, and other factors must be considered. On the one hand, the study needs a deep analysis of concentrations, land covers and P management practices in each sub watershed. In addition, it also needs to be more conservative with extrapolating conclusions, that are not a direct result of the applied methods.

Response: We agree, and we will revise Results part regarding to machine learning-based P loading prediction.

Line 435: indicating or assuming? do you assume the concentration will be the same? or the analysis indicate pollution will persist? Please clarify.

Response: We assumed the agricultural management remain the same as 2021, and we will add more explanation in Materials and Methods section.

Line 444: This is confusing. Is this a decline in flow rates or loading rates. In line 439 is stated that annual flow rates increased.

Response: Before 1990 we saw a decline in loading rates in Maumee River, while loading rates tend to rebound after 1990, and annual flow rate tend to increase.

Line 455: Ensure that conclusions are based on presented analyses. Be cautious with causal language such as “this trend is driven by.”

Response: We agree and have deleted this statement.

Line 464: If reduced atmospheric deposition is influencing the TN:TP ratio, why is this trend not observed across all rivers?

Response: We agree and will delete this statement and try to find more reasonable explanations.

Line 473 to 482: Correlations for nutrient concentrations should be analyzed. Additionally, this section should acknowledge and engage with existing literature on the topic. There is no novelty in these results. This has been discussed and shown in the literature for decades now.

Response: We agree and will add concentration analyses and revise our Results part.

Line 496: This paragraph needs to be merged with the previous paragraph.

Response: We agree and have deleted this statement.

Line 502: You can reach this conclusion if you present an analysis of the concentrations. Did the concentrations increase? Is this for TP or SRP, TP tends to be correlated with TSS hence is attached to flow rate and to extreme events, SRP tends to be more stable.

Response: We agree and will add concentration analyses.

Line 510: Is there any study that analyzed the results of this policy (balanced P fertilization approach)?

Response: Not yet. To address this, we will thoroughly review the literature to identify and discuss studies that have evaluated the outcomes or effectiveness of the balanced phosphorus fertilization approach. We will integrate relevant findings into our revised manuscript to provide a clearer and more comprehensive explanation.