*Response to reviewer comments on manuscript “Assessment of the cumulative effects of restoration activities on water quality in Tampa Bay, Florida”, by M. W. Beck, E. T. Sherwood, J. R. Henkel, K. Dorans, K. Ireland, and P. Varela.*

*We sincerely thank the reviewers for providing helpful comments on our manuscript. Our responses to each of the comments are below. Line numbers match those in the revised manuscript.*

**Reviewer 1 comments:**

Overall comments.

It was enjoyable to consider the substantial contributions made by this original and creative research effort. The spatial analysis appears to have been done carefully and documented well. Mostly, the assumptions and limitations of the study were made clear, which I appreciated; a few exceptions are noted in my detailed comments. The figures are exceptionally clear and communicative throughout, and even highly original and informative in some cases such as Fig. 7. The tables are clean and well-designed.

*We appreciate the substantial comments provided herein and have addressed each concern on a point-by-point basis below.*

Editing is a problem easily rectified by engaging an experienced editor, which will be very important for this paper to communicate the strength of its contributions, as it should. A very short list of examples is provided here but there was not time in a peer review to document the many editorial exceptions:

39-40 "widespread chronic" is awkward

*Changed to “widespread”.*

42 lack of Oxford comma

*Added.*

46 starting sentence with And is pretty informal for a journal publication

*Changed to “Further”.*

50-52 rigorous used twice in one sentence

*Revised to “…can further constrain evaluations of restoration success”*

50-52 tense problem; difficulty and constrain. (Difficulty constrains. Difficulties constrain.)

*Changed to “difficulties”.*

150-154. Awkward. 2 sentences.

*Revised: “We developed and applied a two-level classification scheme that described each restoration project as 1) a habitat or water infrastructure improvement and 2) more specifically as enhancement, establishment, or protection for habitat or as nonpoint or point source controls for water infrastructure.”*

168 included, not include

*Changed.*

176 "matching between" is awk

*Revised to “The matchings began with…”*

Overall: Inconsistent use of bay versus Bay when the word stands alone from Tampa

*All uses were checked and changed to “Bay”.*

Detailed Comments by Line Number and Figure Number:

Introduction. The rationale developed in the opening sections of the Introduction is solid and well cited with a combination of peer-reviewed papers and a small number of necessary reports in the gray literature: both natural and social/institutional challenges to long-term monitoring of restoration projects and to large-scale restoration success; so the unprecedented current opportunity in Gulf following Deepwater Horizon requires careful examination of what has worked before in the region.

Line 76 Given the previously well-developed argument about the challenges of long-term, intensive monitoring, it would be interesting for the reader if the "secrets" to doing so in Tampa Bay were described at this point, even just a single sentence on what made Tampa Bay long-term intensive monitoring "work." eg was it citizen science? Volunteers? A charismatic individual? Etc. (I am guessing it may come down to Holly Greening herself, but even that would be interesting if it has to do with continuity of leadership, for example.) Or maybe it comes down to regional partnerships (Sherwood et al. 2016, Regional Studies in Marine Science)

*The long-term monitoring program in Tampa Bay has facilitated the tracking and recovery of water quality since the 1970s. We agree with the reviewer that its worthwhile pointing out why this program has been so successful. We have added some text to briefly describe some potential reasons: “The establishment of the long-term monitoring program has been instrumental in the development of water quality management targets that have guided restoration efforts beginning in the 1980s.”*

78-79 It is difficult to believe that "no previous efforts have been made to directly quantify potential associations between these activities and estuarine water quality." I think the authors owe readers to follow this statement with a concise lit review, a full paragraph, of eg Sherwood 2016,2017; Greening 2014; and others' published work; stating what they did do, so as to clarify how the authors' work here is unique and different. It may also be worth consulting gray literature because surely there was some "method to the madness" as to which remediation and restoration projects were funded and most importantly why they were prioritized, which was very likely based on past experience of what worked even if it was never published. I note that lines 78-88 are citation free. Deeper consideration of what has come before is warranted here to assure the reader that what's being proposed by the authors is really needed. I believe that it is needed, but that history needs to be documented here in brief.

*We agree that our general statement that “no previous efforts have been made” was imprecise and have revised this paragraph to describe the context more clearly. We have moved some examples from the text on lines 124-131 to this paragraph. The revised paragraph is as follows: “Tampa Bay (Florida, USA) is the second largest estuarine embayment in the GOM and improvement in condition over the last four decades is one of the most exceptional success stories for coastal water quality management (Greening et al. 2006; Greening et al. 2014). Most notably, seagrass coverage in 2016 was reported as 16,857 hectares baywide, surpassing the goal of restoring coverage to levels that occurred in 1950 (Sherwood et al. 2017). Reductions in nutrient loading (Poe et al. 2005; Greening et al. 2014), chlorophyll-a concentrations (Wang et al. 1999; Beck and Hagy 2015), and improvements in water clarity (Morrison et al. 2006; Beck et al. 2018) have also preceded the seagrass recovery. Most of these positive changes have resulted from management efforts to reduce point source controls on nutrient pollution in the highly developed areas of Hillsborough Bay (Johansson 1991; Johansson and Lewis III 1992). These controls allowed nutrient and chlorophyll-a targets to be met by the early 1990s. However, numerous smaller projects, including watershed-focused efforts (Lewis et al. 1998), have likely had a supporting role in maintaining the earlier water quality improvements observed after water infrastructure upgrades. The cumulative effects of over 900 restoration projects, relative to broad watershed-scale management efforts, are not well understood. Understanding how implementation of these projects is associated with adjacent estuarine water quality at various spatio-temporal scales will provide an improved understanding of the link between overall estuary improvements and specific restoration activities.”*

89-90 The argument seems thin. Could probably add several other reasons to this sentence.

*The beginning of this paragraph was revised: “Demonstrating success for restoration activities is challenging for several reasons (Ruiz-Jaen and Aide 2005; Wortley et al. 2013). Success may be vaguely or even subjectively defined because the effects of restoration could be described in different ways depending on project goals (Zedler 2007). For example, site-specific measures of before/after condition are commonly used to measure of success, whereas downstream effects may be more important to consider for baywide conditions (Diefenderfer et al. 2011). More importantly, quantifying success as a measure of environmental improvements is challenged by the variety of factors that affect water quality across space and time. New tools are needed…”*

95-97 Excellent objectives. Challenging and worthwhile.

Methods: Overall, there is a lack of documentation of other published work particularly in regard to the spatial methods which informed those used in the study, which should be added throughout Methods section. It is clear that the team is experienced in spatial modeling and I presume familiar with the literature in GIS methods. That said, the essential work appears to be an important contribution.

*The following was added to the first paragraph of the data synthesis and analysis framework: “Spatial and temporal matching can be accomplished using several methods that vary in complexity. For example, hydrologic distances or other non-Euclidean distance weightings by watershed topology can be used to link measurements to modelled locations in space (Curriero 2006; Gardner et al. 2011). However, we adopted a relatively simple approach with limited data requirements to maximize potential applications in other regions (e.g., no hydrology data are needed, only spatial location).”*

*Additional info was added: “A weighted average by distance (or parametric distance weights, Sickle and Johnson 2008) was used based on the assumption that restoration projects farther from a water quality station will have a weaker association with potential changes in chlorophyll-a.”*

The software name and version number should be mentioned somewhere in Methods.

*The software and version were already mentioned on lines 241-242. No other software was used.*

115 "purportedly" implies you disagree with this conclusion. Instead, suggest removing that word and adding a citation to the fact.

*Citation added for Springer and Woodburn 1960.*

121 Actually, according to Sherwood, the goal was 95% of 1950s level, not 100%. Suggest re-reading Sherwood and citing original literature for the goal if available.

*Text was changed.*

124-131. It would have helped to make this nuance clear at the end of the introduction (see comments lines 78-79). Tampa Bay is such a well known success story that it was not credible to state no cause-and-effect research had been reported. Clarifying that it is the other smaller projects that haven't been evaluated in the Intro would help make the authors believable.

*Please see our response above.*

129-131 "have likely" suggests bias. Instead saying that reports indicate they may have had an effect, which you sought to evaluate, would be a cleaner approach.

*Changed to “may have had”. We are unaware of any reports that explicitly evaluated the hundreds of restoration projects that have occurred in the Bay or its watershed.*

135-148 Please address whether there was overlap in the projects in these datasets.

*There was no overlap between the datasets. A sentence was added to line 134 to clarify: “Each database was unique and no overlap in documented projects was observed.”*

156-158 move down as the topic sentence of the next section on Data Analysis, and remove "methods described below"

*Moved.*

168-170 Why list data you didn't use? Also, previously, it was stated that chlorophyll was the metric. Now two others have been added. Confusing.

*Line 168 – 172 were removed and the following was added: “Chlorophyll-a (ug/L) measurements at each site were used for analysis, totaling approximately 515 observations for each station.”*

The assumption seems to be that the nearest sites have the most effect, which seems to disregard flow direction. However, the time binning does help to ameliorate this concern.

*Please see our response above regarding the citations added for geospatial methods.*

In the absence of any other explanation, I assume that even projects up in the watershed were digitized for hydrologic distance "as the crow flies" and it is rather interesting that despite this apparent weakness, the results were significant. Implications should be treated in the Discussion. Potentially an avenue of future research to build on this?

*We have added a paragraph to the conclusions section that describes future directions: “Our approach is not without limitations and future research could build on the methods to provide an improved assessment of restoration effectiveness. Our geospatial analyses were relatively simple, in that spatial matchings were accomplished through Euclidean distances. Alternative distance measures could also be used that consider hydrologic distances following flow networks in the watershed. The importance of these approaches could provide insight into pollutant dispersal patterns in environments with low elevation gradients, such as Florida. Weighting restoration projects by relative effort could also facilitate an improved assessment of effectiveness, such as considering total restoration area as an important variable to consider for water quality improvements. Some of these data were available in our compiled dataset, although coverage was insufficient for a complete analysis. Finally, the social and human dimensions of different restoration projects were not considered herein but are important factors that can be equally or even more important determinants of success that must be considered when weighing restoration options. Future work should adopt a more comprehensive evaluation of success measures that extends beyond water quality improvements.”*

188-190 Explanations of time windows are not complete enough to understand.

\* Also raises a question of how these were determined. A priori, as a hypothesis? Or after trial and error and discovering which ones yielded results. Important to state this.

\* Also need to state how long projects take to complete, and if this length of time is different for the different categories of projects.

\* Fig 5b seems to indicate much longer time slices than are described in the text.

\* Also raises a question of how much interannual variability there is in hydrologic connectivity in this system. Is one year of after data sufficient to capture effects? I realize the assumption may be that after one year, the likelihood of other changes affecting the results increases.

198-200 I think this should be summations (plural) were (not was) correct? The before and after data were separately divided by the number of projects, correct?

Data synthesis and analysis framework Section: Lack of citations to related methods. Need to contextualize the work with that of others in spatial data processing and water quality analysis.

212-213 Did these hypotheses have a direction or were they as general as is described here?

213-214 Why wouldn't the statements be quantitative, given that this is a quantitative analysis by distance and time? Qualitative may undersell the analytical effort made here.

217-221 This is a promising application.

228-230 A strength of this paper is this multi-scale approach. Others have similarly looked at eg drift cell reaches along shorelines versus watershed-scale effects on coasts, references which could be incorporated in the methods. I think the point here is larger than an "alternative context." Rather each scale provides information about potential stressors and processes controlling the impacts. For stream restoration examples see eg Roni et al., N. Am J of Fisheries Management 2002; Roni et al. 2012 book chapter, available online, "Prioritization of Watersheds and Restoration Projects"; for coastal restoration examples see eg multi-scale analysis of restoration priorities papers by Thom et al., Ecol Rest 2011; Diefenderfer et al. Env Man 2009.

247-249 good idea/verification of assumptions. This is a little bit unclear, "but with random date and location assignments for each restoration project". I assume that both randomized date/location, and actual date/location were analyzed and then compared, correct?

253-58 a bit cryptic

Results. Overall comment. Consider organizing with the key findings and figures first, and lesser/background findings second.

Compare 264-266 with 277-279. These are not in fact "coincident," as stated. The monotonic rate of decline of pollution slowed after 20 years of the 40-year period; but the rate of increase in projects on the ground increased after the year 2000! These are only coincident over the entire 40-year period, not within the 40-year period. In fact, the patterns of rate of pollution reduction and rate of new projects (a project count, not an effect magnitude) appear to be the opposite.

307-320 Very important points that may belong in the Discussion. It's worth adding what these analyses don't say about the value of Protection, as well, because Protection in an area such as Tampa Bay with severe development pressure is very important. (338-341)

368-370 I don't much mind this here but many would call it Discussion material and it might receive more attention there if highlighted as part of the contribution of these Methods.

377-378 not a complete thought (or sentence)

381-382 Very important and unique finding.

383 Effects of random restoration dates and locations Section: Strengthens the paper.

Discussion. Overall: The opening of the Discussion is rather boring in comparison to the strength of what came before. I think the fact that prior findings were supported belongs much further down. In fact, all of that material seems to belong in a "Verification of Approach" type of subsection of the Discussion.

440-442 Sounds a bit naïve relative to cost-benefit. Land acquisition can be extremely expensive in coastal regions with high demand and difficult to obtain grant money to achieve. Still, I support the conclusion in 445-448 based on experience with other conservation/restoration programs.

Adequate consideration should be given to circling back to the objectives of the paper developed in the Introduction, and verbally checking off the boxes you've completed.

Figures

1. Nice figure. Need to credit data sources in caption. "Have been monitored" implies a role by the authors without crediting others.

2. Very useful, especially the maps in b but also a.

3. Also very useful because this was hard to visualize from the text. However, the projects shown are the "easy" ones in terms of digitizing hydrologic distance, because they are in the bay so it was done "as the crow flies." How was hydrologic distance calculated for projects higher in the watershed? Hydrologic distance is usually digitized from the streamlines in a dataset but this was not described. It would be useful to add a second set of panels for the projects in the watershed that reveals this.

4. Reasonable effort. A little hard to understand.

5. clear and useful but Fig 5b seems to indicate much longer time slices than are described in the text.

6. This took a while to understand and is rather original, but ultimately worth the time for the information communicated.

7. Very interesting and informative! This type of information is gold for monitoring program design. May wish to highlight this more in the Discussion.

**Reviewer 2 comments:**

1) The organization might be improved. P23 Assumptions and limitations section (Line 475) should be seperated and the Assumptions should be moved to the section "Data synthesis and analysis framework", 2) The paragraphs (Line 307-320) are suggested to be moved to the Discussion section, 3)Discussion section should be re-arranged with several sub-sections by the major findings, for example, the effectivieness of the proposed framework, limitation of this approach, etc.

2. The language should be concise, particularly (1)the section Introduction is too tedious and the readers cannot catch the key point easily,