Round #2

by Karl Cottenie, 2020-02-05 12:37

Manuscript: https://www.biorxiv.org/content/10.1101/611939v3 version Version 3

This preprint merits a revision

I mirror the comments from the two reviewers who applaud the amount of work done by the author in revising and reworking this manuscript. Reviewer 2, however, provided some additional suggestions. The first paragraph of their review provides some context, but the "Few additions" section provides 5 suggestions that should be addressed. In my mind, their suggestion for using linear regression is not necessary, since I think the author wanted to stay as close as possible to the original analysis approaches.

In addition to these changes, I also have some suggestions.

First a statistical question/suggestion: the author writes: "with stream identity fitted as a random blocking factor in each ANOVA;" Does that mean that the restored and adjacent sample sites within a stream were paired, as they should be? I assume it is, but I would then strongly advice to add a plot of the raw data of the richness and diversity, similar to figures 1 and 2, showing the richnesses of the restored and adjacent sites from the same stream connected by a line. If all those lines going up for the headwaters and relatively 0 for the mainstem streams, that could suggest the lack of power to detect a significant interaction effect? Because figures 1 and 2 show the results of the statistical test, this biological meaning can get lost.
-Response: Adding another figure was a great suggestion and has been added to the manuscript. Also, you are correct that the restored and adjacent sample sites within a stream were paired.

I did not pick up on another issue in the first review round, but there is some ambiguity in the text: The manuscript has a sentence: " I required sites in the re-analysis to have received both the bank stabilization and in-channel manipulations treatments (hereafter "revised" sites), although sites receiving riparian reforestation were also included if they received both the bank stabilization and in-channel manipulations treatments. " The section after "although" seems not necessary, because your criteria are actually "received both bank stabilization and in-channel manipulation treatments". Your selection criteria, I think, do not allow you to say anything about reforestation? This ambiguity is reflected in the first sentence of the Discussion: "I hypothesized that stream-channel manipulations would have a more consistent effect relative to the effect of riparian reforestation between headwaters and mainstems, with stronger effects of restoration in headwaters relative to mainstems. " Again, I think the reference to reforestation should be removed from that statement.
-Response: My hypothesis that drove the inclusion criterion was that channel manipulations would have more consistent effects between headwaters and mainstems, while reforestation would really only have effects in headwaters; this could lead to self-fulfilling study if you hypothesize that headwaters are more affected by environmental factors and apply a treatment that differentially impacts headwaters. More generally, the criterion was designed to reduce variation in the number and type of restoration treatments applied to each site in the study.

I would argue that the selection criterion does allow me to draw inference about reforestation because 7 out of the 9 (~78%) revised sites received reforestation treatments, while only 8 of the 13 (~62%) sites in the original study and full sites analysis received reforestation treatments. The study design precludes explicit evaluation of each restoration treatment, both in isolation and in combination, as does the original study from which the data were obtained. While I am precluded from identifying which restoration treatment(s) affected biodiversity, channel manipulation and reforestation treatments are common in restoration efforts. Given the frequency with which all three restoration treatments were applied in the revised sites analysis, I do think that allows for inference and discussion about the general impacts of restoration, as done by Swan and Brown (2017, 2018).

Once I became aware of that, the next sentence: "As there were no significant effects of restoration on any of the community metrics between headwaters and mainstems,..." became also not correct. Your inference is not about restoration, but it is about "stream-channel manipulations". I would thus strongly recommend to change "restoration" with "stream-channel manipulations", and do this consistently throughout the manuscript. This would explicitly acknowledge throughout this manuscript that the results of this re-analysis are very specific to one type of restoration, and thus explicitly identify the scope of the inference possible. I think that in light of this argument, the second paragraph in "Restoration Ecology & Experimental Design" should be changed considerably. The author equally uses "restoration" maybe not indiscriminately, but at least too generally. Also, I would remove the next sentence "and not to necessarily or strictly compare the effects in-channel manipulations and riparian reforestation treatments on biodiversity in restored streams." Your selection criteria indeed do not allow you to say anything about riparian reforestation, because that was not a selection criterion, the reader does not know anything about this condition and how it affects stream communities.
-Response: I would disagree that my inference is not about restoration, as channel manipulations are common restoration treatments, and sites included in the re-analysis received both channel manipulation treatments and most also received riparian reforestation treatments. Moreover, the impetus for my re-analysis was reducing the variation in applied restoration treatments across sites, when I hypothesized channel manipulations would have more consistent effects between headwaters and mainstems while reforestation would primarily affect headwaters. The original study used sites receiving any combination of restoration treatments, from which they made inferences about the effects of restoration. I would therefore argue that my full sites re-analysis has equal, if not more, weight to Swan and Brown (2017, 2018) regarding inference about broad-sense restoration. For example, only 2 out of the 9 (~22%) sites included in the revised sites analysis did not receive reforestation treatments; in the original study, Swan and Brown had 5 out of 13 (~38%) sites that did not receive restoration treatments. Fewer sites notwithstanding, the revised sites analysis actually has greater representation of reforestation and channel manipulation restoration treatments. Although the inclusion criterion did not focus explicitly on reforestation, reforestation was still represented in the revised sites and, arguably, better represented in the revised sites than in the original study. I have revised the manuscript to better reflect that, although channel manipulations were the focus of the inclusion criterion, reforestation was also accounted for and represented in the analyses (lines 101-105)

In "Restoration Ecology & Experimental Design" you mentioned that this re-analysis did not include "the time since restoration" as a selection criterion. Does that mean that you do not think it is important, or if you included this there would be not enough degrees of freedom to perform a statistical analysis? This is an important point, I think, because the main differences between this author's approach and the original 2 papers are the inclusion criteria and some, important, statistical differences. Since this re-analysis did not provide criteria for time since restoration (maybe because this information is not available in any of the original data?), it does illustrate that all these selection criteria are 1) important, but also 2) subjective to a certain degree, and 3) result in trade-offs with sample size and inference. For instance, I argue above that this re-analysis precludes the use of the word restoration in this manuscript, because it is actually stream-channel modifications. Equally important, if the original data have varying degrees of recovery time, this could very well be more important than the type of stream restoration. Given the importance of selection criteria in this manuscript, this should be included in this manuscript. I also think that the tone in the "Statistical Inconsistencies" can still be toned down. While the first paragraph points out the differences between the text and the R code, the second paragraph is different. I would suggest that the author removes the first ("Finally, and of greatest concern, is the wholesale disagreement between the reported analytical procedure and what was actually conducted when analyzing temporal variability.") and final sentence ("Without consulting the supporting information or if no R code was provided, it would have been assumed the results presented in the erratum (Swan and Brown 2018) were derived from the analytical procedure described in the original study (Swan and Brown 2017), just with the corrected dataset; this assumption would have been incorrect.") of that paragraph. The authors did provide supporting information and R code, actually this whole manuscript benefited from the original authors' willingness for open and reproducible science.
-Response: I agree that time since restoration is an important piece of data, but it was not available to me. In fact, it was not available to the authors of the original study (Swan and Brown 2017), and they did note this in their paper. If the data were available, it would be a useful covariate or even predictor/main effect of community response to restoration; including time since restoration in analyses of taxonomic richness and diversity would most probably be possible, as these analyses had more degrees of freedom to spare; however, spatial dissimilarity and temporal variation would likely not be able to include time since restoration in any statistical analysis due to lower degrees of freedom. I have added a brief discussion of the impact of time on restoration (lines 275-283). I do agree that selection criteria can be subjective, but as long as evidence and hypotheses are provided, criteria can be supported and justified.

Regarding the tone of the “Statistical Inconsistencies” section, I respectfully disagree to an extent. Having discrepancies between the written and published methods and what was actually conducted is concerning, and to ‘tone down’ the section is to not provide the appropriate and requisite response to the problem. There is an assumption that what researchers write in their manuscripts is an honest representation of the study, but that assumption was broken and that implicit trust between reader and researcher was lost. As for providing data and R code, this is commendable but with two important reservations: (1) Ecological Applications required that data were deposited, though credit to Swan and Brown for agreeing; and (2), as noted in the previous round of reviews, the only reason R code was provided was because the editor required it of the authors after I notified both the authors and the editor of irreproducible results (personal communications between myself, the editor at Ecological Applications, and Swan and Brown). Also, R code was only provided for 2 of the 3 analyses because the authors claimed the spatial dissimilarity analyses were unaffected and were given the benefit of the doubt; open practices would have been the deposition of all code for the analyses, not selected deposition of what the authors claimed was consistent or correct. Again, I commend the authors for complying, but acquiescing a requirement is not equivalent to providing the data and R code without request.

My re-analysis was facilitated by the compliance to data deposition requirement of Ecological Applications by Swan and Brown and the R code deposition required by the editor, but compliance to requirements is not necessarily open science. Open science goes beyond sharing data and code and is rather an actual philosophy and methodology of science (Hampton et al. 2015, Ecosphere, 6:120; Nosek et al. 2015, Science, 348:1422-1425; Parker et al. 2016, TREE, 31:711-719; Munafo et al. 2017, Nature Human Behaviour 1:0021; Powers and Hampton 2019, Ecological Applications, 29: e01822). Without institutional data deposition policies, like at Ecological Applications and many journals, it is true that this re-analysis would not have been possible because I would not have had data; however, very few journals currently require deposition of analytical code for non-simulations. In most situations where code is not deposited or required to be deposited, Swan and Brown would have misled the readers, without any presumption of intent, because written methods and statistical code were inconsistent. Code was required to detect the inconsistencies, but this will not always be possible. Instead of needing to compare written methods and statistical code, written methods should be reflected seamlessly into the code. In this context, I feel my criticisms and concerns are fair and valid, but I have made minor revisions to the tone in this section.

These are in my opinion only textual changes that should not be a problem for the author to address, and I am looking forward to the next version of this manuscript that provides an example of "something that should be done more often" (reviewer 1).

Reviews

Reviewed by Eric Harvey, 2020-01-15 16:46

General comments

I am satisfied by the author's answer to my previous concerns. I think the manuscript is now well- balanced and offers a fair criticism of a previously published article. The author here accomplishes something that should be done more often: reproducing results from published studies. The peer review process cannot capture everything. As a whole, the scientific enterprise is based on the implicit assumption that even if a few mistaken results are published, the correct consensus should emerge by weight of evidence from syntheses and/or meta-analyses. This is no excuse to be careless or to use weak statistical evidence as confirmation of our *a priori* ideas of what should or should not be. With a constructive tone, I now feel like the manuscript is motivating for more research on the question to disentangle the different issues and that is exactly what that sort of contribution should do.

-Response: I greatly appreciate your fair and constructive comments. Not only did you catch things I had missed, but also had different thoughts on how to structure and tone the manuscript and what additional metrics (i.e. effect sizes) to include for easier comparison across manuscripts.

This entire project was supposed to be just something I did for my own edification after getting home from waiting tables while exploring graduate school opportunities, but it turned out to be a much more complex and involved process than I would have ever anticipated.

Reviewed by Mariana Perez Rocha, 2020-01-23 23:26

Reviewer 2 This is the second time I'm reviewing this manuscript. I definitely enjoyed the reading and I believe all my previous comments were nicely addressed in this new version. The addition of the effect sizes into the revised version (as partial eta2) was great (even though this was not in my first round of comments) and the changing in the title as well. For this second round of review, I just have minor additions and comments. I sincerely give my compliments to the hard work put on improving the last version of this manuscript. I'm still missing some good references to back up the author's hypothesis and expectations, and 'a good take home message' from the author in terms of ecological meaning for this re-analysis (instead of just stating "because I arrived at my conclusion based on a thorough and robust re-analysis of the data, while Swan and Brown (2017, 2018) based their conclusions on the erroneous reporting and implementation of statistical analyses"). After I read the author's rebuttal letter (from the first round of review), I could understand better what has motivated the author to re-analyze the data used by Swan and Brown (2017, 2018), and I believe that this piece of information is valuable. Perhaps, this could be an addition to the last paragraph of the Intro section? The only time I could spot some of this 'why re-analyzing' was in the beginning of 'Concluding Remarks' which is by the end of the manuscript. After reading couple of time this second version, one question came to my mind: - Would running regression analysis (as a Supplementary Material) help in clarifying the 'why' data has being re-analyzed and the (previous) interpretation of the results?

My reasoning behind this is based on the fact that: if the main aim of this manuscript was to re-analyze data, why not presenting an alternative approach (and perhaps more reasonable one) as well? Also, the use of Anova vs. Regression was also raised by Reviewer 1. Broadly, the ANOVA is a special case of a regression model in which all the predictors are categorical. But there is a difference in the application of "ANOVA" and "regression (analysis)". ANOVA is a tool to check how much the residual variance is reduced by predictors in the models, whereas the regression analysis aims to quantify effect sizes in terms of "how much is the response expected to change when the predictor(s) change by a given amount?" For categorical predictors this reduces to the question to "what is the expected difference in the response between different groups/categories?" For instance, for continuous predictors this is the questions for a slope.
-Response: I agree that this is interesting, but as I want to stay as close to the original papers, this might be something for a later study or experiment. The way the study was conducted, a regression would provide estimates of how broad-sense restoration affects biodiversity, so there would be estimates (that could be used to predict/calibrate future studies!) of how restoration affects biodiversity. My only issue with that here is restoration is only a broad-sense term, because the treatments were not applied consistently to all streams or in a factorial manipulation. I think more value would come from doing a manipulative experiment using regression to evaluate how restoration treatments affect diversity, both the direction and magnitude of the effect, and see if there are difference between treatments. If I was not trying to keep analyses parallel and to keep the story simple (as three paper are being compared), I would absolutely work on this sort of analysis.

Few additions here:
-Lines 24-27: I believe the end of the Abstract should contain a more detailed 'why' the author had all the work to re-analyze the data, in special, emphasizing the ecological meaning behind this.
-Response: This is a good suggestion, and I have added several lines giving more context and detail about the 'why' or motivation behind the re-analysis.

-Lines 53-56: here the author states expectations for the re-analysis (which is great!), but a good set of back up references is needed to support this.
-Response: Thank you for this suggestion, and I have added a few sources. Model papers (i.e. letters to the editor, comment papers) had mixed approaches in the extent to which sources were provided. Given PCI Ecology allows greater freedom, I have revised the manuscript to include more sources to support my hypotheses. Sources are not necessarily in lines 53-56 but are integrated in the following paragraph (lines 65-75).

-Lines 68-77: The author states the hypotheses here. However (as I noted in my previous comment), these lines are lacking of references to back up the author's expectations. Also, I did feel like missing in the end of this paragraph a more 'meaningful ecological reason' regarding the re-analyze of the data used by Swan and Brown (2017, 2018).
-Response: In my response to the previous comment, I have added citations throughout this section to support my hypotheses. I felt the flow of the manuscript was better with citations n this section than at end of the preceding paragraph during the closing/transition sentence.

-Lines 114-118: As far as I understood, the main issue brought up regarding the re-analyses of the data was the use of Type I sums instead of Type III sums (which mainly lead to misleading ecological interpretations). Yet, I could not find in the text why the Type III is more appropriate. But why is more appropriate? It might seems a little too obvious, but me as a reader was missing this information.
-Response: This is a good suggestion, and I have tried to provide more information about the difference between Type I vs. Type III sums of squares. I tried to strike a careful balance between getting into the depths of statistics and focusing on the importance to analysis of an ecological question. Previously, I simply stated Type III sums of squares were more appropriate, which was insufficient because it presumed the reader would know. As I am largely self-taught in statistics, I should have been more cognizant because I would have not known myself only 1-2 years ago. I have added this information on lines 146-149 of the revised manuscript.

-Lines 249-250: perhaps saying 'questionable research practices' is still too harsh?
-Response: I believe that was in an earlier version of the manuscript (version 2 on bioRxiv), but it was removed in version 3 (which was intended for the second round of review on PCI Ecology). I realized shortly after uploading to bioRxiv that I had not incorporated all reviewer comments on the draft, so I had to upload another revised manuscript. "Questionable research practices" does not occur in the main text of the manuscript.