*Response to reviewer comments “Prioritizing management goals for stream biological integrity within the developed landscape context”, by M. W. Beck, R. D. Mazor, S. Johnson, K. Wisenbaker, J. Westfall, P. R. Ode, R. Hill, C. Loflen, M. Sutula, E. D. Stein.*

*We thank the associate editor and reviewers for providing helpful comments on our manuscript. Our responses to each of the comments are in italics.*

**Associate editor:**

I believe the manuscript is of potential interest to many Freshwater Science readers. I like the fact that the paper is more than just another model that predicts biological outcomes using GIS data. If that were the case, I could not recommend the paper for publication. Reviewer 1 thought the manuscript suffers from trying to do too much, and recommended refocusing the paper to either the web tool or the model. Although I agree with Reviewer 1 that both sub-topics need more explanation, having both a model and its application in a single package is what, in my opinion, makes this paper unique.

These strengths notwithstanding, the manuscript suffers from substantial inadequacies. Both reviewers (as did I) had difficulty grasping your objectives and following the methods. Both had difficulty with your interpretations and characterization of how the model (and associated tool) were used in the case study. I describe several general concerns below, followed by detailed comments. Both reviewers also provide many comments that should improve the manuscript.

My decision is to invite the authors to resubmit the manuscript after making the revisions suggested in this review. Additional peer review is likely necessary to ensure the manuscript has been improved, particularly in clarity.

ASSOCIATE EDITOR GENERAL COMMENTS

The Introduction is too lengthy and most of the paragraphs need clarification. Reviewer 2 had the same concern. The first paragraph is about the need for context in interpreting biological assessment data and is rather succinct. The second paragraph is less clear, but I think you’re laying the foundation for the need for managers to prioritize restoration. But the paragraph doesn’t strongly convey that need. Paragraph three is also a bit blurry. It starts with a focus on multiple stressors but appears instead to be making a case for using land use (isn’t it really just land cover?) as predictors of biological condition. Again, the paragraph needs to be brought into sharper focus, although I’m not sure you need to justify using geospatial data to predict biological condition. So perhaps this paragraph can be eliminated outright. Most of the subsequent paragraphs also need more focus.

*Our revisions described below, as well as those in response to the other reviewers, have added clarity to the introduction. In short, the revised paragraphs (in sequence) describe 1) need for a bioassessment context, 2) management challenges in developed landscapes, 3) definition of constrained streams, 4) how our approach differs from previous landscape models, and 5) goals/objectives. Note that the original paragraph three was removed, but additional descriptions of the use of landscape predictors to predict biological condition (including implications for developed landscapes) was added in response to comments from reviewer 1 and your comment herein. Please see our explanations below.*

The Methods section needs substantial work. I don’t think you need two long paragraphs about California’s environment past and present. The reader immediately gets bogged down in extraneous details.

*The first two paragraphs of the methods that described the study location were condensed into a single paragraph.*

More detail is needed about the design of the bioassessment data collection. If the bioassessment data was not collected from a random sample of the NHD population, then you need to discuss the implications of this on your predictions.

*We have added text to describe our survey design in the methods:*

*“We aggregated data collected under more than 20 federal, state, and regional bioassessment programs. Some of these programs employed a spatially balanced probabilistic design (e.g., the statewide Perennial Stream Assessment (Rehn 2015), or the Stormwater Monitoring Coalition’s survey of southern California streams, Mazor et al. 2015), although others used different designs for project-specific purposes (such as the statewide Reference Condition Monitoring Program, Ode et al. 2016). Most of these programs targeted perennial streams, although an unknown number of intermittent streams with flows lasting into the normal sampling period were included (Mazor et al. 2014), particularly in more arid southern California. Because these programs are extensive, most regions and stream-types where perennial wadeable streams are located were represented in the calibration data set.”*

*We have also added text to the discussion that describes potential limitations of applying our model approach to the entire NHD-Plus dataset.*

*“Finally, there are a few concerns applying a landscape modelling approach for bioassessment using the NHD-Plus flowlines as a base layer. We applied our model to the entire network of the NHD-Plus represented in StreamCat, which included a large number of intermittent or ephemeral streams, as well as non-wadeable rivers. Therefore, the application of model results in these stream-types is open to question, valid only to the degree that the CSCI and its response to landscape disturbance can represent more relevant measures of biological integrity. In regions where ephemeral streams are particularly common (e.g., the inland deserts or the South Coast region), estimates of the extent of constrained or unconstrained streams may be inaccurate.”*

The justification for your selection of GIS predictor variables is insufficient, as both reviewers also pointed out. Given the extent and severity of hydrological modification in the state, it seems strange that you would exclude this type of alteration from your model.

*As noted, both reviewers had concerns about how we selected variables for inclusion in the model. We believe that these concerns originate from both a lack of description about how quantile random forest models were developed and our choice of predictors that were descriptive of biological constraints. As noted by the second reviewer, the key description of the rationale behind our model was on line 611 to identify biologically constrained sites as those where “present landscapes were likely to limit CSCI scores that describe macroinvertebrate condition.” Our model predictors were therefore chosen as landscape variables that represented factors that were likely associated with limitations where conventional management options will likely be unsuccessful in achieving reference conditions at a site. There is a distinct difference in this choice between instream stressors associated with landscape development vs those typically beyond management intervention.*

*Further, the presence of channel modification was not chosen as a specific predictor because our goals were to describe landscape effects on biological constraints. While it is true that channel modification can impose constraints, we felt that this categorization was not inclusive as not all modified channels are constrained and not all constrained channels are modified. Our previous description of this distinction was provided on lines 600 – 609.*

*Our selection of variables in Table 1 for inclusion in the model was decided a priori through discussions among co-authors, given the rationale described above. However, a larger model including 99 StreamCat variables was also tested to verify if a simpler model could accurately describe landscape associations with biological constraints. These preliminary analyses showed that our simpler model had comparable performance relative to a more complex model. This demonstrated that biological constraints could be described by a smaller subset of landscape variables that were inclusive of the larger set.*

*We have revised the paragraph describing our choice of variable selection (lines 211 – 226 in original draft):*

*“Expected CSCI scores were modelled using estimates of canal/ditch density, imperviousness, road density/crossings, and urban and agricultural land use for each stream segment (Table 1, Figure S1). StreamCat was used as the only source for predictor variables because of consistent methods and linkage to NHD-Plus flowlines (Hill et al. 2016). Preliminary analyses indicated that these variables adequately described biological constraints relative to a larger model with additional variables. These variables were chosen specifically to describe biologically constrained sites where present landscapes were likely to limit CSCI scores that describe macroinvertebrate condition. Landscape variables were selected rather than more proximal variables (e.g., in-stream water quality) given that constraints were defined relative to potential impacts on biological condition that are typically beyond the scope of management intervention or where costs to mitigate are likely prohibitive. Further, channel modification was not chosen as a predictor because it narrowly described constraints relative to our definition, i.e., urbanization was more inclusive of constraints, whereas modified channels may or may not be constrained. Overall, the model was associative by design and not descriptive of immediate causes of poor biological condition. We assumed that deviation of observed scores from the model predictions (i.e., residuals) could be used to describe in-stream factors associated with condition for follow-up analysis.”*

*We have also added a new figure to the supplementary material (Figure S1) that shows how each predictor (watershed only, the model also included catchment and riparian estimates for the same predictors) relates to the constraint classes and CSCI scores.*

After reading through the paper carefully, it is still unclear to me how the model and associated web tool are supposed to be applied. A diagram may help here. Is the model supposed to provide context for actual CSCI scores from a sample? Or is it meant to provide a prediction of what’s possible biologically at unsampled sites? Figure 6 shows a nice example of how the model puts actual sampled sites into “context,” but it would be more helpful if it also illustrated what inferences were possible as a result of this added context. In your text about the case study, sometimes you discuss all the segments, other times you discuss CSCI scores—many of which were used in model development. Line 439 states that the model was used to help interpret CSCI scores for decision making. But throughout the Methods and Results sections you talk about segments, and draw attention to patterns in predictions of segments. Then in line 489 you say the stakeholder group focused on the entire landscape rather than individual sites. So I get confused about what the emphasis is of the model. I also agree with Reviewer 1 who has some pointed questions on how you are characterizing the SCAPE tool.

*We have made substantial revisions to our description of the case study to more clearly emphasize how the model results can inform management priorities. Please see our response to reviewer 1 for line 297. This includes an additional paragraph that better explains why we included the case study and a new table that explains the priorities derived from the model.*

*Also, please see the new conceptual diagram (figure 2) that graphically describes our approach. This provides a description of how our approach is similar to the DPSIR framework that describes environmental impacts relative to drivers and pressures from societal/economic activities. We have modified the generic framework as appropriate for our landscape model.*

ASSOCIATE EDITOR SPECIFIC COMMENTS

Line 63: This sentence is unclear to me. How do site-specific conditions place limits on scales?

*Sentence was clarified: “…may be challenging if landscape conditions (e.g., watershed imperviousness) place limits…”*

Line 67: This sentence is complicated and unclear. After reading subsequent sentences, it’s still not clear how data accessibility is relevant to the topic of this paragraph, which appears to be the need for context.

*Lines 65-71 were revised for clarity: “Resource management decisions may be difficult to make if information is unavailable that describes these limitations. Context is required that describes how likely a site is to achieve biological integrity and how bioassessment data collected over multiple locations and times can be used to support decisions or identify priorities.”*

Line 73: “…in need of some level of management.” seems like you are advocating here, or at least assuming that government agencies have a recognized duty to actively manage waterways that are in biologically poor condition is a

*Removed “…in need of some level of management.”*

Line 81: It’s not clear why these modifications are “confounding” factors. The subsequent sentences suggest these types of modifications are often the basis for simply redefining management goals. So, perhaps all you need is to improve the connection between these ideas.

*Sentence was revised: “Moreover, extensive modifications to streams for flood control or water conveyance are common in developed landscapes.” The entire paragraph was also restructured for clarity, which better places the content about modified channels.*

Line 87: This sentence is an abrupt transition from the preceding stream of thoughts.

*This sentence was moved to the topic sentence as this is the primary focus of the paragraph, i.e., prioritization in developed landscape is difficult because of challenges that ultimately relate back to the watershed or extensive channel modification.*

Line 104: Will you be using information about historic landscape changes in your models? If not, I’m not sure why you’d use this statement as a topic sentence.

*No, this statement was inaccurate and was removed. Also note that the paragraph starting on this line was thoroughly revised in response to this comment and those below (down to line 121):*

*“**Herein, we define constrained streams as those where present landscapes are likely to limit biological integrity. By describing an expected range of biological conditions due to factors that constrain biointegrity and may be difficult to manage, efforts to improve or protect condition could be prioritized at sites where alternative or more easily managed factors are influencing condition. For example, a monitoring site with an observed biological index score that is above a predicted range could be assigned a higher management priority relative to a site that is scoring within the range that is expected based on landscape development. A predictive model of bioassessment scores that is based on landscape metrics (e.g., imperviousness) could describe constraints on biological integrity, whereas variation of observed scores around a model prediction could suggest other factors at the local scale (e.g., instream physical habitat) are more important. Analysis methods that characterize biotic and abiotic factors that limit assemblage composition have been explored by others (i.e., limiting factor theory, Chessman, Muschal, and Royal 2008, Chessman 2014). Similar concepts have been applied in a landscape context to describe variation in biological communities and metrics at different spatial scales (Waite 2013, Waite et al. 2014), although they have not been developed to describe constraints as defined above.”*

Line 106: Perhaps the sentence would be more clear if phrased as “…where landscape alteration may seriously limit opportunities to restore biological integrity.”

*This sentence was removed.*

Line 107: Your definition of “constrained” streams seems to have confused Reviewer 2. The other reviewer and I didn’t have the same reaction, but consider possible alternatives that could be more clear (e.g., “restoration constrained” or “restoration limited”). Perhaps the reviewer’s confusion was due to their familiarity with other definitions of “constrained” in the context of geomorphology.

*Our definition of constrained streams that was noted by the second reviewer in the discussion was moved to this paragraph.*

Line 107: The definition is also a bit vague. How does “large scale” differ from “landscape alteration?” I know what you’re trying to say, I’m just suggesting you refine the wording.

*This was a bit confusing and we’ve removed “large scale”.*

Line 110: Can we assume that “constrain biology” has the same meaning as “constrain the stream?” This entire sentence is a bit vague. How does a stressor originate from scales? Perhaps an example would help?

*Our revisions to this paragraph should clarify what we mean by constrained. We have also included an example: “A predictive model of bioassessment scores that is based on landscape metrics (e.g., imperviousness) could describe constraints on biological integrity, whereas variation of observed scores around a model prediction could suggest other factors at the local scale (e.g., instream physical habitat) are more important.”*

Line 115: From this sentence to the end of the paragraph is in need of clarification.

*Our revisions prior to this sentence place this paragraph in a more understandable context, i.e., how are constraints defined, why is it useful to describe them, and what theoretical approaches have been used by others (citations therein)?*

Line 117: What is “…variation in bioassessment data?” and “…limits of bioassessment tools?”

*Sentence was revised: “Similar concepts have been applied in a landscape context to describe variation in biological communities and metrics at different spatial scales* *(Waite 2013, Waite et al. 2014), although they have not been developed to describe constraints as defined above.”*

Line 120: What is “bioassessment and management potential?”

*This sentence was removed.*

Line 131: Here, I think you’re attempting to differentiate your effort from previous studies. This point is critically important given the number of previously published modeling papers, but is not given the emphasis in deserves.

*The paragraph was revised to make this clear. In particular, the following was added: “However, past efforts have primarily focused on characterizing condition at unsampled locations, often predicting the most likely condition by estimating averages. Alternative modelling approaches, such as quantile-based methods (e.g., Cade and Noon 2003), could be used to predict a range of expectations for biotic integrity from geospatial data.”*

Line 136: The sentence starts out great but then whimpers out in the end. May I suggest something like “…to classify and prioritize ecologically impaired stream segments based on the likelihood that restoration efforts would be successful given the degree of landscape alteration…” or something like that?

*Changed to “The goal of this study was to present the development and application of a landscape model to classify biological constraints in streams based on the likelihood that an upper expectation of bioassessment scores is limited by landscape alteration.”*

Lines 155-179: Condense this material to a single, shorter paragraph limited to information about California that is critical to the reader’s understanding of your methods.

*These two paragraphs were shortened and retained only relevant information for understanding of the methods.*

Line 190: Is this the only information the reader gets about your GIS predictor variables? Shouldn’t you refer to Table 1 here?

*Our revisions to the paragraph starting on line 211 provides more detail about our predictors, in addition to the added content to the supplementary material.*

Line 209: It is critical that you describe the design of the bioassessment data collection. Were sites selected to be representative of the entire stream network? Or were sites targeted?

*Please see our response to the general comments above. In short, we have added text to the methods and discussion to address these concerns.*

Line 212: Didn’t you already say this?

*Sentence was removed.*

Line 217: If I follow your argument here, you’re saying that you only selected predictor variables that represent human impacts that can’t be mitigated? I don’t believe this argument. If the argument is true, then why not include NPDES dischargers or major dams / reservoir storage, both of which are widely available GIS datasets and represent human impacts that don’t seem to go away? (with the exception of an occasional small dam)

*Please see our response above to your general comment about selecting variables for the model.*

Line 221: If you’re not going to include NPDES discharges and dams, then address the reasons why here.

*Our revisions to this paragraph have addressed this concern. In addition to our description of landscape variables, we have also added the following: “StreamCat was used as the only source for predictor variables because of consistent methods and linkage to NHD-Plus flowlines (Hill et al. 2016).”*

Line 221-223: It’s not clear what this sentence means in the context of statistical modeling.

*This sentence was removed with our revisions to this paragraph.*

Line 223-226: Rephrase sentence to: “…human activity not related to the predictor variables used in the model…”

*This sentence was removed with revisions to this paragraph. However, we have added a more general description with our revisions to the introduction: “A predictive model of bioassessment scores that is based on landscape metrics (e.g., imperviousness) could describe constraints on biological integrity, whereas variation of observed scores around a model prediction could suggest other factors at the local scale (e.g., instream physical habitat) are more important.”*

Line 239: Unclear what these percentiles refer to. It’s relatively easy to envision quantile regression with a single predictor variable, but most readers (myself included) will need more explanation of how the procedure works in the case of several predictors. Is there a separate RF model developed for each 5% increment of CSCI scores? If so, does each model use the full range of all predictor variables?

*Additional text was added: “The statewide validated model (described below) was used to predict percentile expectations of CSCI scores at all stream segments where predictors were available. For example, the 50th percentile prediction was the most likely score for a stream segment given observed values for landscape variables, whereas a lower (e.g., 5th percentile) and upper (95th percentile) bound on the median (and points in between) were also predicted.”*

Line 244-247: The topic sentence makes no sense (typo?). It suggests that the full CSCI dataset was designed so as to be representative of the state’s regions, but this wasn’t explained earlier. But my reading of your methods here is that you randomly selected calibration and validation subsets from the full dataset—which doesn’t by itself guarantee representation of landscape gradients.

*Topic sentence revised as “…gradients across major regions…”. Our earlier addition to the methods that described our sampling design provides clarity. Overall, our modal calibration/validation data were split to represent these regions, while also accounting to variation in development gradients that vary by region. Revision for the next comment provides additional clarity.*

Line 247: What is meant by “this stratification?”

*The beginning of his paragraph was revised: “Calibration data for the landscape model were obtained from a random selection of 75% of segments with observed CSCI scores, where the selection was based on a random draw among sites that were grouped into quartiles defined by increasing watershed imperviousness relative to each region (n = 1965 segments). This ensured that the model was calibrated with data that covered the variation of landscape development between regions (i.e., regions with low development were not under-represented and those with high development were not over-represented).”*

Line 262-276: This is a slightly better definition of “constrained” than was given in the Introduction, but still needs polish.

*These sentences were removed given revisions to our definition of constraints in the introduction.*

Figure 2: Needs clarification. The “Segment type” y-axis label is confusing. Aren’t these just examples of individual segments for which a prediction was made? Make it obvious that the symbols on the right-side panel are actual biological samples. Why use the term “Relative…?”

*The y-axis label was changed to “Example segments”. The plot subtitle for the (d) and the appropriate legend were changed to “Observed site/CSCI score” to emphasize observations relative to model predictions.*

Line 274: Again, not clear what these percentiles refer to.

*Please see revisions above. Sentence was also modified: “...a prediction interval ranging from the 10th to the 90th percentiles of expected CSCI scores for the level of landscape development at each segment.”*

Line 281: “or certainty?” Not clear how you are defining this term and how it was quantified

*A sentence was added: “For the certainty range, this interval does not describe statistical certainty in the traditional sense (e.g., confidence interval), but rather a desired range that is defined as a potentially acceptable lower and upper bound around the median prediction for a CSCI score given landscape development.”*

Line 279-289: I understand, in general why you would want to do this. But why use CSCI thresholds that are essentially meaningless? Isn’t it enough to say that the 0.79 threshold has a clear statistical and management footing and just stick with that? Maybe I’m missing something here.

*Currently, there is no statewide policy in place to guide how these decision-points are chosen. We thought it wise to present a range of options to explore how results vary with changing assumptions about the model.*

Line 291: Why do you need a “relative site score?” Can’t you just say you compared observed bioassessment data to the range of expected scores?

*Yes, this sentence was extraneous and removed.*

Line 305-309: Are these details needed? Did the stakeholders limit their deliberation to segments that had actual bioassessment samples? If not, this information on samples is extraneous.

*Content was removed.*

Lines 409-415: It’s not clear to me how these results are relevant to the Case Study.

*The content was removed and can be inferred from Figures 6 and 7.*

Line 437: Second half of this sentence is unclear. Also, even if there were few degraded streams in California, wouldn’t these two requirements still be true?

*Sentence was revised: “Managing for biological integrity requires the use of 1) assessment tools that can accurately evaluate condition, and 2) tools that can provide a context for evaluating observed conditions.”*

Line 454: What do you mean by “evaluate?”

*Sentence was revised: “The landscape model can place observed scores in an appropriate context relative to their modeled condition within the landscape.”*

Line 458-464: This example REALLY helps, but it needs more detailed explanation. If biological samples from a site revealed that the site met biological objectives, it isn’t intuitively obvious that it could be classified as “constrained.” Include in your example an ecologically plausible explanation of why this could be the case.

*Additional text was added: “Moreover, additional actions could be recommended to determine why these sites score above the constrained expectations, such as causal assessments to identify site-specific characteristics contributing to biointegrity (e.g., intact physical habitat independent of landscape development).”*

Line 481: Very awkward sentence.

*The topic sentence was revised: “The landscape model is primarily an exploratory tool to help identify patterns among monitoring sites where more intensive analyses may be appropriate or assist with decisions of where alternative uses may be warranted.”*

Line 483: Seems like you’ve made this statement a couple times already.

*Sentence was removed.*

Line 501: This sentence doesn’t quite make sense to me.

*Sentence was revised: “The SCAPE application was also critical for demonstrating how results from the statewide model could be used in a regional context.”*

**Reviewer 1:**

Overall Comments for Authors:

This paper describes the application of a statistical learning method to model O/E index scores using a small set of landscape predictor variables to identify which samples may have predicted distributions below reference-based thresholds. The application of landscape predictors to bioassessment scores is not novel and can, indeed, trace its history back to the earliest applications of GIS (E.g., Richard et al. 1997 and Johnson et al. 1997 initial work on biota, water chemistry and habitat and landscape predictors from 20 years ago and all the many studies that followed) and little of this original literature is acknowledged although Allan et al. 1997 (part of that same group that Richards, Johnson, and Host were in), was cited. I think the authors could do more to place their work in the science of landscape modeling of stream characteristics and that would set up the novel application of this quantile regression forests to this end.

*Our revisions to the introduction, primarily described above in response to AE comments, have addressed many of these concerns. Additional literature citations mentioned by the reviewer were also added.*

Right now the paper sits between one that is very management oriented and one that is technique development oriented. As such, I think it under-develops each and it might benefit from deciding to be one or the other. For example, I think the SGR and SCAPE tool are interesting, but they are underdeveloped and more of a distraction from the technical tool description, in my opinion. The details of the SGR exercise are not described in much detail. The SCAPE tool development, testing, and application is also under-described, in my opinion.

*Our revisions throughout, particularly in the methods, have developed a greater distinction between our general approach and our SGR case study. In our perspective, which is also shared by the AE, both components of the manuscript are fundamentally linked and achieving our goals is dependent on presenting both.*

Likewise, the quantile regression tree modeling is under-described. It is my limited understanding that these methods could choose a wide range of predictors that might be ranked in importance, etc. Why did the authors choose the predictors they did? Which were most influential? Were other combinations considered? Why or why not? Did predictors change by quantile? I think more of the technical details of that model selection, application and testing would be interesting to the readers of this journal - who tend towards the technical rather than management side - and many of whom may not have encountered this method before. I think you could remove the whole application story to the SGR and the SCAPE tool and save that for a second article for a management journal (that could get into more detail on the management implications of your predictors) and instead focus on the technical development of the constraint modeling tool - especially what other approaches might have been tried and rejected. I think the readers of this journal would want to know why you might have rejected a lot of other methods so they could learn from that journey rather than just be told this was the one model and these the only predictors used.

*We have added content to better describe model development, including descriptions of model results, that allow readers to better understand our goals and limitations of the approach. In particular, please see the additional figures that were added to supplementary material. Regarding variable selection for the model, please see our comment above to the AE.*

In my opinion, as it stands, this is a very fine and interesting piece of work and I think the writing was generally good (with a few missing words and some sections that seem to vacillate and could benefit from some clearer messaging, see below specific comments). However, I think it is more suited to a management oriented water resources journal at this time because of the reasons stated above. It reads a bit like a technical report that was developed into an article rather than an original scientific contribution to the field of stream ecology developed for that purpose alone. Again, that is not to say this is not publication worthy. I think it definitely needs to be read by the resource management community. I am just not sure FS is the appropriate venue right now (nor one that has that target audience) and I hope that is taken in the constructive light in which it is intended.

*We appreciate the concern that presenting both the technical tool and our case study with SCAPE was not well-described given the potential audiences we are targeting. However, the associate editor notes that both presentations have value for the FWS readership and we have retained both in this manuscript. Our substantial revisions to the introduction and methods have added clarity to both products to provide a better context for what each provides and how they can be used in practice to affect positive change for stream integrity. We thank the reviewer for their insightful comments, which we have addressed below to improve our manuscript.*

Specific comments follow by line number:

Line number Comment

28 Factors constrain in-stream conditions.

*Changed.*

37 …achieve goals for biological integrity…

*Changed.*

Intro first paragraph I think the first paragraph has to have the argument laid out a bit more clearly.

*See edits in response to AE.*

63 You mention site specific characteristics, but you are using landscape predictors.

*Changed “site-specific conditions” to “landscape conditions (e.g., watershed imperviousness)”.*

79-80 May want to reference the national rivers and streams restoration synthesis papers of Bernhardt, Palmer and others on how poorly we understand restoration success

*Revised sentence to “Restoring streams in urban or agricultural settings can be costly, success is not universally defined, and achieving regional reference-like conditions may be costly (Bernhardt et al. 2007, Kenney et al. 2012, Shoredits and Clayton 2013).”*

83/84 This is only CA centric - and this sentence is very vague.

*Sentence was removed.*

85/86 This is allowed by the CWA regulation - so cite to the regulation.

*Added “permitted under section 303(c)(2) of the Clean Water Act”.*

87-89 Why is this a priority? This sentence comes at the end of potential arguments for it - so put it up front.

*Sentence was moved to the beginning of the paragraph.*

94 Have not has

*Paragraph was removed in response to comments AE.*

95-98 I would have expected to see the King and Baker and Cuffney and Qian back and forth arguments on urban thresholds for biointegrity cited here.

*Paragraph was removed in response to comments from AE, but sentence was added to the prior paragraph: “In urban areas, protective thresholds for biological integrity have been debated (Cuffney et al. 2011).”*

102 There are a lot more papers on landscape predictors being used to predict stream condition, dating back 20 years. The question is, what can we do about landscape predictors?? We really can't manage many of them…easily. This needs to be woven in to the introduction a little - that there are limits to what landscape level predictors can help us with at the scale we really can manage streams - reach and segment.

*Additional citations were added to the introduction that more adequately describe the history of using landscape data to predict stream conditions. See also comments above regarding revisions to the introduction.*

104-107 I agree with you, but you have not spelled out why for the reader. Why can these help prioritize actions? For me, it is because local govts have limited budgets and have to decide where they think they can affect the greatest improvements in condition OR protect the most quality waters from degradation. And see comment above on whether "landscape alteration" is a reality. I don't think it is.

*Please see our response to the AE regarding this paragraph. We have substantially modified the content to address these concerns.*

107-109 Not sure even with abundant resources - one may need limitless resources to change "urban land use"…this, again, gets to the issue of how manageable your endpoints are. They are informative, but I would like to have seen an attempt to see if local, reach scale predictors - any - actually mattered. Moreover, we don't know anything about the ranking of these predictors you've used - which are most important? This matters, if riparian zone forest or road crossings are MOST influential, then maybe these are manageable - at least more than something as nebulous as "Urban land use". This is part and parcel of where I think this paper can flex more technical muscle - this is the type of investigative detail I think the FS audience would expect and benefit from.

*Our description of the rationale for model development in the methods addresses these concerns, in addition to our new figures in the supplementary material. But to reiterate, we are not using “manageable” predictors to understand constraints, rather our predictions develop an expectation or how much the bar could actually be moved given factors that may be impractical to manage.*

109-111 I think this can be placed elsewhere in this paragraph

*This sentence was removed.*

120-121 "management potential"? You model really focuses on bioassessment score potential not management potential. You have not measured or included management practices as predictors - so management potential only in a speculative sense.

*We disagree because the application of the model to our case study demonstrates how the results can be used to inform management. However, the use of “management potential” was slightly inaccurate as we are using the model to inform management priorities and not so much to predict a potential for management. We have modified the sentence accordingly.*

122-123 Here is where you need to cite all those papers who have shown that modeling tools for understanding biocondition along landscape gradients HAVE provided support for management. We have 20+ years of this. Hynes paradigmatic paper "The stream and its valley" - prophetic for most stream ecologists - was about this very theme. So, this is not new.

*See comments above regarding revisions to the introduction. As part of this revision, we have added the following sentence: “The relationship between stream condition and watershed characteristics has been a critical concept for ecologists in describing environmental expectations (Hynes1975, Johnson et al. 1997, Richards et al. 1997).”*

139 What different priorities?

*Added “…management priorities (e.g., restoration, protection, monitoring)…”*

144 I think 4 is too much for this paper. I would go back and spend more time on development of 1, research and compare with other approaches, and also detail more of the effects of 3. I also think there is not much or anything in this paper on comparing management decisions or options.

*Our objectives were presented with the intent of describing both the technical and applied aspects of the model, with the latter being demonstrated with the case study. We have reduced the objectives to two, but have also made extensive revisions throughout to better demonstrate both the technical and applied components of the model. The revised objectives are “1) develop the model using statewide bioassessment data to assign streams to different constraint classes and 2) use these results at a regional scale to identify how constraint classes can inform management priorities.”*

152 Overall I think the introduction is a little lengthy and need to focus on the key arguments and specific narrative, which I think can be tailored to a more technical and less management focus if the desire is to publish in this journal.

*See general comments above.*

168-169 Awkward

*Sentence was removed in response to AE comments to shorten this background content about California.*

170-171 Remove showing and remove have

*Removed.*

173 Decide on hydrologic or hydrological; insert pre-development after match

*Hydrologic used throughout. Inserted pre-development.*

176-177 Cite the source of these regionalizations. I think it is PSA?

*Citation for Ode et al. 2011 was added.*

185-186 StreamCat makes these estimates, not you. So, I think rewrite that StreamCat provides estimates of and use at the….

*Changed sentence to “…StreamCat Dataset (Hill et al. 2016) that provided estimates of land use…”*

189 Insert NHD+ between each and segment

*Added.*

193 Insert least disturbed between under and reference and cite to Stoddard et al. 2006

*Added.*

201 I would rewrite as "…and was used herein as a desired target condition". You're not really managing….

*Changed to “…used herein as a potential target condition.”*

206-208 Did sample date not figure into selection? Wouldn't you want samples from a segment closest to StreamCat data?

*Sample dates were chosen randomly for sites with multiple dates so that one sample event was matched to a site. We did not match the closest sample date to the StreamCat data for two reasons. First, the amount of effort required to do so would be excessive relative to the potential gain in model performance as described next. Second, all CSCI scores were calculated from field samples that were collected approximately during the last ten years. This time period was not considered one where rapid land cover change occurred, particularly considering the range of land types that occur statewide. As such, we feel that further screening by dates would not have provided much if any improvement in model precision. Our screen to remove all but the most downstream site on a reach was the more important factor to consider. We added the following sentence to describe our rationale for date selection: “One sample date was chosen randomly for sites with multiple dates so that one CSCI score was matched to a site. This option was preferred relative to selecting sample dates closest in time to StreamCat estimates because land use did not change dramatically during the sample period.”*

212 Add "…using StreamCat predictors/" to end of first sentence.

*Sentence was removed.*

212-214 This sentence is redundant with previous text

*Removed the sentence.*

216-219 So, how do you figure these are beyond the scope of management and why talk about the value of this model for management if the predictors are unmanageable? Wouldn't we be more interested in constraint that is actually manageable? And aren't landscape interventions like LID/GI/riparian management designed to manage these factors? I think you need to explain this more if you believe it is true. Also, what was the reasoning or testing of alternative predictors to come to this list? Was there any analysis to identify how many of these were necessary to the model performance? Model selection needs a bit more development, imo.

*As noted by all reviewers, the variables we selected were critical for developing our model and required additional description. Some of the confusion relates to the rationale for the model, that is, we are building the model to predict locations where largescale factors (e.g., land use) place upper limits on what can reasonably be expected for biological condition at a site. For example, urban streams are unlikely to achieve reference-like conditions because of instream stressors that originate from watershed factors. It is a costly and often impractical management goal to alleviate watershed stressors such as road density or impervious surfaces and it was our goal to identify locations where this may be the case. While it may be possible to address some instream stressors that are symptomatic of watershed factors (e.g., install riparian buffers to mitigate nutrients or sedimentation), the expected range of potential improvement will still be much less than sites with undeveloped or open land cover in the watershed. Our model provides a first step screening tool to begin identifying these locations where further management goals can prioritized given the expectation. It is precisely the range of observed scores around the expectation, or residual variation, that is the manageable portion.*

*Regarding variable selection, please see our comment above in response to concerns of the AE (revisions to lines 211-226).*

219-221 Awkward construct. IS it biological condition that can vary through time or potential effects?

*Sentence was removed in response to previous comment.*

244 Missing word between gradients and major

*Revised to “…gradients across major…”*

264-267 Again, I agree with you but you need to explain more of what you mean here. Give some examples or provide citations to support it.

*Sentences were removed given revisions to the introduction that more carefully explained our definitions and assumptions.*

271 For "most" analyses? Which ones did not use 0.79?

*Removed “For most analyses,”. This statement referred to our sensitivity analysis below, but we agree that it’s confusing.*

278 Remove parentheses

*Removed.*

286-289 I am not clear how this is sensitivity analysis. There is no "truth" (ie typical sensitivity analysis is how a change in a predictor changes calibration success or respond precision;) and it is not clear why the percent in each class is the right response for this.

*We agree that there is no “truth” for this model in the context of conventional sensitivity analyses. Our intent was to demonstrate how changing key decision points in the model affected the extent of stream classification, with the implicit goal of also showing that the model is flexible and not “hard-coded” for 0.79 CSCI thresholds and 10th/90th percentiles. We removed any mention of “sensitivity”.*

297 Not sure a management case study is the right topic for FS.

*We have retained the case study given comments from the AE. However, we have added text that describes how the case study differs from the technical contributions of the statewide landscape model:*

*“Results from the statewide model were used to assign one of four constraint classes described above to every stream segment in the state. Although these classes defined an expectation of biological integrity relative to the landscape, they do not provide guidance on how sites could be managed given observed bioassessment scores relative to the modelled expectation. For example, managers may prioritize sites with bioassessment scores that are above the modelled expectation differently than those that are scoring within the ranges predicted by the model. Alternatively, a site scoring as expected in an unconstrained segment could be prioritized differently than a site scoring as expected in a constrained segment. The statewide model only provides context for an observed score, whereas management priorities relative to modelled expectations must be separately defined.”*

*Also note the addition of table 3 that describes examples of the chosen priorities for the model results as applied to the SGR data.*

357-367 I admit I am surprised the model performs this poorly given your predictors; I would have expected far better prediction than 50% r-squares. That is a lot of O/E score not being resolved to advocate for this being an accurate prediction of constraint. I think more discussion of the implications of this error need to be made. There must be error around the percentiles then as well - so how off are the ranges?!?

Also, from my calculations, 0.85 O/E is your inflection where the predicted to observed agree. Below 0.85, the model predicts a lower value than observed - or more constraint than actually exists. If I have this correct, this needs to be explained to readers since it is actually happening where the threshold is.

*As noted above, relatively poor model fit is not an absolute determinant of an inadequate model. We constructed our model using landscape predictors, which are understandably not the only variables that can explain bioassessment scores. However, these variables are the primary determinants of landscape constraints as defined in the manuscript. As such, deviation of observed values from modelled predictions is representative of variation in bioassessment scores not explained by large-scale watershed characteristics (previously described in lines 223-226). This concept is critical to our statement that the landscape model is a screening tool, whereby sites scoring above or below the expectation should be prioritized differently given that landscape predictors are not the only variables that explain condition. We also provided explanation of why the model may have performed poorly in different regions of the state (i.e., lines 564-576 in the original draft).*

*Regarding potential bias relative to an approximate inflection point of 0.85, we discovered an error in our regression estimates where we were incorrectly evaluating predicted vs. observed rather than observed vs. predicted. This error has been corrected and the appropriate columns for slope and intercept estimates have been updated in the performance table. Note that we have also added figure S3 to the manuscript that shows the direct comparison of observed and predicted values. However, it is additionally noted that some bias existed in the calibration datasets for the Central Valley and Sierra Nevada sites, where the former was over-predicted and the latter under-predicted. We have added text that described this bias.*

372-373 Do you have large river bioassessment models? I thought they were for wadeables?

*Yes, the CSCI was developed for perennial streams. Although we were able to predict stream classes using the model for larger rivers, we agree that it is not entirely accurate to report results for these locations. The sentence was removed.*

374-375 Isn't this self revelatory - I mean you are only using urban and agricultural predictors, which you admit. SO, this should not be a surprise, right?

*Agreed, this sentence was removed.*

382-385 And this is consistent with the model bias (slope and intercept), correct?

*Yes, this could be caused by the model bias in these regions. We have added to the text to make this clear: “…although a slightly larger percentage of sites in the Central Valley were under-scoring compared to the other regions, which may have been caused by a slight bias of over-predicting in this region.”*

387-400 This is where I think you can develop a lot more on model selection and variable selection. How parsimonious was your model? Could you get comparable results with just urban land cover? OR urban and ag land cover? OR just road crossings….

*Please see our comment above regarding the sensitivity analyses and our response to the AE regarding the variables that were used. As correctly noted by the reviewer, this analysis is not a conventional sensitivity analysis. Moreover, our intent with the analysis was to demonstrate how the model results changed with key decision points, unrelated to which predictor variables were used. We see this as a separate objective.*

401 Again, I think this case study is underdeveloped and not really appropriate for the FS audience.

*Please see response above.*

443-447 You seem to be vacillating. I'd stick to just a support tool…

*Sentence was removed.*

447 I am not sure you can say anything about likelihood of achieving management goals. Your model really doesn't apply to where management is applied or has been applied - it is a largescale landscape model. Can you really say that a GI intervention in a subwatershed that reduces stormflow might not improve biotic condition marginally or reduce stress somewhat? Not sure that is the scale of this work. Moreover, not sure using an "exploratory" tool for management is the right language to use…at least not to managers.

*Changed sentence: “The landscape model can inform the interpretation of biotic condition and is a decision-making tool that can help identify where management goals could be focused.”*

448 This is a definite path to pursue - this tool supports TALU and supporting prioritization among watersheds given current knowledge.

*We agree that this tool could support TALU designations and we have added text as explanation (Line 464: “The landscape model could also support the development of Tiered Aquatic Life Uses (TALU, Davies and Jackson 2006), such as identifying locations where tiered uses could apply. However, the model is not intended, nor is it is sufficient, as a standalone tool for defining tiered uses.”). However, we want to make a distinction between our intention of the tool being used as a screening tool for follow-up analysis relative to formalized approaches for TALU. That is, the tool is not sufficient by itself for TALU, but it could be used to start a conversation on where TALU may be considered.*

451-453 You should read the Soranno et al. 2010 bioscience paper and Soranno et al. 2008 L&O paper. There may be a lot of ideas in there that would help here - or at least support.

*The Soranno papers are similar to our approach, where expectations are scaled on a site or regional basis. However, we note that this paper is somewhat the inverse of what we’ve done in that we predict expectations as a function of landscape development, whereas Soranno et al. 2008 predict background nutrient concentrations in the absence of anthropogenic pressures. In essence, the CSCI model developed by Mazor et al. 2016 accomplishes a similar goal by defining deviation of observed communities from expected, where the latter was defined from regional reference pools that capture natural variability. Our landscape model further extends this paradigm by developing an “anthropogenic expectation” on top of the site-level deviation of a bioassessment score that already accounts for natural variation.*

452 I am not sure that selecting appropriate management actions required considering all of these things. Some management actions (e.g., listing, TMDLs) are not dependent on this.

*True, regulatory listings are based on exceedance criteria, but we argue that management actions should be informed by all available data which, in many cases, includes biological, physical, and chemical data. Our landscape model provides context to the biology, whereas additional data could be evaluated to confirm site status or further validate model results (e.g., our recommended “monitor” action identified for the SGR case study).*

455 - Not sure the CWA allows this sort of flexibility. States can't really choose to do nothing.

*Agreed, see our next response.*

461-464 Not considering a site is not an option. What you are doing is providing a tools that helps support TALU structure - which is what should happen for many sites, rather than a life in TMDL purgatory.

*Our prior statement emphasized that the tool is not intended to provide a discount against constrained sites (i.e., doing nothing). Both this paragraph and the following provide descriptions of how the tool can be applied in both a regulatory or planning role, in addition to our case study.*

472-474 That is TALU - get rid of one size thresholding. This tool will help.

*Agreed, see previous responses.*

483-485 So now you are vacillating - either advocate for a regulatory support role for this tool or not. In the paragraphs above, you walk right up to saying this tool can support regulatory actions; but here you demur. I say if you're gonna go out on a limb, don't start chopping….

*We want to make clear that the landscape model can support regulatory activities, but it is not sufficient by itself (see addition on line 464). Thus, the support roles described in the previous paragraph represent options in how the tool can support regulation in parallel with other approaches to defining site-specific targets (e.g., causal assessment, TALU).*

486 UAA's require TALU unless you are saying the ALU should be removed all together, which I do not think you are. Urban streams support ALU - just not a reference one. But if all you have is one ALU, a UAA can only remove it until you establish alternative ALUs, like in TALU.

*This sentence was inaccurate and we are not advocating for removing ALUs with the landscape model. The sentence was revised: “…assist with decisions of where alternative uses may be warranted.”*

505 I agree - and this is TALU and regulatory…so it is a regulatory support tool!

*Yes, see our responses above. However, the statement “alternatives thresholds for biological objectives” referred to the decision point for the CSCI score that determined constrained or unconstrained classes and not site-specific thresholds. We revised the sentence to make this clear: “…specifically related to changing certainties in the CSCI score predictions (e.g., 10th and 90th percentile predictions) and the ability to explore alternative thresholds for biological objectives (e.g., 10th percentile of reference scores that defined constraint classes).”*

523 It'd be nice to back this statement up with data on engineered channels and surrounding land use - I am sure you have the data.

*We have limited data for the South Coast region on channel modification. To address this concern, we verified the number of sites in each of the four constraint classes where the stream segment was defined as hardened (modified) or natural (which included an ‘earthen’ category) – South Coast only.*

|  |  |  |
| --- | --- | --- |
|  | Hardened | Natural |
| likely constrained | 43 | 81 |
| possibly constrained | 7 | 96 |
| possibly unconstrained | 0 | 180 |
| likely unconstrained | 0 | 52 |

*This simple analysis shows that there are many constrained channels in Southern California that are not modified (as for our Tecolote Creek example). Land use as the predictors for constraint classes provided a more accurate discrimination of constraint classes because (line 522) “a constrained channel may or may not be engineered, but an engineered channel may typically be constrained given the surrounding land use”. This is especially evident for the South Coast. However, as noted on line 530, engineered channels may not always be constrained, especially in other regions of California (Stein et al. 2013). We lack the data to fully evaluate this latter statement, but the literature suggests this to be true.*

532 What about your predictors individually? Are there sites with high road density that have good biology, for example?

*See the new figure S1. There is relatively good concordance between the predictors, constraint classes, and the CSCI scores. However, as noted above, there is some variation associated with CSCI scores in relation to the predictions. That is, “good” biology is possible with high road density. Such sites are likely over-scoring, constrained sites (types 13, 14 in Figure S2).*

551-553 Did you test that here? If not, why not given it worked for others?

*No, but please see our response to the AE regarding variable selection.*

566-568 Again, this suggests more work should have been done or could have been done to address these issues and to explain the predictor selection more.

*Our responses above address this concern. However, we also note that the CSCI in general is not as sensitive to stressor gradients in these regions relative to other locations in the state. This is an issue that is well described in Mazor et al. 2016.*

602-603 Consistently indicated engineered channels were constrained? Where are those data? Did I miss them? Apologies if I did.

*See above, but these only apply to Southern California where we have partially reliable data on channel modification.*

605-606 Why the conjecture on habitat limitation and channel modification? Can you cite something or provide supporting data? This has become a paper about modified channels a lot and away from its core focus on constraint in general.

*This statement and the following on line 608-609 reiterates some of our concerns about channel modification and that the model is not a tool to identify these locations. We see constraints and channel modification as separate but confounded issues that can complicate the interpretation of what our model is meant to provide. This is why we have devoted much of the discussion to this issue that is a major concern for managers in southern California and other parts of the state.*

608-609 Doesn't this contradict earlier statements?

*Please see response to above comment.*

628-630 I don't think you can say anything about restoration potential. You can say better biological conditions are unlikely given the current landscape condition. But, you've not really tested whether restoration can help any of these streams or not. How long have we been doing restoration in any of these watersheds, at what scale, and what is the expected scale of recovery? But, you don't really have any predictors on restoration to test this.

*Agreed, we have not explicitly tested restoration potential so perhaps we are overstepping on our claims in this statement. The sentence was revised as “Overall, the model provides a tool to determine how managers can best prioritize limited resources for stream management by understanding landscape factors that might constrain each segment.”*

**Reviewer 2:**

*Note that the second reviewer provided inline comments to our manuscript with an attached pdf. These comments have been transcribed and addressed below.*

Overall, the manuscript addresses an important and practical management question/application and is a strong regional case study with extensive data. However, the Introduction does not clearly present the goals and objectives and needs extensive restructuring and editing (at least it didn't speak to me, perhaps I'm not getting something here). This is key, lines 214-219 give the objectives but it's not clear why these landscape variables were selected and more importantly that they provide a good connection to CSCI, nor why this allows the researchers to make the connection to constrained vs unconstrained. Because of the problems with the Introduction and lack of clear objectives and appropriate theoretical connections, I had issues with the whole manuscript. Then at the very end of the paper a key statement was found - lines 612-613 "Biologically constrained sites were considered those where present landscapes were likely to limit CSCI scores that describe macroinvertebrate condition." This makes a big difference and helps define what is the modeling goal, that constrained is not necessarily urban or Ag structures but is an indication that IF the land use type, amount and management of the land use doesn't change then the CSCI scores are limited and therefore constrained. I can't emphasize this more, this needs to be better described and the connections made stronger in the Intro. This changes everything.

*Thank you for the comments and your concerns about the manuscript as whole have made it clear that we need to more carefully define what we mean by constraints. Your comments on line 612-613 were well taken and we have moved this definition to the introduction. Also, as noted above, we have made substantial changes to the introduction based on your comments and those from the AE and the first reviewer. We are confident that these changes have established a stronger foundation for our work.*

I've made extensive comments within the PDF document, again, many of these can be addressed with a restructuring and more clear and direct Introduction that clearly defines the outcome of the modeling effort -- what is meant by constrained vs unconstrained and how there is a connection between the landscape GIS variables and the CSCI. Together this will make the paper more direct, clear and stronger. I think the procedures, analyses and overall goals are sound and important. I hope my comments are not taken as too harsh, but because I was missing these points at the beginning many of the statements did not make sense to me.

*We have transcribed your comments from the PDF and addressed them below. Again, our lack of clarity in the front material was addressed in revisions throughout.*

Line 64: I assume you are referring to the potential for restoration or applying best management practices and then the limits a site might have based on landscape structures like urban buildings, etc. -- the wording you are using is a bit confusing. Does this work?:

"However, achieving biological reference condition through all management options (i.e., restoration, best management practices) can sometimes be difficult or at least impeded if site-specific stream or landscape structures (i.e. semi-permanent urban structures, agriculture irrigation structures) block the adaption of management actions." Not perfect but perhaps a bit more direct and specific. ???

*Yes, that was the implication of the statement. The sentence was modified to “…may be challenging if landscape conditions (e.g., watershed imperviousness) place limits…”*

Line 66: I'm not sure what you are referring to here -- this sentence needs editing.

"Use of bioassessment information to guide decisions that affect aquatic resources can be challenging ... but what data and what about local stakeholders ? Do the local stakeholders need to have access to bioassessment data to understand that a stream is not in proper condition? Why?

*Lines 65 – 71 were revised for clarity: “Resource management decisions may be difficult to make if information is unavailable that describes these limitations. Context is required that describes how likely a site is to achieve biological integrity and how bioassessment data collected over multiple locations and times can be used to support decisions or identify priorities.”*

Line 69: I'm sorry, I'm trying but I'm not following the thought process here, --

Is this accessibility of the bio data? or -- is the message here, just that managers need to know how likely bio integrtiy is given the potential management actions that can be taken ?

*See above comment.*

Line 77: "whereas" ? would "in addition" ...

followed by "...regulation to achievement of goals." ???

*Changed to “…in addition to upstream…”*

Line 84: An example here would be good.

*Sentence was removed in response to earlier comments.*

Line 94: Remove “has been associated with”

*Paragraph was removed in response to comments from reviewer 1 and AE.*

Line 95: ..."have been implicated (stressors vary by region)." Remove “immediate”

*Paragraph was removed in response to comments from reviewer 1 and AE.*

Line 114: Can this be stated in a more direct manner -- the "Analysis methods have been explored in a bioassessment context" is indirect and confusing

*Revised: “Analysis methods that characterize biotic and abiotic factors that limit assemblage composition have been explored by others (i.e., limiting factor theory, Chessman, Muschal, and Royal 2008, Chessman 2014).”*

Line 118: ?? limits or application ?

*Sentence was revised: “Similar concepts have been applied in a landscape context to describe variation in biological communities and metrics at different spatial scales (Waite 2013, Waite et al. 2014), although they have not been developed to describe constraints as defined above”*

Line 133: I don't understand -- OK, it took me a couple more times of reading it to see what you mean, that the approaches referenced above can calculate min,max ranges and Not just Average. The sentence structure doesn't bring this out well, plus need a reference for Average methods.

*This paragraph was revised to make a better distinction between our methods and those previously used. In particular: “However, past efforts have primarily focused on characterizing condition at unsampled locations, often predicting the most likely condition by estimating averages. Alternative modelling approaches, such as quantile-based methods (e.g., Cade and Noon 2003), could be used to predict a range of expectations for biotic integrity from geospatial data.” Note that the previously cited papers were all based on average predictions.*

Line 139: The first two sentences are in "present" tense, then here you shift to past tense - need to be consistent -- change here to "are" removed "to" add colon:

*Verified tense usage in this paragraph – all verbs changed to past tense.*

Line 150: Run on sentence, needs editing and better punctuation

*Edited: “An interactive software application, the Stream Classification and Priority Explorer (SCAPE), was developed for our case study to help stakeholders choose regional management priorities from the statewide landscape model.”*

Line 207: "further" ? Are you trying to say, that the downstream site was selected for analysis in this study and upstream sites removed from consideration ? Further suggest there were additional analyses focusing on the downstream sites.

*“Further” was meant to imply that these sites were used for model calibration, although all sites were eventually classified by the developed model. The sentence was modified: “…was selected for model calibration under the assumption…”*

Line 225: This doesn't make sense -- "human activity not related to anthropogenic stressors" WHAT? human activity is anthropogenic by definition

ALSO, it's not clear WHY only landscape variables that Can't be easily addressed by management were chosen? This doesn't make sense, I thought part of your goal was to be able to prioritize where management would have the greatest potential benefit. Somethings not right here or there is a mistake or missing statement, sentence.

*Please see our responses to Reviewer 1 for the same paragraph and our response to the AE regarding variable selection. We feel that some of the confusion on model development was caused by insufficient description of why we chose specific variables from StreamCat. Hopefully our revisions clarify these decisions.*

Line 243: Move citation

*Citations were moved to appropriate locations in the sentence.*

Line 244: needs editing, missing a statement or something

*Changed to “…landscape gradients across major regions…”.*

Line 275: I'm not understanding how a ecological condition index = O/E is related to whether a site is landscape constrained or not. The CSCI O/E value of 0.79 only means that there are a relatively low number of Observed taxa and that it is below expectations. This value has Nothing to do directly with evaluated whether the landscape surrounding the stream site and the watershed are in constrained human land use. You could have a 0.79 that is due purely to WQ issues even though the watershed is unconstrained. Agriculture is a good example, there's nothing that stops an watershed that has Ag in it from that being changed, the riparian zone can be replanted, repaired, the Ag can be stopped and native plants put in place. That's just ONE example.

*Our revisions to better define what we mean by constrained sites have made this distinction clear. In short, a site with a CSCI score of 0.79 (low observed relative to expected) could be constrained or unconstrained, where the former is in an undeveloped landscape and the latter is not. Constraints are defined strictly by the landscape and describe a relative likelihood of achieving reference conditions independent of actual CSCI scores.*

Line 278: I fully agree and understand Figure 1 that says sites are altered or unaltered, etc based on values of the CSCI -- I totally disagree with assigning constrained, unconstrained labels based on the CSCI as stated above. I was expecting that you would do this solely by the type of land use in the watershed and/or riparian zone, etc. This makes sense, if the type of land use is of a semi-permanent type, a physical structure like urban buildings, concrete streams, etc. then these can be seen as constrained. Now the tricky part is coming up with thresholds for when enough of the watershed or riparian is of this constrained type (let alone what is considered NLCD land use types that = constrained) that it gets an overall constrained label, maybe it's just levels like: 0-10 in watershed or riparian is unconstrained, 10-25 is partially constrained, 25-50 is mostly constrained and > 50 constrained.

BUT using the CSCI does not work, period. The CSCI can be affected by too many Non-constrained factors so that the connection is not there.

*Our added text in the introduction speaks to this directly, i.e., “A predictive model of bioassessment scores that is based on landscape metrics (e.g., imperviousness) could describe constraints on biological integrity, whereas variation of observed scores around a model prediction could suggest other factors at the local scale (e.g., instream physical habitat) are more important.” As noted above, variation of CSCI scores is a function of multiple factors and we seek to describe a function of this variation due to factors that are difficult to manage (i.e., landscape).*

Line 354: Why aren't these graphs shown, at least in the supplemental files.

*Supplementary figures were added showing the regressions between observed and predicted scores (Figure S3).*

Line 362: Remove “Desert/Modoc, and North Coast”

*These correlations were previously presented as averages among the three regions, but this was unclear. The sentence as revised to show all three: “Model predictions for the Central Valley, Desert/Modoc, and North Coast regions had slightly lower performance compared to the statewide results, with correlations of approximately 0.66, 0.50, and 0.55 with observed values in the calibration dataset and 0.49, 0.55, and 0.55 in the validation dataset.”*

Line 363: Table 2 has values for these last two above 0.7, only CV with low values.

*See above response.*

I would guess that the CV is so low because they are few to no references sites in the region, hurting the gradient.

*This is possible, although the Central Valley does exhibit a wide range of CSCI scores, suggesting that reference conditions are possible despite the lack of reference sites in the region. The maximum observed CSCI score from our model development dataset in the Central Valley region was 0.97, which is approximately reference condition.*

Line 521: I'm not sure the message here -- "failing to recognize ... natural channels"??

*This sentence was revised: “…a regulatory framework based on direct channel modification may be insufficient because constraints are more accurately defined relative to landscape development.” In general, this paragraph describes the distinction between channel development and constraints as defined herein.*

Lines 517-536: Needs editing -- like the use of the examples but some of the sentences are incomplete or unclear -- line 526 (whereas?) need better connection here

and 531 (armored, natural or unnatural, in forest systems, this is very different than concrete channels)

*Changed line 526 to “The CSCI score is 0.61 indicating degraded biological integrity, but the channel is not modified (Rehn, Mazor, and Ode 2018).” For line 531, we agree that this is not the same as a concrete channel but we maintain that armored reaches fall under the “modified channel” category, which describes the broader classification that is often used as a justification for alternative use designations. We conducted a separate analysis of constraint classes as a function of channel alteration (defined through PHAB – physical habitat – field protocols) and found that unconstrained sites were possible even with relatively high channel alteration.*

Line 561 – 564: I want to make it clear, that I like what you are trying to do, but do not agree with the connection between CSCI and estimate of constrained. As stated before: constrained in my opinion should be estimated by GIS landscape based models using NLCD or other remote sensing data (where are structural constraints across watersheds) THEN there should be made a connection between the ecological CSCI O/E scores.

*See above, revisions to the introduction and methods have clarified our approach to avoid this confusion. Also, please see our addition to the methods that described our justification for choosing the landscape predictors for inclusion in the model (paragraph added to lines 211 – 226 in the original draft, see response to AE comments).*

Line 580-581: I'm sorry but I don't get this connection - so you assume if the CSCI deviates from expectation that that is because of stressors that can be mitigated, is that right ? BUT to me that means lots of urban and Ag stressors are ones associated with non-constrained factors OR just saying Urban and Ag are constrained land use doesn't capture all the issues nor stressors that might be able to be mitigated for, this is the part I have problems with.

*Our addition to the introduction and methods have addressed this issue. In short, our modeling approach describes variation in CSCI scores as a function of factors that are difficult to manage, whereas the remainder could be associated with stressors that are more tenable.*

Line 583: the assumptions have been used by other researchers ? I'm not sure some of these papers making the same connection nor assumptions you are here.

*The cited papers describe modelling approaches that have the general goal of describing factors that limit or constrain biological communities at different spatio-temporal scales. Our study is similar in that we are using landscape development gradients as exclusive predictors for reasons mentioned above.*

Line 611 – 613: WOW, this statement is key to your whole theory and application, why is it only now coming forth? This I might be able to get behind, the key is "considered those where PRESENT landscapes were likely to limit CSCI scores" This was not described in previous description and methods of what constrained equates to, big problem.

*This statement was moved to the introduction as our basis for defining constraints. Terminology in the statement was added throughout the manuscript to make this distinction clear.*