Response to Reviewer 1 Comments

The authors present results from a well designed study investigating the influence of winter flooding on bird-mediated pathogen transport in two different rice production system types. Overall, I liked this study but feel the presentation of the results could be improved. There are a few larger issues that if resolved will make this paper suitable for publication. I list them below, followed by detailed comments.

1. Both in methods and discussion the lead author's thesis is cited. I am not sure this is appropriate. The data on bird densities and estimated fecal loads seems highly relevant to this study considering it is the backbone of what is driving the response variables. In the methods, details on how fecal loads were estimated from camera trap data need to be included rather than just referencing the thesis.

   Response: Agreed and changed accordingly. This manuscript is part of a companion study where we assessed the potential of migratory birds to contribute to soil health in rice fields. Bird densities and fecal matter loads were detailed in the other manuscript. At the time of submission to Water, the companion study was under review in another journal, but had not been accepted for publication. We cited the thesis as a place holder. The companion paper has since been accepted for publication and we were able to change the citations to reference the peer-reviewed status.

   Per your suggestion, we further described the fecal matter estimates and added more detail, while still referring readers to the companion paper.

   Companion study:

   Firth, A.G.; Baker, B.H.; Brooks, J.P.; Smith, R.; Iglay, R.B.; Davis, J.B. Low external input sustainable agriculture: Winter flooding in rice fields increases bird use, fecal matter and soil health, reducing fertilizer requirements. *Agric. Ecosyst. Environ.* 2020, *300*, 106962.

   DOI: 10.1016/J.AGEE.2020.106962

   It is not clear why you chose to use PERMANOVA to assess seemingly univariate data. PERMANOVA was designed for distance-based analysis of primarily ecological community data. You are not assessing the whole bacterial community, just two species. I don't think PERMANOVA is providing any additional info that the ANCOVA are not. If you are concerned about assumptions there is a myriad of variance structures and data distributions that could be applied to count data to address issues. See "Mixed Effects Models and Extensions in Ecology with R (Zuur et al. 2009)" or similar references.

Response: Thank you for your insight. We agree, the PerMANOVA wasn't adding any additional information to the study and it has been removed from the manuscript

2. The results are extremely brief. Please expand your description of your results some, including moving the no detect info for some pathogens from the stats methods to results, and include bird and fecal load estimates across treatments. You mention them and even present a figure in the discussion. This is results.

   Response: Agreed. We moved the no detection *of Campylobacter spp., Salmonella spp.,* and *Enterococcus* to the results. We also moved the analysis of the fecal matter and bird abundance to the statistical methods and the results to the results section. We cite the statistics and results from Firth. et al. 2020. (Lines 190-202)

3. The discussion is lacking. There is little effort to link the results of this study to other literature. I have made a couple specific suggestions for expansion but there are many opportunities beyond that to broaden the discussion and include literature.

   Response: Agreed and we expanded the discussion significantly from the original. We discuss multiple ways pathogens could be removed from the environment (sedimentation, biofilms, predation, etc.), speculate on why LN and CF had similar *C. perfringens* detection, and link the detection in soils back to bird abundance and behavior. (Lines 233-294)

Specific comments.

L2: The title suggests you found potential for pathogenic transport. I would frame the title more like a headline that conveys the actual message of your findings.

Response: Thank you for your suggestion, and we agree. We changed the title to "Investigation of pathogenic bacterial transport by waterbirds: a case study of flooded and non-flooded rice systems in Mississippi" (Lines 2-4)

L19: but only significantly greater than conventional non-flooded. Consider writing this in terms of the observed gradients with the endpoints that are significantly different.

Response: We changed the abstract to reflect the significant difference in LF vs CN and added the sentence "There was an observed decreasing trend of *C. perfringens* detection associated with lower bird abundances among treatment groups" (Lines 19-21)

L32: change "have long been" to "are"

Response: Agreed and we made the correction. (Line 33)

Line 40-42: This sentence is awkward. I think the "long been recognized" language here and elsewhere could be dropped. Try and write more directly.

Response: We agree. We changed the sentence to: “The Lower Mississippi Valley Joint Venture, a component of the North American Waterfowl Management Plan (NAWMP), advocates winter-flooding of agricultural fields to create habitat for waterfowl using the MAV” (Lines 42-44)

Line 49-50: Is it collected, translocated, or recycled? N and P bound up in organic matter is processed by waterfowl, and what is not used for growth and maintenance is excreted. How much consumption occurs outside of field and is translocated in, or vice versa, or recycled from within field? I don't expect you to answer that, just recognize it is more complicated. Reference you should consider: https://doi.org/10.1016/j.scitotenv.2014.12.068

Response: Thank you for the reference. We re-worded accordingly to capture the complexity of nutrient recycling by birds and cited your suggested reference. However, we wanted to avoid excessive discussion of N and P pathways, as the focus of the paper is bacterial pathogens.

The sentence now reads, “Migratory birds harboring in flooded rice fields recycle nutrients in fields, depositing nitrogen and phosphorus directly via fecal matter or by transporting nutrients from adjacent habitat” (Lines 53-55)

Line 52-55: I would split this into two sentences, 1 on potential benefits, 1 on potential detriments.

Better yet start this paragraph out with a contrast statement about potential agronomic benefits vs. potential environmental impacts then back up the statement with this info.

Response: We appreciate your suggestion for more clarity and conciseness. We reworded the paragraph, beginning with a contrasting statement about agronomic benefits vs pitfalls:

“Flooded rice fields used by waterbirds, particularly over-abundant snow geese (*Anser caerulescens*) [8], have the potential to provide an agronomic benefit to soil via deposited fecal matter [9–11]. However, waterbirds can act as vectors of certain pathogenic microorganisms that could detrimentally affect humans and other wildlife [12–17]”

And then followed it up with supporting information. (Lines 63-75)

Line 55: Is the term "free living" necessary? Is this opposed to caged?

Response: “Free living waterbirds” typically differentiates between non- or semi-domesticated birds, such as city pond ducks, and wild birds. As you suggest, we deleted for saliency. (Line 57)

Line 67: Is the term indicators neccessary? Are the indicators a cause of impairment or just a way to measure things?

Response: There are two different types of bacteria important to public health: “Frank pathogens,” which are bacteria that are always pathogenic to its host, and “fecal indicator

bacteria," which are bacteria that are present as gut microbiota of warm-blooded animals but indicate the presence of fecal pathogens. It's important that we differentiate between the two because one positive detection of *E. coli*, an indicator bacterium, does not pose the same threat as a positive *Salmonella* detection, a frank pathogen. However, the detection of *E. coli*, above a threshold level, indicates an increased risk for exposures to frank pathogens, like *Salmonella*. Most utilities and regulatory agencies only measure fecal indicator bacteria (FIB) because they're easier to measure and often present at greater number than frank pathogens. Regulatory guidelines and rules are based on FIBs as a proxy for frank pathogens.

We added clarification on these terms in the introduction. (Lines 64-70)

Line 74: Another way to say "of pathogenic impairment to" might be "increasing the likelihood of pathogen transport to"

Response: Thank you for your suggestion, the sentence has been changed to, "This management practice may create a situation where pathogens are incidentally accumulated in rice fields from wintering birds', increasing the likelihood of pathogenic transport to surrounding waterways" (Lines 77-80)

Line 75: The second half of the sentence requires you change "detect and characterize" to "characterize and compare". You are making comparisons among treatments not just random observations.

Response: Thank you for your insight. We changed accordingly: "Therefore, the objective of this study was to characterize and compare waterborne pathogens and fecal indicators between conventional and LEISA farms, and the flooded and unflooded rice fields at each farm." (Line 80)

Line 77: Change The primary… to "Our primarily…"

Response: We made the suggested change. (Line 82)

Line 78: Consider changing "resulting from" to "associated with"

Response: We made the suggested change. (lines 82-82)

Line 78-81: The way your hypothesis is currently structured makes it sound like you actually controlled bird densities. A better way to do this would be to structure a hypotheses about bird densities associated with different flood/management treatments coupled with a hypothesized linkage with pathogen levels/transport. Put another way are you actually keeping flooded fields devoid of birds?

Response: Thank you for pointing this out. We were not discouraging birds from any of the fields. Therefore, per your suggestion, we revised for correctness: "It was hypothesized that post-harvest flooded rice fields managed specifically for overwintering waterbirds would have

a greater incidence of waterborne pathogens or indicators compared to flooded or unflooded fields devoid of birds." (Lines 83-85)

Line 112: Why just March? It would be good to know the timing of potential pathogen transport.

Response: We had two reasons for only sampling water in March: First, assays for pathogen detection in water are expensive, time consuming, and extremely time sensitive. Water samples have to be processed within 24 hrs of their sampling time, otherwise the targeted bacteria die, increasing the risk of a false negative. This study was part of a master's thesis and funds and time limited our ability to take samples over time. Second, because the largest potential for pathogen transport from rice fields occurred when producers drained their fields to prepare for planting, it would cause a large quantity of potentially contaminated water to move into downstream systems. Thus, we mention this in lines 263-265; 283-284. Sampling in March maximized our likelihood of detecting an environmental risk if there was one.

Line 117-123: I am assuming that this whole study came out of the lead author's thesis so I am not sure why you are citing it as a previously published entity here. Why not just say "We estimated fecal matter inputs....And also give us the details on how you went from camera trap info to fecal matter mass.

Response: Thank you and the thesis was merely used as a place holder for an impending publication. The citations have been changed and more details of the fecal matter quantification were added.

Line 161: What are the units for the data? Are all species on the same scale?

Response: We agree with your suggestion. We added the units into the methods section (CFU/100ml or g of soil) and stated that these values were log transformed for analysis (Lines 155-156; 171; 173)

Line 167-169: Should these statements be in the results somehow? I wouldn't mention it here unless it changes how you would have analyzed the data.

Response: We moved the no detection statements to results. (Lines 190-192)

Line 169: Can you better justify this. Do you have a reference you can cite to justify higher alpha?

Response: There is a growing school of thought that $p<0.05$ is an arbitrary threshold and meaningless unless accompanied by measures of effect size and statistical power. P-values encourage scientists to only look at "significant" results, while ignoring non-significance. Furthermore, p-values can be misleading when there is a small sample size, an unbalanced design or non-normal data distributions (as was the case with this study).

Increasing the alpha threshold (or even throwing out the p-value completely) is becoming more common and accepted, particularly in ecological studies:

1. Gaudarta, J.; Huiart, L.; Milligan, P.J.; Thiebaut, R.; Giorgi, R. Reproducibility issues in science, is P value really the only answer? *Proc. Natl. Acad. Sci. U. S. A.* 2014, *111*, E1934–E1934.
2. White, J.W.; Rassweiler, A.; Samhouri, J.F.; Stier, A.C.; White, C.; White, J.W.; Stier, - A C Ecologists should not use statistical significance tests to interpret simulation model results. *Oikos* 2013, *000*, 1–004.
3. Amrhein, V.; Korner-Nievergelt, F.; Roth, T. The earth is flat (p > 0:05): Significance thresholds and the crisis of unreplicable research. *PeerJ* 2017, *2017*.
4. Lemoine, N.P.; Hoffman, A.; Felton, A.J.; Baur, L.; Chaves, F.; Gray, J.; Yu, Q.; Smith, M.D. Underappreciated problems of low replication in ecological field studies. *Ecology* 2016, *97*, 2554–2561
5. Wasserstein, R.L.; Schirm, A.L.; Lazar, N.A. Moving to a World Beyond "p < 0.05." *Am. Stat.* 2019, *73*, 1–19.

We wanted to focus more on trends and less on significance testing in this study. However, we recognize that many schools of thought require p-values for publication. In order to accommodate this, we raised the p-value threshold in order to draw attention to observed trends and acknowledge the uncertainty in environmental data.

This discussion of p-values is highly relevant to our study because there were different results in group differences depending on if we log-transformed the *C. perfringens* data or analyzed it with the raw numbers. The raw numbers showed LF significantly different than CN and CF at a $p<0.05$ level, while in the log-transformed model LF was different from CN only. Log-transformation is a standard practice when working with microbial data and thus those are the results we presented, but we doubt anyone would have raised questions had we reported the model results from the raw data, especially when you examine Figure 1A. In the log transformed data, LF vs CF had a p-value of 0.13. So, is LF actually different from CF?

This is all a moot point, however, because the only p-value we found to be significant was significant at the $p<0.05$ level. We removed the statement because it isn't relevant.

Line 171: Why are you using PERMANOVA? This analysis was designed for community data. You are only looking at 2 bacteria not the entire community? What does this approach gain you that the individual pathogen load ANCOVAs does not? If you are using PERMANOVA you should cite Anderson's paper/papers on PERMANOVA but I am not sure she would agree with you using it here.

Response: You are correct in your assessment. The ANCOVA was sufficient and PERMANOVA was not really necessary so we removed it.

Line 174: Previously you said Enterococci was excluded due to detection in only 1 sample.

Response: Indeed, it was found in only one sample. This has been removed.

Line 179: Need a description of what your group variable was with Mann-Whitney-Wilcoxon Rank sum test. I don't think it was the same as previous analyses.

Response: The group variables were the flooded fields: CF and LF. We added this to lines 183-184.

Line 181: I am all for brevity but surely you can describe your results with a little bit more detail. See later comments on bird density/fecal load data too. Why is this not in the results?

Response: We moved the suggested material (no detection and fecal inputs) to the Results. We couldn't expand further than this, however, because we didn't test any other hypotheses. (Lines 197-292)

Table 2: The table is a bit hard to read. Is there a way for the authors to delineate columns associated with treatments better? Or better yet, scrap table 2 and 3 and provide bargraphs which would be easier for the reader to quickly interpret your results. Given color is free with this journal why not jazz it up a bit.

Response: Good suggestion! We removed all tables and added 3 Figures: Figure 1: C. *perfringens* loads in soil and fecal matter inputs, Figure 2: *C. perfringens* and *E.coli* loads in water, and Figure 3: monthly means of goose abundance by treatment group. We appreciate your suggestion; the figures are much more effective. (Lines 204-214)

Simply bolding the two treatments that were significantly different from each other is not intuitive nor does it tell the whole story. Please use traditional approach of letters (A, AB, AB, B) to show how treatments significantly differed or did not differ.

Response: Thank you for the suggestions; we used letters to denote significance in the Figures.

Discussion in general: There is not much effort to link results to literature in this discussion. There are very few new references incorporated into the paper through the discussion which suggests the discussion needs to be expanded.

Response: We heeded your suggestion and attempted to create a more informative Discussion. We added 15 references to the paper, where most incorporated into the Discussion of the possible phenomena of bacterial removal from water and soil. We also expanded on how management practices could affect bird densities and connected bird habitat choice with pathogen detection. (Lines 230-287)

Line 205-207: I am having a hard time understanding why some of the methods and results are referencing a thesis. Your results section is 3 sentences long. Why not incorporate some of the other related results on differences in birds/fecal inputs, and bacterial measures into this manuscript to beef up the info. Is it planned for a second manuscript?

Response: Yes, this information is part of another manuscript. See the response above. The citation has been changed to reflect the published work.

Line 213: Reference the reader back to table 2 or future bar graph.

Response: Table 2 has now become a Figure.

Line 218-220: Maybe describe this as a gradient with LF having the highest and CN having lowest and significant differences.

Response: We explain this as a trend- in fields that had more birds, we see greater pathogen detection. We also placed the soil *C. perfringens* detection figure next to the fecal matter figure for ease of comparison. (Lines 233-247; Figure 1)

Line 222-224: Need to discuss possible mechanisms for higher birds and bacteria levels in LN vs CN? What is different about these treatments? Are there potential temporal (flooding history) or spatial (Proximity to LF) at play here? This is one example where the reader is wanting more thoughts about what your results mean.

Response: We agree. We expand here: Geese forage in a variety of field conditions, from dry to shallowly-flooded, but typically roost nocturnally in flooded fields or other deeper water bodies less accessible by potential mammalian predators See discussion lines 235-240

Line 232 and Figure 1. Whoa. You are presenting new results in the discussion. Please move this to results.

Should have been presented first in results before referencing in discussion.

Response: The Figure has been moved to results.

Line 242: Are there other ways microorganisms are removed from the water column. Microbial food webs, etc. Another opportunity to discuss your data more.

Response: We agree and thus expand on ways microorganisms are removed from the water column. We address sedimentation, predation, biofilms and temperature induced mortality. (Lines 248-287)

Line 247: Do you have water samples through time? These riser board systems are pretty leaky and still release pulses if storm events push more water through than the system can handle. What if pathogen levels were higher earlier and water was flushed out during storm events?

Response: As previously mentioned, funding and time to process samples limited our ability to obtain more samples through time. However, we believe the period that these fields would pose the greatest environmental risk was when the fields were drained for planting, which occurred in March. We address this briefly in lines 263-265; 283-284.

Storm events could flush contaminated water out of the system, but, again, we were limited on time and funding. Future studies could look at pathogen loads over time.

Line 254-257: Given your results and the current literature, would it be possible to develop a model that predicts pathogen levels with bird density and estimate flock sizes that would be potentially hazardous within your system? Something worth considering or at least mentioning as a future research effort. I think the goal is not to do regular monitoring but develop tools that tell managers when to be concerned and start monitoring.

Response: We agree that this is an excellent idea for a future research. We mention it in lines 304-305.