PEER REVIEW OF
Improving Tools to Link Nutrients to Adverse Effects on Stream Ecosystem Services in California

The "Improving Tools to Link Nutrients to Adverse Effects on Stream Ecosystem Services in California" report (hereafter refers as the Report) has the:

Objectives of:

1. Estimating the natural background and ambient concentrations of nutrients and candidate indicators of primary producer abundance in California wadeable streams;

2. Exploring relationships and identify thresholds of adverse effects of nutrient concentrations and primary producer abundance on aquatic life use indicators in California wadeable streams;

3. Evaluating the Benthic Biomass Spreadsheet Tool (BBST) for California wadeable streams using existing data sets, and recommend avenues for refinement.

With intended use of:

1. Providing science based information for policy decisions for the California State Water Resources Control Board (SWRCB) for developing nutrient water quality objectives for the State's surface waters.
2. Providing knowledge to SWRCB staffs who are using the cause-effect approach to develop Nutrient Numeric Endpoint (NNE) framework. A newly available statewide database was used to:
    a. Characterize the relationship and quantifying thresholds between nutrients and benthic algae abundance;
    b. Provide context for these thresholds by summarizing available data on reference and ambient concentrations of candidate nutrient and algal abundance indicators
    c. Evaluate the existing BBST model performance in predicting algal abundance.

The Report, focusing on such objectives and intended uses, conducted "exhaustive" analyses to evaluate relationship among nutrients, benthic chlorophyll a, benthic diatom, benthic non-diatom algae, and benthic AFDM; to identify thresholds at which nutrient changes result in dramatic algal abundance changes, and to evaluate models that use nutrients to predict algal abundance.

Overall, the Report presented a strong report with conclusions that are well supported by the results of "exhaustive" analyses using well developed statistical methods. The objectives and the approaches, especially the many indicators and stressors and the many statistical methods, used are very ambitious.

The Report does a good job in describing the needs, identifying and assembling needed data, conducting extremely intensive data analyses, and summarizing and explaining the results in a useful way. The Report may need some editing, especially it has too many places with sentences that are way too long.

Comment Questions on Overall Report:

1. *Is the report organization optimal for a document of this length? Would it be better if each of the report sections was a stand-alone piece?*

The organization of the report is proper for a document of this length. It would not be better if each of the report sections was a stand-alone piece. This is because the sections of background, methods, threshold identification, and predictive models are highly linked. If each of them becomes a stand-alone piece, much of the information needs to be repeated in each piece.

The current Report already repeated some of the information, which needs to be removed. For example, Nutrient Numeric Endpoint (NNE) is already defined in Chapter 1 (page 1) and it is again defined in Chapter 2 (page 4)

2. *Is the cited literature sufficiently comprehensive? Are there any key references that have been omitted?*

The cited literature is sufficiently comprehensive to cover the current knowledge in the study field. There is no need to include all references in the study field for a report that is already this long.

3. *Are any limitations of the datasets and monitoring approaches employed to support the analyses in this report adequately addressed?*

The limitation of the datasets and monitoring approaches that collected the data for conducting the analyses and drawing conclusions are adequately addressed in the report.

Executive Summary

1. *Does the executive summary adequately capture the major findings of the report?*

The executive summary could be improved by:

a. Clearly stating how the study findings can help the SWRCB's effort in the development of water quality objectives. Currently, the executive summary only states "… improved data from statewide stream probabilistic and targeted bioassessment surveys can strengthen the scientific basis for policy decisions …". The objectives of the study were not clearly linked to the intended uses.

b. The executive summary can be improved by more concise writing and choose of more proper words. For example, "nutrient pollution" (line 2) and "management control" (line 4) could be replaced by words like nutrient enrichment and nutrient management or nutrient enrichment regulation. A few of the sentences are way too long.

2. *Are the summary statements adequately supported in the body of the report?*

For most part of the summary statements are adequately supported by the body of the report. However, some statements are inaccurate. For example, "Nearly 66% of stream kilometers had benthic chlorophyll *a* and 59% has TN and TP values below the 75th percentile of benthic chlorophyll *a* at Reference sites statewide" statement is inaccurate. This is because the authors did not conduct a statewide assessment for all stream length and this statement is for sampled sites only. Therefore, the statement should be "Nearly 66% of THE SAMPLED stream kilometers had benthic chlorophyll a and 59% has TN and TP values below the 75th percentile of benthic chlorophyll *a* at Reference sites statewide. Also, the summary of the findings could be presented more precisely and clearly.

Chapter 1

1. *Does Chapter 1 provide sufficient background information to put the rest of the report into context with respect to information needs for the state of California?*

The Chapter provides sufficient background information to put the rest of the report into context with respect to information needs for the state of California, but it could be improved by clearly stating how the study objectives and findings can help the SWRCB's effort in the development of the water quality objectives.

Chapter 2

1. *Are the methods used to estimate reference and ambient values for stream eutrophication indicators pooling data across multiple monitoring programs scientifically valid?*

The methods used to pooling data across multiple monitoring programs are scientifically valid. However, the description of those methods can be improved substantially.

For example:

a. The statement of "What is the distribution of nutrient and algal abundance indicators …?" (page 4) does not refer to the distribution of indicators, but refers to the distribution of the values or status of the indicators.

b. The statement of "The probabilistic survey design for the California ambient surveys…" (1st line, page 5) does not clearly indicate if the California ambient surveys apply to all the 3 data sources.

c. Last paragraph from the top on page 5 – what is the spatial stream unit (what is considered as a stream?) for the "probability sites"?

d. Paragraph 2 from bottom of page 7 – It is not clear how macroinvertebrates and algae were sampled. Were they sampled from the same location?

e. Table 2.2 caption– "1-km buffer of the sample point" and "5-km buffer of the sample point" are unclear. Are they using radius? If yes, do they exclude the portion outside of the watershed? If they are buffer parallel to the channel, do they exclude the portion outside of the watershed?

f. Table 2.2 – Reference threshold column should have <, and Stressed threshold should have > signs.

g. Table 2.4 caption – "SE: standard error of the mean; CI: confidence interval (95%)" are not in the Table.

h. Table 2.5 – May consider change column 2 from "to" to "-"; for example change "0.01 to 26" to "0.01-26".

i. Page 14 - Fig 2.3 should be Fig 2.2.

Chapter 3

1. Overall

The introduction of Chapter 3 needs more thorough explanation on what ALUs mean here. For example, does this include benthic algae, macroinvertebrates, microinvertebrate, macrophytes, fish, and/or anything else? The model cited in the Report (Davies and Jackson 2006) includes different biological communities, especially focuses on macroinvertebrates and fish. Because nutrients effects on the different biological communities are different, the use of specific biological communities has strong implication on the Nutrient Numeric Endpoint (i.e., different biological communities will have different Nutrient Numeric Endpoint).

Although Section 3.2.2 presented different measures of macroinvertebrates, diatom, and soft algae that are used as ALU indicators, this needs to be mentioned earlier. Most importantly, the list of measures of macroinvertebrates, diatom, and soft algae are exhaustive and their responses to nutrients enrichments will be different. This exhaustive list of indicators and the later on analyses on those indicators obscured a clear conclusion of the Report. At least it needs a clean explanation why all the measures of the indicators on the list are needed.

Also, the intended uses of the Report are to provide science based information for policy decisions for the SWRCB to develop nutrient water quality objectives and for the SWRCB staffs

using the cause-effect approach to develop the Nutrient Numeric Endpoint framework. The description on how the findings of the report can be linked to those uses are unclear. For example, the Report not only examined relationships between nutrients and macroinvertebrates, diatom, and soft algae measures, but also examined relationships between macroinvertebrates and chlorophyll a, macroinvertebrates and macrophytes, and macroinvertebrates and AFDM. How the relationships between macroinvertebrates and other biological measures will be used or helpful for the establishment of Nutrient Numeric Endpoint need to be explained clearly. This is because the traditional methods of establishing nutrient criteria are based on relationships between nutrients and biological indicators, but NOT based on relationships between macroinvertebrates and chlorophyll *a*, and/or relationship between macroinvertebrates and macrophytes.

2. *Have the different methods for evaluating response thresholds of primary producer biomass and nutrient effects been described adequately so that someone previously unfamiliar with these methods can understand the approach and the method strengths and weaknesses and interpret the results?*

Yes, the different methods for evaluating response thresholds of primary producer biomass and nutrient effects have been generally described adequately. It is probably helpful if the Report provides clear reasons why all the response thresholds identification methods are needed.

Page 27 – subheading 3.2.3 should be 3.2.2.

3. *Are the methods used to estimate response thresholds scientifically valid? Have statistical assumptions been adequately tested?*

The methods used to estimate response thresholds are scientifically valid and their statistical assumptions have been reasonably adequately tested in the literature.

4. *Are the conclusions of this chapter adequately supported by the analyses and results?*

Overall, the Report does a good job in drawing conclusions from the analyses and the conclusions of this chapter are adequately supported by the analyses and results. Some of the conclusions could be presented more clearly. Presently, it takes a lot of efforts for a reader to wade through the materials and figure out what each of the analysis for and what the conclusions are. It would be helpful if the Report provides a sentence of purpose of the analysis and a concise summary of results at the very beginning of each section.

Some specific comments:

a. NMS and CART results – Results presented in Figs 3.3 and 3.4 indicate that the NMS method is not useful. This is because the Y axis lumps all macroinvertebrate indicators

together, and some of them are positively while others are negatively related to the X-Axis measures. The CART analysis used NMS axis 1, hence CART analysis has the same issue. The results of such a mix of the response variables may be misleading.

I am not sure how the thresholds were identified using NMS without CART. The thresholds on Figs 3.3 and 3.4 are also confusing. For Fig 3.3, the MBI vs TN has 2 thresholds. The 1st one is at a location where BMI increases when TN increases; the 2nd one is at a location where BMI decreases when TN increases. The Report selected the 2nd one. Similarly, Fig 3.4 Diatom vs TN or Diatom vs TP have the same issue. The other plots do not show a clear threshold.

Y-axis label on Fig 3.5 is unclear. I suggest labeling the actual Y-axes instead of "Gradient" and moving the legend into the boxes.

b. TITAN and nCPA results – Figs 3.9 is unclear. I suggest labeling the actual Y-axes instead of "Gradient" and moving the legend into the boxes.

c. Piecewise and SiZer – I like the way of the results are presented in Figs 3.10-3.12. The lower panels of Figs 3.14-3.18, each needs an arrow to indicate where the threshold is. As it is presented now, readers have to guess where it is.

5. *Does this chapter do a good job of synthesizing the results of multiple analyses contributing to a weight-of-evidence approach that could be used to support numeric nutrient endpoint development? Can you suggest any improvements?*

The Report generally does a good job in synthesizing the results of the analyses. Some areas could be improved. Fig 3.19 is not particularly helpful. Fig 3.20 needs clearer labels. For example, it is not clear what the colors for the "predictor ordered" mean.

The Section 3.3.4 could provide a precise summary on what the major thresholds are. Currently, this Section does not provide a conclusion and leaves readers to interpret Figs 3.25 and 3.26.

The discussion section stated that "This study found evidence for a range of thresholds for benthic chlorophyll, AFDM, TN, and TP…". I still question why the Report wants to identify thresholds for macroinvertebrates, diatom, and algae in response to benthic chlorophyll a and AFDM. The report could not find many studies of the influence of benthic chlorophyll a and AFDM on macroinvertebrates, diatom, and benthic algae is because they should not be considered as stressors for establishing nutrient criteria.

Chapter 4

1. *Is the evolution of the NNE benthic biomass spreadsheet tool adequately explained to allow the reader to understand its use and potential strengths and weaknesses?*

The Report does a reasonable job in explaining the NNE benthic biomass spreadsheet tool to allow the reader to understand its potential strengths and weaknesses.

However, it is not clear how the spread output is related to NNE, which needs some explanation.

Also, the QUAL2K models are not clearly explained. For Equation 5, K_p (the rate of photosynthesis), K_r (the rate of algal respiration), and K_d (the rate of algal death) are not field monitoring measures. How those parameters are determined from monitoring data?

Page 95, bottom paragraph explains that the users provide nutrients (what kind?), water depth and velocity, radiation, and others. How these input variables are linked to the parameters in Equation 5 needs to be explained.

Page 95, bottom paragraph and Fig 4.2 state that the outputs of the spread sheet are maximum algal density and chlorophyll *a*. The output of equation 5 is BAFDM. How are the maximum algal density and chlorophyll *a* converted into the maximum algal density and chlorophyll *a* values?

2. *Are the methods used to evaluate the performance of the NNE benthic biomass spreadsheet tool scientifically valid? Have statistical assumptions been adequately tested?*

The BRT and B-CART are well documented statistical methods for establishing predictive models. Their uses here are appropriate.

However, the section 4.2.4 is unclear whether the evaluation of the spread sheet models was conducted using the exact regional models with California data. If it is, this needs to be stated clearly. Otherwise, the original spread sheet models (at least the Dodds models) are parametric regression models, while the BRT and B-CART models are Non-parametric models. The predictive models and the models used to evaluate them are two different approaches, which may have some influences on the evaluation results and their interpretations. The BRT and B-CART methods may be more useful for developing new models for California, but they may not be suited for evaluating the spread sheet models.

Section 4.2.5 needs to explain how the RandomForest method is used to identify bias. In another word, it needs to explain how the relationship between "predicted-observed" and predictors can be used for identifying bias. Scientists understand this, but policy makers may not; hence it needs to be explained.

Section 4.2.6 needs to clearly state that this section is to develop potential new predictive models for California, but is not for evaluating the spread sheet models. Also, the Report used both BRT and B-CART to develop predictive models. It may help if it explains why both methods are needed and the strength and weakness of each method.

3. *Are the conclusions of this chapter adequately supported by the analyses and results?*

The conclusions are generally adequately supported by the analyses and results.

Table 4.4 is unclear. Are the 2 columns on the left number of sites?
The 1st sentence below Table 4.4 is incomplete.

It may be helpful if the model performance parameters (R2, intercept, slope) of the original models from their development data sets are added to Table 4.6. This way the readers can judge how good the original models are.

For Section 4.3.2, it may be more meaningful if the Report examines how much variance of the "predicted-observed" is explained by the predictors instead of examining ranking of predictors. This is because predictors ranked high could explain only small portion of variance, hence contribute little in improve the predictive models. In contrast, predictors ranked low could explain a large portion of the variance, hence makes significant contribution to improve the predictive models.

4. *Does the analysis of residuals for model predictions presented in this chapter help to guide future improvements in these models?*

The analysis of residuals for model predictions does help to guide future improvements in these models since it identifies parameters that could contribute significantly to the model, but they have not been included in the current spread sheet models.

The analyses on PCT_MAP, PCT_MCP, PCT_MIAT1, and Soft algal total volume (listed in Table 4.8) do not belong to here because this Chapter is to evaluate the spread sheet and the spread sheet does not include those response variables.

For Chapter 4, it does not explain how the BBST is used for identifying NNE. This needs to be explained clearly at the beginning of the chapter and the discussion needs to be closely linked to such uses.