Peer Community In Ecology

# A stronger statistical test of stream restoration experiments

***Karl Cottenie*** *based on peer reviews by* ***Mariana Perez Rocha*** *and* ***Eric Harvey***

David Murray-Stoker (2020) On the efficacy of restoration in stream networks: comments, critiques, and prospective recommendations. Missing preprint_server, ver. Missing article_version, peer-reviewed and recommended by Peer Community in Ecology. 10.1101/611939

Submitted: 21 September 2019, Recommended: 12 May 2020

**Cite this recommendation as:**
Cottenie, K. (2020) A stronger statistical test of stream restoration experiments. *Peer Community in Ecology*, 100052. 10.24072/pci.ecology.100052

Published: 12 May 2020



---

The metacommunity framework acknowledges that local sites are connected to other sites through dispersal, and that these connectivity patterns can influence local dynamics [1]. This framework is slowly moving from a framework that guides fundamental research to being actively applied in for instance a conservation context (e.g. [2]). Swan and Brown [3,4] analyzed the results of a suite of experimental manipulations in headwater and mainstem streams on invertebrate community structure in the context of the metacommunity concept. This was an important contribution to conservation ecology. However, David Murray-Stoker [5] was not satisfied with their statistical analyses, and recreated, and more importantly, improved their original analyses in the peer-reviewed article. The new analyses are based on a combination of a more consistent site selection, checking the model assumptions, using different estimation procedures, and focusing more on effect size calculations versus statistical significance. This peer-reviewed article is thus the perfect example of the advantages of open research: the original authors making available both the data and their R script files, initially first updating the analyses and results themselves, followed by more in-depth analyses of the original data and question. This peer reviewed went through a very in-depth process itself, with several rounds of questions and feedback that addressed both the statistical analyses, the interpretation of the results, and the conclusions. It also, however, addressed something that is often harder to provide feedback on, for instance the tone of the argument. I hope that scientists interested in these issues will not only read the final manuscript, but also the different steps of the peer review processes. These are very informative, I think, and provide a more complete picture of mainly the raison for certain decisions. Not only does this provide the reader interested in stream conservation with the opportunity to make up their own mind on the appropriateness of these decisions, but it could potentially lead to more analyses of this important data set. For instance, maybe a formal meta-analysis that starts with the effect sizes of all the original studies might bring some new insights

into this question?

***References:***


[1] Leibold, M. A., Holyoak, M., Mouquet, N. et al. (2004). The metacommunity concept: a framework for multi-scale community ecology. Ecology letters, 7(7), 601-613. doi: [10.1111/j.1461-0248.2004.00608.x](https://dx.doi.org/10.1111/j.1461-0248.2004.00608.x)

[2] Heino, J. (2013). The importance of metacommunity ecology for environmental assessment research in the freshwater realm. Biological Reviews, 88(1), 166-178. doi: [10.1111/j.1469-185X.2012.00244.x](https://dx.doi.org/10.1111/j.1469-185X.2012.00244.x)

[3] Swan, C. M., and Brown, B. L. (2017). Metacommunity theory meets restoration: isolation may mediate how ecological communities respond to stream restoration. Ecological Applications, 27(7), 2209-2219. doi: [10.1002/eap.1602](https://dx.doi.org/10.1002/eap.1602)

[4] Swan, C. M., and Brown, B. L. (2018). Erratum for: Metacommunity theory meets restoration: isolation may mediate how ecological communities respond to stream restoration. Ecological Applications 28:1370–1371. doi: [10.1002/eap.1738](https://dx.doi.org/10.1002/eap.1738)

[5] Murray-Stoker, D. (2020). On the efficacy of restoration in stream networks: comments, critiques, and prospective recommendations. bioRxiv, 611939, ver. 7 peer-reviewed and recommended by PCI Ecology. doi: [10.1101/611939](https://dx.doi.org/10.1101/611939)


# Reviews

## Evaluation round #4


DOI or URL of the preprint: https://doi.org/10.1101/611939
Version of the preprint: Version 4


### Authors' reply, 14 April 2020

**Download author's reply**
**Download tracked changes file**

### Decision by Karl Cottenie, posted 02 April 2020

**Final statistical issue**

It is obvious from the response from the author that my marble analogy did not clarify the statistical issue associated with the interpretation of the results. That is why I was glad to see the table in the response, because I think that will help me to explain the issue more clearly. If I understand the table correctly, the Revised Sites portion shows the streams that were selected for the new analyses. If that is correct, then

1) it clearly shows that the 2 treatments all replicates have in common are Bank Stabilization and In-Channel Manipulation. This thus means that the inference can only be applied to other streams that have received similar treatments. It does not matter if Riparian Forestation is present in a lot of them, because that treatment was not part of the selection regime. All replicate in an experiment should be as similar as possible to each other, except for the factors of interest (in this case whether they are Headwaters or Mainstems, and whether they received the treatments Bank Stabilization/In-Channel Manipulation or not). They are not similar in the Riparian Reforestation category.

2) the table also shows that Riparian Reforestation is confounded within the Headwaters and Mainstems category, since Reforestation is presence in all Mainstem streams and if absent, only absent in the Headwaters category. So any difference (or lack thereof) between Headwaters and Mainstems could potentially be caused the lack of Reforestation in the Headwaters streams.

I recommend that this table should be included in the manuscript, and the author should also discuss these limitations associated with their approach in re-analyzing the data. I don't think that this recommendation diminishes the value or novelty or impact of the manuscript. I just think that this is necessary to be clear and explicit about the correct inferences associated with the analyses.

## Evaluation round #3

DOI or URL of the preprint: https://www.biorxiv.org/content/10.1101/611939v4
Version of the preprint: Version 4

### Authors' reply, 19 March 2020

**Download author's reply**

### Decision by Karl Cottenie, posted 08 March 2020

**This preprint merits a revision**

The author addressed most of the comments satisfactorily. There is only one sticking point left, regarding the effect of site selection on the inference possible.

At the end of the introduction, the author writes: "I used this hypothesis to guide and inform site selection in my re-analysis. 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."

This means that the population from which the samples are drawn from, and thus what influences the inference, are all sites that received both bank stabilization and in-channel manipulations, period. Since reforestation is not part of the site selection criteria, the author cannot make any inference about reforestation.

This is similar to a bowl of a 1000 marbles that have 3 colors (red, blue, and black), and that are either small or big. If you randomly select 10 red or blue marbles (but not black ones) from the bowl, and 8 of the selected red or blue marbles are small, your inference will be about the population of red and blue marbles, not about the black marbles, or about the small marbles. Maybe the black ones are big marbles, but because you did not select these, you will not be able to test that.

I don't know what the range of restoration practices are, but out of all the possible ones, the author only selected the bank stabilization and in-channel manipulations restoration practices. That means, similarly to the marbles, that the population this analysis allows inference for is the population of bank stabilization and in-channel manipulation experiments. To call that subset "restoration" is not statistically correct, I think. Or that it reflects reforestation practices.

Thus, the first sentence of "Effectiveness of local restoration" that reads: "I hypothesized that restoration would have stronger effects in headwaters relative to mainstems." is incorrect. I do think that the author should replace every instance of "restoration" with "bank stabilization and in-channel manipulations restoration practices". Further in the discussion, the author can then potentially make the case that these are the most common restoration practices, or the most effective ones, or some other argument to convince the reader that the analysis applies to restoration in general, but the author will have to make it very explicit that this is speculation, that is not necessarily supported by the data analysis.

When the author writes, correctly: "the hypothesis was intended to guide criteria for site selection and reduce variation in restoration treatments among sites and not to necessarily or strictly compare the effects

in-channel manipulations and riparian reforestation treatments on biodiversity in restored streams." The flip side of the reduced variation in restoration treatments among sites is that the population sampled is different, and smaller. I think that this has to be explicitly acknowledged in the manuscript, and that the correct words have to be used.

Finally, I think that the section "Statistical inconsistencies" has indeed been toned down enough. The author writes in the response to the comment to "tone it down little" that " 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." I have an almost completely opposite view on the original authors' actions: they were extremely open about what they did and how they did it. They did provide all the code and all the data. That they provided this information when asked does not diminish that they did provide all of this, without which this manuscript would not have been possible. Did they maybe make some mistakes, yes, could they have analysed the data better, yes (see the resulting manuscript). But I am pretty sure that is the case for a lot of published papers, mine included, without there being any . I benefited from reading some of the blog posts from Stephen Hearst on the function of the Methods section in scientific papers, especially https://scientistseessquirrel.wordpress.com/2015/02/27/reproducibility-your-methods-section-and-400-years-of-angst/, and maybe that might help the author too.

## Evaluation round #2

DOI or URL of the preprint: https://www.biorxiv.org/content/10.1101/611939v3
Version of the preprint: Version 3

### Authors' reply, 28 February 2020

**Download author's reply**

### Decision by Karl Cottenie, posted 05 February 2020

**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.

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.

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.

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.

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).

## Reviewed by Eric Harvey, 15 January 2020

**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.

### Reviewed by Mariana Perez Rocha, 23 January 2020

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.

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. -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. -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). -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. -Lines 249-250: perhaps saying 'questionable research practices' is still too harsh?

## Evaluation round #1

DOI or URL of the preprint: https://www.biorxiv.org/content/10.1101/611939v1.abstract

### Authors' reply, 29 November 2019

**Download author's reply**

### Decision by Karl Cottenie, posted 10 November 2019

**This preprint merits a revision**

I have now read the two reviews of the preprint "On the efficacy of restoration in stream networks: comment and critique" by Murray-Stoker. Based on their comments, I have again read the preprint, and here is my recommendation. I think, though, that my recommendation might not be very satisfying for all parties involved. I hope, though, that I provide enough justification to convince the reader that a new approach would be beneficial.

Both reviewers point out the dual objectives of this preprint: on the one hand it provides a reanalysis of existing data from a previous study based on a different opinion on study design and analysis decisions, and on the other hand a critique of the motivation or intent of these differences. I agree with both reviewers that re-analysis of existing data in papers show the strength of the current practice of open and reproducible research. However, I also agree with one of the reviewers that the manuscript lacks some structure to easily convey the differences of the three papers in question. And their recommendation of a table with three columns, one for the original paper, one for the erratum, and one for this preprint would be useful. Each row would be a comparison item, which could be what studies were included, what design decisions were made, what statistical decisions were made, what results and conclusions were made for each of these 3 approaches. That will make it easier for the reader to compare and contrast, and then also decide where they stand on each of these decisions.

I also agree with one of the reviewers, though, that attributing negative intent to the original authors seems harsh and even unnecessary. They shared their data to make all this re-analysis even possible, so I think it is very unlikely that they intentionally made decisions that would bias their results and thus their conclusions. I think that what we are seeing here is that open science exposes more of the messiness of scientific practice with subjective decisions that have always been present, but now are out in the open and can be questioned and potentially corrected. This productive debate can lead to better or different analyses and conclusions, and it is up to the reader to decide which one is the most convincing one.

This is important for me, because after reading all the different opinions, I think that there is an inherent flaw with the current analysis method. If I understand it correctly, for most of the restored streams, there is a comparison between the restored stream and an adjacent stream. (I tried to find the original data mentioned in the preprint, but for some reason the link did not work.) If this is in essence a comparison between an experimental and control condition, and the study wants to combine results from different experiments, I think that the appropriate statistical approach should be a meta-analysis. This has several advantages. For each experiment, you will be able to directly compute an effect size (which was also advocated by one of the reviewers) for each study (in essence treatment - control, potentially scaled by sample size and/or standard deviation). This direct comparison will ensure that the reach dependency is included in the analysis but not as a random effect, and it will also reflect in the figures the true power of the statistical design. Right now when you look at the figures, this paired nature of the design is not at all represented. Secondly the meta-analytical approach would avoid one of the points raised by the author: what studies to include and which ones not to include. If you work with effect sizes in a meta-analysis, one of the assumptions being made by the different authors (whether bank stabilization and in-channel manipulations treatments are strong enough) will actually not be an assumption but a test. In the meta-analysis, you could test whether the stronger impacts result in

larger effect sizes, or whether the different types of treatments (reforestation etc) have different effect sizes. Meta-analysis techniques were developed to exactly address these types of questions. Finally, by computing the effect size, you will avoid having to include the random stream identity, and since this is a difference, not only is the interpretation more directly, but also more likely to result in a statistical model with easier model assumptions.

I realize that this recommendation for an additional or different analysis is one of the reasons that started this debate and series of manuscripts. However, I hope that my arguments outlined above will convince somebody that the meta-analytical approach is superior to the analyses from the previous articles and preprint, and that this might provide a more detailed and useful analysis and associated conclusions.

## Reviewed by **Eric Harvey**, 21 October 2019

**General comments**

The author offers here a criticism and re-analysis of the data from a previous study published by Swan and Brown. The author findings seem to be strikingly different from the one in the original study casting doubts on the quality of the data processing and analyses performed in the original study. In the original study, the authors had found an overall positive effects of habitat restoration on diversity and temporal stability in headwater but not mainstem sites in accordance with their expectations. Here the author show, after correcting for some issues with the analyses, that there are in fact no detectable effects of habitat restoration on any of the diversity metrics regardless of whether the site is located up or downstream. The author then conclude on the potential causes of this absence of effect.

The exercise proposed here by the author is on its own definitely valuable. Reproducing results from published studies is an important way to validate results and improve reproducibility of scientific studies. At best, it should initiate discussions or debates and eventually lead to a more informed consensus on the state of our understanding and knowledge on a specific issue (here the efficiency of habitat restoration in river systems). It should also help us to identify issues with the way we sometime use and interpret statistics so that everyone end up being better off from the process (i.e., a constructive process). Here, however, I would argue that some of the technical issues found should clearly have been found during the reviewing process, but that is another question. Sadly, despite all this, as I will expose in more details below, the article does give an impression of a public *vendetta*, making important accusations against specific authors. At the very least, PCI should invite Brown and Shawn to write a reply.

As is probably obvious from what I wrote so far, I enjoyed reading this manuscript. I also think that it is very clearly written. However I have a few main questions and concerns before it can published and some minor concerns below that I hope the author, editor and PCI will find useful:

1) The first criticism exposed by the author in the Introduction is more philosophical and open to debate I think than the other ones which are more technical. The author argues that in comparing restored vs. unrestored sites, Swan and Brown should have taken into account that not all restoration treatments should be expected to lead to the same outcome or to influence headwater versus mainstem sites the same way. I must admit that I am on the fence here. While I agree that not all habitat restoration treatments are the same, I still see values in making a broad hypothesis such as the one made by Swan and Brown: All else being equals, headwaters will be more affected by habitat modifications than mainstems because of fundamentals X and Y. To me what the author is getting at here is the idea of context-dependency. It ultimately would suggest that Swan and Brown were simply lucky to find results matching their main hypotheses because of the specific treatments they were looking at (with their specific effects not cancelling each other out) and that re-sampling the general pool of the different restoration types could lead to different results (i.e., "identity effect" matters). In that sense the author brings here an important note of caution about making sweeping generalizations or inference from a limited dataset. Basically here the author is building his own set of expectations for a different, but still interesting study (or more like a follow-up). If the goal was to make a more general point, then why a focus on one study and not just a more general concept and synthesis (or Forum) sort of paper?

2) The tone: I am not arguing that the author should have just said nothing when he found those issues (of course, not!) - but why going all-in public and writing a whole article clearly targeting specific authors? As a reviewer I am missing background information here to assess the situation and so it's very hard for me to understand some of the accusations. The author goes as far as accusing them of *Questionnable Research Practices*, which is an important offense because it could suggest that they manipulated data or analysis to mislead readers on purpose. This accusation also comes with no evidences of this. I don't know all the details but did the author contact Swan and Brown first? The concluding message of the article upon re-analyzing the data is quite interesting, but is loss in what seems to me like a public *vendetta*. As mentioned above I think it is essential that Swan and Brown are given the opportunity to submit a reply.

3) Several of the criticisms made by the author rely on a hard stance on statistics that is not always completely justified I think. That Brown and Swan should have used a Type III analysis is clear and sound, but many are highly skeptical of data-transformations and would even argue that linear models tend to be quite robust to violations of their assumptions (or then why not using a generalized linear model rather than a transformation?). I am not taking a side here as this is outside my field of expertise. But if a simple log transformation changes the results so much, could it simply be that fundamentally the main treatment effects are statistically significant but biologically meaningless? On this, it would be interesting for the author to present effect sizes for each set of models to give the reader an idea. This could open up on a more general discussion, I think, in ecology about how we tend to interpret statistical outputs based only on a p-value. I feel like perhaps if the signal was that strong, the changes made by the author would not have changed the results that much. Perhaps overall this is more a constructive tale of caution about how we interpret the statistical significance vs. the biological importance of the processes we attempt to measure and quantify.

**Minor comments**

*Introduction*

The Introduction is clear and well written. The author gives enough information so that a naive reader can get and idea of Swan and Brown hypotheses. The author also explains clearly his perceived issues with the way data were analyzed in Swan and Brown.

*Methods*

[91] does the author meant a PERMANOVA? Otherwise it's not clear how these multi-variate dissimilarity analyses were performed.

[88:95] Why a Gower for spatial dissimilarity and a Bray-Curtis for temporal dissimilarity? (I understand that the author cannot respond for Swan and Brown). It's also not clear to me why no transformations were imposed on the abundance matrix. It's been shown many times that some transformations (e.g., hellinger) can really improved detections of patterns in the data by weighting the disproportionate effects of rare and very abundant species.

[113] would not ln-transforming the Gower base 5 dissimilarity values totally alter the interpretation of the dissimilarity index?

[123] Here the "nlme" package is cited - unless I am missing something I think the author should then use the more specific "linear mixed effect models" rather than ANOVA.

[160:168] The idea suggested at the end is quite interesting but rather speculative in light of what the author says about his capacity to test his hypothesis.

*Results*

The difference in the results, even for the same 'full sites' data are quite stricking. Unless I missed something, it seems like for the 'full sites' data, the only differences between the original study by Swan and Brown and the re-analysis here is that 1) Response variables were transformed to meet model assumptions and 2) a Type III analysis was used. Am I correct? Those transformations were performed "to better meet the assumption" - did the transformation actually made the data fit the assumptions? Also how does a square-root transformation on diversity, an ln transformation on richness and Gower dissimilarity index actually affect interpretations of those response variables? Why not using a generalized-linear model instead? I am asking

because those transformations might be one of the main reasons for that stricking difference in the results with the original study, but sometimes transformation are known to have undesirable impacts on variable distribution and interpretation. If a simple transformation changes the results so much, could it simply be that fundamentally the main treatment effects are statistically significant but biologically meaningless? On this, it would be interesting for the author to present effect sizes for each set of models.

### Reviewed by **Mariana Perez Rocha**, 04 October 2019

**Download the review**