# Freshwater Science

## Prioritizing management goals for stream biological integrity within the developed landscape context

--Manuscript Draft--

| | |
|---|---|
| **Manuscript Number:** | 2019026 |
| **Full Title:** | Prioritizing management goals for stream biological integrity within the developed landscape context |
| **Short Title:** | Stream priorities in developed landscapes |
| **Article Type:** | Regular |
| **Corresponding Author:** | Marcus Beck<br>Southern California Coastal Water Research Project<br>Costa Mesa, CA UNITED STATES |
| **Corresponding Author's Institution:** | Southern California Coastal Water Research Project |
| **First Author:** | Marcus W Beck |
| **Order of Authors:** | Marcus W Beck |
| | Raphael D Mazor |
| | Scott Johnson |
| | Karin Wisenbaker |
| | Joshua Westfall |
| | Peter R Ode |
| | Ryan Hill |
| | Chad Loflen |
| | Martha Sutula |
| | Eric D. Stein |
| **Order of Authors Secondary Information:** | |
| **Manuscript Region of Origin:** | UNITED STATES |
| **Abstract:** | Stream management goals for biological integrity may be difficult to achieve in developed landscapes where channel modification and other factors constrain in-stream conditions. To evaluate potential constraints on biological integrity, we developed a statewide landscape model for California that estimates ranges of likely scores for a macroinvertebrate-based index that are typical at a site for the observed level of landscape alteration. This context can support prioritization decisions for stream management, like identifying reaches for restoration or enhanced protection based on how observed scores relate to the model expectations. Median scores were accurately predicted by the model for all sites in California with bioassessment data (Pearson correlation r = 0.75 between observed and predicted for calibration data, r = 0.72 for validation). The model also predicted that 15% of streams statewide are constrained for biological integrity within their present developed landscape, particularly for urban and agricultural areas in the South Coast, Central Valley, and Bay Area regions. We worked with a local stakeholder group from the San Gabriel River watershed (Los Angeles County, California) to evaluate how the statewide model could support local management decisions. To achieve this purpose, we created an interactive application, the Stream Classification and Priority Explorer (SCAPE), that compares observed scores with expectations from the landscape model to assign priorities. We observed model predictions that were consistent with the clear land use gradient from the upper to lower watershed, where potential limits to achieving biological integrity were more common in the heavily urbanized lower watershed. However, most of the sites in the lower watershed scored within their expected ranges, and were therefore given a low priority for restoration. In contrast, two low-scoring sites in the undeveloped upper watershed were prioritized for causal assessment and |

| | |
|---|---|
| | possible future restoration, whereas three high-scoring sites were prioritized for protection. The availability of geospatial and bioassessment data at the national level suggests that these tools can easily be applied to inform management decisions at other locations where altered landscapes may limit biological integrity. |
| **Suggested Reviewers:** | Andy Rehn<br>California Department of Fish and Wildlife<br>Andy.Rehn@wildlife.ca.gov |
| | Terry Flemming<br>USEPA<br>Fleming.Terrence@epa.gov |
| | Kai Chen<br>Nanjing University<br>ckai2005@gmail.com |
| | William Bouchard<br>MN Pollution Control Agency<br>will.bouchard@state.mn.us |
| **Opposed Reviewers:** | Lester Yuan<br>USEPA<br>yuan.lester@epa.gov<br>already reviewed |

**SOUTHERN CALIFORNIA COASTAL WATER RESEARCH PROJECT**
*A Public Agency for Environmental Research*

March 3rd, 2019

Dr. Charles Hawkins
Chief Editor
Freshwater Science

I am pleased to resubmit our manuscript, "Prioritizing management goals for stream biological integrity within the developed landscape context," to be considered as an original research article in Freshwater Science.

We appreciate the substantial comments on our previous submission provided by our handling editor, Dr. Dan Carlisle, and those from two reviewers. We have made substantial revisions to the manuscript to address these concerns and a point by point response is provided with our re-submission. In short, we have revised the paper to better describe the rationale and assumptions underlying the landscape model, with substantial revisions to the introduction, methods, and discussion.

It is also noted that this manuscript has undergone additional review by a science advisory panel that was convened to support the development of biointegrity/biostimulatory policies by the California Water Quality Control Boards. Comments provided by the advisory panel were also addressed in this revision and we have included a separate response with this re-submission. Many of the revisions that were made in response to the initial review from FWS also apply to concerns provided by the science advisory panel.

We are also submitting this paper with an example application that is described in an additional paper currently in review with FWS ("The importance of open science for biological assessment", manuscript no. 2019020). The previous submission uses the case study herein to demonstrate open science principles that benefited the application of our landscape model, whereas the current submission describes the theory and direct application as complementary examples. There is little overlap between the two papers given that the current describes our approach for bioassessment in "developed landscapes" and the previous focuses on the elements of open science that were used.

Our organization also agrees to submit payment for page charges if the paper is published.
We are confident that readers of FWS will find this information useful and appreciate the opportunity to publish our work in this venue.

Marcus W. Beck (marcusb@sccwrp.org), 714-755-3217
S. California Coastal Water Research Project
3535 Harbor Blvd. Suite 110
Costa Mesa, CA 92626

***Response to FWS 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 $10^{th}$/$90^{th}$ 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.*

***Response to Science Advisory Panel (SAP) 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 greatly appreciate the feedback received from the SAP on the technical aspects of this project. Our response to these comments is outlined below. These responses are accompanied by a separate document that describes the extensive revisions that were made in response to review comments received from our initial submission to Freshwater Science. In many cases, our revisions to comments from FWS reviewers apply to concerns from the SAP and we reference these revisions below when appropriate.*

This document describes a modeling approach to estimate the likely degree that 'large-scale, historical impacts from landscape alteration' constrains the condition of aquatic life at individual waterbodies defined here as reaches within the National Hydrography Dataset Plus. The goal of the project was to 'present the development an application of a landscape model to classify and prioritize stream monitoring sites based on probable ranges of bioassessment scores relative to landscape alteration'. Such a model could be off considerable value in that it could provide a rapid and inexpensive way of identifying streams that have been so greatly modified that they can no longer achieve biological integrity.

The charges to the science advisory panel were to:

- Comment on the adequacy of the data set and the analytical approaches to predict ranges of biointegrity scores associated with landscape development.
- Comment on the evaluation of performance and findings of the Channels in Developed Landscape Tool, including applicability of the Tool to the range of constructed or hydro-modified channels.
- Consider if there are technical ways to address stakeholder concerns?

Conceptual background

As acknowledged in the document, water resource managers need a way of estimating the degree to which individual waterbodies can support or achieve biological integrity goals. Specifically, we need objective ways of identifying waterbodies that are constrained in achieving biological integrity by historical land use and waterway alteration. This knowledge can then be used by policy makers to set practical goals for the protection or restoration of individual waterbodies and to prioritize management actions across waterbodies.

The conceptual framework underpinning the approach is not wholly new, and the document would be improved by at lease briefly pointing out that the approach is conceptually similar to ideas that have been discussed extensively elsewhere – e.g., the Driver-Pressure-State-Impact-Response (DPSIR) conceptual framework (e.g., Smeets and Wethering, 1999).

Smeets, E. and R. Weterings. 1999. Environmental indicators: typology and overview. European Environmental Agency Technical Report No. 25. 19 p

*The conceptual model underpinning the approach was clarified in our response to reviewer comments from FWS. We have also added a brief explanation of the DPSIR framework at appropriate points in the manuscript to clarify the connection with our approach. Also note the addition of Figure 2 that provides a conceptual diagram for our approach within the DPSIR framework.*

*Introduction, fourth paragraph: "These approaches can be conceptualized within the Driver-Pressure-Stress-Impact-Response (DPSIR) framework that describes relationships between the origins and consquences of environmental problems (Smeets and Weterings 1999)."*

*Methods, second paragraph: "A general assumption was that water quality issues could be conceptually linked to societal and economic drivers (Smeets and Weterings 1999), reflected through the link between land use and stream biotic integrity."*

**Clarity of the document**

The document was generally well written and clear, but the clarity of some terms and sections could be improved. The need for improved clarity was noted by members of the science advisory panel, the external reviewers of the submitted manuscript, and the stakeholders. The authors define their use of the term 'constrained' to mean situations in which "reference conditions for the biological community may be difficult to achieve with limited resources because of large-scale, historical impacts from landscape alteration". In isolation, this definition is straightforward, but the term 'constrained' is used in the general stream ecology literature to mean that the location and morphology of the stream channel is physically constrained by natural landscape elements (e.g., canyon walls). The authors acknowledge this historical use of the term in the discussion, but it should be addressed early on. The authors should also strive to minimize the use of jargon and ensure that their use of terms is not ambiguous. For example, the authors use the term 'context' throughout the document, but its specific meaning is not always clear, and its meaning seems to vary.

*Our revision for FWS included a more thorough and pointed definition of "constrained" in the introduction.*

*"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."*

*We have also made a note in the above paragraph to distinguish between "constrained" in the current paper and as defined in general stream ecology: "This definition describes a biological expectation and is distinct from the classical definition used in the general stream ecology literature (e.g., a physically constrained channel in the morphological sense)."*

*All uses of "context" in the current manuscript were also reviewed for clarity. All instances now refer exclusively to the "landscape context" for evaluating biological condition provided by the model.*

**The modeling approach**

The objective of the modeling was to estimate the range of CSCI values likely to occur at individual waterbodies as a function of site-specific landscape alteration. The authors developed a statewide quantile random forest model to estimate upper and lower quantiles (and hence prediction intervals) that bracket the mean (or median) predicted values. The upper and lower prediction quantiles can be set by the user depending on how certain the user wants to be that the actual CSCI value falls within the prediction interval (e.g., 95%, 90%, 80%, etc.). The panel thought this was a technically sound approach to estimating site-specific prediction intervals for random forest models, which often describe complex, non-linear relationships. The data on which the model was calibrated and validated were robust (3,252 sites) stratified across 6 regions and 3 levels of watershed imperviousness. Analyses of model performance indicate that the model is reasonably precise (pseudo-$r^2 = 0.62$) and unbiased.

Although we understand why modeling prediction intervals is useful, it seems like this approach would complement a simpler approach of comparing observed CSCI values with the values predicted from a regular random forest regression based on these same predictors. The simple difference between observed and expected CSCI values represent the potential scope of improvement that could be achieved.

*There are certainly alternative approaches that could be used for evaluating the observed scores relative to the modelled results. The approach described above (observed minus median prediction) would be simpler and, in fact, this information is currently provided on the SCAPE website as an option to view the mapped results. However, we opted for the former approach for the final model given the flexibility offered by choosing different prediction intervals. This was noted as a useful feature in discussions with our stakeholder group during the case study. This is currently described in the manuscript as "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."*

Miscellaneous questions in need of answers include:

1. Do the same set of predictor variables provide the same level of performance at all quantiles? If not, how would this information be used when interpreting the 'contraints' on upper and lower CSCI values.
   *This is an interesting question that deserves some attention. Predictive performance can be evaluated for conditional quantiles using a goodness of fit measure described in Koenker and Mochada 1999. This measure has a similar interpretation as R-squared and a follow-up analysis that evaluates how fit varies at different quantiles for different predictors could provide additional insight into constraints. We make a note in the discussion that this analysis could be informative: "Additional analyses that evaluate how different predictors influence model performance at different quantiles could provide insight into how landscape factors relate to constraints (e.g., Koenker and Mochado 1999)."*

   *Koenker, R., Mochado, J.A.F. 1999. Goodness of fit and related inference processes for quantile regression. Journal of the American statistical Association. 94(448):1296-1310.*

2. Can you tease out what constraints can be mitigated and which ones cannot? This seems like a central issue vis-à-vis UAA or tiering decisions.
   *As stated in our definition of constraints, we describe expected ranges of biological condition as a function of factors that are difficult to manage, e.g., watershed imperviousness, road density, etc. This design was intended and we suggest that variation around the predictions could potentially be related to factors that can be mitigated. An addition to the introduction describes this point: "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."*

*We also provided additional text about how the model can support UAA or TALU in this context, but we note that the model is not sufficient by itself for doing so: "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"*

**Comprehensiveness and limitations of the model**

The current model uses 6 predictors: canal density, % impervious surface, % urbanized area, % area in agriculture, road density, and density of road stream crossings. The panel wondered if the model's predictive accuracy and hence its utility may be compromised by environmental alterations that are not captured by STREAMCAT – e.g., hydrological modifications. The panel was somewhat confused regarding what breadth of land use factors (predictors) should ideally be included in the model. The degree of permanence of the land uses included in the model varies. To most accurately estimate the effects of truly permanently modified landscapes, the model should perhaps only include measures associated with urbanization and dams. But, if the intent is to better understand the relative effects of the suite of land use alterations that exist in California, it makes sense to include measures of agriculture, riparian condition, and forest management. As presented, the document seems to want the model to serve both functions, and the overall goal of its application and how it could be used for a variety of purposes needs more through development and explanation. The panel wondered how well the STREAMCAT coverage adequately represented alteration in riparian land cover, and whether it would be better measured by photo-imagery. We also thought STREAMCAT land use condition estimates should be validated with field (or Goggle Earth) observations to assess how much error in predictions could be associated with error in quantifying predictor values.

*The choice of predictors was a main point of concern of the reviewers for FWS and we have made extensive revisions to justify our choices, in addition to describing alternatives in the discussion that could be explored as needed.*

*A revision to the methods in response to FWS reviewer comments describes our rationale for selecting predictors.*

*"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."*

The panel could not tell how predictions, and hence interpretations and applications, might be affected by whether streams are perennial or nonperennial. This is an important issue because you say the model was developed for all streams and rivers in California, but differences in aquatic life associated with water permanence is not addressed in any formal way. This seems like it is a serious issue that needs resolving by better specifying the domain of streams the model applies to. If it truly applies to all streams and rivers, you need to specifically show that predictions and interpretations do not vary between perennial and nonperennial streams.

*An addition to the discussion highlights the concern about permanence.*

*"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."*

In general, we agree the model might be best used as a screening tool, but we still need to know how often it might mislead.

Miscellaneous questions related to model comprehensiveness and limitations:

1. What are the model limitations in both in terms of predictive performance and how it could inform policy in its current state of development?
   *Additional content was added that provided more information about model performance.  Figure S4 in the supplement, in particular, was added to provide a more thorough evaluation of regional performance.*

2. Can the model be used in conjunction with the nutrient-stressor response models?
   *We envision that landscape constraints identified by the model could be implemented in conjunction with the nutrient-stressors response models.  However, because the latter is currently under active development, we can not comment specifically on how this may be approached.  It will be considered as we move forward with the biostimulatory products.*

**Value added and other uses of the tool**

Uses for the on-line application beyond trying to predict biological condition are of interest. The tool might have utility in linking to flow-inundation models and identifying groundwater/surface water zones of gaining and losing stream reaches.
*We agree that additional functionality for the online tool could be useful, particularly as it relates to flow regime.  We feel the current application is generally sufficient for what is outlined in the current draft.  However, we also recognize its limitations (many of which are outlined in the discussion) and we are actively working with the SGR stakeholders on the second phase of this project that will expand the current application (e.g., incorporation with causal assessment models, additional lines of evidence, etc.).*

Can you add TN and TP to the suite of predictor variables in the model and examine partial dependence plots to help inform nutrient criteria? Perhaps classify stream reaches according the level of constraint and fit TN/TP models within each class.
*As noted above, our justification for including landscape predictors rather than instream measures is in line with our assumptions for what the landscape model is meant to provide. However, we do see value in how these models can inform nutrient criteria, as for the biostimulatory response models currently under development. We will consider how the landscape model can inform this work as we move forward.*

**Stakeholder concerns**

Most of the stakeholder concerns involved how the tool would be used within a policy context. Those questions are formally outside of the purview of the science advisory panel, but we note the following questions are important to address.

1. Where along the perennial-nonperennial continuum does the model not apply?
   *Please see response above.*
2. Will this tool ultimately be used for use attainability assessments? If so, how would it best be used for that purpose?
   *Please see response above. We see that this tool could help support UAA but it is not insufficient by itself for doing so. Policy guidance in this area will inform exactly how the landscape model could be used in this role.*

**Miscellaneous**

Some panel members thought the case study was interesting but may not be appropriate for journal article. However, the handling editor, while asking for significant revision, thought the combination of model development and case study was a strong aspect of the paper.
*Our response to comments from the handling editor states our agreement and we have retained the case study in the paper. However, substantial revisions to the current draft have added clarity regarding the purpose of the model and how the case study provides a regional demonstration of this purpose.*

# Running head: Stream priorities in developed landscapes

# Prioritizing management goals for stream biological integrity within the developed landscape context

Marcus W. Beck (marcusb@sccwrp.org)[1*]

Raphael D. Mazor (raphaelm@sccwrp.org)[1]

Scott Johnson (scott@aquaticbioassay.com)[2]

Karin Wisenbaker (karin@aquaticbioassay.com)[2]

Joshua Westfall (jwestfall@lacsd.org)[3]

Peter R. Ode (peter.ode@wildlife.ca.gov)[4]

Ryan Hill (hill.ryan@epa.gov)[5]

Chad Loflen (Chad.Loflen@waterboards.ca.gov)[6]

Martha Sutula (marthas@sccwrp.org)[1]

Eric D. Stein (erics@sccwrp.org)[1]

[*] Corresponding author

[1] Southern California Coastal Water Research Project, 3535 Harbor Blvd., Costa Mesa, CA 92626

[2] Aquatic Bioassay & Consulting Laboratories, Inc. 29 North Olive, Ventura, CA 93001

[3] Sanitation Districts of Los Angeles County, 1955 Workman Mill Road, Whittier 90601

[4] California Department of Fish and Wildlife, Office of Spill Prevention and Response, 2005 Nimbus Road, Rancho Cordova, CA 95670

[5] Department of Forest Engineering, Resources, and Management, Oregon State University c/o Western Ecology Division, 140 Peavy Hall, 3100 SW Jefferson Way, Corvallis, OR 97333

[6] San Diego Regional Water Quality Control Board, Healthy Waters Branch, Monitoring and Assessment Research Unit, 2375 Northside Drive, San Diego, CA 92108

# Abstract

Stream management goals for biological integrity may be difficult to achieve in developed landscapes where channel modification and other factors constrain in-stream conditions. To evaluate potential constraints on biological integrity, we developed a statewide landscape model for California that estimates ranges of likely scores for a macroinvertebrate-based index that are typical at a site for the observed level of landscape alteration. This context can support prioritization decisions for stream management, like identifying reaches for restoration or enhanced protection based on how observed scores relate to the model expectations. Median scores were accurately predicted by the model for all sites in California with bioassessment data (Pearson correlation $r = 0.75$ between observed and predicted for calibration data, $r = 0.72$ for validation). The model also predicted that 15% of streams statewide are constrained for biological integrity within their present developed landscape, particularly for urban and agricultural areas in the South Coast, Central Valley, and Bay Area regions. We worked with a local stakeholder group from the San Gabriel River watershed (Los Angeles County, California) to evaluate how the statewide model could support local management decisions. To achieve this purpose, we created an interactive application, the Stream Classification and Priority Explorer (SCAPE), that compares observed scores with expectations from the landscape model to assign priorities. We observed model predictions that were consistent with the clear land use gradient from the upper to lower watershed, where potential limits to achieving biological integrity were more common in the heavily urbanized lower watershed. However, most of the sites in the lower watershed scored within their expected ranges, and were therefore given a low priority for restoration. In contrast, two low-scoring sites in the undeveloped upper watershed were prioritized for causal assessment and possible future restoration, whereas three high-scoring sites

were prioritized for protection. The availability of geospatial and bioassessment data at the national level suggests that these tools can easily be applied to inform management decisions at other locations where altered landscapes may limit biological integrity.

*Key words*: Bioassessment, biotic integrity, streams, urbanization, modified channels, landscape stressors, random forests, prioritization, data visualization, stakeholder group

# Introduction

The widespread use of bioassessment data to assess ecological condition of aquatic environments is a significant advance over chemical or physical methods of assessment, yet managers and stakeholders require contextual information for synthesizing and interpreting biological information. The reference condition concept that is built into many biological indices provides a broad context for observed condition relative to unaltered habitats for a particular region (Reynoldson et al. 1997; Stoddard et al. 2006). However, achieving a reference condition of biological integrity (i.e., having structure and function comparable to natural habitat for the same region, Karr et al. (1986)) may be challenging if landscape conditions (e.g., watershed imperviousness) place limits on spatial and temporal scales that can be effectively managed (Chessman and Royal 2004; Chessman 2014). 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.

Prioritizing among sites that are affected by landscape alteration is a critical challenge for managers in urban and agricultural settings (Walsh et al. 2005; Beechie et al. 2007; Paul et al.

2008). In developed landscapes, the majority of stream miles are in poor biotic condition (USGS 1999; Finkenbine, Atwater, and Mavinic 2000; Morgan and Cushman 2005). Managing streams in urban or agricultural settings can be costly, success is not universally defined, and achieving regional reference-like conditions may not be feasible (Bernhardt et al. 2007; Kenney et al. 2012; Shoredits and Clayton 2013). Conventional approaches to protect and restore biological integrity have commonly focused on direct improvements at the site level to mitigate instream stressors (Carline and Walsh 2007; Lester and Boulton 2008; Roni and Beechi 2012; Loflen et al. 2016), in addition to upstream preventive measures that may be incentivized or enforced through regulation. Although these approaches can lead to improvements in ecological condition, there is no universal remedy for achieving biological integrity in streams. In urban areas, protective thresholds for biological integrity been debated (Cuffney et al. 2011). Moreover, extensive modifications to streams for flood control or water conveyance are common in developed landscapes. For biological integrity, several states have implemented a tiered aquatic life use or alternative use designations to account for baseline shifts in ecosystem condition from channel modification (e.g., FDEP 2011; USEPA 2013; MBI 2016; permitted under section 303(c)(2) of the Clean Water Act).

Herein, we define constrained streams as those where present landscapes are likely to limit biological integrity. This definition describes a biological expectation and is distinct from the classical definition used in the general stream ecology literature (e.g., a physically constrained channel in the morphological sense). 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.

The relationship between stream condition and watershed characteristics has been a critical concept for ecologists in describing environmental expectations (Hynes 1975; Johnson et al. 1997; Richards et al. 1997). Consistent and empirical links between land use thresholds and poor biotic integrity have been identified in many cases (Allan, Erickson, and Fay 1997; Wang et al. 1997; Clapcott et al. 2011) and previous modelling efforts have successfully used geospatial data to predict stream condition at regional or national scales using geospatial data (Vølstad et al. 2004; Carlisle, Falcone, and Meador 2009; Brown et al. 2012; Hill et al. 2017). These approaches can be conceptualized within the Driver-Pressure-Stress-Impact-Response (DPSIR) framework that describes relationships between the origins and consequences of environmental problems (Smeets and Weterings 1999). 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. This approach differs fundamentally from previous efforts of estimating average condition by providing an estimate of the minimum and maximum scores that are likely for the landscape context. Once the responses of macroinvertebrate communities to landscape changes at large spatial scales are understood, expectations can be compared to field samples and sites can be prioritized by local managers based on deviation from the expectation.

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. Our specific objectives were to 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. The model was developed and applied to all streams and rivers in California, specifically focusing on the potential of urban and agricultural land use to constrain biological condition. Our case study demonstrated how the statewide model could be used to classify and prioritize at the regional scale using guidance from a local stakeholder group from a heavily urbanized watershed where obstacles for achieving biological integrity have been encountered. 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.

# Methods

## Study area and data sources

California covers 424,000 $km^2$ of land with extreme diversity in several environmental gradients, such as elevation, geology, and climate (Figure 1a, Ode et al. (2016)). Temperate rainforests occur in the north (North Coast region, NC), deserts and plateaus in the northeast and southeast (Deserts and Modoc Plateau region, DM), and Mediterranean climates in coastal regions (Chaparral and South Coast regions, CH and SC). The Central Valley region (CV) is largely agricultural and drains a large mountainous area in the east-central region of the state (Sierra Nevada region, SN). Urban development is concentrated in coastal areas in the central (San Francisco Bay Area, Chapparal region) and southern (Los Angeles, San Diego metropolitan area, South Coast) regions of the state. Landscape alteration has been relatively recent, with one estimate that developed lands increased in California by 38% from 1973 to 2000 (Sleeter et al. 2011). Silviculture and logging activities have also occurred in forested regions (SN, NC). For analysis, the state was evaluated as a whole and by the major regions described above (Ode et al. 2011).

The landscape model was developed using land use data, stream hydrography, and biological assessments. A general assumption was that water quality issues could be conceptually linked to societal and economic drivers, reflected through the link between land use and stream biotic integrity (e.g., under the DPSIR framework, Figure 2, Smeets and Weterings (1999)). Stream data from the National Hydrography Dataset Plus (NHD-plus) (McKay et al. 2012) were used to identify stream segments in California for modelling biological integrity. The NHD-plus is a surface water framework that maps drainage networks and associated features (e.g., streams, lakes, canals, etc.) in the United States. Stream segments designated in the NHD-plus were used as the discrete spatial unit for modelling biological integrity. Here and throughout, “segment” is defined based on NHD-Plus flowlines. Hydrography data were combined with landscape metrics available from the StreamCat Dataset (Hill et al. 2016) that provided estimates of land use at the riparian zone (i.e., a 100-m buffer on each side of the stream segment), the catchment (i.e., nearby landscape flowing directly into the immediate stream segment, excluding upstream segments), and the entire upstream watershed for each NHD-Plus segment. Many of the metrics in StreamCat were derived from the 2006 National Land Cover Database (Fry et al. 2011).

The California Stream Condition Index (CSCI) (Mazor et al. 2016) was used as a measure of biological condition in California streams. The CSCI is a predictive index that compares the observed taxa and metrics at a site to those expected under least disturbed reference conditions (Stoddard et al. 2006). Expected values at a site are based on models that estimate the likely macroinvertebrate community in relation to factors that naturally influence biology, e.g., watershed size, elevation, climate, etc. (Moss et al. 1987; Cao et al. 2007). The index score at a site can vary from 0 to ~ 1.4, with higher values indicating less deviation from reference state. Because the index was developed to minimize the influence of natural gradients, the index scores have consistent meaning across the state (Mazor et al. 2016). A CSCI threshold of 0.79, based on the tenth percentile of scores at all reference calibration sites, has been proposed to identify the likelihood of biological alteration (SDRWQB 2016) and was used herein for applying the landscape model.

Benthic macroinvertebrate data were used to calculate 6270 individual CSCI scores at nearly 3400 unique sites between 2000 and 2016 (Figure 1b). 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 (2015)), although other programs 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.

Field samples were collected during base flow conditions typically between May and July following methods in Ode, Fetscher, and Busse (2016). Bioassessment sites were snapped to the closest NHD-plus stream segment in ArcGIS (ESRI 2016). In cases where multiple sites were located on the same segment, the most downstream site was selected for model calibration under the assumption that the landscape data in StreamCat was most relevant to this site. 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. This created a final dataset of 2620 unique field observations used to calibrate and validate the landscape model.

## Building and validating the landscape model

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.

The model was developed using quantile regression forests to estimate ranges of likely CSCI scores in different landscapes (Meinshausen 2006, 2017). Random forests are an ensemble learning approach to predictive modelling that aggregates information from a large number of regression trees and have been used extensively in bioassessment applications (Carlisle, Falcone, and Meador 2009; Chen et al. 2014; Mazor et al. 2016; Fox et al. 2017). Random forest models provide robust predictions by evaluating complex, non-linear relationships and interactions between variables relative to more commonly-used modelling approaches, such as multiple regression (Breiman 2001; Hastie, Tibshirani, and Friedman 2009). Quantile models, such as quantile regression forests, evaluate the conditional response across the range of values that are expected, in contrast to conventional models that provide only an estimate of the mean response (Cade and Noon 2003). This modelling approach allows use of prediction intervals to describe the range of likely scores, which can be used to identify sites where that range includes management targets. Quantile regression forests were used to predict CSCI scores in each stream segment at five percent increments (i.e., 5$^{th}$, 10$^{th}$, etc.) from the 5$^{th}$ to 95$^{th}$ percentile of expectations. 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 50$^{th}$ percentile prediction was the most likely score for a stream segment given observed values for landscape variables, whereas a lower (e.g., 5$^{th}$ percentile) and upper (95$^{th}$ percentile) bound on the median (and points in between) were also predicted. The quantregForest package (Meinshausen 2017) for the R Statistical Programming Language (RDCT 2018) was used to develop the landscape model using the default settings, with the exception that out of bag estimates were used for model predictions.

We stratified sample data to ensure sufficient representation of landscape gradients across major regions in the state (Figure 1). 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). The remaining sites were used for model validation (n = 655). Where multiple samples were available at a single site, one sample was selected at random for both calibration and validation purposes. Model performance was assessed for the statewide dataset and within each major region by comparing differences between observed CSCI scores and median predictions at the same locations. Differences were evaluated using Pearson correlations and root mean squared errors (RMSE); high correlation coefficients and low RMSE values indicated good performance. Regression analysis between observed and predicted scores was used to assess potential bias based on intercept and slope values differing from 0 and 1, respectively. Collectively, the performance metrics were chosen to evaluate both predictive ability of the landscape model and potential for bias which may vary depending on different land use gradients across the state.

## Statewide application of the landscape model

We applied the landscape model to 138716 stream segments statewide to estimate the extent of streams in one of four different constraint classes: likely unconstrained, possibly unconstrained, possibly constrained, and likely constrained (Table 2). The classification process is described in Figure 3a through c. Classifications were based on the comparison of a CSCI threshold representing a management goal and the predicted range or predicted median score at a segment. These two decision points (i.e., the threshold and the size of the predicted range) were critical in defining segment classifications. We used a CSCI treshold of 0.79 (i.e., the $10^{th}$ percentile of reference calibration sites) following previous examples (Mazor et al. 2016; SDRWQB 2016) and a prediction interval ranging from the $10^{th}$ to the $90^{th}$ percentiles of expected CSCI scores for the level of landscape development at each segment. Stream segments with the range of CSCI score expectations entirely below the threshold were considered likely constrained, whereas those with expectations entirely above were considered likely unconstrained (Figure 3c). The remaining sites were classified as possibly unconstrained or possibly constrained, based on whether the median expectation was above or below the threshold respectively (Table 2).

The influence of the key decision points on the extent of segment classifications created by the landscape model was evaluated. Stream segment classifications depend on the chosen range of score expectations (or certainty) from the landscape model (Figure 3b) and the CSCI threshold for evaluating the overlap extent (Figure 3c). 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. Eight different ranges of values for the score expectations from wide to narrow were evaluated at five percent intervals, i.e., $5^{th}$-$95^{th}$, $10^{th}$-$90^{th}$, …, $45^{th}$-$55^{th}$. Different CSCI thresholds were also evaluated using values of 0.63, 0.79, and 0.92, corresponding to the $1^{st}$, $10^{th}$, and $30^{th}$ percentile of scores at reference calibration sites used to develop the CSCI (Figure 1b) (Mazor et al. 2016). The percentage of stream segments in each class statewide and by major regions were estimated for each of the twenty-four scenarios (width by threshold combinations).

Sites were further classified by comparing observed CSCI scores from biomonitoring data to the range of expected scores (Figure 3d). Sites with observed scores above the upper limit of the segment expectation (e.g., above the $90^{th}$ percentile of expected scores) were considered “over-scoring” and sites below the lower limit (e.g., $10^{th}$ percentile) were considered “under-scoring”. If neither “over-scoring” nor “under-scoring”, the site was considered as “expected” within the context of the landscape model.

## Defining management priorities in the San Gabriel River watershed

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 landscape context for an observed score, whereas management priorities relative to modelled expectations must be separately defined.

A regional application of the statewide results allowed a local stakeholder group to develop a framework for evaluating data from a watershed monitoring program to prioritize management actions. The San Gabriel River (SGR) Regional Monitoring Program (Los Angeles County, California) includes stakeholders from water quality regulatory agencies, municipalities, and non-governmental organizations that cooperatively work to manage aquatic resources in the watershed and improve coordination of compliance and ambient monitoring efforts. A strong land-use gradient occurs in the SGR watershed that creates challenges for managing stream condition (Figure 4). The upper watershed in the San Gabriel mountains is largely undeveloped or protected for recreational use, whereas the lower watershed is in a heavily urbanized region of Los Angeles County. The SGR is dammed at four locations in the upper watershed for flood control. Spreading grounds in the middle of the watershed are used to recharge groundwater during high flow. As a result, the upper and lower watersheds are hydrologically disconnected when annual rainfall is normal. Nearly all of the stream segments in the lower half of the watershed are channelized with concrete or other reinforcements. The majority of flow in the lower watershed is provided to the mainstem and major tributaries of the SGR by wastewater treatment plants releasing tertiary treated effluent. Approximately half of the monitored sites in the watershed are in poor biological condition, nearly all of which are in the lower watershed.

Stakeholders identified their relevant priorities by evaluating the different site types that were possible from the landscape model relative to the stream classes. The priorities defined by the group were generalized into three categories (Table 3):

- Investigate: Conduct additional monitoring or review of supplementary data (e.g., field visits, review aerial imagery);
- Protect: Recommend additional scrutiny of any proposed development and/or projects;
- Restore: Pursue targeted action for causal assessment and/or restoration activity.

A template that showed the possible site scores relative to the segment classifications was given to the stakeholders (Figure S2, left side). The three priorities were then assigned a low, medium, or high importance for the scoring possibilities that could occur from the landscape model (Figure S2, right side). The assignments were made with the explicit recognition that any priority recommendations were in addition to baseline monitoring and maintenance that is currently provided by existing management programs. The final assignments were then mapped to each monitoring site in the watershed. Table 3 shows examples of the priority recommendations and sites for which they applied.

The outcomes of these assignments were visualized in an interactive and online application, the Stream Classification and Priority Explorer (SCAPE, Figure S3, http://shiny.sccwrp.org/scape/)(Beck 2018b). The application allowed stakeholders to provide input on the two key decision points for classifying stream segments (i.e., choice of a threshold and a prediction interval), as well as to assign priorities to each management action described above. The application then allowed stakeholders to see the outcomes of these decisions. Specifically, SCAPE created maps showing the classifications for segments in the watershed, deviation of observed CSCI scores from the expectation, and maps of recommended priority actions that were assigned to each of the scoring possibilities. In addition, the application tabulated the extent of streams in each class, as well as the number of sites prioritized for each management action. Crucially, SCAPE allowed the stakeholders to modify key decisions points in the model and rapidly evaluate how these changes propagated to changes in recommended priorities for each site.

# Results

## Model performance

Model performance statewide indicated generally good agreement between observed CSCI scores and the median prediction for the associated stream segment (Table 4, Figure S4). Agreement between observed and predicted values for the entire calibration dataset was r = 0.75 (Pearson) and RMSE = 0.17. The intercept and slope for a regression between observed and predicted values were 0.04 and 0.93, suggesting minimal bias of predictions. The statewide validation data showed similar results, with slightly smaller correlation (r = 0.72) and larger RMSE (0.18) estimates.

Overall, the model performed well in regions with a mix of urban, agricultural, and open land (e.g., South Coast and Chaparral regions), whereas performance was weakest in regions without strong development gradients (e.g., Sierra Nevada region) (Table 4, Figure S5). Performance for the Chaparral and South Coast regions were comparable or slightly improved compared to the statewide dataset for both the calibration (r = 0.71, 0.75, respectively) and validation (r = 0.74, 0.72) datasets. 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. Model performance was weakest for the Sierra Nevada region, where timber harvesting, rather than urban or agricultural development, is the most widespread stressor. A slight bias in model predictions was observed for the Central Valley and North Coast, where the former was over-predicted and the latter was under-predicted (Figure S4).

## Statewide patterns in stream constraints

Statewide patterns in stream constraints were apparent from the results of the landscape model that were consistent with land use (Figure 5). A majority of stream segments statewide were classified as possibly constrained (11% of all stream length) or possibly unconstrained (46%), whereas a minority were likely constrained (4%) or likely unconstrained (39%) (Table 5). Likely unconstrained streams were common in the Sierra Nevada (50%), North Coast (46%), and Desert/Modoc (46%) regions, whereas likely constrained were relatively abundant in the Central

Valley (22%) and South Coast (15%) regions. However, constrained and unconstrained streams were both found in every region (Figure 5)

Observed CSCI scores were within the predicted range as often as expected (i.e., 80% statewide, based on the $10^{th}$ and $90^{th}$ prediction interval), and over-scoring sites were roughly as common (9%) as under-scoring sites (10%) (Table 6). Similar patterns were observed within regions, 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. Over-scoring sites were slightly more common in certain regions (i.e., the South Coast and Sierra Nevada regions) than others (i.e., the Chaparral, Central Valley, and Desert/Modoc regions).

Changing key decision points of the landscape model affected the estimates of the extent of streams in each class (Figure 6). Unsurprisingly, decreasing the certainty of predictions from the landscape model by narrowing the prediction interval ($5^{th}$-$95^{th}$ to $45^{th}$-$55^{th}$) shifted a number of streams from the possible to likely category in both constrained and unconstrained segments. Similarly, changing the CSCI threshold from relaxed to more conservative (0.63 to 0.92) increased the number of streams classified as possibly or likely constrained and decreased the number of streams as possibly or likely unconstrained. However, the effects of these decision points varied greatly by region. For example, over 80% of segments in the Central Valley were classified as likely constrained using a high CSCI threshold with the narrowest range of predictions, whereas less than 1% of segments were in this category using a low CSCI threshold with the widest range of predictions. Opposite trends were observed in regions with reduced land use pressures. For example, almost all stream segments in the North Coast and Sierra Nevada regions were classified as likely unconstrained using a low CSCI threshold and narrow range of predictions.

## San Gabriel River Case study

Application of the landscape model results to the CSCI scores provided biological expectations consistent with the strong land use gradient in the watershed (Figure 7). Stream segments in the upper watershed were a mix of likely and possibly unconstrained (40% and 28%), whereas stream segments in the lower watershed were classified as likely and possibly constrained (25% and 7%). Several segments in the lower watershed had median CSCI scores that were very close to the $10^{th}$ percentile (i.e., right-skewed) consistent with extreme landscape pressures (bottom left, Figure 7b).

The SCAPE application was effectively used to select management priorities for all monitoring sites in the SGR watershed. In general, the stakeholder group assigned high priority recommendations to over- and under-scoring sites in likely unconstrained segments or those below the biological threshold with possibly unconstrained classification (Figure S2, Table 3). Continuing current practices (e.g., routine monitoring) were generally recommended at constrained sites or restoration actions were recommended as a lower priority despite low CSCI scores. Recommended actions to investigate were more common for both over-scoring and under-scoring sites, protect was given a high priority exclusively at over-scoring sites, and restore was more common at under-scoring sites.

The SCAPE application also allowed the stakeholders to identify spatial patterns among the watershed priorities. For example, a clear distinction between low and high priority actions was observed on the watershed map (Figure 8). Sites in the lower watershed were lower priority if an action was recommended, whereas the five high priority sites were in the upper watershed (multiple recommendations were assigned to the sites). The distinction between lower and higher priorities between the lower and upper watershed was driven exclusively by the segment classifications, where constrained segments were in the lower watershed and unconstrained segments were in the upper watershed. Several sites that were scoring as expected for likely and possibly unconstrained segments in the upper watershed were recommended as medium priority for protection.

# Discussion

Managing for biological integrity requires the use of 1) assessment tools that can accurately evaluate condition, and 2) tools that can provide a landscape context for evaluating observed conditions. The landscape model was developed with these needs in mind to better inform application of the CSCI for decision-making relative to landscape constraints on biological condition. Statewide application of the model demonstrated where streams are likely constrained on a regional basis, whereas application to the SGR watershed demonstrated how the model can be used by local stakeholders to prioritize management actions that are informed by landscape context. 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.

Results from our analysis could be used for managing the biological integrity of streams under state or federal water quality mandates (e.g. “biological criteria” under the Clean Water Act). Management activities for biological integrity could involve the protection of sites meeting biological objectives or the restoration of sites that do not meet biological objectives. The selection of appropriate management actions for streams requires the consideration of the physical and chemical condition of streams concurrent with biological monitoring results. The landscape model can place observed scores in an appropriate context relative to their modeled condition within the landscape. This information could provide flexibility in the selection of regulatory or management actions at specific sites or watershed scales (e.g., hydrologic subareas), and to further prioritize where and when actions should take place based on the temporal and spatial scale needed for protection or restoration actions. For example, for sites that meet biological objectives but where the models predict some degree of constraint (e.g., Figure S2, site types 5, 9, 10, or 13), regulatory actions may be associated with protecting that condition and could be implemented in the short-term to prevent degradation. 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). This flexibility is not intended to exclude sites from consideration that are less likely to achieve biological objectives, but rather to facilitate the decision-making process through a more transparent application of the model in a regulatory application. 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.

Non-regulatory applications of the landscape model are also possible by identifying where additional restoration, monitoring, or protection may have the most benefit. For example, landscape models could be used to support conservation planning, particularly at the watershed scale where land use practices can be a critical factor for decision-making. Ongoing work in California has focused on setting priorities for managing biodiversity that focus on watersheds within a conservation network (Howard et al. 2018). Results from the landscape model could be used to enhance this network by providing supporting information on constraints in an assessment framework. More generally, these applications could represent a novel use of bioassessment data beyond the pass/fail paradigm in the regulatory sense, for example, as tools for land use planning (Bailey et al. 2007). In many cases, including California, bioassessment indices have been sufficiently developed to allow large-scale condition assessment across regions, yet they are rarely used as planning tools to guide decisions on where resources should be focused (Nel et al. 2009). Our landscape model makes bioassessment data in California more accessible and identifies an appropriate expectation for the information, enabling the potential for both regulatory and non-regulatory applications.

## The landscape model is a tool for exploring options

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. This application was effectively demonstrated through engagement of our local stakeholder group. Rather than identifying individual sites in need of specific management actions, the group used the landscape model to characterize patterns on the landscape that were consistent with the recommended management priorities. In doing so, the group was able to explore and discuss potential management actions relative to the landscape characteristics of the watershed. The final decision by the group to prioritize management actions for the different sites in broad categories of protect, restore, and investigate was based on an iterative process where ideas were discussed and shared freely among stakeholders. This approach ensured that stakeholders were generally in agreement with the final product and, therefore, potentially more likely to adopt the recommendations provided by these tools in formal decision-making (Stein et al. 2017). The recommended actions have relevance only in the interests of the SGR Regional Monitoring Program. Localized applications of the statewide model must engage stakeholders in a similar process to develop recommendations that are specific to regional needs at the watershed scale (Brody 2003; Reed 2008).

The SCAPE application was also critical for demonstrating how results from the statewide model could be used at a regional scale. The application demonstrated core concepts of the model and allowed stakeholders to explore the key decision points that affect the model output, specifically related to changing certainties in the CSCI score predictions (e.g., $10^{th}$ and $90^{th}$ percentile predictions) and the ability to explore alternative thresholds for biological objectives (e.g., $10^{th}$ percentile of reference scores that defined constraint classes). This functionality allowed the stakeholders to develop recommendations that were completely independent of the model, i.e., decisions were not hard-wired into the model nor SCAPE. Because of this application, this stakeholder group has a better understanding of the potential impacts of biointegrity policies currently under review in California. Additionally, the SCAPE application provided assurance to the prioritization process by correctly identifying sites where discrepancies between CSCI scores and other measures of stream condition had been observed. Without this information (i.e., Figure 7a), stakeholders struggled to prioritize among sites, particularly for restoration activities. For example, some advocated that the lowest scoring sites should be prioritized, whereas others prioritized sites that scored just below the CSCI threshold. Conflicting priorities were common in the absence of information about the range of scores typical for these urban settings.

Several states have implemented alternative use designations for applying bioassessment criteria in modified channels (FDEP 2011; USEPA 2013; MBI 2016). Although our results generally support the link between impacted biology and channel modification, a regulatory framework based on direct channel modification may be insufficient because constraints are more accurately defined relative to landscape development. As defined for the model, a constrained channel may or may not be engineered, but an engineered channel will typically be constrained given the surrounding land use. For example, Tecolote Creek (San Diego County, USA) was identified by our model as a constrained channel in an urban landscape (Figure 9). The CSCI score is 0.61 indicating degraded biological integrity, but the channel is not modified (Rehn, Mazor, and Ode 2018). Other stressors originating at the landscape scale (e.g., water or sediment chemistry) have likely constrained the biological community at this site independent of the physical habitat quality. Furthermore, channel modification does not always result in biological degradation, particularly if the contributing watershed is largely undeveloped. For example, Stein et al. (2013) observed reference-like bioassessment index scores in armored reaches within national forest lands in southern California. A classification framework for biological constraints using only channel modification would provide incomplete and potentially misleading information on streams with limited biological potential. Ideally, context for evaluating biological condition from a landscape model, in conjunction with reach-specific data on channel modification, should be used to determine where aquatic life uses may be limited.

Our approach to assessing constrained streams is readily transferable outside of California. The landscape model could be applied to other bioassessment methods, such as a multi-metric index (the most common bioassessment approach within the US; Buss et al. (2014)), O/E assessments (Moss et al. 1987), biological condition gradients (Davies and Jackson 2006), or with other biological endpoints (e.g., fish or diatoms). More importantly, our use of national geospatial datasets (i.e., NHDPlus, McKay et al. (2012); StreamCat, Hill et al. (2016)) means that these methods could be applied across the United States. National bioassessment indices have been developed and the landscape model could be developed as a national-scale product of constraints on biological condition to complement recent work that predicted probable biological conditions with the National Rivers and Streams Assessment (Hill et al. 2017). Global geospatial datasets of freshwater-specific environmental variables are also available and could be used to develop similar models outside of the United States (Domisch, Amatulli, and Jetz 2015).

Extension of the landscape models beyond California should also consider landscape stressors that are predictive of biotic condition in other regions. For example, urban and agricultural gradients were sufficient to characterize constraints in many regions of California, whereas Hill et al. (2017) found that the volume of water stored by dams was an important predictor of biological condition in the Northern Appalachian and Northern Plains regions of the US. In their paper, Hill et al. (2017) provided an example of how predictive models could be used to identify potential sites for restoration or conservation, however, their illustration did not explicitly identify sites that were over- or under-scoring relative to a biological endpoint. Doing so in California provided stakeholders with relative information that helped establish management priorities, demonstrating the potential utility of this approach in other states.

## Model assumptions and limitations

There are several characteristics of the landscape model that could affect its performance when applied outside of urban and agricultural settings. First, the model was developed with a focus on the needs of managers that apply bioassessment tools in developed landscapes where conditions are presumably constrained. As such, landscape variables were chosen to capture the effects of development on CSCI scores in these areas (Table 1). Application of the model in regions where different stressors have strong impacts on stream condition should consider the relevance of urban and agricultural stressors and if an alternative model that better captures other stressor gradients is needed. For example, our results suggest that streams in the North Coast and Sierra Nevada regions are largely unconstrained, but the landscape model was a poor predictor of CSCI scores in these areas. The dominant stressors likely to affect stream condition in these regions originate from sources that are less common in developed landscapes, such as silviculture and cannabis cultivation. The current landscape model does not adequately capture these impacts outside of urban and agricultural environments. Moreover, poor model predictions are compounded by low sensitivity of the CSCI to relevant stressor gradients in these regions (Mazor et al. 2016). Accurate data for quantifying these potential stressors are not available in StreamCat, but this is an area where investments in improving spatial data could yield significant improvements in further development of bioassessment indices and tools for their interpretation.

An additional assumption is that the landscape model can adequately discriminate between intractable constraints on biology that are spatially and temporally pervasive relative to more manageable constraints. That is, we assumed that the impacts of stressors included in the model, such as urbanization, are not manageable in the short term, whereas stressors associated with deviations from model predictions can be mitigated. These assumptions are not unique to our model and have been used in other applications that have evaluated biological potential (Paul et al. 2008; Chessman 2014; Waite et al. 2014). However, many stressors excluded from the model can have long-lasting impacts, leading to potentially irreversible degradation or management scenarios where long-term recovery may only be possible with sustained and costly application of resources. For example, logging activities can impact benthic macroinvertebrate communities for a decade or more after harvesting activities have stopped (Stone and Wallace 1998; Quinn and Wright-Stow 2008). In urban areas, pervasive and profound alteration to groundwater and hydrology is common and stream communities in groundwater fed systems may require substantial time and resources for restoration. The potential legacy impacts of large-scale alterations of the natural environment are not well-captured by the current model, neither from a spatial nor temporal perspective. A more refined application of the landscape model would be necessary to evaluate different scales of impact, which could include developing separate models for each region, as well as more careful selection of model inputs to capture scales of interest for potential impacts on stream condition.

The landscape model is associative by design and does not identify mechanistic links between biological constraints and proximal causes. The model describes constraints at scales larger than instream characteristics as a necessary approach to accurately predict bioassessment scores. Additional analyses that evaluate how different predictors influence model performance at different quantiles could provide insight into how landscape factors relate to constraints (e.g., Koenker and Machado 1999). Further, a distinction between constraints on biological condition and channel modification is implicit such that indication of the former by the model does not explicitly indicate presence of the latter. As noted above, our results consistently indicated that engineered channels are biologically constrained, but the model is based on an a priori selection of land use variables to predict biotic integrity. A correspondence between habitat limitations and channel modification is likely in many cases but data are insufficient to evaluate biological effects statewide relative to land use constraints. Moreover, bioassessment scores can be similar in modified channels compared to natural streams independent of watershed land use, i.e., concordance between degraded stream condition and channel modification may not always be observed (Stein et al. 2013). More comprehensive assessments at individual sites may be needed to diagnose the immediate causes of degraded condition.

An additional consideration in using the landscape model is the meaning of biologically constrained relative to whole stream communities. Biologically constrained sites were considered those where present landscapes were likely to limit CSCI scores that describe macroinvertebrate condition. In many cases, poor biotic condition of the macroinvertebrate community translates to poor stream condition. However, a constrained macroinvertebrate community does not always mean other biological attributes of stream condition (e.g., fish assemblages) are also constrained. Urban streams sometimes support diverse algal assemblages such that algal-based measures of biotic condition may alternatively suggest good biotic condition relative to macroinvertebrate-based indices (Brown et al. 2009; Mazor, Beck, and Brown 2018). Broadening the landscape model to include multiple taxonomic assemblages or endpoints would allow a more complete assessment of how condition relates to landscape alteration.

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.

# Summary

The landscape model can be used to characterize the extent of biologically constrained channels in urban and agricultural landscapes. Our application to the SGR watershed demonstrated how the results of the model can be used at a spatial scale where many management decisions are implemented through close interaction with a regional stakeholder group with direct interests in the local resources. 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. The approach leverages information from multiple sources to develop a context for biological assessment that provides an expectation of what is likely to be achieved based on current land use development. This can facilitate more targeted management actions that vary depending on the landscape context and can also inform decisions on extent and effort for future monitoring locations.

# Supplement

Geospatial data of model results mapped to stream reaches in California is provided at Beck (2018a). The SCAPE model application website is available at http://shiny.sccwrp.org/scape/, full source code accessible at Beck (2018b). Additional figures and tables are available in the supplement.

## Author contributions

MB, RM, SJ, KW, JW, PO, RH, CL, MS, and ES performed the research and analyzed the data. MB, RM, SJ, JW, PO, RH, and CL wrote the paper. RM, SJ, KW, and PO provided data. All authors discussed the methods and results and contributed to the development of the manuscript.

## Acknowledgments

The authors acknowledge support from the San Gabriel River Regional Monitoring Program and the California State Water Resources Control Board. We thank Phil Markle and Lester Yuan for reviewing an earlier draft of the manuscript. The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency. Any mention of trade names, products, or services does not imply an endorsement by the U.S. Government or the U.S. Environmental Protection Agency. The EPA does not endorse any commercial products, services, or enterprises.

## References

Allan, D., D. Erickson, and J. Fay. 1997. The Influence of Catchment Land Use on Stream Integrity Across Multiple Spatial Scales. *Freshwater Biology* 37 (1): 149–61. https://doi.org/10.1046/j.1365-2427.1997.d01-546.x.

Bailey, R. C., T. B. Reynoldson, A. G. Yates, J. Bailey, and S. Linke. 2007. Integrating Stream Bioassessment and Landscape Ecology as a Tool for Land Use Planning. *Freshwater Biology* 52 (5): 908–17. https://doi.org/10.1111/j.1365-2427.2006.01685.x.

Beck, M. W. 2018a. Constrained streams for biological integrity in California. Knowledge Network for Biocomplexity. urn:uuid:75411f50-32ed-42a5-bbfd-26833c7a441f.

———. 2018b. SCCWRP/SCAPE: v1.0 (Version 1.0). Zenodo, http://doi.org/10.5281/zenodo.1218121.

Beechie, T., G. Pess, P. Roni, and G. Giannico. 2007. Setting River Restoration Priorities: A Review of Approaches and General Protocol for Identifying and Prioritizing Actions. *North American Journal of Fisheries Management* 28 (3): 891–905. https://doi.org/10.1577/M06-174.1.

Bernhardt, E. S., E. B. Sudduth, M. A. Palmer, J. D. Allan, J. L. Meyer, G. Alexander, J. Follastad-Shah, et al. 2007. Restoring Rivers One Reach at a Time: Results from a Survey of U.S. River Restoration Practitioners. *Restoration Ecology* 15 (3): 482–93. https://doi.org/10.1111/j.1526-100X.2007.00244.x.

Breiman, L. 2001. Random Forests. *Machine Learning* 45: 5–32.

Brody, S. D. 2003. Measuring the Effects of Stakeholder Participation on the Quality of Local Plans Based on the Principles of Collaborative Ecosystem Management. *Journal of Planning Education and Research* 22 (4): 407–19. https://doi.org/10.1177/0739456X03022004007.

Brown, L. R., T. F. Cuffney, J. F. Coles, F. Fitzpatrick, G. McMahon, J. Steuer, A. H. Bell, and J. T. May. 2009. Urban Streams Across the USA: Lessons Learned from Studies in Nine Metropolitan Areas. *Journal of the North American Benthological Society* 28 (4): 1051–69. https://doi.org/10.1899/08-153.1.

Brown, L. R., J. T. May, A. C. Rehn, P. R. Ode, I. R. Waite, and J. G. Kennen. 2012. Predicting Biological Condition in Southern California Streams. *Landscape and Urban Planning* 108 (1): 17–27. https://doi.org/10.1016/j.landurbplan.2012.07.009.

Buss, D. F., D. M. Carlisle, T. -S. Chon, J. Culp, J. s. Harding, H. E. Keizer-Vlek, W. A. Robinson, S. Strachan, C. Thirion, and R. M. Hughes. 2014. Stream Biomonitoring Using Macroinvertebrates Around the Globe: A Comparison of Large-Scale Programs. *Environmental Monitoring and Assessment* 187: 4132. https://doi.org/10.1007/s10661-014-4132-8.

Cade, B. S., and B. R. Noon. 2003. A Gentle Introduction to Quantile Regression for Ecologists. *Frontiers in Ecology and the Environment* 1 (8): 412–20.

Cao, Y., C. P. Hawkins, J. Olson, and M. A. Kosterman. 2007. Modeling Natural Environmental Gradients Improves the Accuracy and Precision of Diatom-Based Indicators. *Journal of the North American Benthological Society* 26 (3): 566–85. https://doi.org/10.1899/06-078.1.

Carline, R. F., and M. C. Walsh. 2007. Responses to Riparian Restoration in the Spring Creek Watershed, Central Pennsylvania. *Restoration Ecology* 15 (4): 731–42. https://doi.org/10.1111/j.1526-100X.2007.00285.x.

Carlisle, D. M., J. Falcone, and M. R. Meador. 2009. Predicting the Biological Condition of Streams: Use of Geospatial Indicators of Natural and Anthropogenic Characteristics of Watersheds. *Environmental Monitoring and Assessment* 151 (1-4): 143–60. https://doi.org/10.1007/s10661-008-0256-z.

Chen, K., R. M. Hughes, S. Xu, J. Zhang, D. Cai, and B. Wang. 2014. Evaluating Performance of Macroinvertebrate-Based Adjusted and Unadjusted Multi-Metric Indices (MMI) Using Multi-

Season and Multi-Year Samples. *Ecological Indicators* 36: 142–51. https://doi.org/10.1016/j.ecolind.2013.07.006.

Chessman, B. C. 2014. Predicting Reference Assemblages for Freshwater Bioassessment with Limiting Environmental Difference Analysis. *Freshwater Science* 33 (4): 1261–71. https://doi.org/10.1086/678701.

Chessman, B. C., M. Muschal, and M. J. Royal. 2008. Comparing Apples with Apples: Use of Limiting Environmental Differences to Match Reference and Stressor-Exposure Sites for Bioassessment of Streams. *River Research and Applications* 24 (1): 103–17. https://doi.org/10.1002/rra.1053.

Chessman, B. C., and M. J. Royal. 2004. Bioassessment Without Reference Sites: Use of Environmental Filters to Predict Natural Assemblages of River Macroinvertebrates. *Journal of the North American Benthological Society* 23 (3): 599–615. https://doi.org/10.1899/0887-3593(2004)023%3C0599:BWRSUO%3E2.0.CO;2.

Clapcott, J. E., K. J. Collier, R. G. Death, E. O. Goodwin, J. S. Harding, D. Kelly, J. R. Leathwick, and R. G. Young. 2011. Quantifying Relationships Between Land-Use Gradients and Structural and Functional Indicators of Stream Ecological Integrity. *Freshwater Biology* 57 (1): 74–90. https://doi.org/10.1111/j.1365-2427.2011.02696.x.

Cuffney, T. F., S. S. Qian, R. A. Brightbill, J. T. May, and I. R. Waite. 2011. Response to King and Baker: Limitations on Threshold Detection and Characterization of Community Thresholds. *Ecological Applications* 21 (7): 2840–5. https://doi.org/10.2307/41416699.

Davies, S. P., and S. K. Jackson. 2006. The Biological Condition Gradient: A Descriptive Model for Interpreting Change in Aquatic Ecosystems. *Ecological Applications* 16 (4): 1251–66.

Domisch, S., G. Amatulli, and W. Jetz. 2015. Near-Global Freshwater-Specific Environmental Variables for Biodiversity Analyses in 1 Km Resolution. *Scientific Data* 2: 150073. https://doi.org/10.1038/sdata.2015.73.

ESRI (Environmental Systems Research Institute). 2016. ArcGIS v10.5. Redlands, California.

FDEP (Florida Department of Environmental Protection). 2011. Development of aquatic life use support attainment thresholds for Florida’s Stream Condition Index and Lake Vegetation Index. DEP-SAS-003/11. Tallahassee, Florida: FDEP Standards; Assessment Section, Bureau of Assessment; Restoration Support.

Finkenbine, J. K., J. W. Atwater, and D. S. Mavinic. 2000. Stream Health After Urbanization. *Journal of the American Water Resources Association* 36 (5): 1149–60. https://doi.org/10.1111/j.1752-1688.2000.tb05717.x.

Fox, E. W., R. A. Hill, S. G. Leibowitz, A. R. Olsen, D. J. Thornbrugh, and M. H. Weber. 2017. Assessing the Accuracy and Stability of Variable Selection Methods for Random Forest Modeling in Ecology. *Environmental Monitoring and Assessment* 189: 316. https://doi.org/10.1007/s10661-017-6025-0.

Fry, J., G. Xian, S. Jin, J. Dewitz, C. Homer, L. Yang, C. Barnes, N. Herold, and J. Wickham. 2011. Completion of the 2006 National Land Cover Database for the Conterminous United States. *Photogrammetric Engineering and Remote Sensing* 77 (9): 858–64.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer.

Hill, R. A., E. W. Fox, S. G. Leibowitz, A. R. Olsen, D. J. Thornbrugh, and M. H. Weber. 2017. Predictive Mapping of the Biotic Condition of Conterminous U.S. Rivers and Streams. *Ecological Applications* 27 (8): 2397–2415. https://doi.org/10.1002/eap.1617.

Hill, R. A., M. H. Weber, S. G. Leibowitz, A. R. Olsen, and D. J. Thornbrugh. 2016. The Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the Conterminous United States. *Journal of the American Water Resources Association* 52: 120–28. https://doi.org/10.1111/1752-1688.12372.

Howard, J. K., K. A. Fesenmyer, T. E Grantham, J. H. Viers, P. R. Ode, P. B. Moyle, S. J. Kupferburg, et al. 2018. A Freshwater Conservation Blueprint for California: Prioritizing Watersheds for Freshwater Biodiversity. *Freshwater Science* 37 (2): 417–31. https://doi.org/10.1086/697996.

Hynes, H. B. N. 1975. The Stream and its Valley. *SIL Proceedings, 1922-2010* 19 (1): 1–15. https://doi.org/10.1080/03680770.1974.11896033.

Johnson, L., C. Richards, G. Host, and J. Arthur. 1997. Landscape Influences on Water Chemistry in Midwestern Stream Ecosystems. *Freshwater Biology* 37 (1): 193–208. https://doi.org/10.1046/j.1365-2427.1997.d01-539.x.

Karr, J. R., K. D. Fausch, P. L. Angermeier, P. R. Yant, and I. J. Schlosser. 1986. Assessing Biological Integrity in Running Waters: A Method and its Rationale. Special Publication 5. Champaign, Illinois: Illinois Natural History Survey.

Kenney, M. A., P. R. Wilcock, B. F. Hobbs, N. E. Flores, and D. C. Martínez. 2012. Is Urban Stream Restoration Worth It? *Journal of the American Water Resources Association* 48 (3): 603–15. https://doi.org/10.1111/j.1752-1688.2011.00635.x.

Koenker, R., and J. A. F. Machado. 1999. Goodness of Fit and Related Inference Processes for Quantile Regression. *Journal of the American Statistical Association* 94 (448): 1296–1310.

Lester, R. E., and A. J. Boulton. 2008. Rehabilitating Agricultural Streams in Australia with Wood: A Review. *Environmental Management* 42 (2): 310–26. https://doi.org/10.1007%2Fs00267-008-9151-1.

Loflen, C., H. Hettesheimer, L. B. Busse, K. Watanabe, R. M. Gersberg, and V. Lüderitz. 2016. Inadequate Monitoring and Inappropriate Project Goals: A Case Study on the Determination of Success for the Forester Creek Improvement Project. *Ecological Restoration* 34 (2): 124–34. https://doi.org/10.3368/er.34.2.124.

Mazor, R. D. 2015. Bioassessment of Perennial Streams in Southern California: A Report on the First Five Years of the Stormwater Monitoring Coalition's Regional Stream Survey. 844. Costa Mesa, California: Southern California Coastal Water Research Project.

Mazor, R. D., M. W. Beck, and J. Brown. 2018. 2017 Report on the SMC Regional Stream Survey. 1029. Costa Mesa, California: Southern California Coastal Water Research Project.

Mazor, R. D., A. C. Rehn, P. R. Ode, M. Engeln, K. C. Schiff, E. D. Stein, D. J. Gillett, D. B. Herbst, and C. P. Hawkins. 2016. Bioassessment in Complex Environments: Designing an Index for Consistent Meaning in Different Settings. *Freshwater Science* 35 (1): 249–71.

Mazor, R. D., E. D. Stein, P. R. Ode, and K. Schiff. 2014. Integrating Intermittent Streams into Watershed Assessments: Applicability of an Index of Biotic Integrity. *Freshwater Science* 35 (2): 459–74. https://doi.org/10.1086/675683.

MBI (Midwest Biodiversity Institute). 2016. Identification of predictive habitat attributes for Minnesota streams to support tiered aquatic life uses. MBI Technical Report MBI/OHPAN1518840. Columbus, Ohio: Midwest Biodiversity Institute, prepared on behalf of the Minnesota Pollution Control Agency.

McKay, L., T. Bondelid, T. Dewald, J. Johnston, R. Moore, and A. Reah. 2012. NHDPlus Version 2: User Guide.

Meinshausen, N. 2006. Quantile Regression Forests. *Journal of Machine Learning Research* 7: 983–99.

Meinshausen, N. 2017. *QuantregForest: Quantile Regression Forests*. https://CRAN.R-project.org/package=quantregForest.

Morgan, R. P., and S. E. Cushman. 2005. Urbanization Effects on Stream Fish Assemblages in Maryland, USA. *Journal of the North American Benthological Society* 24 (3): 643–55.

Moss, D., M. T. Furse, J. F. Wright, and P. D. Armitage. 1987. The Prediction of the Macro-Invertebrate Fauna of Unpolluted Running-Water Sites in Great Britain Using Environmental Data. *Freshwater Biology* 17 (1): 41–52. https://doi.org/10.1111/j.1365-2427.1987.tb01027.x.

Nel, J. L., D. J. Roux, R. Abell, P. J. Ashton, R. M. Cowling, J. V. Higgins, M. Thieme, and J. H. Viers. 2009. Progress and Challenges in Freshwater Conversation Planning. *Aquatic*

*Conservation: Marine and Freshwater Ecosystems* 19 (4): 474–85. https://doi.org/10.1002/aqc.1010.

Ode, P. R., A. E. Fetscher, and L. B. Busse. 2016. Standard Operating Procedures (SOP) for the Collection of Field Data for Bioassessments of California Wadeable Streams: Benthic Macroinvertebrates, Algae, and Physical Habitat. SWAMP-SOP-SB-2016-0001. Sacramento, California: California State Water Resources Control Board Surface Water Ambient Monitoring Program. https://www.waterboards.ca.gov/water_issues/programs/swamp/bioassessment/docs/combined_sop_2016.pdf.

Ode, P. R., T. M. Kincaid, T. Fleming, and A. C. Rehn. 2011. Ecological Condition Assessments of California's Perennial Wadeable Streams: Highlights from the Surface Water Ambient Monitoring Program's Perennial Streams Assessment (PSA). Sacramento, California: A collaboration between the State Water Resources Control Board's Non-Point Source Pollution Control Program (NPS Program), Surface Water Ambient Monitoring Program (SWAMP), California Department of Fish; Game Aquatic Bioassessment Laboratory, the U.S. Environmental Protection Agency. https://www.waterboards.ca.gov/water_issues/programs/swamp/docs/reports/psa_smmry_rpt.pdf.

Ode, P. R., A. C. Rehn, R. D. Mazor, K. C. Schiff, E. D. Stein, J. T. May, L. R. Brown, et al. 2016. Evaluating the Adequacy of a Reference-Site Pool for Ecological Assessments in Environmentally Complex Regions. *Freshwater Science* 35 (1): 237–48.

Paul, M. J., D. W. Bressler, A. H. Purcell, M. T. Barbour, E. T. Rankin, and V. H. Resh. 2008. Assessment Tools for Urban Catchments: Defining Observable Biological Potential. *Journal of the American Water Resources Association* 45 (2): 320–30. https://doi.org/10.1111/j.1752-1688.2008.00280.x.

Quinn, J. M., and A. E. Wright-Stow. 2008. Stream Size Influences Stream Temperature Impacts and Recovery Rates After Clearfell Logging. *Forest Ecology and Management* 256 (12): 2101–9. https://doi.org/10.1016/j.foreco.2008.07.041.

RDCT (R Development Core Team). 2018. R: A language and environment for statistical computing, v3.5.1. R Foundation for Statistical Computing, Vienna, Austria.

Reed, M. S. 2008. Stakeholder Participation for Environmental Management: A Literature Review. *Biological Conservation* 141 (10): 2417–31. https://doi.org/10.1016/j.biocon.2008.07.014.

Rehn, A. C. 2015. The Perennial Streams Assessment (PSA): An Assessment of Biological Condition Using the New California Stream Condition Index (CSCI). SWAMP Management Memorandum, SWAMP-MM-2015-0001. Sacramento, California: California State Water Resources Control Board Surface Water Ambient Monitoring Program. https://www.waterboards.ca.gov/water_issues/programs/swamp/bioassessment/docs/psa_memo_121015.pdf.

Rehn, A. C., R. D. Mazor, and P. R. Ode. 2018. An Index to Measure the Quality of Physical Habitat in California Wadeable Streams. SWAMP Technical Memorandum, SWAMP-TM-2018-0005. Sacramento, California: California Water Boards, Surface Water Ambient Monitoring Program, California Department of Fish; Wildlife, Southern California Coastal Water Research

Project. https://www.waterboards.ca.gov/water_issues/programs/swamp/bioassessment/docs/physical_habitat_index_technical_memo.pdf.

Reynoldson, T. B., R. H. Norris, V. H. Resh, K. E. Day, and D. M. Rosenberg. 1997. The Reference Condition: A Comparison of Multimetric and Multivariate Approaches to Assess Water-Quality Impairment Using Benthic Macroinvertebrates. *Journal of the North American Benthological Society* 16 (4): 833–52. https://doi.org/10.2307/1468175.

Richards, C., R. Haro, L. Johnson, and G. Host. 1997. Catchment and Reach-Scale Properties as Indicators of Macroinvertebrate Species Traits. *Freshwater Biology* 37 (1): 219–30. https://doi.org/10.1046/j.1365-2427.1997.d01-540.x.

Roni, P., and T. Beechi. 2012. *Stream and Watershed Restoration Guide: A Guide to Restoring Riverine Processes and Habitats*. First. Hoboken, New Jersey: John Wiley & Sons.

SDRWQB (San Diego Regional Water Quality Control Board). 2016. Clean Water Act Sections 305(B) and 303(D) Integrated Report for the San Diego Region. Sacramento, California: California Environmental Protection Agency. https://www.waterboards.ca.gov/sandiego/water_issues/programs/303d_list/docs/Staff_Report_101216.pdf.

Shoredits, A. S., and J. A. Clayton. 2013. Assessing the Practice and Challenges of Stream Restoration in Urbanized Environments of the USA. *Geography Compass* 7 (5): 358–72. https://doi.org/10.1111/gec3.12039.

Sleeter, B. M., T. S. Wilson, C. E. Soulard, and J. Liu. 2011. Estimation of the Late Twentieth Century Land-Cover Change in California. *Environmental Monitoring and Assessment* 173 (1-4): 251–66. https://doi.org/10.1007/s10661-010-1385-8.

Smeets, E., and R. Weterings. 1999. Environmental Indicators: Typology and Overview. No. 25. Copenhagen, Denmark: European Environmental Agency.

Stein, E. D., M. R. Cover, A. E. Fetscher, C. O'Reilly, R. Guardado, and C. W. Solek. 2013. Reach-Scale Geomorphic and Biological Effects of Localized Streambank Armoring. *Journal of the American Water Resources Association* 49 (4): 780–92. https://doi.org/10.1111/jawr.12035.

Stein, E. D., A. Sengupta, R. D. Mazor, K. McCune, B. P. Bledsoe, and S. Adams. 2017. Application of Regional Flow-Ecology Relationships to Inform Watershed Management Decisions: Applications of the ELOHA Framework in the San Diego River Watershed, California, USA. *Ecohydrology* 10 (7): e1869. https://doi.org/10.1002/eco.1869.

Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson, and R. H. Norris. 2006. Setting Expectations for the Ecological Condition of Streams: The Concept of Reference Condition. *Ecological Applications* 16 (4): 1267–76. https://doi.org/10.1890/1051-0761(2006)016[1267:SEFTEC]2.0.CO;2.

Stone, M. K., and J. B. Wallace. 1998. Long-Term Recovery of a Mountain Stream from Clear-Cut Logging: The Effects of Forest Succession on Benthic Invertebrate Community Structure. *Freshwater Biology* 39 (1): 151–69. https://doi.org/10.1046/j.1365-2427.1998.00272.x.

USEPA (US Environmental Protection Agency, Region 10). 2013. Technical Support Document for EPA's Action on the State of Oregon's Revised Water Quality Standards for the West Division Main Canal. USEPA Region 10, Seattle, Washington.

USGS (US Geological Survey). 1999. The quality of our nation's waters: nutrients and pesticides. Reston, Virginia.

Vølstad, J. H., N. E. Roth, G. Mercurio, M. T. Southerland, and D. E. Strebel. 2004. Using Environmental Stressor Information to Predict the Ecological Status of Maryland Non-Tidal Streams as Measured by Biological Indicators. *Environmental Monitoring and Assessment* 84 (3): 219–42. https://doi.org/10.1023/A:1023374524254.

Waite, I. R. 2013. Development and Application of an Agricultural Intensity Index to Invertebrate and Algal Metrics from Streams at Two Scales. *Journal of the American Water Resources Association* 49 (2): 431–48. https://doi.org/10.1111/jawr.12032.

Waite, I. R., J. G. Kennen, J. T. May, L. R. Brown, T. F. Cuffney, K. A. Jones, and J. L. Orlando. 2014. Stream Macroinvertebrate Response Models for Bioassessment Metrics: Addressing the Issue of Spatial Scale. *PLOS ONE* 9 (3): e90944. https://doi.org/10.1371/journal.pone.0090944.

Walsh, C. J., A. H. Roy, J. w. Feminella, P. D. Cottingham, P. M. Groffman, and R. P. Morgan. 2005. The Urban Stream Syndrome: Current Knowledge and the Search for a Cure. *Journal of the North American Benthological Society* 24 (3): 706–23. https://doi.org/10.1899/04-028.1.

Wang, L., J. Lyons, P. Kanehl, and R. Gatti. 1997. Influences of Watershed Land Use on Habitat Quality and Biotic Integrity in Wisconsin Streams. *Fisheries* 22 (6): 6–12. https://doi.org/10.1577/1548-8446(1997)022%3C0006:IOWLUO%3E2.0.CO;2.

## Figure captions

*Figure 1 Urban and agricultural land use (a) and distribution of observed stream CSCI scores (b) in California. Cover of urban and agricultural land use in stream watersheds was used to develop a landscape model for stream segment expectations of bioassessment scores. Breakpoints for CSCI scores are the $1^{st}$, $10^{th}$, and $30^{th}$ percentile of scores at least-disturbed, reference sites throughout the state. Altered and intact refers to biological condition (Mazor et al. 2016). Grey lines are major environmental regions in California defined by ecoregional and watershed boundaries, CV: Central Valley, CH: Chaparral, DM: Deserts and Modoc Plateau, NC: North Coast, SN: Sierra Nevada, SC: South Coast.*

*Figure 2 Conceptualized response and management pathways captured by the landscape model under the Driver-Pressure-Stress-Impact-Response (DPSIR) framework (Smeets and Weterings 1999). Landscape predictors provided in StreamCat (Hill et al. 2017) were used to describe pressures from urban and agricultural development that could impact the macroinvertebrate community in streams by altering physical and chemical habitat. Biological response was measured using the CSCI (Mazor et al. 2016) as an impact indicator and then evaluated relative to ranges of CSCI scores that were expected at each site provided by the landscape model. Observed CSCI scores and context from the landscape model provide a basis for informing management actions that could address environmental impacts at different points in the response pathway, where the management pathway could address causes at different scales and efficiencies.*

*Figure 3 Application of the landscape model to identify site expectations and bioassessment performance for sixteen example stream segments. A range of CSCI scores is predicted from the model (a) and the lower and upper limits of the expectations are cut to define a certainty range* *for the predictions (b). Overlap of the certainty range at each segment with a chosen CSCI threshold (c) defines the stream segment classification as likely unconstrained, possibly unconstrained, possibly constrained, and likely constrained. The observed bioassessment scores are described relative to the classification as over scoring (above the certainty threshold), expected (within), and under scoring (below) for each of four stream classes (d).*

*Figure 4 San Gabriel River watershed in southern California. Land cover is shown in plot (a) and the predicted median CSCI scores at each stream segment and observed CSCI scores are shown in (b).*

*Figure 5 Statewide application of the landscape model showing the stream segment classifications. Major regional boundaries are also shown (see Figure 1).*

*Figure 6 Changes in stream segment classes by region and statewide for different scenarios used to define biological constraints. Twenty-seven scenarios were tested that evaluated different combinations of certainty in the CSCI predictions (nine scenarios from wide to narrow prediction intervals as identified by the tail cutoff for the expected range) and potential CSCI thresholds (three scenarios from low to high). The percentage of total stream length for likely unconstrained and likely constrained is shown for each scenario. Stream classifications as possibly unconstrained or possibly constrained are not shown but can be inferred form the area of white space above or below each bar. The solid black line indicates the percentage division between unconstrained and constrained classifications. CV: Central Valley, CH: Chaparral, DM: Deserts and Modoc Plateau, NC: North Coast, SN: Sierra Nevada, SC: South Coast.*

*Figure 7 Application of the landscape model to stream segments in the San Gabriel River watershed, Los Angeles County, California. CSCI scores with (a) no biological context from the model are on the left and (b) scores with context from the model are on the right. Relative site* *scores as under-scoring, expected, or over-scoring are based on observed scores given the segment class as likely constrained, possibly constrained, possibly unconstrained, and likely unconstrained. Segment classes are based on overlap of the expectations with a biological threshold for the CSCI (0.79, dashed lined) and location of the median expectation (white ticks).*

*Figure 8 Relative site scores and recommended management actions for locations with CSCI scores in the San Gabriel River watershed. Relative site scores as under scoring, expected, or over scoring are based on observed scores given the segment class as likely constrained, possibly constrained, possibly unconstrained, and likely unconstrained. Recommended management actions were defined by a local stakeholder group (see Figure S2, Table 3) and are ranked by priority for actions to investigate, protect, and restore a site. No recommended actions assume baseline maintenance and monitoring is sufficient.*

*Figure 9 Tecolote Creek (San Diego County, USA) is a constrained channel in an urban landscape (a, Source: 32.81736, -117.19986. Google Earth. November 8, 2016. Accessed July 20, 2018.). Physical habitat (b, Source: R. Mazor) at the sample site suggests no channel alteration. The CSCI was scored at 0.61 indicating degraded biological integrity.*

# Tables

*Table 1 Land use variables used to develop the landscape model of stream bioassessment scores. All variables were obtained from StreamCat (Hill et al. 2016) and applied to stream segments in the National Hydrography Dataset Plus (NHD-plus) (McKay et al. 2012). The measurement scales for each variable are at the riparian (100 m buffer), catchment, and/or watershed, scale relative to a stream segment. Combined scales for riparian measurements (e.g., riparian + catchment, riparian + watershed) are riparian estimates for the entire catchment or watershed area upstream, as compared to only the individual segment. Total urban and agriculture land use variables were based on sums of individual variables in StreamCat as noted in the description. Rp100: riparian, Cat: catchment, Ws: watershed*

| Name | Scale | Description | Unit |
|---|---|---|---|
| CanalDens | Cat, Ws | Density of NHDPlus line features classified as canal, ditch, or pipeline | km/sq km |
| PctImp2006 | Cat, Ws, Cat + Rp100, Ws + Rp100 | Mean imperviousness of anthropogenic surfaces (NLCD 2006) | % |
| TotUrb2011 | Cat, Ws, Cat + Rp100, Ws + Rp100 | Total urban land use as sum of developed open, low, medium, and high intensity (NLCD 2011) | % |
| TotAg2011 | Cat, Ws, Cat + Rp100, Ws + Rp100 | Total agricultural land use as sum of hay and crops (NLCD 2011) | % |
| RdDens | Cat, Ws, Cat + Rp100, Ws + Rp100 | Density of roads (2010 Census Tiger Lines) | km/sq km |
| RdCrs | Cat, Ws | Density of roads-stream intersections (2010 Census Tiger Lines-NHD stream lines) | crossings/sq km |

*Table 2 Stream class definitions describing potential biological constraints. Classes are based on the overlap of the range of likely bioassessment scores with a potential threshold for a biological objective. Identifying stream classes requires selecting the cutoff range of likely scores from the landscape model and a chosen threshold for the objective.*

| Class | Definition | Example |
|---|---|---|
| Likely unconstrained | Lower bound of prediction interval is above threshold | $10^{th}$ percentile > 0.79 |
| Possibly unconstrained | Lower bound of prediction interval is below threshold, but median prediction is above | $50^{th}$ percentile > 0.79 |
| Possibly constrained | Upper bound of prediction interval is above threshold, but median prediction is below | $50^{th}$ percentile < 0.79 |
| Likely constrained | Upper bound of prediction interval is below threshold | $90^{th}$ percentile < 0.79 |

*Table 3 Recommended management actions defined by a local stakeholder group for application of results from the landscape model to prioritize stream reaches. Actions were assigned to stream types based on observed CSCI scores relative to the stream expectation from the landscape model (see Figure S2). Actions were recommended in addition to baseline monitoring and maintenance that occurred at all sites.*

| Action | Example activity | Example high priority site | Example low priority site |
|---|---|---|---|
| Investigate | Higher frequency of sampling, evaluate additional data (e.g., habitat) | Sites scoring outside prediction interval | Sites scoring as expected |
| Protect | Extra scrutiny of proposed impacts | Unconstrained sites | Constrained sites |
| Restore | Make funding recommendations, prioritize TMDL development | Low-scoring unconstrained sites | Low-scoring constrained sites |

*Table 4 Performance of the landscape model by calibration (Cal) and validation (Val) datasets in predicting CSCI scores. The statewide dataset (Figure 5) and individual regions of California (Figure 1) are evaluated. Averages and standard deviations (in parentheses) for observed and predicted CSCI values of each dataset are shown. Pearson correlations (r), root mean squared errors (RMSE), intercept, and slopes are for comparisons of predicted and observed values to evaluate model performance. All correlations, intercepts, and slopes are significant at alpha = 0.05. CV: Central Valley, CH: Chaparral, DM: Deserts and Modoc Plateau, NC: North Coast, SN: Sierra Nevada, SC: South Coast.*

| Dataset | Location | n | Observed | Predicted | r | RMSE | Intercept | Slope |
|---|---|---|---|---|---|---|---|---|
| Cal | Statewide | 1965 | 0.82 (0.26) | 0.83 (0.20) | 0.75 | 0.17 | 0.04 | 0.93 |
| | CH | 512 | 0.76 (0.27) | 0.79 (0.21) | 0.71 | 0.19 | 0.03 | 0.92 |
| | CV | 116 | 0.51 (0.18) | 0.57 (0.15) | 0.66 | 0.15 | 0.05 | 0.81 |
| | DM | 86 | 0.87 (0.22) | 0.91 (0.14) | 0.50 | 0.20 | 0.15 | 0.79 |
| | NC | 208 | 0.92 (0.20) | 0.94 (0.13) | 0.55 | 0.17 | 0.12 | 0.86 |
| | SC | 631 | 0.79 (0.24) | 0.78 (0.21) | 0.75 | 0.16 | 0.11 | 0.87 |
| | SN | 412 | 0.98 (0.18) | 0.98 (0.09) | 0.45 | 0.16 | 0.12 | 0.88 |
| Val | Statewide | 655 | 0.82 (0.25) | 0.84 (0.20) | 0.72 | 0.18 | 0.07 | 0.90 |
| | CH | 172 | 0.76 (0.27) | 0.81 (0.21) | 0.74 | 0.19 | -0.04 | 0.98 |
| | CV | 40 | 0.52 (0.19) | 0.59 (0.16) | 0.49 | 0.19 | 0.16 | 0.60 |
| | DM | 28 | 0.84 (0.17) | 0.93 (0.11) | 0.55 | 0.17 | 0.07 | 0.83 |
| | NC | 71 | 0.94 (0.19) | 0.96 (0.11) | 0.55 | 0.16 | 0.00 | 0.98 |
| | SC | 208 | 0.80 (0.24) | 0.78 (0.21) | 0.72 | 0.17 | 0.17 | 0.81 |
| | SN | 136 | 0.97 (0.17) | 0.98 (0.09) | 0.21 | 0.17 | 0.57 | 0.41 |

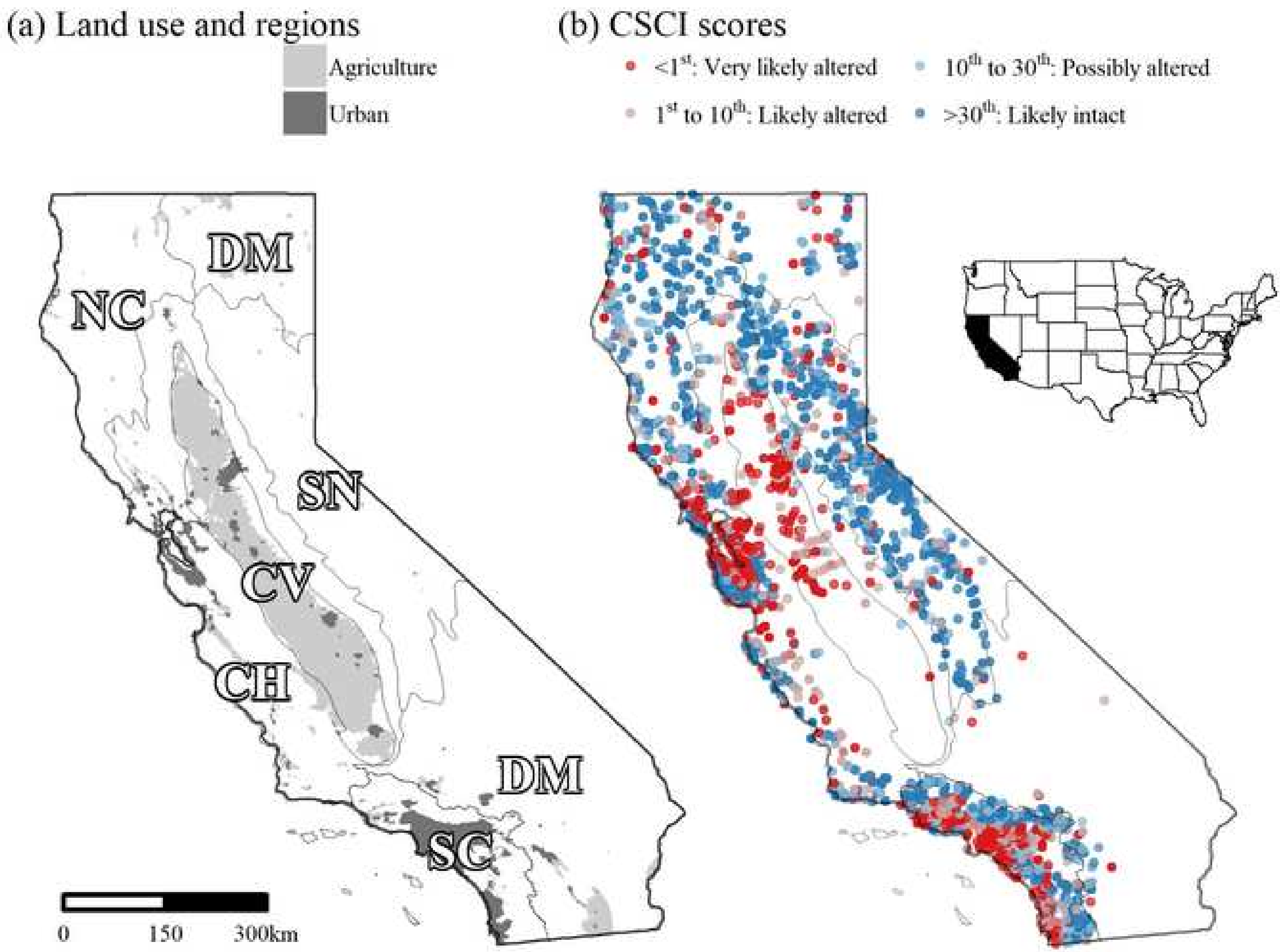
*Table 5: Summary of stream length for each stream class statewide and major regions of California (Figures 1, 5). Lengths are in kilometers with the percentage of the total length in a region in parentheses. All lengths are based on a CSCI threshold of 0.79 and the 10$^{th}$ to 90$^{th}$ percentile of expected scores from the landscape model. CV: Central Valley, CH: Chaparral, DM: Deserts and Modoc Plateau, NC: North Coast, SN: Sierra Nevada, SC: South Coast.*

| | constrained | | unconstrained | |
|---|---|---|---|---|
| Region | likely | possibly | possibly | likely |
| Statewide | 8150 (4) | 24735 (11) | 101591 (46) | 85317 (39) |
| CV | 3356 (22) | 8010 (52) | 3202 (21) | 951 (6) |
| CH | 1642 (3) | 7840 (13) | 30693 (50) | 21206 (35) |
| DM | 255 (0) | 3395 (6) | 27194 (47) | 26479 (46) |
| NC | 108 (0) | 1442 (5) | 14152 (49) | 13286 (46) |
| SN | 20 (0) | 1067 (3) | 18228 (48) | 19032 (50) |
| SC | 2770 (15) | 2981 (16) | 8122 (45) | 4363 (24) |

*Table 6: Summary of CSCI scores by relative expectations for each stream class statewide and in each major region of California (Figures 1, 5). Average CSCI scores (standard deviation) and counts (percent) of the number of monitoring stations in each relative score category and region are shown. Sites are over-scoring if the observed scores are above the range of expectations at a segment, expected if within the range, or under-scoring if below the range. CV: Central Valley, CH: Chaparral, DM: Deserts and Modoc Plateau, NC: North Coast, SN: Sierra Nevada, SC: South Coast.*

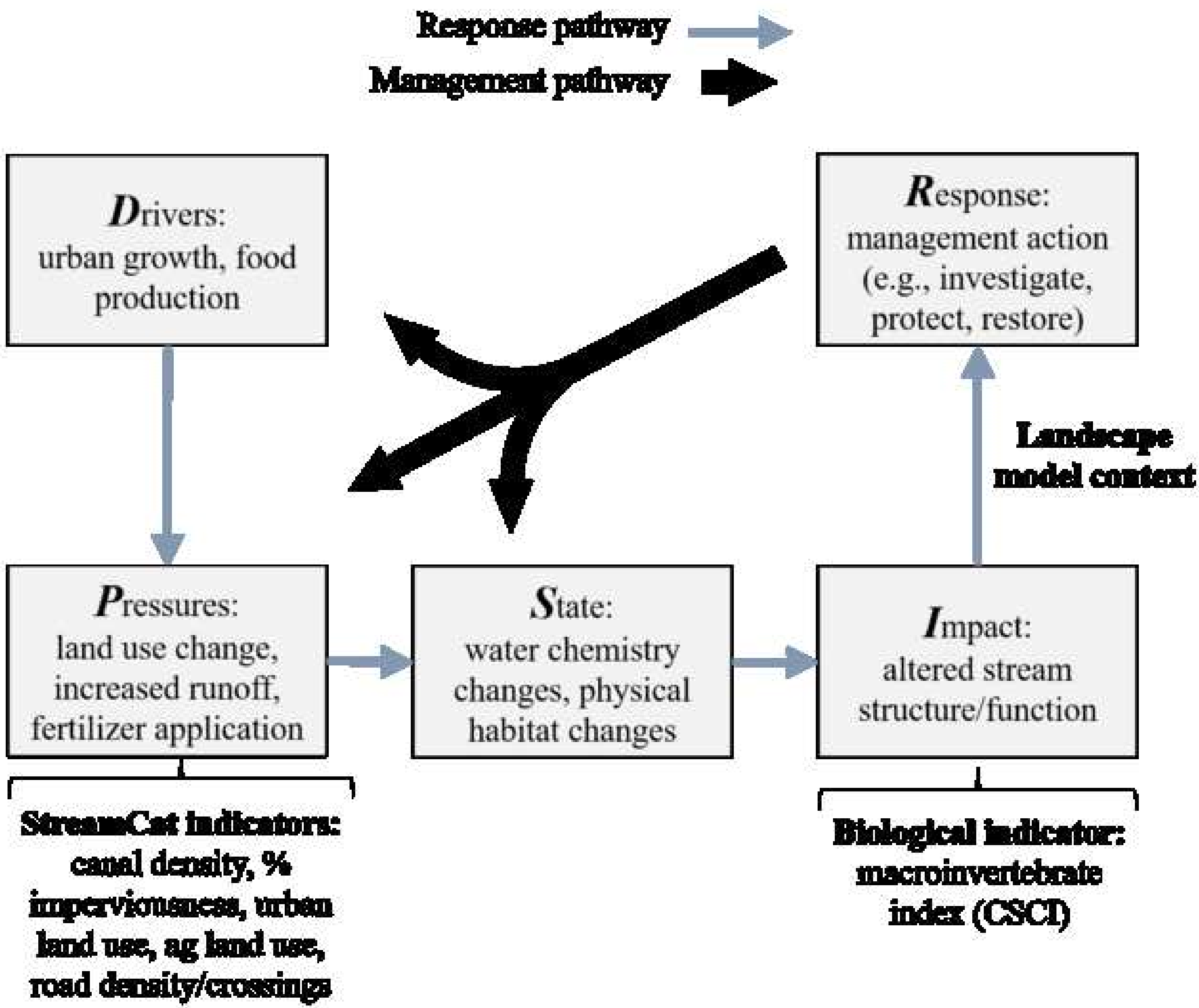
| | under-scoring | | expected | | over-scoring | |
|---|---|---|---|---|---|---|
| Region | CSCI | n (%) | CSCI | n (%) | CSCI | n (%) |
| Statewide | 0.54 (0.21) | 267 (10) | 0.83 (0.23) | 2041 (80) | 1.08 (0.17) | 242 (9) |
| CH | 0.47 (0.18) | 89 (13) | 0.79 (0.24) | 535 (80) | 1.08 (0.17) | 45 (7) |
| CV | 0.34 (0.12) | 25 (17) | 0.54 (0.17) | 118 (81) | 0.63 (0.25) | 2 (1) |
| DM | 0.6 (0.17) | 15 (14) | 0.9 (0.17) | 89 (80) | 1.15 (0.08) | 7 (6) |
| NC | 0.66 (0.17) | 28 (10) | 0.93 (0.16) | 228 (82) | 1.15 (0.08) | 22 (8) |
| SC | 0.54 (0.22) | 56 (7) | 0.78 (0.22) | 656 (81) | 1.02 (0.2) | 97 (12) |
| SN | 0.67 (0.16) | 54 (10) | 0.99 (0.11) | 415 (77) | 1.16 (0.06) | 69 (13) |

Figure1



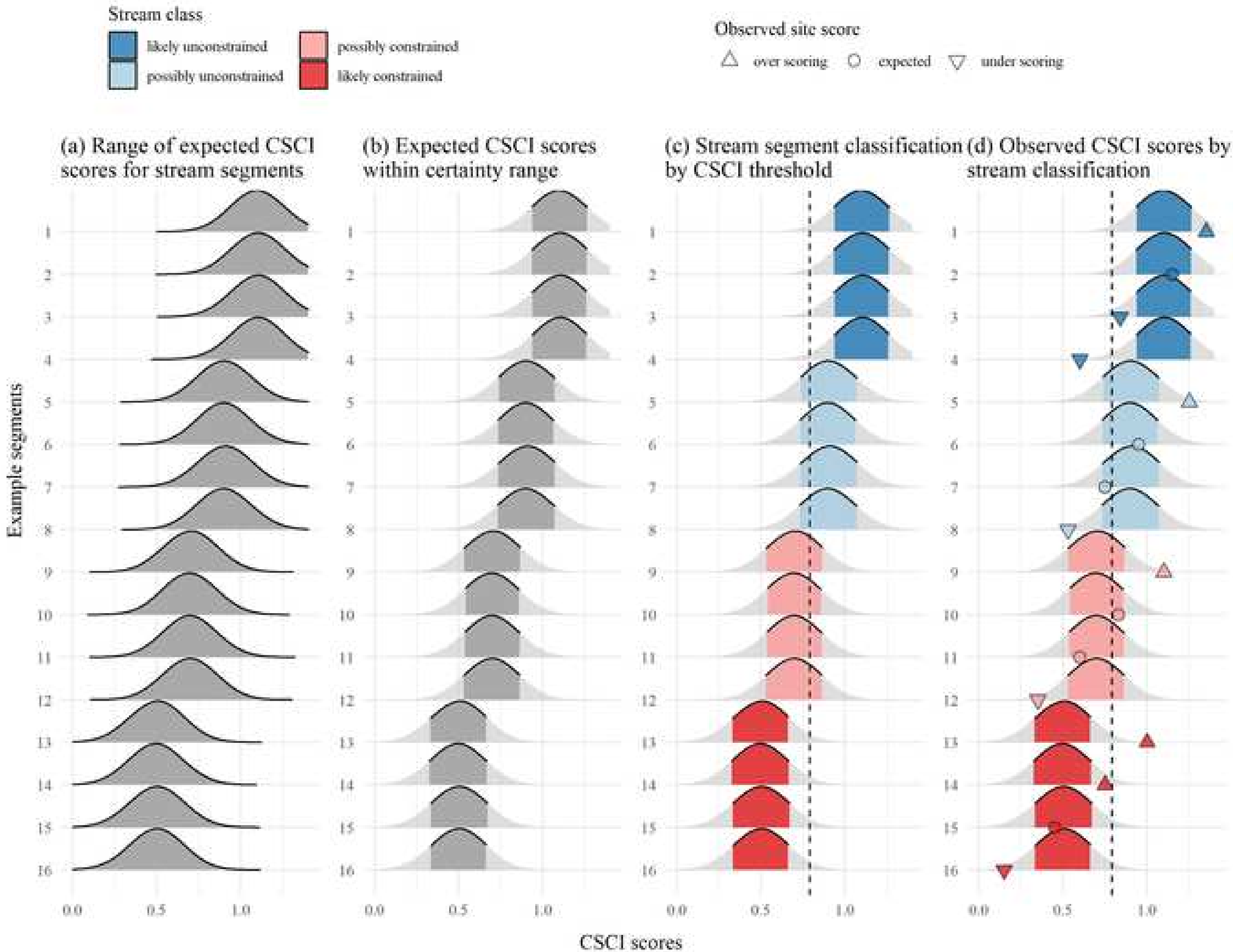
(a) Land use and regions
Agriculture
Urban
DM
NC
SN
CV
CH
DM
SC
0
150
300km
(b) CSCI scores
<1st: Very likely altered
1st to 10th: Likely altered
10th to 30th: Possibly altered
>30th: Likely intact

Figure2



Response pathway
Management pathway
Drivers:
urban growth, food production
Response:
management action (e.g., investigate, protect, restore)
Landscape model context
Pressures:
land use change, increased runoff, fertilizer application
State:
water chemistry changes, physical habitat changes
Impact:
altered stream structure/function
StreamCat indicators:
canal density, % imperviousness, urban land use, ag land use, road density/crossings
Biological indicator:
macroinvertebrate index (CSCI)

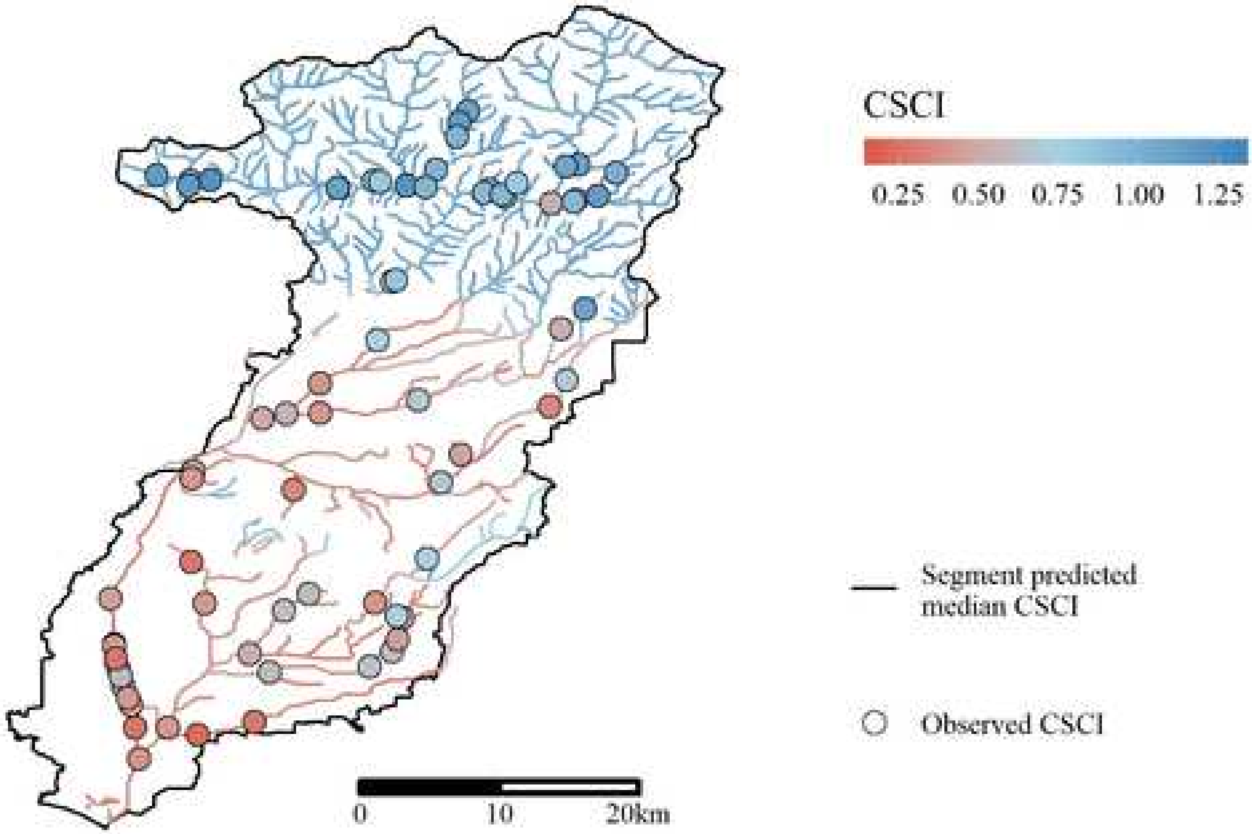
Figure3



Stream class
likely unconstrained
possibly unconstrained
possibly constrained
likely constrained
Observed site score
over scoring
expected
under scoring
(a) Range of expected CSCI scores for stream segments
(b) Expected CSCI scores within certainty range
(c) Stream segment classification by CSCI threshold
(d) Observed CSCI scores by stream classification
Example segments
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
0.0
0.5
1.0
CSCI scores

Figure4

(a) Segment median predictions and observed scores for the CSCI




(b) Land cover in the San Gabriel watershed

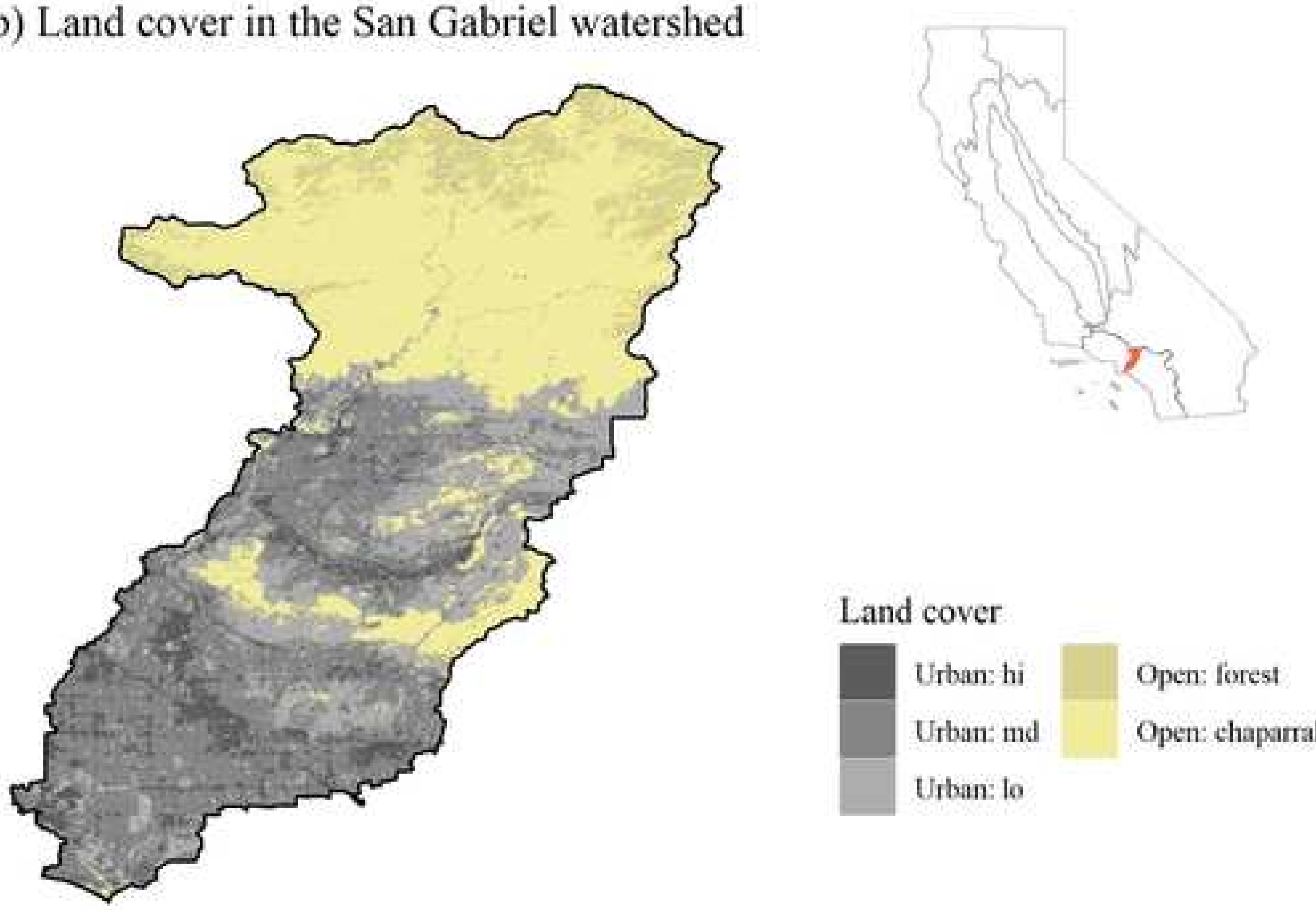


Figure5

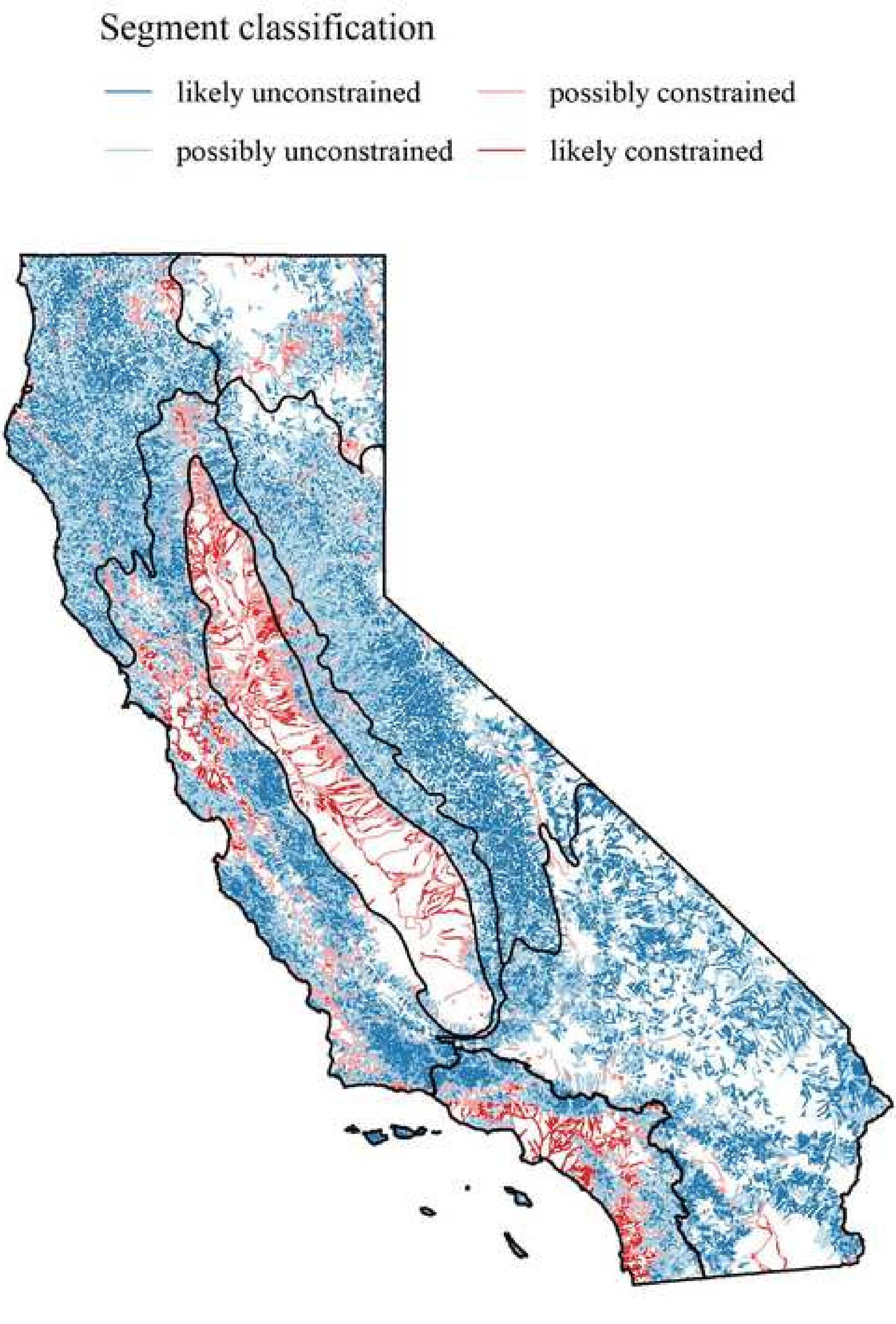
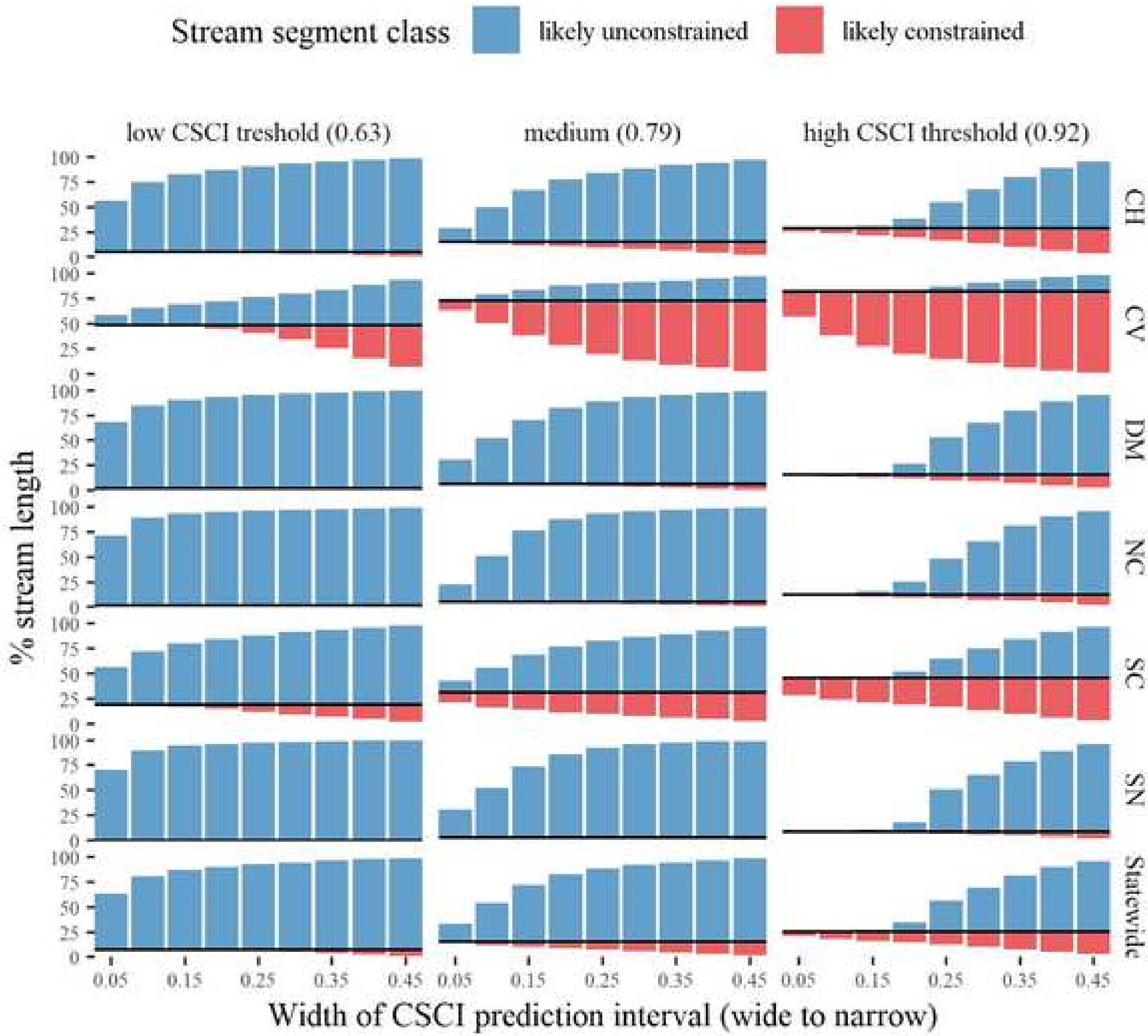
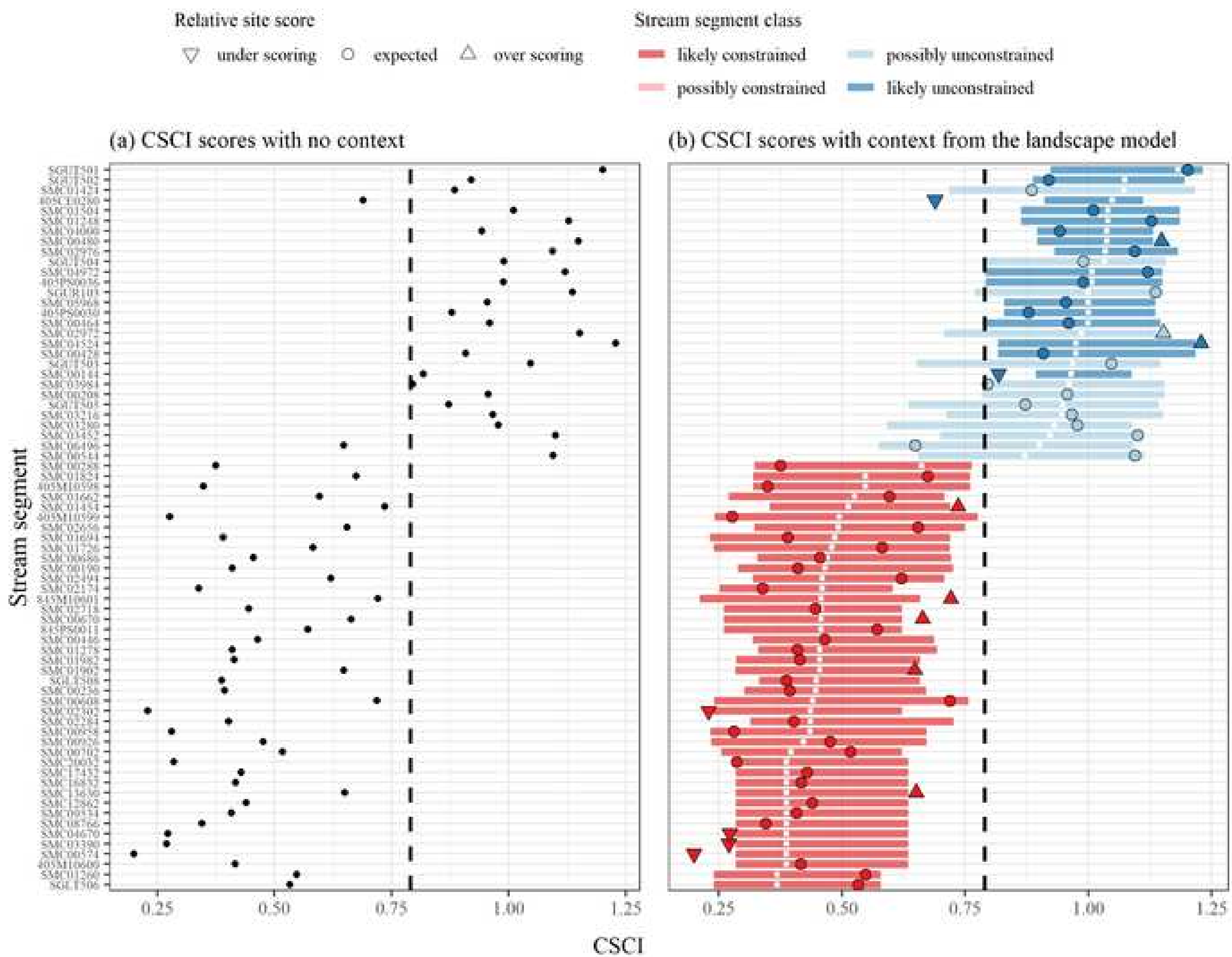

Segment classification
likely unconstrained
possibly constrained
possibly unconstrained
likely constrained

Figure6



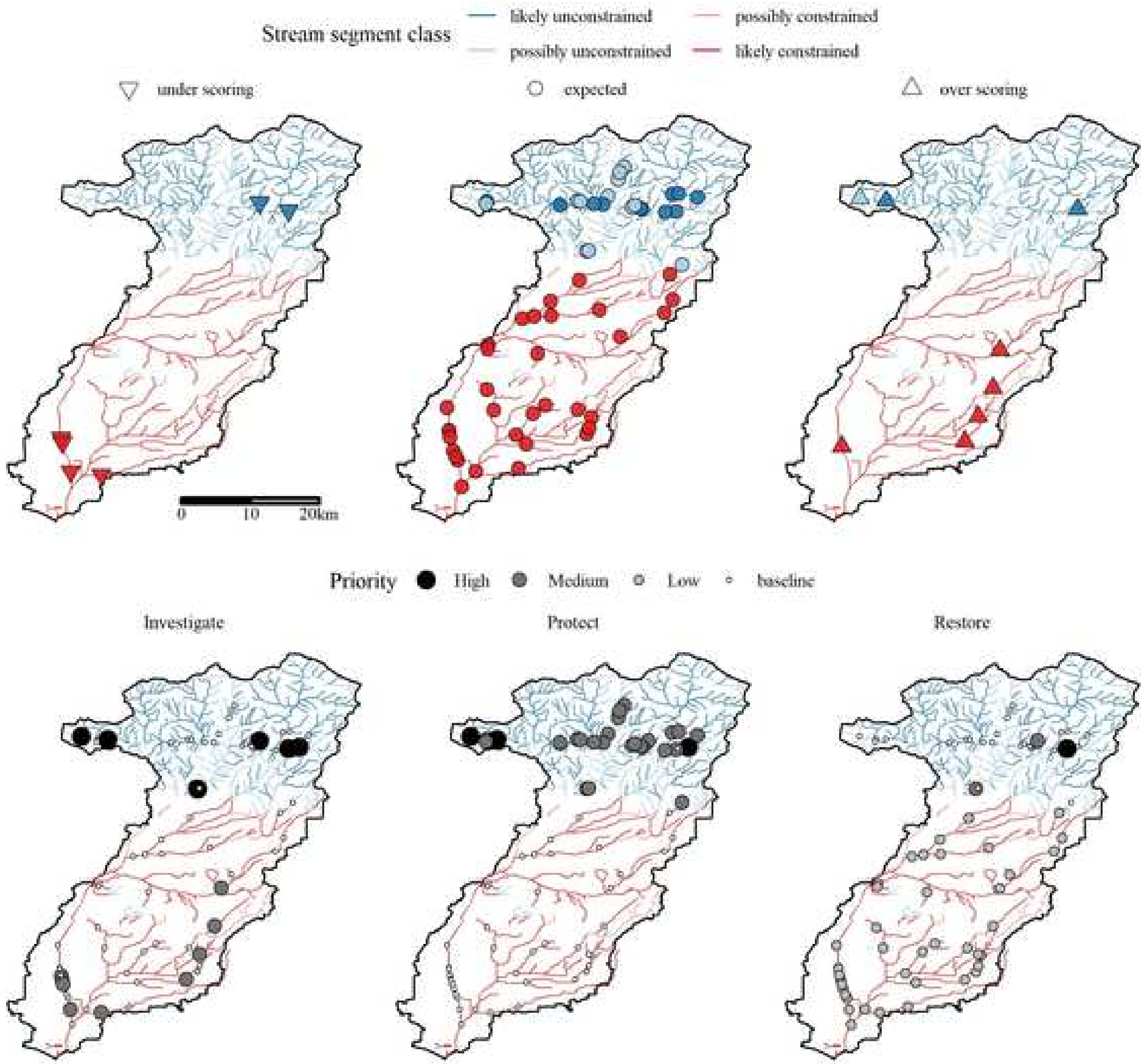
Stream segment class
likely unconstrained
likely constrained
low CSCI treshold (0.63)
medium (0.79)
high CSCI threshold (0.92)
CH
CV
DM
NC
SC
SN
Statewide
% stream length
100
75
50
25
0
0.05
0.15
0.25
0.35
0.45
Width of CSCI prediction interval (wide to narrow)

Figure7



Relative site score
under scoring
expected
over scoring
Stream segment class
likely constrained
possibly constrained
possibly unconstrained
likely unconstrained
(a) CSCI scores with no context
(b) CSCI scores with context from the landscape model
Stream segment
CSCI
0.25
0.50
0.75
1.00
1.25

Figure8



Stream segment class
likely unconstrained
possibly unconstrained
possibly constrained
likely constrained
under scoring
expected
over scoring
0
10
20km
Priority
High
Medium
Low
baseline
Investigate
Protect
Restore

Figure9

(a) Satellite view of Tecolote Creek and surrounding land use, San Diego County, USA


(b) Physical habitat conditions at Tecolote Creek