Comments: Eutrophication Indicator Thresholds Protective of Biological Integrity in California Wadeable Streams

Lester L. Yuan

In my review, I first considered the written comments from different stakeholder groups and broadly grouped them into four categories (Appendix 1). My own comments echo those of the stakeholders and provide suggestions for analytical solutions to some of the issues.

Accuracy of estimated relationships

The causal relationship between biostimulatory factors and changes in the stream macroinvertebrate and algal assemblages has been established by numerous manipulative experiments (as indicated by the references cited in the manuscript). However, the accuracy of the estimated relationship (i.e., the quantitative agreement between the estimated relationship and the "true" causal relationship) is of critical importance for the application of thresholds derived from this analysis. Accuracy of the estimated relationships is one aspect of threshold derivation that can be addressed via targeted analysis. More specifically, the accuracy of the estimated relationship can be improved by accounting for strong covariates. Strong covariates can be identified as other stressors or natural factors that are strongly correlated with nutrients (TN and/or TP) and with the response variable. The effects of these covariates can be controlled by stratifying the dataset into groups with similar values of each covariate. The EPA Stressor-Response guidance document (US EPA, 2010) provides a systematic approach for identifying and controlling for the effects of covariates, and would be a useful resource to consult for this exercise.

Eliminating all sources of uncertainty regarding the accuracy of the estimated relationships is unlikely, but it would be good to provide transparency regarding the effects of this uncertainty on the final management decisions.

Logistic vs. continuous regression

The current analysis categorizes the biological response into two classes, one that exceeds some biological threshold and one that is lower than the threshold. Subsequent analysis is bases on this two-class response variable. When continuous response variables (such as biological index scores) are categorized into a binary outcome, information is lost. It may be useful to consider whether the advantages of using a binary outcome outweigh the loss of information.

Threshold and constraint models

The interaction between these threshold models and Beck's landscape models needs careful consideration. Beck's models predict constraints on biological condition based on landscape characteristics, but these landscape characteristics can also give rise to increased nutrients. For example, the proportion of agricultural land use is one predictor variable in Beck's model, and one pathway by which agricultural land use can affect stream biota is via increased nutrients. If Beck's model is used to identify the potential range of biological conditions, it seems possible that stream reaches that could benefit from reductions in nutrients will be a priori be designated as constrained.

My understanding is that Beck's models are intended to predict constraints on biological condition that are not related to pollutant load, but as they are currently formulated, pollutant load is included as one important pathway linking the predictor variables in the model to biological condition. Consideration of the full conceptual model, and thinking through the relationships between land cover, pollutant loads, and biological condition may help design a system of models that captures both constraints and pollutant effects.

Uncertainty in eutrophication variables

Much of the difference in the performance of different eutrophication variables in predicting biological condition can likely be attributed to differences in measurement uncertainty for each of the eutrophication variables. A variable such as percent algal cover would be expected to much more temporally variable and measured with much more uncertainty than TN, and therefore, models estimated with percent algal as a predictor will yield broader confidence intervals than TN. It would potentially be useful to control for the effects of uncertainty in the predictor variable in these models, so that comparisons of model performance would be more meaningful. If revisit or resample data are available, it would be good to quantify the temporal and sampling variability of different eutrophication variables and incorporate this knowledge in subsequent models.

Metric sensitivity

Evaluating the relative sensitivity of different metrics (e.g., ASCI vs. CSCI) to nutrient enrichment is challenging from a conceptual perspective. How do we quantitatively define metric sensitivity? Steepness of initial response? Most precise relationship? Think about metric sensitivity with regard to final management decision, and then potentially incorporate this thinking in the comparisons of different thresholds for different metrics.

Implementation issues

Strictly speaking, implementation issues should be resolved by environmental managers, but scientific insight can help inform the management discussion. Below are some implementation issues that could benefit from scientific insights.

1. **Reference site threshold vs. stressor-response thresholds**: There is no expectation that thresholds derived by these two approaches would be the same. Reference thresholds speak to the best conditions that are currently observable in a particular study area, whereas stressor-response thresholds provide insight into the values of eutrophication indicators that are associated with a particular biological condition. Differences between these two thresholds can inform discussions of final management targets, especially when the meaning of the thresholds is clear.
2. **Allowable frequency of exceedance**: The current logistic regression approach defines nutrient threshold based on a targeted value for the biological index and a targeted exceedance rate. The allowable exceedance rate (and possibly the targeted biological

index score) are management decisions. Tools (e.g., Rshiny) that allow managers to interactively adjust these management decisions help increase the transparency of final decisions.

3. **Criterion expression**: The accuracy and precision of the final estimated relationships can influence how any biostimulatory objective is expressed. In final estimated relationships are relatively inaccurate or imprecise, additional metrics (in addition to nutrient concentrations) that support assessment decisions may be considered (see https://www.epa.gov/nutrient-policy-data/guiding-principles-integrated-nutrient-criteria-bioconfirmation). In cases in which estimated relationships are highly accurate and precise, biostimulatory objectives can be expressed more simply.
4. **Regional flexibility:** Clearly identifying sources of uncertainty and quantifying this uncertainty in the estimated relationships can help each regional board determine how best to implement thresholds in their programs. That said, comments from stakeholders identifying biases in statewide biological objective models based on differences in index values associated with different biological clusters seem misguided to me. The biological groups are defined only as an intermediate step for specifying predictive bioassessment models, and are not intended to be used for assessing performance of the final index scores.

References

US EPA (2010). *Using stressor-response relationships to derive numeric nutrient criteria*. Office of Water, U.S. Environmental Protection Agency, Washington, DC.

Appendix 1: Summary of stakeholder comments

1. Concerns with the data:
    a. How does sampling variability associated with different field measurements affect the results?
2. Are the estimated relationships accurate?
    a. Do correlative relationships indicate causality?
    b. What are the effects of covarying stressors?
    c. What are the effects of natural factors?
    d. Are there important regional differences among the models?
3. Are the relationships precise enough?
    a. Can useful thresholds be derived given the uncertainty of the relationships?
4. Implementation questions
    a. Will these thresholds support regional variability?
    b. How are these thresholds implement with causal assessment?
    c. How do these thresholds compare with reference thresholds?