American Ecological Engineering Society

Journal of Ecological Engineering Design
journals.uvm.edu/jeed

Peer Review File

Assessing Environmental Drivers of Denitrification in Restored Riverine Floodplains

Danielle Winter Lay[1], Sara W. McMillan[2], Jacob D. Hosen[3], Sayan Dey[4], Gregory B. Noe[5]

[1]Purdue Extension, Purdue University, West Lafayette, Indiana, United States
[2]Agricultural and Biosystems Engineering, Iowa State University, Ames, Iowa, United States
[3]Forestry and Natural Resources, Purdue University, West Lafayette, Indiana, United States
[4]Taylor Geospatial Institute, Saint Louis University, St. Louis, Missouri, United States
[5]U.S. Geological Survey, Florence Bascom Geoscience Center, Reston, Virginia, United States

Editors
Mauricio E. Arias, Editor in Chief
University of South Florida

Eric D. Roy, Associate Editor
University of Vermont

Reviewers
Anonymous[1]

Jonathan Czuba[1,2]
Virginia Tech

Nate Jones[1,2]
University of Alabama

[1]Round 1
[2]Round 2

Reviewer A
Nate Jones

Reviewer summary to be shared with the author and editors:
This study examines potential denitrification rates across four floodplain sites: agricultural, restored prairie (n=2), and restored wetland sites. Denitrification was measured using DEA assays, with both ambient and potential measures. (However, it's important to note both assays are technically measures of potential denitrification). Then authors measure soil properties, plant communities, and hydrogeomorphic position to examine drivers of measured denitrification. They do this using an ANOVA, fixed effect models, and boosted regression trees. Across sites, the highest denitrification potential rates were observed in the restored wetland and lowest denitrification potentials rates were observed in the agricultural floodplain. Within sites, soil properties and plant communities played a significant role in predicting potential denitrification rates. However, hydrogeomorphic variables were not significant.

This is an interesting study that adds to our understanding of biogeochemical processing and nitrogen removal in restored floodplains. The document is well written, the study design appears to be robust (except one notable exception), and the analyses appear to be appropriate. Below are several suggestions to make the manuscript stronger. Most of these comments are aimed at making the document more accessible to JEED readership; however please note the important points about characterizing plant communities. My suggestion is to ask for revisions.

We appreciate the thoughtful comments from the reviewer. We have made changes throughout that improve the manuscript. We have numbered the comments from each reviewer to allow us to reference them throughout the document.

Detailed reviewer notes to be shared with the author and editors:
Broad Comments:

1. My one major concern – there is very little information provided about the plant characterization at the site. The authors use a combination of onsite plant identification and supervised classification of remotely sensed imagery. There are no results provided for the supervised classification, and it's unclear why the remotely sensed data was needed (i.e., instead of simply measuring plant communities at sampling locations). Perhaps authors could omit this component and/or provide more information in the results.

We collected plant species data in Fall 2019, approximately a year into the study. We used ground-level site photographs to confirm that the dominant vegetation species at each site did not change over time and found limited changes in vegetation communities between years. We used remote *sensing data to verify that vegetation communities were stable. We have revised the text as suggested and provided a summary table with the results from the supervised classification of vegetation (Supplemental Table 1). Text in the methods has been revised as follows (Lines 273-275): "Classification using remote sensing was consistent with survey observations and classification accuracy for each growing season was consistently high (Supplemental Table S1)"*

2. Given the readership of JEED is largely composed of engineers (and not floodplain biogeochemists!), I would love to see this study's rate measurements put in the context of other measurements reported in the literature. There is a good amount of literature on floodplain denitrification rates using DEA assays (including previous work from coauthors!) and other measurement techniques (see https://doi.org/10.1007/s10533-004-6016-4, https://doi.org/10.1007/s10533-014-9993-y, and https://doi.org/10.1021/acs.est.5b02426)

We agree that this context is helpful. We have added the following text to the discussion (Lines 485-493) to add this information: "Our results suggest that practitioners can design floodplain restorations to include wetlands to maximize reduction of downstream N transport. This is consistent with other studies that demonstrate the importance of using floodplain wetlands when designing strategies for NO_3^- retention in fluvial systems (Forshay and Stanley, 2005, Roley et al. 2012, Wollheim et al. 2014, Hanrahan et al. 2018). These studies identified higher soil organic matter (Roley et al. 2012) and greater water residence time (Wolheim et al. 2014) as key elements driving increased uptake of NO_3^- floodplain wetlands. Both soil conditions and water residence time within a floodplain wetland are key. When floodplain wetland restorations are situated in areas where water residence time is low, for example in smaller streams and rivers, NO_3^- retention was reduced (Jones et al. 2015)."

3. Given the readership of JEED, it may be helpful to frame the results around potential design parameters that could be used to optimize N removal. Specifically, N removal is a function of N flux, residence time, and denitrification rate. This study is obviously focused on the "rates" component. The authors hint at this throughout the document, but I would encourage them to make this a stronger part of their discussion.

We appreciate this suggestion. Connecting relative differences in denitrification rates to design and management factors that are actionable from an engineering perspective was a major goal of our study. We added a statement to the introduction (Lines 83-84) stating that N removal is a function of N flux, residence time, and denitrification capacity.

Residence time is difficult to parameterize in floodplains, especially given the complex hydrology. We used HEC-RAS (which is frequently used by designers) to estimate flooding frequency as a surrogate. Unfortunately, residence time is not an output. We added a discussion of residence time *as a potential driver of high denitrification potential in wetland features in these floodplains that was not reflected in the hydrogeomorphic metrics that we assessed in this study (Lines 654-664).*

4. In the discussion, a greater distinction should be made between floodplains and floodplain wetlands. (See https://doi.org/10.1002/hyp.11153 for a more nuanced discussion.) When thinking about soil assay experiments, sampling in a wetland (or area that is wetter and has great OM accumulation) it is somewhat intuitive to expect greater DNF rates. Thus, it makes complete sense that the WR-W site would have higher DNF than the prairie sites.

We revised our discussion to make this distinction in Section 4.2 (Lines 485-503); please also see Response #2.

5. I would love to see (i) time series of streamflow -- even if it's a nearby gage, it would just be good to get an understanding of the flow regime, (ii) time series of river water chemistry (especially nitrate), and (iii) time series of the sampled DNF rates. Perhaps these could be incorporated as boxplots into figure 3?

i.) We added time series plots of river discharge during our study period to illustrate the flow regime and included thresholds at which we anticipate inundation at each sampling location. We also note that these are not co-located with our study sites and should be interpreted with that in mind (Figure S2). We added a table of inundation frequency by sampling location (Table S3) using HEC-RAS results during the study period.

ii.) In our original manuscript, we included a table of river chemistry by sampling date in the supplemental information (Table S2). We do not have a time series of river chemistry at a finer temporal resolution.

iii.) We added boxplots of denitrification enzyme assays that were incubated with river water and glucose- and nitrate-amended water parsed by season and year to the supplemental data (Figure S3).

Detailed line-by-line comments:

6. L26 The first few lines of the abstract focus on hydrologically connected vs hydrologically disconnected wetlands. This isn't what the study is about and is distracting to the reader. Consider revising to help focus the reader.

We agree that this framing in the abstract was not needed and was removed. Given that the concept of hydrologic connectivity is foundational to characterizing floodplain function and targeting restoration strategies, we revised text in the introduction (Lines 67-71) to clarify and address how we estimated this characteristic. We added a statement to our methods section

(2.3.2) stating the metrics we used for hydrologic connectivity (Lines 224-226). Specifically, we included height above the nearest drainage area (HAND), horizontal distance to nearest drainage (HDND), slope to the nearest drainage, and cumulative days of flooding based on HEC-RAS modeling results.

7. L46 The abstract states "denitrification will be maximized in riverine floodplains with enhanced surface or subsurface hydrologic connectivity." While this may be true in many cases, I don't think the results support this statement. At best, authors can say that different patch types experience different denitrification rates. However, hydrologic connectivity was not explicitly measured in this study.

We assessed floodplain connectivity in several ways in our study. We included height above the nearest drainage area (HAND), horizontal distance to nearest drainage (HDND), and cumulative days of flooding based on HEC-RAS modeling results. We added a statement to our methods section (2.3.2) explicitly highlighting what metrics we used for hydrologic connectivity (Lines 224-226). Please also see the response above.

We agree that subsurface connectivity is not characterized explicitly in this paper, but we argue that the metrics above provide a reasonable estimate of surface hydrologic connectivity. We have updated the final sentence of the abstract (Lines 47-48) to focus on the contrast of patch types as suggested. We have replaced "with enhanced surface or subsurface hydrologic connectivity and diverse native vegetation" to "that are restored as wetlands with diverse native vegetation and enhanced hydrologic connectivity."

8. L73 For some reason, it seems like 10% N removal is a common estimate for N removal by floodplains in the literature. I'm not sure this is a problem (as posed), because floodplains provide a suite of other ecosystem services (flood retention, biodiversity, etc). Thus, I would suggest rephrasing this as "an opportunity to increase N removal."

We agree that this could be stated more clearly. We aim to highlight how restoration/conservation projects can increase return-on-investment in terms of N removal through optimized design and management. Thus, we have used the following edit in the discussion to highlight this perspective (Lines 442-443): "This highlights that the capacity of these sites to improve downstream water quality is not being fully realized."

9. L97: In general, I'm not a big fan of using broad terms like hydrologic connectivity, sustainability, or resilience. They are important because they can be unifying concepts (see https://doi.org/10.1111/1752-1688.12739), but they also cause confusion because they lack specificity. Typically, I like to introduce hydrologic connectivity and broadly define it (i.e., the water mediated movement of materials and energy across watershed components), and then be specific about what is meant by connectivity (i.e., the

timing, duration, magnitude, and mode of hydrologic fluxes). In this case, what varies over hourly to decadal time scales – overbank flow, hillslope fluxes, hyporheic exchange flows?

We agree that such terminology needs to be properly defined if used. For the reasons described above (please refer to Responses #6 and #7), we have kept the term hydrologic connectivity in the manuscript. To improve clarity as suggested by the reviewer, we have clarified the definition of hydrologic connectivity in the introduction (Lines 66-67) and have identified the metrics that we used to estimate hydrologic connectivity in the methods section (Line 224-225).

10. L102: "Light-touch restoration strategies" is a new term for me. If that's not commonly used in the field (or contemporary literature), my suggestion is to omit the term from the document. consider using something other than restoration… or define it. incorporate some concepts of self-design here. We are hoping that readers will be able to differentiate

We have taken the reviewer's recommendation to remove the term "light-touch strategies." Here, we were referring to restoration projects that do not involve geomorphic alterations. We have opted to explain these as passive and active restoration strategies – terms that are commonly used in the restoration literature. We modified our introduction (Lines 108-111) and methods (Lines 144-146) sections to clarify our terminology and added references accordingly.

11. Figure 1: I would love to see a hydrograph from the region in this figure. It would just help me understand the flow regime. Also, it would be great to have a callout of where this site is at the regional/continental scale. Finally, I LOVE site pictures. Given this study occurs across three very different patch types, it would be informative to have pictures of the different ecosystems.

We added a hydrograph to the supplemental information (Figure S2). This hydrograph includes a gage along the Wabash River, which is downstream of the sites, and a gage along the Tippecanoe River, which is upstream of TR-P.

We revised our site map to include a callout where sites are at a regional scale (Figure 1). We also included site pictures in the supplemental information (Figure S1).

12. Figure 3 – intuitively, it appears there's a mistake at TR-P with the alphanumeric classes. Also, see note above. It would be great to see boxplots of nitrate concentrations and DNF rates incorporated into this figure.

We order our x-axis based on design/management strategy, rather than alphabetic order. The order WR-A, TR-P, WR-P, and WR-A represents a gradient of intervention and helps the reader group the prairies.

Figure 3 depicts ΔDEA (DEA_{CN}-DEA_{River}), which is the difference in DEA rates for experiments conducted with either optimal or ambient (riverine) concentrations of nitrate. Further, in each experiment the nitrate concentration is consistent as described in the methods (Lines 193-196): "We measured DEA in slurries with river water (DEA_{River}) and C- and N-amended water (DEA_{CN}) (47.2 mg $C_6H_{12}O_6$ /L and 60.7 mg $NaNO_3$/L) to represent ambient and optimized concentrations of NO_3^- and labile organic C, respectively."

13. Figure 4. The x axis on Figure 4A is tough. I'd love to have the actual factors spelled out.

We cannot list the full factors because of space limitations. We do include all abbreviations in the caption for reference.

14. L403 - The results from this study do NOT completely support the statement that "the floodplain is limited by supply of nitrate." Yes, I agree the dDEA measurements suggest there is substrate limitation. However, if you conceptualize the floodplain as bioreactor, there are other factors that control N removal. Notably, short residence in the floodplain could limit N removal. Thus, authors should be very careful about describing their results here. (Note my suggestion above about design parameters.)

We agree that limitation is a multifaceted issue. Here, we designed the dDEA experiments to specifically isolate the effect of substrate limitation. DEA rates (redox optimized) at NO_3^- concentrations in the river water were well below potential maximum rates (redox optimized, no C or N limitation) and assert that this is evidence for substrate limitation. We agree that our language was not clear. Thus, we amended our language as follows (Lines 440-442): "Despite excess riverine N export from agricultural production throughout the study sites' watersheds, experiments showed that for all floodplain sites studied, incoming NO_3^- was not high enough to support maximum rates of soil denitrification". This highlights that we are focused specifically on how rates are controlled by substrate, rather than other sources of limitation.

We anticipate that we can design and manage these systems to operate closer to this demonstrated capacity by increasing the quantity and concentration of NO_3^- delivered to floodplain soils. To clarify this point, we amended the latter half of this statement as follows (Lines 442-445), "This highlights that the capacity of these sites to improve downstream water quality is not being fully realized. Thus, greater annualized N removal can be supported by maximizing hydrologic connectivity and increasing residence time when concentrations of NO_3^- from the river are highest."

15. L446 – This study does not support the statement that heterogeneity and N removal via DNF are correlated. The study does not quantify heterogeneity. The authors may be able to make this argument by combining their results with contemporary literature. However, this study's results highlight differences in DNF rates across three different landscape patches.

This is an excellent point. To resolve, we removed this statement and replaced it with the following (Lines 485-486): "Our results suggest that practitioners can design floodplain restorations to include wetlands to maximize reduction of downstream N transport."

16. L473: See caution above about "connectivity." It's unclear what "enhanced" connectivity means here – more details are needed.

Consistent with the prior responses above about connectivity (6, 7, 9), we have modified this language to be more precise. Text now reads as follows (Lines 493-495): "We measured the highest denitrification rates using river water in the wetland floodplain where hydrologic connectivity, based on geomorphic characteristics, was estimated to be greatest in terms of frequency of overbank flooding and restored wetland hydrology."

17. L506 – This is a good point! Thanks for including.

18. L593 – It's not super surprising the HEC-RAS results were not useful. I'm not familiar with the Dey study, but it appears to be a network scale modeling exercise (based on the title). Given that velocity distributions are already a major point of uncertainty for hydrodynamic models, it appears the model lacks specificity needed to produce useful covariates. My suggestion would be to remove the HEC-RAS results from the study.

We appreciate this perspective. However, there is value in including the HEC-RAS finding because of the generalized use by practitioners. If the goal of modeling is to support design for water quality, our results show the limitation of HEC-RAS for that purpose and suggest that practitioners consider other models. Thus, we argue that this result may be of interest to some of our target audiences.

19. L602 – It's interesting that HAND was not a useful metric for predicting DNF rates. HAND is becoming a popular metric for inundation mapping, and it is readily available across most of CONUS. I think its important to highlight that it was NOT a good predictor of DNF rates in this study. However, as the authors point out, there still may be connection between hydrogeomorphic features (see https://doi.org/10.1029/2018WR023527) and N removal processes. It may be worthwhile for authors to further explore this connection intext.

We highlight that HAND is not a predictor by amending the following statement in the discussion (Lines 635-636), "Yet, quantitative metrics that directly measure hydrologic connectivity, including HAND, HDND, flood velocity, and days inundated, were not strong predictors of denitrification."

We amplified the connections between hydrogeomorphic features and N removal (648-650): "Neglecting groundwater connections may have underestimated the overall degree of connectivity of primarily groundwater-fed geomorphic features, thereby limiting the explanatory power of hydrologic connectivity metrics included in our study."

Reviewer B
Anonymous

Reviewer summary to be shared with the author and editors:
The authors provide a well-organized, easily interpretable manuscript that addresses an important issue in the field of soil science and ecological design. There are a few areas in the introduction that could be expanded upon to enhance the paper's impact. The paper is presented through the lens of the importance of floodplains in their ability to denitrify excess N that causes hypoxic conditions in the Gulf. Reading the manuscript from this perspective, I was left wondering about how frequently the floodplains were actually inundated, as this may impact both wetland and non-wetland areas' denitrification capacity. Denitrification capacity was limited by the the lack of N within the soil system, and inundation from floodwaters could ultimately lead to higher denitrification rates by supplying this substrate that is lacking. Thus, the authors stated that "restoring floodplains as wetlands maximizes denitrification potential." The authors might consider also commenting on how inundation patterns themselves might affect denitrification capacity of non-wetland floodplains.

Detailed reviewer notes to be shared with the author and editors:

1. Line 57: Add "has" between sediment and diminished

Thank you. We made this correction.

2. Line 72: Consider expanding on this point. You're suggesting that average restoration efforts will not result in a meaningful reduction of hypoxia in the Gulf – is there any existing research that supports the idea that soil denitrification capacity (or just NO3 removal rates in general) in particular could be enhanced in restored floodplains using ecological engineering methods?

We referenced a study focused on enhancing N removal through more intentional site selection for wetland restorations after Lines 103-107.

---

3. Line 77: I also wonder if there is existing research that addresses potential N retention via denitrification versus physical retention of potentially N-rich sediment. If not, you could clarify that there is also another way N could be reduced (via sediment retention).

This is a good point that has been clarified in the text as follows (Lines 57-60): "Floodplains temporarily retain high levels of riverine N, which can be subsequently transformed and permanently removed from the system via denitrification "

4. Lines 81-88: I think this is a really interesting point about the actionability of soil science research, which fits in nicely with the aims and scope of JEED. Consider clarifying the apparent disconnect between soil science research and what is actionable from an engineering perspective: we know which soil conditions lead to the highest denitrification capacity, but this is sometimes difficult to achieve from an engineering perspective, which occurs on less granular scales (i.e., plot, field, catchment).

We agree that this is a good point, we have modified the text as follows (Lines 94-101): "Requisite soil conditions for denitrifiers to remove excess NO_3^- are well-understood: a supply of labile organic matter or other reduced compounds and anaerobic conditions (Reddy and DeLaune 2008). However, how ecological engineers can design floodplain restoration practices that maximize these desirable soil environmental conditions is less clear. Assessing how plot-, field-, and catchment-scale characteristics that are more directly manipulated by design and management relate to denitrification is necessary to generate more actionable strategies from an ecological engineering perspective to increase NO_3^- removal via denitrification."

5. Line 120: Consider clarifying which sites were mitigated with intensive vs. light-touch here.

Rather than using intensive and light-touch restoration approaches, we introduced "active" and "passive" restoration strategies, which should be more familiar terms for readers. We have made that change in the introduction (Lines 108-113): "Passive restoration strategies focus on minimizing or eliminating disturbances, such as agricultural land retirement or grazing exclusion. Active restoration strategies aim to accelerate the rate of ecosystem recovery after the disturbance is addressed through additional design and management actions. Active restoration approaches span a continuum of effort and investment, with more expensive, intensive interventions typically resetting ecosystems closer to pre-disturbance conditions (Atkinson and Bonser 2020; Jones et al. 2018)."

We also made that change in the methods (Lines 144-146): "The two prairie sites underwent minimal geomorphic modification whereas the wetland site, WR-W, was geomorphically modified *to restore wetland hydrology, representing more effort and investment along the continuum of active restoration approaches."*

6. Line 149: Is there any existing information on fertilizer application for your sites?

We do not have quantitative information on fertilizer applications for these sites. We state that these sites are farmed with conventional practices typical for Indiana and apply fertilizer based on recommended rates (Lines 175-177).

7. Line 161: Was there one sampling event each season?

Yes. We revised our text (Lines 191-193) to indicate this: "We collected soils with a 2-centimeter diameter stainless-steel soil probe from 0-10 cm depth at the nine sampling locations within the four sites in one sampling event during each season."

8. Line 229: Please clarify what ERDAS Imagine is.

ERDAS Imagine is a software package to help process and analyze raster-based imagery. We have added this description to the text in the methods (Lines 266-268): "We used our field surveys as training data for the supervised classification of vegetation communities in the ERDAS Imagine remote sensing software package (Hexagon AB, Stockholm, SE) based on surface reflectance raster images."

9. Line 272: How many total observations were available to use in the model? It seems like there are a relatively limited number of observations – if so, could you comment on how well-suited your model is for this type of dataset?

The number of observations for each site are stated in the caption of Table 3. We added the total number of observations for each type of denitrification enzyme assay (i.e., ambient (DEA_{River}) and potential (DEA_{CN})) to our description of training and testing datasets (Line 311). We had 284 samples for DEA_{River} and 251 samples for DEA_{CN}. We used a low learning rate to account for the sample size based on guidance from Elith et al. 2008 on modeling samples with 250 observations who assessed BRT performance with samples of 250, 500, 1000, 2000, and 6000.

10. Line 295 & 338: Consider deleting "the" from "the site"

Thank you. We incorporated this suggestion.

11. Line 403: Consider re-framing this line/logic. Since the introduction framed the study within the context of reducing the downstream N mobilization, this reasoning was confusing to me. In order for the floodplains to function as floodplains, they must be

inundated with floodwaters. This line is essentially saying that floodwaters would need to be delivered to the floodplains in order for them to have higher denitrification capacity to remove excess N from those floodwaters. I think this could be re-framed as: the study sites are inherently N-limited, which may result in lower denitrification capacity, which lessens their effectiveness for floodwater N removal. The literature suggests higher baseline N availability in a soil environment can result in higher N-removal capacity. So, over time, more frequent connectivity leading to N substrate build up could potentially prime the soil for higher N removal.

This is a good point. We made the following adjustment based on this comment (Lines 440-445): "Despite excess riverine N export from agricultural production throughout the study sites' watersheds, experiments showed that for all floodplain sites studied, incoming NO_3^- was not high enough to support maximum rates of soil denitrification. This highlights that the capacity of these sites to improve downstream water quality is not being fully realized. Thus, greater annualized N removal can be supported by maximizing hydrologic connectivity and increasing residence times when concentrations of NO_3^- from the river channel are highest."

12. Line 489: You may want to reconsider highlighting dominant vegetation type as the most important predictor of denitrification, given that the model performed just as well without this variable. Alternatively, consider adjusting how this finding is presented (the current way is a bit conflicting). You could first highlight that vegetation could have a direct or indirect effect, and then mention how the model performed without it.

Our intent in the discussion was to make clear that, even though vegetation was the top predictor in our model, vegetation may be having direct or indirect effects. Given this comment, this was not conveyed effectively. We have modified the text in the discussion as follows to more explicitly discuss direct and indirect effects of plant species on denitrification rates (Lines 536-543): "We observed clear differences in denitrification capacity across soils underlying different vegetation species. Such differences in denitrification across vegetation types are frequently observed in riparian and wetland ecosystems (Seitzinger et al. 2006; Alldred and Baines 2016), but the degree to which plant species is a driver of higher denitrification rates versus simply an indicator of soil conditions that are more conducive to denitrification is unclear. Similar explanatory power of boosted regression tree models with and without vegetation type as a predictor because vegetation type covaries with soil properties that are key predictors of denitrification, particularly soil organic matter and moisture."

Reviewer C
Jonathan Czuba

Reviewer summary to be shared with the author and editors:
Review of "Assessing environmental drivers of denitrification in restored riverine floodplains" by Lay et al. for the Journal of Ecological Engineering Design (JEED).

This paper evaluated denitrification rates in four floodplains around the Wabash/Tippecanoe River confluence in Indiana. The authors used statistical analysis (primarily) to assess how various factors influenced denitrification rates. Overall, the paper was well written and of broad interest as an "original research paper."

1. The one major comment that I have is that I would recommend that the authors explicitly return to their objectives and hypothesis in the discussion to reiterate how their findings explicitly relate to their hypothesis. I think the content is there, but an added sentence or two would strengthen this connection. I also have a few additional minor comments below.

We now discuss the results in light of our initial hypothesis as follows (Lines 628-630): "We found indirect evidence to support our hypothesis that hydrogeomorphic characteristics would explain variability in denitrification. Specifically, we identified indirect evidence of the role of hydrogeomorphology on modulating denitrification."

Detailed reviewer notes to be shared with the author and editors:
Detailed line comments:

2. Line (L) 72: Indicate whether these 8,000 km2 are of restored floodplain areas or wetlands on the hillslopes or something else. Not clear. I would also suggest that the 8,000 km2 be put into a broader context. The sentence seems like 8,000 km2 is a large area, but only provides a 12% reduction. Maybe add what the total watershed area is and what percent of the watershed is being restored as floodplains(?). That way it will be clearer if 2% of the land area is being restored as floodplains, which provides 12% reduction, which seems like a more balanced statement.

We contextualized the 8,000 km^2 with respect to the total drainage area of the Upper Mississippi River Basin as follows (Lines 78-81): "Yet, even if all 8,000 km^2 potential restorable non-floodplain wetlands in the Upper MRB are rehabilitated, these restoration efforts would only reduce annual loads of NO_3^- by 12%, falling short of the target reductions needed in that sub-basin to reduce hypoxia in the Gulf to less than 5000 km^2 (Evenson et al 2021)."

3. L 77-80: These two statements… "restoration approaches should focus on creating optimal conditions for denitrification to boost NO3– areal removal rates (Olde Venterink et al. 2006; Lutz et al. 2020; Li and Twilley 2021). Requisite conditions for denitrification are the presence of NO3–," suggest that one restoration approach to improve denitrification is to increase nitrate concentrations. While this is technically accurate, the important point is that the overall restoration goal should be to remove nitrate not simply increase denitrification, otherwise I think that can send the wrong message.

We have revised the text to highlight that the restoration goal is to ensure that excess nitrogen is delivered to the right place, rather than increasing overall fluxes of nitrate simply to increase denitrification. We have made the change as follows (Lines 91-94): "Thus, floodplain restoration and conservation projects can increase NO_3^- removal rates by promoting the formation of environmental conditions that are optimal for denitrification and ensuring that as much excess NO_3^- is delivered to these restoration sites as possible."

4. L 110: Suggest changing "anticipated" to "hypothesize" so that you have an explicit hypothesis to support this "original research paper."

We made this change as suggested.

5. Figure 1. I recommend labeling the rivers in panel B. Also, because your panels are awkwardly rotated relative to north and are including the grid tick marks, it makes for cluttered and difficult to follow lat/long labels, particularly in the second panel in A. Instead, I recommend just including 1 or 2 lat/long cross hairs in panel B. Also, scale bars are needed for all 4 panels in A because some are resized relative to B.

We added labels to the rivers in Figure 1.

We agree that the lat/long labels generate some suboptimal formatting. USGS guidelines require that Lat/long crosshairs on all maps. Co-author Noe is employed by USGS and the document must conform to USGS style for him to be included as a co-author. Thus, we are reluctantly keeping these map elements. We moved the site maps in Panel A of Figure 1 to the supplemental document to reduce the clutter and added scale bars to each site map (Figure S1).

6. L 322: Should "bc" be a value or should it either be "b" or "c"? Unclear what "bc" would mean in this case, I think it is statistically similar to both b and c? Maybe point this one out specifically in the caption.

Using compact letter display, 'bc' means that level is both not statistically different than other 'b' levels and not statistically different than other 'c' levels. By convention, commas are not added *when using compact letter display. To clarify these points, we modified our description in the caption as follows (Lines 358-360): "Compact letter display designates which sites had comparable denitrification rates such that sharing a letter indicates there was no statistically significant difference between a given site and all other sites that share this letter based on post hoc pairwise testing."*

7. L 379: Rather than UV, so as not to confuse it with ultraviolet light, I recommend "bare soil."

We agree this could be confusing. We have changed UV to bare soil (BS) in all plots with vegetation in the manuscript and supplemental document.

8. Figure 4: Why are there so many x-axis tick marks? Does there have to be?

The tick marks are a rug plot with each tick mark representing the density of observations across the range. We include the rug plot because this is an important detail for interpreting machine learning models. We have made our description of rug plots clearer in the Figure 4 caption (Lines 431-432): "The x-axis includes rug plots that show the marginal distributions (observation density) of each predictor variable."