**Prioritizing Management Goals for Stream Biological Integrity Within the Developed Landscape Context.**

*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.*

*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-r2 = 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.*

1. 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.*

1. 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.*

1. 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.*