**Responses to ORD External Peer Reviewer Responses to Charge Questions (Questions repeated in shaded rows)**

| order | Overall comments | Response (revision made) |
|---|---|---|
| 1 | This is generally a well-prepared document. There has been considerable thought that has gone into the approaches given here, and generally the approaches that have been taken are good and defensible. A major issue that cannot be solved is the inherently variable nature of periphyton and the dynamic nature of multiple factors that control biomass. Some of the work is a bit difficult to follow, but these are complex analyses and models, and mostly the authors have done as well as could be expected in explaining the approaches. | No response needed. |
| 2 | Table of Contents: Section 4.4 is in bold faced font. Is this correct? | Fixed heading style in text. |
| 3 | 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. | No response needed. |
| 4 | 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. | Additional editing has been done to reduce the length of run-on sentences. |
| 5 | **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?** | |
| 6 | The report organization is fine, making them free-standing pieces would just require repetition of much information. It might be helpful to break the 3rd chapter into two, as it is pretty long and complex. | There was only one reviewer who suggested breaking Chapter 3 down into two sections. Since there is no logical way to break the Chapter down into two parts, we are choosing not to. |
| 7 | I think the document length and organization is appropriate. The document length is long, but I do think the pieces belong together for a report. It is possible that Ch 3 and 4 might be stand alone manuscripts. | At minimum, we intend to submit Chapter 3 as a manuscript |
| 8 | 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. | No response needed. |
| 9 | 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) | Fixed. |
| 10 | The report is organized well into chapters of appropriate length and topics. | No response needed. |
| 11 | Yes. The report is very well organized into a logical framework and gives a good platform for the final recommendations. I do not recommend separating it into stand alone pieces. | No response needed. |
| 12 | **Is the cited literature sufficiently comprehensive? Are there any key references that have been omitted?** | |
| 13 | Generally the literature has been well represented. There are a few that would have helped, I have noted suggestions for additions and listed them at the bottom. | Suggested references have been added to the document where appropriate. |
| 14 | It is sufficient. | No response needed. |
| 15 | 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. | No response needed. |
| 16 | Recommendations for some literature have been added in comments on the manuscript. Often, only book chapters reviewing primary literature were cited, rather than the actual primary literature. More citations of primary literature documenting research and results should be cited, rather than book chapters. | References have been checked and there are only a few (other than methods references) that cite chapters in books rather than the primary literature. |
| 17 | There are two areas, and associated literature, that don't appear to have been adequately canvassed for such an analysis: | See below. |
| 18 | 1. The form of nutrients being measured and associated scale of the processes (time and space variability) that result in the end-point concentrations being measured in the streams. How do nutrient concentrations vary during the algal 'growing period' (eg. following scouring)? What is the relevance of 'point in time' samples? What is the best integrative measure of nutrient concentration (or flux) that the algae have access to? What does a total nutrient measure represent vs. the soluble fraction. These questions are fundamental to developing nutrient management objectives and controlling stream eutrophication at the watershed scale. | With the exception of a few field duplicates, the data available for the CA state monitoring programs do not include time series of stream nutrient concentrations or measurements of the full suite of nutrient forms. Most of the nutrients are in dissolved form. To assess the representativeness of point samples of nutrients during the growing season, a summary of data for 60 California NAWQA stations sampled at a fixed frequency over the year has been added to Appendix 5. Maximum total N and P tend to occur in January - February, followed by declines, then an increase during the growing season. TP tends to peak in July, while TN slowly increases from May through September. There is a fair amount of variability among biweekly TN and TP values over the growing season, which contributes to the noise in the relationships we have assessed. The predictive power of relationships could probably be improved by averaging values for multiple water quality samples over the months preceding collection of benthic algal samples, or by using nutrient values inferred by diatom taxa composition (Pan and Stevenson 1996). However, in most cases, annual maxima for TN and TP increase linearly in proportion to growing season averages and thus growing season values should represent the relative trophic condition of streams. |

| | | |
|---|---|---|
| 19 | 2. Appropriate ways of classifying the stream environments that are relevant to the scale of the individual sampling sites and better reflect local upstream conditions and the hydrological regime (eg, a hierarchical controlling factors approach). Sites could then be partitioned accordingly (at a less coarse scale than done in the present analysis) to help understand the variability in algal biomass and environmental drivers of this. Ecoregions are too broad. | The current analyses do in fact include controlling factors at a range of hierarchical scales. California ecoregions are tested as one approach, but co-factors influencing algal abundance that vary at finer scales than ecoregions (e.g., slope, temperature, shading, substrate) are also included. Finally, we have explored refining geographic boundaries of regions with similar nutrient-biomass responses through Bayesian CART analyses, using geographic coordinates as classification variables. |
| 20 | There are many published studies on these two issues. | See responses above |
| 21 | **Are any limitations of the datasets and monitoring approaches employed to support the analyses in this report adequately addressed?** | |
| 22 | The authors are very careful to note the limitations of their data sets and how they analyzed them. Using multiple approaches is a great way to strengthen inference and the authors are to be commended for their comprehensive approach. | No response needed. |
| 23 | Yes. Limitations are addressed. | No response needed. |
| 24 | 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. | No response needed. |
| 25 | Although the sample size is relatively large, the role of confounding factors and temporal variability in conditions is likely limiting ability to observe relationships among some variables. Relationships between algal biomass and nutrients, for example, are notoriously challenging, and often require repeated sampling of a site and 75 or more sites to resolve nonlinear relationships between algal biomass and nutrients in streams when natural variability is controlled. Given the great variability in climatic, geologic, hydrologic, topographic, and riparian conditions of streams in California, we would expect challenges in establishing biomass-nutrient relationships. Variability in substratum and current velocity among habitats in streams also affect algal biomass. The EPA/CA protocols calling for sampling random locations along 11 transects in a reach does not provide sufficient sample size to sample stratification for within-stream habitats, such as riffle, pools, and runs, to provide sufficiently precise haracterizations of algal biomass in streams to develop good algal biomass-nutrient relationships without enormous sample sizes. Alternatively, the NAWQA approach of sampling richest target habitat or a habitat specific sampling protocol for characterizing algal biomass would improve assessments given the great problem with spatial and temporal variability in algal biomass in streams. Given relative costs of sampling versus field travel and the importance of algal biomass as an indicator of ecosystem services, perhaps more intensive sampling, stratified sampling, and separate composite samples for different habitats is warranted. Could data from the National Rivers and Streams Assessments be used to supplement CA's data? | We agree that, due to the patchy nature of benthic algae in streams, any rapid protocol is unlikely to yield high-precision biomass data. The suggestion to supplement our data set would not be wise due to inconsistencies between our sampling protocol and that of National Rivers and Streams Assessment. Using a richest-targeted habitat approach, as suggested, has some drawbacks for estimating benthic algal biomass because it may not be representative of overall biomass levels across the stream (nor is it necessarily repeatable, due to the potential for subjectivity associated with determining "richest" among operators). Although we have information about substratum types and microhabitats in the project data set, and we use some of this information in the BRT models to try to account for the influence of that type (as well as other types) of natural variability, we do not have those measures associated with specific benthic algal subsamples, so there is a limit to how well such confounding influences can be factored out at site scale. Given all that, we have fairly extensively discussed the limitations of the rapid bioassessment-type data that we had available for this report, particularly with respect to the low preceision in biomass estimates it yields. To reiterate, future sampling strategies would do well to seek to improve precision through higher-effort (and thus costlier) sampling, when nutrient management planning, or assessment of success of nutrient-management efforts, is an express goal of the sampling campaign. Repeat sampling at sites where nutrient impacts are being evaluated would also be desireable for inclusion in any such protocol. |
| 26 | The datasets and BBST models have significant limitations. The authors generally recognize these and explain them well. However, there is little they can do to address these problems which I believe suggests that the results of this study have limited application. See further comments below. | No response needed. |
| 27 | **Executive Summary** | |
| 28 | **Does the executive summary adequately capture the major findings of the report? Are the summary statements adequately supported in the body of the report?** | |
| 29 | The executive summary does adequately support the body of the report. The authors may wish to alter it a little based on my comments in the document. | No response needed. |
| 30 | I think the majority of it is adequate. I placed comments in areas where I had a question. | No response needed. |
| 31 | The executive summary could be improved by: | See below. |
| 32 | 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. | The executive summary and introduction clearly state that the intended outcome of this project is research, but that the results may be used as a line of evidence in the SWRCB's policy decisions on nutrient criteria. Additional analyses and scientific syntheses are underway to specifically support the SWRCB in this endevour (M. Sutula, personal communication). Minor modifications to the text in the executive summary and introduction have been made to further clarify this point. |
| 33 | 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. | We have changed "nutrient pollution" to "nutrient overenrichment". Sentence lengths have been reviewed and reduced through the report where appropriate. |

| | | |
|---|---|---|
| 34 | 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. | Because the surveys were conducted based on probabilistic designs, it is possible to infer the percentage of stream kilometers associated with different levels of nutrients and algal biomass. However, these statements are valid only for the population of streams sampled (perennial wadeable streams). **This point has been highlighted in the text and conclusions reworded to reflect this fact.** |
| 35 | The findings of the report provide a good overview and support conclusions from chapters. Recommendations 1 and 3 are well supported by the results. Recommendation 2 provides a complementary approach for developing nutrient management targets, but not an alternative approach unless ecosystem services that are affected by algae, other than benthic macroinvertebrates, are not important. | No response needed. |
| 36 | Yes. A good summary. | No response needed. |
| 37 | **Chapter 1** | |
| 38 | **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?** | |
| 39 | The background is a little thin in this chapter, but most of the required background is given in the introductions to the following chapters, so this is not really a very important point. As stated later, I would prefer a broader view of eutrophication based on potential shifts of heterotrophic state as well as autotrophic state with respect to biotic integrity and water quality. Some have referred to this as green versus brown food webs. This would set up much of the work that follows in a way that is closer to a more holistic view of eutrophication in streams. The first approaches were mainly based on the example of lake eutrophication management, planktonic algae and nutrients control being of main concern. Biotic integrity is an endpoint that can be altered through heterotrophic (allochthonous) food web pathways. The heterotrophic state may also tie to other water quality issues such as dissolved oxygen levels. | The introduction to this chapter was intended to be brief and to mainly set up the management context for this research; we are thankful that the reviewer recognizes that the needed literature review is given in subsequent chapters. We've added changes to DO and biotic communities to place additional focus on the endpoints of concern in the introduction. Review of green versus brown pathways is a helpful suggestion and has been added to chapter 3. |
| 40 | Yes | No response needed. |
| 41 | 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. | See response above. |
| 42 | The title of the report "IMPROVING TOOLS TO LINK NUTRIENTS TO ADVERSE EFFECTS ON STREAM ECOSYSTEM SERVICES IN CALIFORNIA" indicates tools will link nutrients to stream ecosystem services. Ecosystems services are not related well to the endpoints measured in the research. Ecosystem services are related to beneficial uses in the introduction, but only benthic algal biomass is those uses. And all that linkage takes place in the first paragraph of the introduction. Ecosystem services are not mentioned throughout the rest of the document. Maintaining high levels of biological condition is also beneficial use and ecosystems service, even though its relationship to valuation by most people may be debated. It is a goal of the Clean Water Act. Since many endpoints measured in the project are related to taxonomic metrics, they should be explicitly related to ecosystems services if it is to take a prominent position in the title of the report. | Thanks for pointing this out. This research projects has several components, only one of which is addressed in this report. One component of this project is an element specifically linking stream condition to beneficial uses. We have changed the title of the report to better reflect the focus of this particular element. |
| 43 | Cite primary literature that documents that relationships exist between measurable stream attributes (algal blooms in wadeable streams) and ecosystem services and human well-being exist rather than reviews that state these relationship exist (e.g. Biggs 2000, Lembi 2003). Suplee et al. 2009 is an important reference for linking benthic algal biomass to valuation of aesthetics, and thus establishing benthic algal biomass is an appropriate measure of aesthetics, which is an important ecosystem service. It should be used in the first paragraph of the introduction. Do you have evidence that wadeable stream algal blooms negatively impact human health and other ecosystem services or beneficial uses through toxin-forming harmful algal blooms, proliferation of pathogenic bacteria, and taste/odor problems in municipal drinking water supplies, or are problems in downstream lakes and coastal zones? | This report is part of a larger project investigating the potential impact of nitrogen on ecosystem services in wadeable streams in California. However, relatively little has been done in this part of the study to relate nitrogen to specific ecosystem services. Therefore we have changed the title of this report to refer to impacts on aquatic life . Additional references have been added to the introduction to reflect the potential effect of nitrogen on ecosystem services related to recreational use (Suplee et al. 2009) and human health. Dog deaths have in fact been recorded due to exposure to cyanotoxins in California rivers (http://www.cdph.ca.gov/healthinfo/environhealth/water/pages/bluegreenalgae.aspx). A 2011-12 survey of southern California streams detected benthic cyanotoxins in 41% of sites sampled (B. Fetscher, pers. comm.) |
| 44 | Aquatic life use, as measured with metrics using taxonomic information, an important endpoint in analyses, is not integrated into the introduction. The introduction is about benthic biomass and does not relate biodiversity elements of biological condition as measured with taxonomic metrics to ecosystem services, beneficial uses, and goals of the research and report. | Thanks for pointing this out. We've intentionally left the introduction to chapter 1 brief and management oriented, given the fact that we have two audiences for this report (scientists and managers). Per your suggestion, we've added changes to DO and biotic communities to place additional focus on the endpoints of concern in the introduction. We've used your suggestion to a fuller extent in Chapters 3 and 4 to improve the conceptual basis for our approach. |
| 45 | Adequate. However, while it is good to start with a wider international context I was expecting to see a good summary of water use/ecological issues caused by nutrient enrichment and associated algal growth in streams of California. I would have thought that demonstrating a clear State-wide problem, and then a commentary on where the most significant detrimental effects occur (and what these are), would be fundamental to justifying further work on this topic. Maybe this is well documented elsewhere and could be covered by another citation or two??? | In Chapter 1, we've intentionally left the introduction to chapter 1 brief and management oriented, given the fact that we have two audiences for this report (scientists and managers). Per your suggestion, we've added a sentence in the introduction to point to well-documented examples of eutrophication in California wadeable streams. |
| 46 | **Chapter 2** | |
| 47 | **Are the methods used to estimate reference and ambient values for stream eutrophication indicators pooling data across multiple monitoring programs scientifically valid?** | |

| | | |
|---|---|---|
| 48 | The data in table 2.2 suggest that feedlots were not considered distinct from agriculture, that urbanization was considered equivalent to agriculture, and provides no real justification for the thresholds. | Existing data from the ambient and reference surveys were used in this analyses. A complete description of the approach to selecting reference sites can be found in the Ode et al. (under review). This level of detail was not appropriate to include in the report. |
| 49 | The authors may want to consider other methods for assessing reference. Smith et al (2003) have published reference distributions for the United States. Dodds and Oakes have published a method for backing out all human influence. Such an approach might be useful in the Central Valley were reference sites are few and far between. | As noted above, existing data from the ambient and reference surveys were used in this analyses. That being said, this is an interesting suggestion and will be passed along to the managers of the stream bioassessment program. Results from the current study are comparable to results of the alternative approach to describing reference condition described in Smith et al (2003); this has been added to Chapter 2. |
| 50 | It is important to set the upper 75% correctly and be certain that the values for reference indicate true reference. For example atmospheric nitrogen deposition can be considerable in California (e.g., Bytnerowicz and Fenn 1996). Smith et al. (2003) clearly indicate that this is an important and substantial correction to reference nutrients in many parts of the United States. Also, lumping all agriculture might be problematic. A proportional correction from the Smith et al (2003) article could be applied to the current numbers. | We agree that the definition of reference is important. A second component of this project, not included in this report, is focused on measuring the influence of atmospheric deposition on Southern Califronia reference sites. We hope to provide additional clarity to the issue with this study component. As noted above, existing data from the ambient and reference surveys were used in this analyses; suggestions re: treatment of reference sites will be passed along to the managers of the stream bioassessment program. |
| 51 | The expectation is that 25% of the sites will exceed 75%, so maybe the document should report the percentages over that in table 2.8. The executive summary stresses that 50% of all streams fall below the 75% reference level, which sounds ok phrased that way, **but another way of looking at it is that there has been a doubling in the number of high nutrient streams.** | We think that this suggestion makes the table more complicated and would prefer to add a sentence to the table reminding folks that, by definition, 25% of reference sites exceed this value. |
| 52 | I applaud the caveat that more detailed seasonal sampling could be important. Are the authors certain that an analysis of variance, or analysis of covariance that factors out intensity of human impacts cannot dissect out ecoregional differences? Data presented in chapter 4 sort of get to this, but not explicitly. | An analysis of variance will not do a good job of capturing significant interactions and nonlinear effects such as those bewteen temperature and nitrogen. |
| 53 | I do think it is scientifically valid providing you are completely honest about the proportions of data collected using a similar laboratory and field methodology. I placed some comments within the text where this could be a bit more clear. | No response needed. |
| 54 | The methods used to pooling data across multiple monitoring programs are scientifically valid. However, the description of those methods can be improved substantially. | There was no trick to the pooling. As noted in the report, standard protocols were used for the data-collection efforts. |
| 55 | For example: | |
| 56 | 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. | We have inserted "values of" into the question. |
| 57 | 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. | Similar designs apply to the PSA and SMC surveys (RCMP is not probability-based). Language has been added to clarify. |
| 58 | 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”? | The survey design was based on a "linear" resource *sensu* Kincaid and Olsen (2008). The reporting unit for this type of survey is in terms of length (e.g., stream kilometers). Language has been added to clarify. |
| 59 | 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? | The two assemblages were collected in tandem at each of the 11 subsampling locations described at each study site; first bugs, then algae, slighly offset so that sampling locations did not interfere with one another. Language has been added to clarify. |
| 60 | 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? | Only the land within the catchment contributing to the sampling site is included within the indicated radius (i.e., the area is clipped at the watershed boundaries). Language has been added to clarify. |
| 61 | f. Table 2.2 – Reference threshold column should have <, and Stressed threshold should have > signs. | Fixed |
| 62 | g. Table 2.4 caption – “SE: standard error of the mean; CI: confidence interval (95%)” are not in the Table. | “SE: standard error of the mean; CI: confidence interval (95%)” have been deleted from the caption. |
| 63 | | Fixed |
| 64 | i. Page 14 - Fig 2.3 should be Fig 2.2. | Fixed |
| 65 | What was the variation in quality (e.g. % undisturbed watershed) in reference sites among ecoregions? Without that qualification, comparison across ecoregions is less informative because quality of reference sites in addition to natural factors remains in important variable. Is there a standardized scale for reference condition across ecoregions? | As noted in the report, the same standard screens for establishing "Reference" were applied across all ecoregions. We have looked at the distribution of % open space within the contributing watersheds for the Reference sites and found that there is good congruence among them except for within the South Coast (and particularly the xeric portion of the south coast), which does have a lower overall percentage of open space. South Coast xeric was 96% whereas the regions ranged from 98 to 99%). We will qualify this in the Results. **The attached PowerPoint file shows analysis output.** |

| | | |
|---|---|---|
| 66 | It is difficult to answer this question as not enough detail is given in the description of the different programs and their methods. Major differences in data can occur amongst different operators using the same methodologies, let alone amongst programs using variations or quite different methods and habitats (eg, runs vs. riffles) for stream benthic algal sampling/monitoring. There needs to be confidence that the choice of the sampling reaches, then the choice of the sampling points, were consistent among the operators and programs (eg, how does the methodology deal with macro-algal patchiness?). What was the overarching QA system and statistics to ensure comparability of data and therefore that the datasets can be pooled? Unless inter-operator variability can be understood and factored out, then I suggest that there will be significant variance in the dataset that is not driven by environmental variation. | The state and regional monitoring programs through which the data used in the report were collected all use the same sampling SOP, as described in the report. That methodology is a "multihabitat" approach (very similar to the EMAP 11-transect multihabitat method upon which it is based). Unlike NAWQA, we do not use a targeted-habitat approach precisely because it can be more subjective and less likely to be representative of any absolute-value response variable, such as biomass. Because of the objective nature of how sampling points are "chosen" in the CA sampling protocol, we are not highly concerned about inter-operator errors. We have acknowledged in the report that macroalgal patchiness can create a great degree of spatial heterogeneity in the stream, and for this reason, we have incorporated the 105-point, point-intercept method for estimating macroalgal percent cover (PCT_MAP), in order to complement the 11-subsample biomass samples use for chlorophyll *a* and AFDM analysis (which yields data with suboptimal levels of precision, as we have noted multiple times throughout the report). |
| 67 | A further issue that is correctly identified in the Discussion (section 2.4) is the limitation of using one-off sampling. Major temporal variability in algal abundance in streams is well documented in the literature, as well as asynchronous seasonality amongst stream types. Biomass can vary by orders of magnitude over periods of several months. A minimum of monthly sampling for at least a year is required to get a reasonable estimate of maximum and mean biomass for any given site, and thus a true indication of the biological response to watershed scale nutrient stressors. Nutrient leaching and transport processes, as an output of landscape processes, varies over weekly to seasonal to annual time scales so it is important to characterize the biological response over similar periods to understand the true nature of the instream effects. The algal – BMI relationships can occur over different time scales as well, which point sampling will not correctly typify (ie, 'duration' of any given high algal biomass event is very important in determining BMI effects). | We have acknowledged in the text that one-time sampling is inadequate to estimate average or maxumum seasonal values of biomass at any one location. However, this does not preclude substituting space-for-time to evaluate the response of macroinvertebrate communities to variations in benthic algal biomass. Several states (summarized in Evans-White et al. 2009, 2014 and in the Chapter 3 literature review) are using bioassessment data to set nutrient criteria protective of aquatic life uses, which are typically sampled once seasonally during a fixed window. We have also added recommendations for increased temporal sampling of algal biomass in more focused studies. |
| 68 | As a result of the two major points above, I question the ability of these datasets and analyses to provide authoritative guidance on wadeable stream algae and nutrient criteria, and fulfil the Objectives of the study. Given the huge potential impact on economic uses of the streams, and of the needs of the instream communities if insufficient protection is given, I suggest that a great deal of caution should be used in applying results from analysis of these pooled datasets. A study that is specifically designed to meet the objectives of developing criteria is really required here. | See above. |
| 69 | **Chapter 3** | |
| 70 | Overall, I am very impressed by this chapter. I think that the authors should be encouraged to submit a streamlined version of this chapter in the peer-reviewed literature. In particular the application of multiple methods to determine thresholds on multiple indicators is really great, while it is not completely innovative, it really represents the state of the environmental science and pulls together a number of disparate approaches. I enjoyed this chapter. | No response needed. |
| 71 | It might be important in the introduction to state that increased nutrients can also lower biodiversity of indicator organisms and depress biological integrity. This is not necessarily related to algal biomass, as enriched detrital pools may have altered food value (work by Wang et al (2007) and by Evans-White et al, 2009), again the point of a more holistic view of stream eutrophication. It gets mentioned eventually, but more up front would be better. | A conceptual model has been added that includes multiple pathways for nutrient effects. |
| 72 | 1. Overall | |
| 73 | 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). | We have added some clarification of this in the introduction of Ch. 3. |
| 74 | 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. | A large number of BMI and algal IBI metrics were evaluated in order to better understand the complexity of the community response to the chosen stressor gradients. We have revised the introduction and methods of chapter 3 to: 1) provide a conceptual model of nutrients, stream co-factors, ecological response and link to ecosystem services, 2) highlight in the literature review the need for studies that characterize thresholds in the response surface of different, and 3) provide a more description of the AL indicators used and the rationale for their approach. |

| | | |
|---|---|---|
| 75 | 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. | In response to this comment, we emphasize that this study is focused on research; the translation of this and other research will occur in a synthesis document that the State Water Board has contracted SCCWRP to conduct for that intended purpose. We revised the introduction to chapter 3 to include a conceptual model to explain better the rationale for looking at response indicators rather than nutrients per se for condition assessment. |
| 76 | **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?** | |
| 77 | This is a fairly complex chapter and someone that had a basic training in science and statistics could probably get it, but it the chapter is not for the lay person. A section that defined key terms might be useful (say an inset box in the introduction that defines terms such as ALU, BMI, and WQO in addition to the table of acronyms at the beginning). I cannot see how one would write this chapter so someone completely unfamiliar with the methods would get it, but the authors have done as good of a job as could be expected. | There is an appendix which defines many of the important concepts. We have added links from those terms to the Appendix rather than increase the length of the document with text boxes. |
| 78 | Section 3.2.2 does not demonstrate an understanding that nutrient enrichment can alter heterotrophic state, and nutrient effects in streams are not always linked directly to increases in algal production. Thinking in terms of autotrophic and heterotrophic state can clarify this issue (e.g., Dodds 2007). The document eventually makes this point (page 80) but up to then the effort seems to almost entirely assume the nutrient-algae-invert link as cause and effect. | A conceptual model has been added that includes multiple pathways for nutrient effects. |
| 79 | Additional types of threshold other than the example diagrams can exist that indicate the areas where systems move out of a manageable state (Dodds et al. 2010). An important one in this case is the level of nutrients above which chlorophyll no longer increases as a function of nutrients, and this represents a level of nutrients above which nutrient control is pointless unless the nutrient can be brought way below those levels. This threshold plays into the models of Chapter 4, but does not pop up till there. | The example noted by the reviewer represents a special case of "exhaustion thresholds" which are discussed in general terms in Chapter 3. However, none of the final BCART models discussed in Chapter 4 included a significant second-order term for TN (i.e., TN2) or for TP (TP2) which would indicate a saturation effect. |
| 80 | It is good that this document describes 1) multiple response variables to eutrophication and 2) multiple methods to determine thresholds. Both of these approaches, while perhaps adding some increased uncertainty over specific numbers, avoid reliance on a single approach when each approach can give different results because they are applied to naturally variable environmental data. They will give the regulated community ammunition to criticize, but they represent the scientific uncertainty in a broader sense. Individual methods applied separately can give relatively narrow confidence bands for threshold points, but the approach taken here is more honest. | No response needed. |
| 81 | Table 3.4 gives a really nice summary as does the following text. | No response needed. |
| 82 | The use of the arcsine transformation is falling out of favor for some (Warton and Hui 2010). | Warton and Hui make arguments against the use of arcsine transformations in two cases: 1) cases in which the response is binomial and the statistical approach of logistic regressions is appropriate (not the case here), and 2) cases in which a proportion is being predicted (nonbinomial response). In the latter case, Warton and Hu argue that there is no a priori reason to choose an arcsine transformation over any other transformation to stabilize the variance. They argue that use of an arcsine transformation yields results (e.g., slopes) that are "difficult to interpret" and that sometimes yields unrealistic (e.g., negative) prediction values. They argue instead for the use of a logit-transformation, suggesting that a slope coefficient related to the change in the "odds of the response" (e.g., log10(loss in relative proportion of taxa X divided by remaining proportion of community)) is intuitive. What is important for our parametric analyses is that the residuals are normally distributed (by whatever transformation). In this particular study we are not focusing on the interpretation of the value of slopes. In this study, we have found no evidence of unrealistic prediction values because use of the arcsine **square root** transformation (like the logit transformation) prevents the prediction of negative values. |
| 83 | Table 3.7 and throughout the document PO4 should be soluble reactive P (or DRP) since the assays don't exactly give phosphate. | Fixed |
| 84 | 1st Paragraph at 3.3.4. This is a very excellent point that is often misunderstood by managers. | No response needed. |
| 85 | Table 3.11 could include values from Dodds et al. 2006. | We did not include algal biomass thresholds in this table. |
| 86 | Yes | No response needed. |

| | | |
|---|---|---|
| 87 | 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. | We provided a justification for using a variety of methods: "Analytical techniques differ in terms of whether and how confounding factors such as other sources of stress can be taken into account, and also differ in their susceptibility to outliers. As such, we used a variety of techniques to attempt to mitigate these challenges and seek consensus in results among different techniques (Dodds et al. 2010; Smucker et al. 2013a,b)." |
| 88 | Page 27 – subheading 3.2.3 should be 3.2.2. | No, correct. |
| 89 | The methods are described really well. I would recommend that results be described more thoroughly. Often there seems to be a disconnect between the results and highlights developed in the discussion. Results should be more thorough reported in the text and linked to referenced tables and figures. Currently substantial amounts of the results that are important were not reviewed well in the text. | Some additional detail has been added. |
| 90 | Yes. The associated citations give more detail, which is appropriate. | No response needed. |
| 91 | **Are the methods used to estimate response thresholds scientifically valid? Have statistical assumptions been adequately tested?** | |
| 92 | Yes, the application of multiple methods makes this a very strong analysis | No response needed. |
| 93 | Yes | No response needed. |
| 94 | The methods used to estimate response thresholds are scientifically valid and their statistical assumptions have been reasonably adequately tested in the literature. | No response needed. |
| 95 | Yes, the methods used to estimate response thresholds scientifically valid, and assumptions for analyses have probably been met in most cases. However, I prefer transforming independent variables to get an even distribution to provide equal weight to values along the stressor gradient,rather than the common right skew with most data in the lowest levels of the stressor gradient. | Although not specified, we assume that this comment refers to the piecewise regression analyses. There is no requirement that the independent variable be normal/uniformally distributed. If there are clear outliers that are highly influential, then there could be a problem (as suggested), but the mere fact of an even distribution is not a requirement for the statistics. Furthermore, the bootstraping should have mitigated the potential effects of any true outliers. |
| 96 | The piecewise linear regression results show the problem with the highly skewed results, and potential problems with loosing pattern recognition of assimilative capacity at really low nutrient concentrations and levels of productivity. | We have run a subset of the piecewise regressions (i.e., those corresponding to Figs. 3.16 through 3.20) using the dataset only up to the threshold in the first draft of the report (which is based on the full data set) in order to bore down and see if an additional (i.e., resistance) threshold can be discerned, thus suggesting the point at which assimilative capacity is reached. We found that for 4 of the 5 relationships, the piecewise regression actually revealed the signature of an exhaustion threshold (again), and for the 5th, while a resistance-style threshold was indicated, it occurred at nearly 0 along the gradient, and as such may not be very meaningful. **The attached PowerPoint file shows analysis output.** We have added a note in the report that ecologically meaningful resistance thresholds may not always exist (or may be so low as to be undectable with available methods/data.) |
| 97 | The use of sample weights in analysis of stressor-response relationship should be explained. I can see how weights help in stressor-response analyses to make them more representative of the state. But, aren't weights related to stratification variables which are natural classification variables? So when results with weighted and unweighted data agree, then would natural variables used in stratification probably have little effect on stressor-response relationships? When they don't agree, then stratification factors do affect stressor-response relationships and the current analysis that doesn't account for natural factors (e.g. piecewise regression) should e downweighted relative to results from BRT because BRT did account for natural factors in the partial dependence plots. | It is correct that weights are a function of the stratification scheme used in the surveys' designs. However, several factors contributed to the development of stratification across the combined surveys, not only natural gradients (which were imperfectly captured). The discrepancy, when it occurs, between the weighted and unweighted results is due to a few outlier sites that happened to have unusually high weights. We don't believe there is justification (or a defensible, objective way) to downweight certain analyses based on the potential for weights to covary with natural gradients. |
| 98 | | To address the first point, we have added box plots to show the distribution of the ALI response variable among those samples with a gradient value below the threshold vs. those with values above in order to show that beyond the threshold there is limited variability in ALI values relative to below the threshold. To address the second point, CART analyses already test for the significance of a cut-point; as we stated in the text we used a cutoff of $p = 0.05$ for significance. We have increased the robustness of those analyses by doing a bootstrap analysis to show how tight the thresholds are. CART will detect points along a gradient at which a difference in values can be detected below the threshold (e.g., "reference condition") and above the threshold ("altered condition"). In this sense, it can be considered to be finding the lowest detectable effect level. It does not (nor should it) substitute for value judgements by stakeholders. |
| 99 | | ---------------------------------------------------------------------------------------------------------------------- |
| 100 | | ---------------------------------------------------------------------------------------------------------------------- |
| 101 | | ---------------------------------------------------------------------------------------------------------------------- |
| 102 | | ---------------------------------------------------------------------------------------------------------------------- |
| 103 | | ---------------------------------------------------------------------------------------------------------------------- |
| 104 | Thresholds with assimilative capacity focus stakeholder consensus on a specific level of stressor that would protect ecosystem services, above which great losses would be expected. Linear responses do not provide that focus on a stressor level that provides stakeholder consensus. | No response needed. |

| | | |
|---|---|---|
| 105 | Some responses of biological condition to productivity and nutrient gradients did illustrate assimilative capacity, such as in Figure 3.4 from the report. | The few cases of documented resistance thresholds were noted in the text. |
| 106 | Positively and negatively asymptotic responses have similar problems as linear responses with respect to policy because there is not a clear stressor level with a tipping point before the attribute is essentially lost or severely degraded. Figure 3.14 from the report provides an example of such a response illustrated with SiZer and piecewise regression. The delineated threshold is too high to protect the illustrated attribute. Yet these thresholds are used in the results to guide policy with no apparent modification. Some states have used the stressor level at the mid-point of the relatively linear decrease as a management target to protect at least half of the condition that would be lost. | We disagree that the results of the thresholds are being used to direct policy without any modification. We are clear in the chapter 1 (introduction) that the study is focused on research and not intended to be construed as recommendations for policy. We have further highlighted this language. We have also provided a conceptual model and rationale in chapter 3 explain that we are interested in evidence of breakpoints across the response surface. We have added emphaiss in the text and additional detail in Appendices to reflect the fact that most thresholds identified are exhaustion thresholds rather than resistance thresholds. Additional analyses are planned that will use methods such as those suggested by this and other reviewers to identify regions along the response surface that are "protective" of aquatic life. |
| 107 | Another problem with the threshold analysis is that statistic analyses are not well designed to isolate the lower bound of threshold response ranges. For example, the results highlight the upper bounds of the stressor range where the attributes respond most sensitively: "Furthermore, the scatterplots show that the most pronounced relationship between the diatom community (NMS axis 1), and TP occurs <0.1 mg/L, whereas for TN, the most pronounced relationship occurs <1 mg/L. These observations are corroborated by the results of the CART analyses of diatom and BMI NMS axis 1 scores (Table 3.5, Figure 3.5), in which median cut point values for TP and TN were consistently <0.1 and <1 mg/L, respectively. All median cut points for chlorophyll a were <31 mg/m$_2$, and for AFDM were <42 g/m$_2$." | We agree that we should have pointed out both the lower and upper thresholds evident in the NMS scatterplots, and have added mention of this in the Results. |
| 108 | In Figure 3.4 above, I've marked the upper bounds of the stressor range with assimilative capacity and the lower bounds of the stressor range with greatest response with the vertical dashed arrow. These are not the thresholds highlighted in the results or by the statistical analyses. CART often finds the upper change point, rather than the lower change point in curves with assimilative capacity. I would have highlighted the following. Below 0.01 mg TP/L and 0.1 mg TN/L for nutrients and below 2 mg chl/m2 and 2 g AFDM/m2 for algal biomass, little response was observed in valued attributes (response variables). So these thresholds are the good protection levels because there is assimilative capacity below 0.01 mg TP/L, 0.1 mg TN/L, 2 mg chl/m2 and 2 g AFDM/m2 and unwanted response above those levels. Thus, these are the kinds of responses that develop stakeholder consensus for management targets. The CART and Piecewise Regression analyses are not well designed for threshold analyses of assimilative capacity. | We agree that we should have pointed out both the lower and upper thresholds evident in the NMS scatterplots for nutrients, and will add mention of them in the Results. Data density below what the reviewer is suggesting is "assimilative capacity" thresholds for the biomass variables is so low, however, that we don't find these values to be particularly convincing. Therefore we are not highlighting them in the text, contrary to the reviewer's suggestion. In comparing CART results to perceived resistance and exhaustion thresholds in smoothed NMS response plots, we noted that CART cutpoints fell slightly above the perceived resistance thresholds but well below the exhaustion thresholds. We have added this to the text. |
| 109 | Probably. I don't have a working knowledge of all the methods so I can't comment on this with confidence. However, I do note that while many of these methods appear 'analytically elegant', the most useful and meaningful analyses were the simple bivariate 'dose – response' plots (eg, Figures 3.14 – 3.17). Use of integrator indices is questionable in the current context due to the number of assumptions and the need to show clear cause – effect stressor relationships to water managers and the public. This type of work needs to focus on such approaches and then become more sophisticated analytically at a later stage, but only if this is required to clarify relationships and process understandings. We should always try to keep it simple when developing recommendations for regulators. | There are no existing bioassessment targets to rely on for CA, in order to pursue the suggested approach. However, when we revise the Bayesian CART analytical approach, we'll end up with classified dose-response plots that will be simpler to visualize, thus addressing concerns raised. |
| 110 | I have a query about the use of TN and TP. How much of these measures are just re-suspended benthic algal cells?? I suspect that there is a significant degree of auto-correlation inherent in these ALU vs total nutrient results. | The index period during which the samples were collected throughout the state represents largely warm-season base flow. The streams that were sampled are all wadeable, as per the requirements of the monitoring programs, and in general tend to be shallow and dominated by benthic, as opposed to planktonic, algal forms. There is a limited amount of data available for dissolved (in addition to whole-water) TN and TP, but among sites with both fractions analyzed, when the two fractions are compared via linear regression, they are found to be very closely related, with slopes close to 1 for both the nitrogen and phosphorus versions of the analysis. Thus, we are quite confident that the possibilty of auto-correlation to which the reviewer refers is not applicable due to a substantial amount of stream nutrients being tied up in plankton. In addition, most of the total nutrients are in dissolved form. |
| 111 | **Are the conclusions of this chapter adequately supported by the analyses and results?** | |
| 112 | Yes it does. | No response needed. |

| | | |
|---|---|---|
| 113 | 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. | The introduction to the Results refers readers to the final section that provide an overall summary of the thresholds that were determined across all the analyses, which should address part of the reviewer's comment. As regards the other part of the comment, i.e., what each of the analyses are for, we have now added a one-sentence summary of what each analyses does at the beginning of each subsection in the Results, and moved the main points of the results for each subsection to the first paragraph in that subsection. |
| 114 | Some specific comments: | |
| 115 | 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. | It is true that the NMS axis may be one of the less sensitive indicators tested as it is not constructed to be responsive to stressor gradients. However, if the NMS scores do respond to nutrient or algal biomass gradients, this is more ecologically significant as it suggests that the whole community is responding and that much of the regional variation in community composition can be explained by a nutrient gradient. NMS scores were successfully used to demonstrate community-wide shifts in composition in response to eutrophication in the Everglades. |
| 116 | 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. | We now discuss both thresholds in those figures, for all panels in which two thresholds are clearly apparent. |
| 117 | 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. | We have added clarification to the figure legend that the "gradients" along the y-axes refer to the biomass or nutrient variables labeled at the top of each panel in the figure. That should ameliorate the problem referenced in the comment. |
| 118 | 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. | We have added clarification to the figure legend that the "gradients" along the y-axes refer to the biomass or nutrient variables labeled at the top of each panel in the figure. That should ameliorate the problem referenced in the comment. |
| 119 | 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. | We have added arrows as requested |
| 120 | The analyses of results do not distinguish whether thresholds mark the upper bounds or lower bounds of a response range and whether the response is below the threshold is okay or bad. See above. | See response above. In short, we have added box plots for a subset of the relationships to show the distribution of the ALI response variables among those samples with a gradient value below the threshold vs. those with values above in order to show that beyond the threshold there is generally limited variability in ALI values relative to below the threshold. In addition, we have added (to new Appendix 3D) the interquartile ranges of the distributions of ALI values for sites with gradient values below vs. above the threshold detected, for each analysis. |
| 121 | They appear to have been, or the tests are robust to non-conformities (eg, to requirement for normal distributions) Yes. The conclusions are appropriately very cautious. | No response needed. |
| 122 | **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?** | |
| 123 | Some of the logical flow could be improved, as there were places (noted above) where the authors eventually got to an important point, but if they set things up better in the beginning, then they results and their discussion would follow better. However, the specific syntheses of analyses is good. | Summaries of important results have been moved to the beginning of each result subsection to highlight these. |
| 124 | This objective to employ and compare different methods to provide a weight of evidence approach is a daunting task. I have placed some suggestions for improvement within the original document. | No response needed. |
| 125 | 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. | We have removed "ordered" (it was an artifect from the coding of data for the analyses) |
| 126 | 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. | We have provided in tabular form (in a new Appendix, 3D), the values from Fig. 3.26 |
| 127 | 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. | We state in both the introduction to the report (chapter 1) and in chapter 3 that the management context for this study is that the State of California is interested in establishing nutrient criteria based on: 1) assessments of condition based on ecological response indicators and 2) use of models to establish watershed or site-scale nutrient targets. Our research aimed to understand what is the nature of the biological condition gradient with respect to algal abundance and nutrients (as stressors) and aquatic life indicators (response). The decision of how to establish nutrient criteria is a policy decision that will be made by the State Water Board. This study is focused on science. We have reempahized these points in the introduction and in chapter 3. |

| | | |
|---|---|---|
| 128 | Medians of thresholds in responses of ecosystem services should not be used to guide environmental policy. Policy should protect known elements of the ecosystem services of water resources. In the Everglades, some researchers suggested a similar approach and it was rejected because sensitive changes in the Everglades occurred at stressor levels below median response thresholds. Part of this rationale could be related to the idea that we are just using indicators of change in systems, not changes that we really care about. However, the focus of this report is ecosystem services and many of the responses to nutrients can be related to ecosystems services. | The executive summary and introduction clearly state that the intended outcome of this project is research, but that the results may be used as a line of evidence in the SWRCB's policy decisions on nutrient criteria. Additional analyses and scientific syntheses are underway to specifically support the SWRCB in this endevour (M. Sutula, personal communication). Minor modifications to the text in the executive summary and introduction have been made to further clarify this point. |
| 129 | We should better understand the progressive degradation in water resources with increasing stressor levels and multi-stressor effects, and then set management targets to provide appropriate levels of protection given regional extents of landscape alteration and tiered uses that protect ecosystems services within a region and not necessarily at all sites. | This reviewer suggestion is appropriate, but again, we are focused on science, not on policy. Additional technical products will be produced to support policy decisions on nutrient criteria in California. This report is but one piece. |
| 130 | Figure 3.25 does a good job of synthesizing results, but many of these thresholds do not account for the distinction of whether thresholds mark the lower or upper stressor bounds of the sensitive responses and whether they would protect ecosystem services (e.g. the mid-point of linear range of non-linear asymptotic responses). Update the figure with management targets based on statistically derived breakpoints in the data versus just the statistically derived breakpoints in the data. This additional correction is important for application in policy development. | Biocriteria objectives for macroinvertebrates and periphyton have not yet been established for the state of California. The choice of management objectives in lieu of pre-specified targets is beyond the scope of the current report, which was not designed to make policy decisions. This report is one line of evidence being used to support the development of nutrient objectives for CA. Additional analyses are planned once management targets have been finalized. |
| 131 | Yes. However, I have significant concerns about the dataset, as described earlier. I suggest that the relationships (eg, thresholds) for algal biomass and ALU condition metrics are likely to be more robust than the nutrient – algal biomass relationships due to the stronger influence of temporal factors on both nutrients and algal biomass. | Limitations of the dataset have been discussed in the text. |
| 132 | The dataset should probably have been partitioned to remove obvious non-relevant sites (eg, those from urban areas where local issues occur which are specific to those sites such as sediments, toxic pollutants and concrete lining of channels). | Censoring the dataset to remove urban areas would be at cross-purposes to the analysis. Urban areas, while admittedly posing a number of potentially confounding, non-nutrient-related stressors, nonetheless also can contribute significant nutrient loads and thus must be kept in the equation. Furthermore, eliminating such sites would likely unnaturally truncate the gradient of nutrient concentrations to which streams across the state are exposed, and this could result in us failing to detect ecologically relevant thresholds (note that most thresholds detected exceeded the natural background nutrient concentrations measured in "Reference" sites). |
| 133 | **Chapter 4** | |
| 134 | **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?** | |
| 135 | There were some confusing bits in there, as noted below. I wonder if there is some clearer way to explain this? Maybe a flow chart or something instead of the screen shots? It really took me a while to figure this out and there were some parts I am still not clear on. | We think it's important to show a screenshot of the spreadsheet interface where data are entered and output is visualized in order to give the reader an idea of the tool. However, to address the reviewer's concern, we provide some more clarification (e.g., that not all of the input parameters are necessary for running all the models--for example, only the TN/TP inputs are required for estimates based on Dodds' models, whereas the other nutrient types and the environmental data are required for the QUAL2k estimates). |
| 136 | In general I think that ways to smooth very variable data (e.g. seasonal means) may remove much of the variance and lack of predictability in the models. Even the nice total nutrient vs chlorophyll plots for lakes are based on summer mean epilimnetic vales. Another approach would be a probability of exceeding some set value. | Unfortunately we do not have repeated measures of the relevant parameters at most sites, making it difficult to act on this suggestion. |
| 137 | Yes | No response needed. |
| 138 | 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. | No response needed. |
| 139 | However, it is not clear how the spread output is related to NNE, which needs some explanation. | The spreadsheet is a scoping tool to help users determine, given nutrient concentrations and other environmental co-factors at play at a given site, whether the biomass growth potential is likely to lead to an exceedence, based on the proposed NNE endpoints (i.e., 100, 150, or 200 mg/m2 chlorophyll a, depending upon the beneficial use designation (WARM vs. COLD) of the stream and whether the BURC I/II or II/III threshold is being evaluated). We have added some language to make sure this is clear in the text. |

| | | |
|---|---|---|
| 140 | 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? | More text has been added for QUAL 2K. Equation 5 has been changed for clarity. **Excerpt from report**---<br>Versions of the River and Stream Water Quality Model (QUAL2K) in the BBST are a parametric representation of the benthic algal component of the mechanistic steady state model developed by (Chapra and Pelletier, (2003). This simple parametric representation was adapted to provide initial estimates of benthic algal responses to availability of light and nutrients, and can be adjusted to achieve general agreement with the empirical relationships developed by Dodds et al. (1997, 2002).<br>The model calculates the steady state algal growth as<br>$B=(K_{pmax} . \phi_{Nb} . \phi_{Lb})/(K_{rb}+K_{db})$ Eq (5)<br>Where, Kpmax is the maximum photosynthetic rate at a reference temperature of 20°C, (ϕNb ) is the benthic algae nutrient attenuation factor represented by the Michaelis-Menten nutrient limitation equation for inorganic nitrogen and phosphorus. ΦLb light limitation factor with a benthic algae light parameter, Krb is the temperature-dependent bottom algae respiration rate, and kdb is the temperature-dependent bottom algae respiration rate. The prediction of biomass uses only the sum of respiration and death, and the model is unable to distinguish the processes independently. Equations to estimate individual components of equation 5 are not provided in this report, but can be found in the Tetratech 2006 report. QUAL2K used default parameter values provided in the model, however the revised version uses modified parameter values which were derived by fitting the QUAL2k performance to the Dodds 2002 performance for the Ecoregion 6 dataset. |
| 141 | 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. | The text is now on page 133 not 95. Text has been modified to show that the user needs to provide ammonia, nitrite, nitrate as N, total Kjeldahl Nitrogen, phosphate as P, and total phosphorus for nutrient information. |
| 142 | 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? | The conversion factor between AFDW and chlorophyll a assumed in TetraTech 2006 has been added to the text. |
| 143 | I understood its use. I'm not sure what the "potential" strengths and weakness were, but the strengths and weaknesses before and after this report are evident and clear. | No response needed. |
| 144 | Yes. | No response needed. |
| 145 | **Are the methods used to evaluate the performance of the NNE benthic biomass spreadsheet tool scientifically valid? Have statistical assumptions been adequately tested?** | |
| 146 | Note the Dodds et al. 2006 paper presents corrected values for the 2002 paper. | We have pointed this out in the text. |
| 147 | I am curious why the Dodds model screen shots have canopy and light components as those are not presented in the original equations? How was the Biggs model included when it is based on inorganic nutrients but the Dodds model based on total nutrients? OK I see the later explanations, maybe the screen shots would be better after the detailed description? | We think it's important to show a screenshot of the spreadsheet interface where data are entered and output is visualized in order to give the reader an idea of the tool. However, to address the reviewer's concern, we provide some more clarification (e.g., that not all of the input parameters are necessary for running all the models--for example, only the TN/TP inputs are required for estimates based on Dodds' models, whereas the other nutrient types and the environmental data are required for the QUAL2k estimates). We have also moved the screen shots relative to the text, as suggested. |
| 148 | How do nodes in figure 4.13 correspond to USEPA proposed nutrient ecoregions, by guess is that there is something here to start delineating ecoregions even though earlier it was stated there are not enough data to do so. | The nodes do not correspond to existing ecoregions but suggest that we may be able to derive "empirical" nutrient ecoregions based on sensitivity of response to nutrients. As stated in the text, this approach is being refined in future analyses. |
| 149 | Could the heavy influence of ammonium in the models be related to the ion indicating anoxic conditions? It seems likely to be an indirect indicator, even if it is a nutrient source. However, ammonium is the most available inorganic nutrient. | Unfortunately, dissolved oxygen data are not available to test this hypothesis. |
| 150 | Page 147. "Inclusion of explanatory variables that are integrative over time and space" is a key point as well as the idea that periphyton biomass is intrinsically variable. This is exactly why the seasonal means used by Dodds et al. yielded better correlations with nutrients than the spottier USG measurements. A central point in all this is that there is lots of natural variance, and control of nutrients lowers the probability that algal biomass will exceed a certain amount. | No response needed. |
| 151 | Have the authors considered time-lagged models for P? This is what Lohman and Priscu (1992). found for the Clark Fork in Montana where luxury consumption of P in spring alleviated P limitation in the summer. | Unfortunately we do not have time series of nutrient concentrations. An analysis of CA NAWQA time series suggest that nutrient levels typically peak in January to February, which would be quite a lag even if luxury consumption of P is occurring given the turnover rates for algae. NAWQA plots (now in Appendix 5) show that TN and TP levels are actually increasing over the course of the growing season. NAWQA data also show that the growing season averages are correlated with annual peak values. |
| 152 | Page 149. How do you know that filamentous algae are responsible for higher biomass accrual? They are more obvious to the eye, but often the chlorophyll is the same for them and for diatoms. | We did not mean to imply that, volume for volume, chlorophyll *a* is inherently higher for soft algae than for diatoms. We simply meant that soft algae are more likely to proliferate to nuisance conditions than diatoms, as measured by chlorophyll a. We have clarified this in the text. |

| | | |
|---|---|---|
| 153 | Yes. I have highlighted some areas for clarification. | Clarifications added in text where indicated. |
| 154 | The BRT and B-CART are well documented statistical methods for establishing predictive models. Their uses here are appropriate. | No response needed. |
| 155 | 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. | The BRT and Bayesian CART techniques were used to determine which factors were related to unexplained variation in the spreadsheet models (BBST) and to explore whether classification could help to improve the fit of nutrient-biomass models. We have added text to clarify the objectives of the different analyses in Chapter 4. |
| 156 | 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. | new text has been added to explain the bias analysis better. **Excerpt**---Bias is the deviation of predicted values from the observed values (chlorophyll a for this study) resulting from usage of poor explanatory variables in the model or just incorrect choice of models. In the bias analysis process we try to examine the impact of a single or a group of explanatory variables on the prediction abilities of a given model. |
| 157 | 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. | See p. 118: "Boosted regression trees allow nonlinear relationships and variable interactions to be represented in model predictions. Bayesian CART analyses provide a simplified set of regression models to predict algal biomass by site class, along with a set of classification rules to define groups." |
| 158 | The methods used to evaluate the performance of the NNE benthic biomass spreadsheet tool are scientifically valid. There was little discussion of statistical assumptions for tests, nor transformation of variables and subsampling data to meet assumptions, or qualifying interpretation of results based on lack of assumptions being met. | Additional detail has been added to Chapter 4 to cover tests of statistical assumptions. Most of the methods applied were nonparametric, with limited assumptions. The following text was added to explain violation of assumptions for regression comparisons of observed versus model results: "Results of linear regressions are shown to illustrate the poor match between observed values and model predictions. However, because model fits were so poor, model assumptions for linear regressions were not met even after multiple standard transformations (log, square root, inverse, power) and higher order equations were applied. In general, both variance and residuals tended to decrease with the mean, and residuals were not normally distributed." |
| 159 | Were response variables transformed to reduce skewness in the data and overemphasis on high nutrient and biomass levels? Although I have heard it argued that CART and BRT like analyses are robust to assumptions of normality in variables, given the importance of variance analyses, I'm not sure how that could be the case. Some simple runs of analyses using transformed and untransformed variables would provide a valuable evaluation of this potential problem if variables were not transformed. | We re-ran two test cases for BRT with the input data transformed as necessary and found that the results were nearly identical to those in the report (which had used raw data). We feel this adequately confirms that the way we conducted the analyses in the report is acceptable. We have made a short note in the report that we conducted a sensitivity analysis confirming no need to transform. **The attached PowerPoint file shows analysis output.** |
| 160 | The Bayesian Cart models included factors that could be affected by human activities (e.g. $NH_4$ and conductivity), and even some measures of human activities (e.g. urban5K). Most classification of ecological data to develop relationships between ecosystem services (endpoints) and stressors only includes naturally varying factors that humans affect relatively little, like precipitation, temperature, geology, soils, ecoregion….. This confounding of the types of predictor variables is probably really only in assessment, whereas here the BBST tool is really to set site-specific nutrient management targets. So distinction between these variable types is probably not an issue. | Classification for application of nutrient criteria is usually done to explain background variation in reference condition. In the case of the Bayesian CART analyses, we are trying to explain differences in sensitivity of nutrient-response relationships that could be affected by either natural or anthropogenic factors. |
| 161 | There are many ways that this analysis could be done. The approach does appear to be valid. | No response needed. |
| 163 | Yes, the conclusions are accurately supported. They authors of this document are careful to point out the limitations of the modeling approaches employed here, and there application to management. | No response needed. |
| 164 | Yes. I appreciate the way the authors specifically stated limitations of the analyses. | No response needed. |
| 165 | The conclusions are generally adequately supported by the analyses and results. | No response needed. |
| 166 | Table 4.4 is unclear. Are the 2 columns on the left number of sites? | Yes. We have clarified this in the text. |
| 167 | The 1st sentence below Table 4.4 is incomplete. | We have fixed this. |
| 168 | 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. | R2 values have been added for Dodds 97, and 2002. They are not available for QUAL 2K models. |
| 169 | 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. | The report does report the range of variance for the predictor-observed as explained by all predictors (0-53%, see excerpt below). The ranking is additional to see if some predictors are more important than the rest.----(from the report)-----The random forest regression ranks the explanatory variables that account for the variance and bias. These explanatory variables can be divided into three major categories: 1) water chemistry variables, 2) other site-specific parameters, such as physical habitat and 3) land use (Table 4.7). Recurring key explanatory variables from these categories are observed for both chlorophyll a and AFDM. The variance explained by the models ranges from 0-53%, signaling that model fit could be improved if refinements are made for how these variables are currently used in the models. |
| 170 | Yes. The models show promise but they need more work and simplification if possible. | Future work is planned to try to simplify models. This is beyond the scope of the original project. |
| 171 | Yes. | No response needed. |

| | | |
|---|---|---|
| 172 | **Does the analysis of residuals for model predictions presented in this chapter help to guide future improvements in these models?** | |
| 173 | Yes, it should help. This is actually sort of related to another approach that has been taken for some statistical models using ANCOVA to separate out categories such as nutrient ecoregions, and improve models within regions or allow combinations of some regions to increase the amount of data. One of the key refinements and issues that has come up all along in these suggestions for nutrient criteria is how long of an average do you need, and how should you gauge what is acceptable deviation from the target nutrient amounts. Unfortunately, the natural variance of streams may preclude models with a high degree of predictive ability. | We agree that these are important issues to address in future work. |
| 174 | Yes | No response needed. |
| 175 | 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. | Additional variables were not included in the spreadsheet models because it was determined that a "one-size-fits-all" approach is probably not useful for an area with as much variation in environmental conditions as California. |
| 176 | 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. | It is appropriate to include those response variables because they were being tested as potential biomass indicators (in addition to, or potentially in lieu of, chlorophyll *a* ) of nutrient impacts to streams. It's true that those variables are not currently included in the NNE framework or the BBST spreadsheet model, but the chapter wasn't exclusively about validating the BBST; it was also about exploring the potential for improved models for application in California. |
| 177 | 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. | The spreadsheet is a scoping tool to help users determine, given nutrient concentrations and other environmental co-factors at play at a given site, whether the biomass growth potential is likely to lead to an exceedence, based on the proposed NNE endpoints (i.e., 100, 150, or 200 mg/m2 chlorophyll a, depending upon the beneficial use designation (WARM vs. COLD) of the stream and whether the BURC I/II or II/III threshold is being evaluated). We have added some language to make sure this is clear in the text. |
| 179 | This could be strengthened by, for example, a better portioning of the dataset to allow for the different environmental domains the sites came from. | We conducted some analyses (BRT) where we did in fact include other factors reflecting different environmental domains. |
| 180 | One issue that has not been canvassed very well is the strength and validity of the original BBST models used for this analysis. For example, the datasets used for the Dodds models encompass point values (incl. seasonal maxima) through to mean monthly values from year-long sampling campaigns. The sampling protocols are all completely different in his datasets and so are the analytical methods (eg, different extraction methods can result in 2 – 3 fold differences in chlorophyll concentrations on the same samples!). | Text has been added to note sources of variation in Dodds model results. |
| 181 | Given the issues with the Californian data (as appropriately recognized by the authors) and the problems with the BBST models, I suggest that it would not be prudent to attempt to develop nutrient criteria or objectives from the current study. A purpose-designed sampling program that includes a strong temporal component and a full range of water quality and other environmental parameters (particularly flow), is required to develop the understandings and models necessary to achieve the desired management objectives. I suggest that this should be a high priority for the State. | The results from the current study are being considered as only one line of evidence in the discussion of the development of stream nutrient criteria for California. |
| 183 | **Suggested reference additions** | |
| 184 | Bytnerowicz, A. and M. E. Fenn. 1996. Nitrogen deposition in California forests: a review. Environmental Pollution **92**:127-146. | added |
| 185 | Dodds, W. K. 2007. Trophic state, eutrophication and nutrient criteria in streams. Trends in Ecology & Evolution **22**:669-676. | added |
| 186 | Dodds, W. K. and R. M. Oakes. 2004. A technique for establishing reference nutrient concentrations across watersheds affected by humans. Limnology and Oceanography Methods **2**:333-341. | not added |
| 187 | Dodds, W. K., V. H. Smith and K. Lohman. 2006. Nitrogen and phosphorus relationships to benthic algal biomass in temperate streams (Vol 59, pg 865, 2002). Canadian Journal of Fisheries and Aquatic Sciences **63**:1190-1191. | added |
| 188 | Lohman, K. and J. C. Priscu. 1992. Physiological indicators of nutrient deficiency in Cladophora (Chlorophyta) in the Clark Fork of the Columbia River, Montana. J. Phycol. 28:443-448. | added |
| 189 | Smith, R. A., R. B. Alexander, and G. E. Schwarz. 2003. Natural background concentrations of nutrients in streams and rivers of the conterminous United States. Environmental Science and Technology **37**:3039-3047. | added |
| 190 | Warton, D. I. and F. K. C. Hui. 2010. The arcsine is asinine: the analysis of proportions in ecology. Ecology **92**:3-10. | No - see comments above. |
| 191 | Riseng CM, Wiley MJ, Stevenson RJ. 2004. Hydrologic disturbance and nutrient effects on benthic community structure in midwestern US streams: a covariance structure analysis. Journal of the North American Benthological Society 23:309-326. | added |
| 192 | Stevenson, R.J., B.J. Bennett, D.N. Jordan, R.D. French. 2012. Phosphorus regulates stream injury by filamentous green algae, thresholds, DO, and pH. Hydrobiologia 695:25-42. | added |
| 193 | Wootton JT, Parker MS, Power ME. 1996. Effects of disturbance on river food webs. Science 273:1558-1561. | added |