Wadeable Streams Nutrient Objectives

Science Panel Report

Cliff Dahm

Charles Hawkins

Paul Stacey

Ken Reckhow

R. Jan Stevenson

Lester Yuan

Meeting: April 19-20, 2017

The Science Panel met and compiled answers to the charge questions following presentations and discussions at SCCWRP in Costa Mesa, CA. The answers to the charge questions are organized below by question and Science Panel member. We did not attempt to reach consensus on issues in the short time that we had in Costa Mesa or afterward. We provide answers to the charge questions by question and Science Panel member to provide detail about variation in ideas, opinions, and priorities among panel members.

Appropriateness and suggested refinements to the Biostimulatory-Biointegrity Science Plan (Sutula et al., 2017, Appendix1)

1. **The Science Plan focuses on work needed to support policy decisions on a statewide set of "default" ASCI and CSCI assessment endpoints and numeric targets for nutrients et al. causal eutrophication indicators.**
    a. **Are refinements or additional elements to the Science Plan needed to improve scientific support for the Biostimulatory-Biointegrity Project and if so, what refinements?**

Cliff: *My feedback on this document and the presentation associated with the document is generally positive to the conceptual model and the use of the Biological Condition Gradient (BCG) as a communication tool to present the results from the CSCI and ASCI indices. I think that BCG can assist when the time comes to present the results from these studies to the State Water Resources Control Board and the nine regional boards. I support BCG use.*

I have two areas for consideration that could be classified as refinements or additional elements. The first area for consideration comes from my background and expertise in measuring functional attributes of streams and rivers. The two indices (CSCI and ASCI) do a good job at looking at structural elements of these wadeable streams for benthic macroinvertebrates and for benthic algae both as diatoms and soft bodied forms. Both indices integrate a certain period of time. The measurements of the stressors do not. As the document states "when chemical grab *samples are taken, they are really a snapshot of the water at that moment, that can change rapidly, but the algae and BMI live there all the time and provide an integrated snapshot of environmental quality of their habitat." Aren't there more integrative measures of the stressors that would strengthen the x-axis in the correlations with CSCI and ASCI? Nutrient load may not be appropriate but some longer term average or median would be worth investigating for sites where water quality time series are sampled. Another example is dissolve oxygen. A single grab sample provides limited information, but a diurnal survey provides insight into rates of primary production, the likelihood for hypoxia or anoxia at night, and the level of ecosystem respiration. I'd really like more effort on characterizing the stressors as the indices do for the bioassessment tools. "Multiple stressors interacting" is the mantra I use to think about the panoply of factors affecting the two indices.*

The second area for consideration is whether the multiple conditions that are characterized as a snapshot at each site might be looked at as another type of index. Light availability, hydrologic alteration, geomorphological alteration (e.g. canalization or concrete), total nitrogen, total phosphorus, chlorophyll a, ash free dry mass, temperature, pH, dissolved oxygen, and coverage are some of the potential indicators of stressors that get measured at each site. Can some of these variables be utilized to produce a stressor index? How might some stressor indices graph up against CSCI and ASCI?

Paul: The Science Plan presents a plausible approach for providing scientific support for the Project. The conceptual model incorporates relevant nutrient and biostimulatory (NBS) conditions (i.e., stressors) that should be considered in developing endpoints and numeric targets for nutrients under a NNE approach. The relationship between biointegrity indices (BI) and nutrient concentrations can be tenuous and require more analysis because response indicators (RI) reflect the gross entanglement of multiple-stressor effects caused by "pollution" not necessarily the target "pollutant". This is likely the cause for weak statistical relationships between nutrient concentrations and the RIs exhibited in the analyses thus far. Additional analysis may better support the utility of nutrient concentrations for setting viable and managerially effective endpoints under an improved "dose-response" concept. Appropriate emphasis is being placed on an algal indicator (ASCI), which is more responsive to nutrients than other biological community indicators such as macroinvertebrates or fish (e.g., CSCI).

However, even a potentially tighter relationship between nutrients and primary producers can be complicated by the nature of nutrient pollution in stream environments. Dose-responses relationships in the field are complex in and of themselves even without confounding stressors. Nutrients not only cycle in and out of biomass (a primary RI for assessing nutrient impact) but spiral through the system making causal relationships questionable and potentially misleading due to non-representativeness of the data in a statistical approach based on empirical data. Nutrient concentrations may be higher when biomass is lower if light or substrate limitations are not considered, for example. And nutrient concentrations are usually low during peak bloom or productivity conditions when plants scour the nutrients out of the system. Better measures of nutrient loading paired with productivity along with a multimetric RI will help confirm concentration-effect relationships in the field.

Continued exploration of a reference approach to assess biointegrity will add support for setting potential ecosystem health targets. Additional statistical analysis of nutrient concentrations may *provide better benchmarks against watershed stressors, as was done in the current analysis. The presumption that higher concentrations lead to reduced structural and functional integrity is plausible but, as noted above, other stressors may confound the relationships. Evaluation in the context of reference conditions is a traditional approach for single pollutant analyses and the outcome could be reasonably justified as protective and shore up use of a multimetric RI. Some refinements might include stronger emphasis on regional/bioregional attributes, an improved watershed condition index (WCI) applied over a range of watershed stress conditions to support a biocondition gradient (BCG) approach that could be tested with site-specific applications to account for anticipated differences shown in some of the analyses and assess predictive capacity to verify relational understanding. This provides a plausible cause-and-effect relationship between watershed structural and functional losses that perhaps proportionally impact receiving water structural and functional integrity, i.e., healthy watersheds = healthy waters.*

Finally, much of the focus of the science plan has been on developing those relationships to good effect, but the strength of those relationships weakens when translated to a single pollutant, i.e., nutrients. So, refinements to improve this work would be to continue to use the disturbance gradient as the integrated stressor index (e.g., a WCI) that would match the RIs, which are indicators of that integrated stressor effect, not just nutrients. The approach should also be tested in point-source dominated streams separately from streams with nonpoint source/stormwater stressors only. This would provide a better fit for the BCG approach and the tiered-aquatic life use (TALU) management framework, which were developed specifically for integrated stressor effect assessment and management, respectively. It establishes the desired platform and benchmarks for quantifying recovery potential and expectations and setting management targets effective at meeting biointegrity targets, rather than just nutrient pollutant targets while incorporating nutrient stress and direction for nutrient management planning in a TMDL, which separates wasteload allocations from load allocations.

b. What specific refinements or elements would aid in directly addressing stakeholder concerns or issues?

Paul: Stakeholders have expressed strong concerns over how science would be translated into policy and, specifically, how the nutrient numeric objectives would be set and implemented. Their focus emphasizes point source discharges, which are probably more effectively regulated using numeric nutrient criteria. However, they did express support for using the BCG and TALU approaches. As argued above, those approaches do not readily lend themselves to point source (sources exclusive of land-based nonpoint and stormwater sources) NPDES management of nutrients, but do provide a better link between pervasive and diffuse pollution sources from the landscape (primarily from development and agricultural drivers) that stakeholders seek. A bifurcation of the two source categories to facilitate the discussion and allow the science to be more properly, and effectively, applied would provide attentiveness to stakeholder concerns over point source effects. This separation could lead to more targeted and effective point source management through traditional methods for setting limits.

Nonpoint sources and stormwater represent more intractable challenges, but show a wide range of BIs given similar coarse-assessment conditions (e.g., percent urban cover in the watershed).

Developing and refining a WCI that captures the causes of those BI ranges would not only allow better targeting of best management practices (BMPs) effective for nonpoint and stormwater sources, but provide valuable insight into planning development with less impact, and how to more effectively distribute BMPs and conservation (e.g., buffers) in the landscape. An improved WCI does not need to be overly complex, but should quantify land covers that retain structural integrity (i.e., "natural" landscapes) with presumed functional integrity and, probably more importantly, the distribution of natural landscapes in two divisions: core natural areas and fragmented natural areas in the landscape. Location is important for buffering effects of disturbed landscapes, not only at the water's edge, but throughout the landscape. Concentrated disturbance close to the water has a direct and concentrated effect, for example. This would support stakeholder interest in applying BCG and TALU to management efforts.

With the information, plausible relationships between integrated stressor effects and stream BIs can be drawn in the BCG approach. Quite often, these will be impaired systems that lend themselves to a TALU management approach with a rational target of improvement to the predicted recovery potential, not beyond. This produces a more cost-effective and productive approach and can be formalized in a UAA, and is an application supported by stakeholders for ALUS beneficial use improvement in areas where attainment is improbable.

Stakeholders also questioned site-specific, not only in terms of target-setting, which might be best done on a site-specific basis, but as to scale of application and implementation. Use of the watershed BCG assessment is scalable and, thus, quite site-specific in implementation, defined by watershed area relationships and condition. TALU provides the link to site-specific implementation, which puts the management actions close to the target management area, which is likely to yield the highest and most cost-efficient benefits with ALUS is fully attained or not. But, this cannot be applied for nutrient pollution alone, and, in any case, land management practices (BMPs) that reduce nutrient loads are often beneficial for multiple stressors but not necessarily in a predictable manner.

Ken: The statistical models that will be a part of the development of nutrient criteria for wadeable streams provide an estimate of prediction error (which most process models do not). This presents an opportunity for conducting an uncertainty analysis, which is an issue raised by stakeholders. If an estimate of the uncertainty in a nutrient criteria is estimated (based on the selected value of the ASCI/CSCI), how might this be used? The uncertainty estimate for each nutrient criterion might simply be presented, so that decision makers and stakeholders become aware of the level of uncertainty. Alternatively, the uncertainty could be part of a risk analysis. The current EPA approach to water quality standard setting and TMDLs is based on point values for water quality standards, despite the fact that water quality is naturally variable and predictions of attainable water quality levels are uncertain. So natural variability and prediction uncertainty are compatible with a risk analysis and a probabilistic approach to standard setting. This may be done in several ways, with the most basic approach is to set each nutrient criterion as a point value with a predictive distribution for that nutrient based on the statistical model for the selected ASCI/CSCI level. This distribution, combined with the point-valued nutrient criterion, will yield a probability of exceedance of the nutrient criterion.

The BCG is part of the Wadeable Streams Science Plan to assist stakeholders and decision makers understanding of the ASCI and the CSCI. Neither of these is a direct statement of the beneficial uses to be protected. The absence of a quantitative linkage between beneficial uses and nutrient criteria is a concern of both stakeholders and the Water Boards. This could be addressed in subsequent work. One way to quantitatively link the ASCI and the CSCI (and nutrients) to beneficial uses is to conduct a user survey (focused on user perception of beneficial use attainment) at selected wadeable stream sites and simultaneously take measurements for the ASCI and CSCI (and nutrients). These data can then provide the basis for development of a statistical model with users survey results as the dependent variable and the ASCI/CSCI (nutrients) as the predictor variables.

Lester: The application of the landscape constraints work wasn't clear to me. None of the potential uses of the model addressed constrained systems (Slide 23). Why?

Watershed vs. statewide approach - Watershed prioritization approach is described in EPA memo. What added value is sought by state board with regard to state defaults?

Implementation - Distinguish between water column nutrient concentration limits, and loading limits. Current effort should focus on the former, but specifying load limits to achieve desired concentrations is a watershed-specific activity.

Site-specificity - Models that are currently contemplated are inherently site-specific because data characterize conditions at individual sites.

Beneficial use tie in - Consider describing the loss of species (as expressed by O/E) values as a way to describe different thresholds to stakeholders, in addition to BCG narratives.

Chuck: **Communicating and justifying reference condition concepts and thresholds** – From the presentations given, there seems to be confusion regarding how reference site data are used to infer if the aquatic life observed at a waterbody deviates from its desired condition (i.e., designated use). The issue appears to be related to how a statistical distribution of reference site assessment (index) scores is interpreted to inform an assessment. Confusion regarding this issue is likely related to changes over time in how the reference condition has been defined, described, and estimated.

As described in Stoddard et al., (2006)*, the reference condition for any given waterbody should be thought of as dynamic and not static. The reference condition is therefore most accurately described as a 'range of naturally occurring conditions' specific to an assessed site. The value of an index used to estimate the condition of an ecological endpoint of interest is thus expected to vary across years and seasons within some limits (Fig. 1). The idea still applies if reference is not anchored in pristine condition but instead in a more typical least-disturbed condition. **Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson, and R H. Norris. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. Ecological Applications 16:1267-1276.*

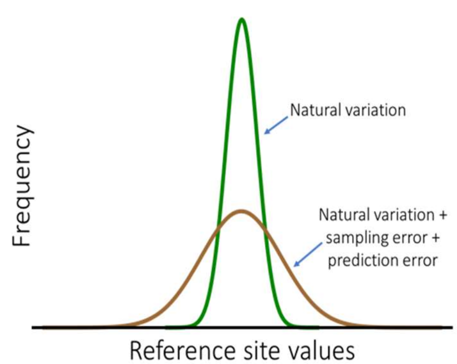

Figure 1. Hypothetical frequency distributions of reference site scores (range of natural variation) for an individual waterbody associated with only naturally occurring forces (green distribution) as well as sampling error and prediction error (brown distribution).

Ideally, an assessment will determine if the condition (e.g., index value) at an assessed site falls outside of the range of naturally occurring conditions expected at that site. In practice, though, assessments were historically based on concepts of classic experimental design in which treatment (assessed) sites are compared with control, or reference, sites. This approach does not directly incorporate estimates of the temporal dynamics of either control or assessed sites in comparisons. Instead, some number of matched control/reference sites are used to test the hypothesis that conditions at the test site are not the same as those at the control sites. Conceptually, this approach is problematic in at least two ways: (1) the statistics are generally designed to test if the mean condition of reference and assessed sites are the same, not if the condition at one test site is outside of the range of control group variability, and (2) it assumes that control/reference sites are true replicates of one another. Interval and equivalency tests can be used to more appropriately test if a single site is outside the range of conditions described by the reference sites, but the second assumption is more difficult to justify. It is neither obvious nor readily intuitive how observed variation among control and assessed sites relates to the Stoddard et al (2006) concept of reference condition.

Early attempts to estimate the distribution of reference site values applicable to an individual waterbody were based on the *regional reference condition approach* (*Hughes, R. M., D. P. Larsen, and J. M. Omernik. 1986. Regional reference sites: a method for assessing stream potentials. Environmental Management 10:629-635*). In this approach, inferences regarding biological condition are based on statistical distributions of biological index values derived from a group of reference sites assumed to be representative of the biota expected at an assessed site. However, the ranges of biological index values observed across reference sites were often large, and thresholds for inferring biological alteration were often set well within the observed range of reference site values (e.g., 10^{th} or even 25^{th} percentile of reference site values). Such thresholds were set ostensibly to balances type I (false positive) and II (false negative) errors of inference. However, the lack of both a rigorous justification for the selected threshold values (which logically implied that 10 or 25% of reference sites would be flagged as being in non-reference condition) and an intuitively clear biological interpretation of index scores were often confusing to scientists and stakeholders alike. Moreover, this approach did not couch the reference condition concept in context of the range of variation expected at individual waterbodies.

Modeling is now often used to adjust for the potentially confounding effects of naturally occurring environmental gradients among references sites on assessments. Such modeling is especially critical when standardized assessments must be made across an environmentally heterogeneous region. Modeling both improves index precision, and may also provide a means of estimating the temporal variation expected at individual waterbodies. The residuals from modeling include both (1) variation associated with incomplete sampling in space and time and (2) model (prediction) error. Sampling variability has two components: (1) variation due to imperfect characterization of the biota on the date of sampling and (2) the fact that a sample taken in a single year provides no information on year-to-year variation in the biota at a site. That combined variation can be used within a traditional control-treatment analysis context, but it also potentially estimates the range of natural variation expected at an individual site over time. This property may aid in explaining the reference condition to stakeholders. If modeling accounts for much of the spatial variation observed across reference sites, the remaining residual variation observed after modeling should be largely associated with unaccounted for temporal variation in biota. In these cases, we assume that a space-for-time substitution accurately characterizes the range of natural variation representative of individual sites. This assumption is more likely to hold if the dynamics of individual reference sites used in the modeling are not synchronous. The interpretation further assumes that the range of natural variation is similar across waterbodies.

Explicitly defining the reference condition as the range of natural variability expected at a site and explaining how that variation is estimated could help avoid confusion regarding how assessment thresholds are estimated and why they are reasonable and appropriate. Regardless of the approach used to estimate the range of expected conditions, it may always be challenging to explain why a specific threshold is appropriate when those thresholds are based on statistical distributions alone. Stakeholders will need to understand the concepts of type I and type II errors of inference and why they need to be balanced, but setting thresholds primarily based on biological reasoning (supported by appropriate statistical tests) rather than purely on statistical distributions could minimize the confusion regarding the appropriateness of the thresholds used in biological assessments.

Interpreting assessments – The current BCG effort seems to be partly designed to address confusion among stakeholders regarding use of statistical distributions of reference site scores to infer biological status and interpret the biological significance of the thresholds selected for use. However, the effort seems misplaced in that it is unlikely to improve understanding of the reference condition as estimated via site-specific modeling and could potentially lead to misleading and inaccurate interpretations that could conflict with the more refined interpretations that can be directly derived from the State's recently developed macroinvertebrate-based stream condition index (CSCI). For every assessed site, the CSCI provides a list of both the specific taxa and metric values expected to occur under reference conditions – and hence what taxa have been lost from or added to the assessed waterbody and how specific metrics have been altered. These specific alterations in taxa composition and metric values provide a direct and much more quantitative means to interpret the specific nature and significance of the biological changes that have occurred at assessed sites without needing to refer to an overly generalized BCG. The generalized, abstract stressor gradient inherent in the BCG cannot capture either the inherent natural variability in reference conditions that exist across the state or the differential sensitivity/vulnerability that different communities can exhibit in response to stress. For example, Hawkins et al. (*Hawkins, C. P., Cao, Y., & Roper, B. 2010. Method of predicting reference condition biota affects the performance and interpretation of ecological indices. Freshwater Biology 55:1066-1085*) showed that communities in different reference-condition streams responded differently to the same simulated stress (Fig. 2) implying that the specific trajectory of community alteration to stress depends on either initial (reference) taxonomic composition or local environmental conditions. Even greater differences among communities are expected with respect to their differential responses to different types and combinations of stressors. This level of understanding and interpretation can be extracted from the CSCI indices but not from the BCG as developed and presented. Furthermore, these site-specific alterations can be communicated in the same types of standard, intuitive biological terms used in the BCG but with much more specificity thus negating the need for the BCG as a stakeholder translation tool.

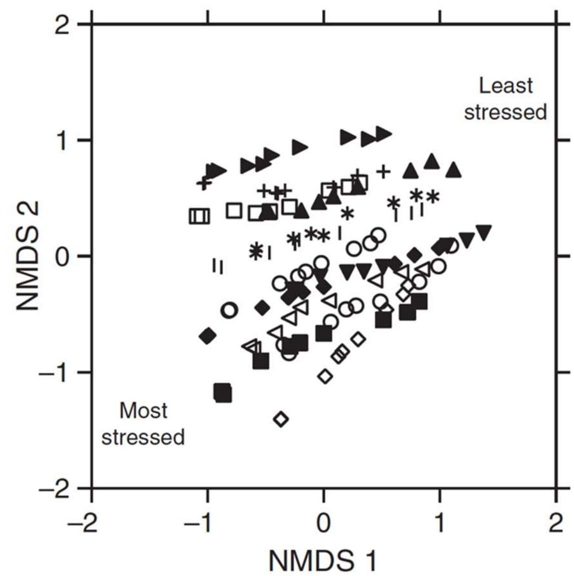

Figure 2. Differential trajectories of changes in community composition for 13 reference sites in response to the same simulated stress. Trajectories are projected in non-metric multidimensional ordination space.

In general, the project could also reduce confusion regarding what reference condition means by casting 'alteration' as departure from the range of naturally occurring biological condition expected at a given site while acknowledging that all such assessments have inherent, but quantifiable, uncertainty. The program should also focus on directly interpreting deviation from expected CSCI scores in terms of expected taxa composition (O/E increasers/decreasers) and metric values rather than overly simplified generalizations based on the BCG. Consider using the experts assembled to produce the BCG to improve our understanding of what type and amount of biological alteration, as assessed with the CSCI, represents an unacceptable change in biology rather than solely relying on statistical distributions to set thresholds.

Stressor identification – The diagnostic potential of the indices and their component metrics is overly optimistic and needs to be acknowledged as likely a long-term research need, not something that can be accomplished over 2-3 years.

Modified channels – The specific modeling objectives and methods need to be better articulated – e.g., are you trying to estimate biological potential of altered landscapes/waterways or something else. Also, don't neglect the value in mapping likely biological condition in helping develop and evaluate management options. Restoring a stream in poor condition when its surrounded by other streams in poor condition is unlikely to be as successful and restoring a stream in poor condition that is surrounded by streams in better condition. These kinds of maps can inform both local and regional management strategies.

Lester: Defining reduced biological objectives for constrained systems may raise regulatory issues (this seems like a policy question rather than one that can be addressed by analysis).

JAN: Develop a more detailed theoretical construct (conceptual model) of what you expect the progressive degradation of beneficial uses of streams and rivers to be along a nutrient gradient, and then relate that to data with specific response levels at specific nutrient concentrations, with specific probabilities. Repeat this general model for different types of streams, maybe by hydrologic classification, which was recently done for California streams. Hydrologic class can affect stream nutrient responses, for example the top-down and bottom-up regulation of assemblages as they differ in KY and MI (e.g. Riseng, C. M., M. J. Wiley, and R. J. Stevenson. 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).

Develop a Living Guidance Document to increase consistency in implementation of numerical nutrient endpoints across the state. The Living part indicates a document that is regularly updated with new information and examples of how the different regional boards implement rules.

I don't like the idea of classifying sites into BCG categories and then setting nutrient criteria/NNEs based on 75th percentiles of frequency distributions of nutrient concentrations at sites for each category. Use models and default numbers and adaptive management for TALU/BCG levels and numeric nutrient endpoints (NNEs). Here's a couple slides illustrating the process from the BCG working group meeting that I proposed (Figure 3). Here, you set NNEs to protect important elements of biological condition, not a multimetric index, and thresholds (tipping points) in responses help develop stakeholder consensus for NNEs because risk of losing valued attributes changes so much at those NNEs.

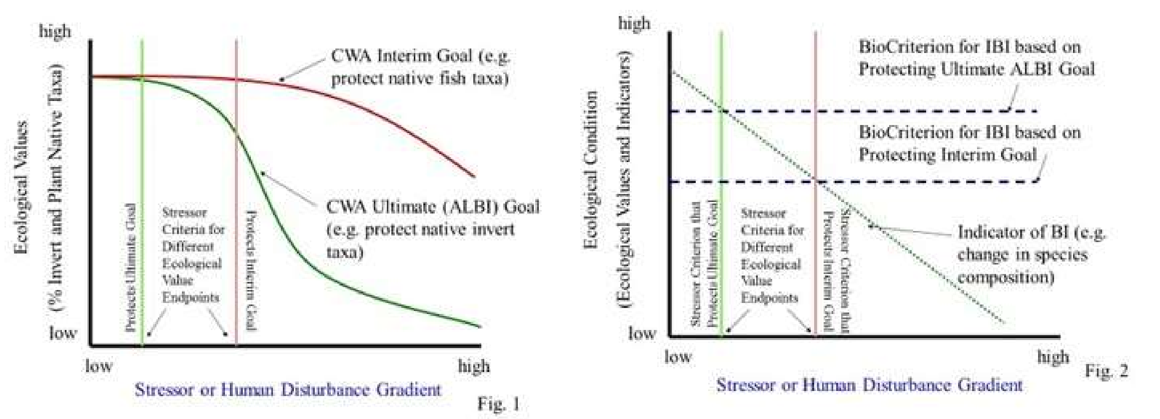

Figure 3. A figure with two slides from a presentation showing the use of non-linear and linear responses to stressors to set stressor management targets (left figure – labeled Fig. 1) and to assess ecological condition (right figure – labeled Fig. 2). The two slides (labeled in Figure 3 as Fig. 1 and Fig. 2) were from a presentation by Stevenson to the BCG workgroup. Non-linear responses help develop stakeholder consensus for levels of stressor that protect management goals (left panel). Linear responses are sensitive along all increments of the stressor gradient, and therefore are best for assessing condition because there are no stressor conditions in which they do not respond to changes.

Especially avoid using reference variability for lower BCG tiers. That seems ill-conceived.

Risk and Uncertainty – NNEs should consider risk of not meeting goals and different risk should be allowed for different beneficial uses – but they should be applied consistently so differences in risk do not add to issues of inconsistency and limit accounting for performance.

Risk should be related to resilience of the beneficial use to nutrient stress, which is likely relative great for streams versus lakes and wetlands, and great for algae and fish, versus invertebrates, and especially mussels.

Develop separate indicators or MMIs for each beneficial use, if possible, e.g. recreation (% macroalgal/macrophyte cover, % toxic cyanobacteria (e.g. appropriate for Lyngbya in FL springs)).

Some of the concerns of stakeholders are beyond the scope of the technical group to address individually, but the tools developed by the technical team, or adaptations of those tools using the collected data, should enable addressing stakeholder concerns.

Appropriateness and suggested refinements to the ASCI Work Plan (Theroux at al., 2016, Appendix 2)

2. **The approach to develop the California algal stream condition index (ASCI; Theroux et al. 2016 ASCI work plan) follows that used for the California Stream Condition Index (CSCI; Mazor et al. 2016) for benthic macroinvertebrates.**

Cliff: I found the approach described in the document to be appropriate. I'll let the algal experts address the question of refinements. My comments on the first panel charge question might be considered for the question of differentiating chemical, physical, habitat, and hydrological impacts on the ASCI. I do support looking at the algal index at a regional scale using the six regions described for the CSCI as an initial cut at a regional analysis. These six regions are North Coast, Central Valley, Chaparral, South Coast, Sierras, and Deserts plus Modoc.

   a. **Is this approach described in Theroux et al. (2016; Appendix 2) appropriate for the ASCI? Are there additional refinements that you suggest that will improve on the proposed approach?**

Paul: Theroux et al. present a viable approach provided it is applied to landscape sources and integrated stressor effects. Although it could be used to set nutrient endpoints, it is a more robust indicator of integrated stressors and should be used that way to aim towards ALUS, which are impacted by integrated stressors, not nutrients alone. Suggested improvements are presented above, but should be subject to additional discussion, exploratory analysis and refinement.

Chuck: The approach appears to be generally sound and based on methods that have been successfully applied previously (e.g., CSCI) and elsewhere.

The technical team needs to explicitly recognize that neither an algae based O/E index nor a pMMI may perform as well as their invertebrate predecessors / equivalents, and the team should not give the impression that the final index (indices) will necessarily be a merged O/E + pMMI index.

The team needs to carefully check the algae data for taxonomic consistency, which is critical for O/E index development and application.

Lester: Explore how well the algal O/E performs (without additional O/E has a direct interpretation in terms of aquatic life.

Explore the effects of taxonomists in ordinations that underline algal O/E.

Jan: The pMMI is a good step forward and a state of the art tool that accounts for natural variation among sites for bioassessment.

The data collected is a great resource for further model development to meet needs emerging during adaptive management.

Consider combining O/E into the pMMI as one metric.

O/E may not work with algae. Compare % reference taxa to 0/E. % ref is also a metric of percent of reference taxa lost that allows for lottery processes in community development and may be more relevant for diatoms than invertebrates, because of the higher taxonomic resolution for most diatom assays.

Taxonomic robustness in species swarms may help diatom metrics be robust to taxonomic "error" or the inability to be taxonomically consistent – because the species are not well defined biologically. The small *Achnanthidium* clade may be a good example of a species swarm, where most taxa are found in relatively good ecological conditions.

Do % reference taxa by major taxonomic group, with is analogous to EPT type metrics? See Tang et al. 2016 (Tang, T., R.J. Stevenson, D.M. Infante. 2016. Accounting for regional variation in both natural environment and human disturbance to improve performance of multimetric indices of lotic benthic diatoms. Science of the Total Environment 268:1124-1134), for an example. This was used in models for the 2008-2009 NRSA periphyton. I think they were good even though they were not included in the report.

Develop indicators specifically for structural and functional attributes that stakeholders care about to the extent possible; then stakeholders will want to have as low a risk of non-attainment as possible. Functional groups and functions could both be included, and some, like N2 fixers, clearly are. This develops stakeholder buy-in and it does relate to BCG attributes.

% non-native/tolerant taxa is a good biodiversity metric. % stalked, % different major taxonomic group (e.g. green, diatom, cyano) by growth form can be used to define functional groups.

Consider using both diatom and non-diatom algae in assessments are better than using diatoms for discriminating, but is there an less expensive method that provides information that is cost effective. Then use the all algae counts for some times or stream types and a simpler, e.g. diatom only count, for other stream types where that is sufficient.

Use additional metrics in the ASCI if it does not diminish the pMMI performance and those metrics characterize different and important elements of biological condition.

Do not include stressor indicators in pMMI. That allows some independence for their use in stressor diagnosis.

b. Can biological indices or metrics be developed that distinguish between chemical, physical, hydrological and habitat impacts to biological integrity from altered water quality versus altered hydrological regime or physical habitat (substrate, cover, flow, etc.)?

Paul: This is a "can" question, and much of the preceding discussion points out how a multimetric BI integrates multiple chemical, physical, hydrological and habitat conditions and stressors making separation of single or even suites of stressors challenging. Nevertheless, with point sources excluded, *integrated management can more effectively be applied to the suite of stressors represented in a BI through the application of BMPs on the landscape and pollution prevention/conservation planning practices to aim towards an appropriate management outcome or target as a multimetric BI using a BCG approach. If management is guided by a robust WCI gradient that relates to the BI, this could provide an effective remedy for integrated watershed management (IWM) rather than application of pollutant specific BMPs. The presumptive outcome of attaining ALU goals is best applied through a multi-source and stressor IWM approach to broadly capture the complexities of the ecosystems. More natural structural and functional watershed integrity, to the extent it can be attained, will likely have concomitant, and perhaps proportional, beneficial effects predicted by the BCG approach for stream biointegrity and beneficial uses.*

Chuck: See comment under section 1. This aim is the holy grail of bioassessment. Partial success has been achieved with invertebrate indices (i.e., stressor-specific, diagnostic indices for pesticides, fine sediment, temperature, salinity), but signals are often somewhat confounded, and multiple stressors can interact in complex and unknown ways to influence biological responses. Such indices should be pursued, but significant effort will be needed to develop and test these types of indices, and the research team should not count on having such tools in the short- to medium-term.

Lester: Possibly, but I haven't seen much success in this area. I would lower the priority of this work given the low probability of success

Jan: Develop separate diagnostic indicators for stressors using biota. Interpret deviation from natural based on which biota changed and traits of biota (tolerance and sensitivity to stressor gradients). Weighted average models work well. Consider BRT or RF models, also.

I have found that in highly variable environments like streams, biological indices of stressor conditions may be more accurate than one-time sampling as proposed in guidance document and shown elsewhere.

c. What are the tradeoffs in designing an index to be tuned to a generalized stressor gradient versus a specific gradient (e.g. water quality)?

Paul: *The arguments above address this concern. It bears repeating that a key relationship is how well a WCI index predicts the receiving water biointegrity index. And, clearly, there are exceptions for habitat-constrained systems and point-source dominated systems that can't be addressed by the conceptual model and described by the BCG that warrant special management attention that falls outside the BCG and TALU protocol. The nutrient gradient alone does not lend itself to as robust an analysis, a huge tradeoff in understanding integrated impacts, and prospects for integrated watershed management.*

Chuck: The technical team needs to carefully consider what they desire of an index. Is it meant to be a measure of overall biological integrity or an indicator that is responsive to specific stressors? O/E indices are agnostic to type of stress in the sense that stressed sites are not used in their calibration. MMIs can be either intentionally or unintentionally tuned to specific stressor gradients because stressed sites are used to calibrate the indices. Decisions regarding how to calibrate MMIs will therefore affect their general utility and transferability.

Lester: Why not just use an O/E that is not tuned to any gradient? Would provide a more direct linkage to aquatic life use protection. Ideally, we would like to know how changes to the algal assemblage influences higher trophic levels, but I don't think the science for this linkage is well developed.

Consider specifying an acceptable loss of species to define threshold (e.g., 20%).

Metrics tuned to specific gradients will provide an estimate of the levels of the stressor (e.g., algal indices tuned to TP will provide an estimate of TP concentration), but then extra work is required to link this information to aquatic life.

Jan: Again, why just have one index? In general, individual metrics will be more sensitive and have non-linear responses more often than when using using multimetric indices, whether they are for biological condition (MMIs) generalized stressor gradient.

d. What are the tradeoffs in design of an index for statewide applicability versus application at a regional scale?

Paul: I can see a viable application at multiple scales, as noted above. A statewide index, conservatively drawn to accommodate the more sensitive systems would in application provide a "red flag" for additional evaluation and assessment to diagnose causes. Follow up application of the BCG and WCI might be used to assure site specific attributes are not traded-off, natural excursions are identified, and beneficial use goals are not compromised. This is especially important for unimpaired and antidegradation situations where protection is the goal.

Jan: Consistency is really valuable and maintained if you use the same method in different regions/typologies. I have not conceptually liked fine tuning MMIs among regions/typologies by selecting the most sensitive metrics independently for different regions/typologies. That seems to change the definition of biological condition. In the simplest case, let's say % sensitive reference and % tolerant species are used in one region/typology, because both sensitively respond to human disturbance (or GSG). In a second region/typology, only % tolerant species is used in an MMI because % sensitive reference species don't respond to human disturbance. In my mind, the second region does not cause as great an alteration in biological condition because % sensitive reference species don't respond to human disturbance, and that should be reflected in smaller decreases in MMI scores. However, different metrics are used in different ecoregions in the national assessments for some biota. So a statewide consistent index may not be as sensitive to change, but it is consistent and appropriate in my mind.

Chuck: Indices developed over smaller versus larger spatial extents tend to be more precise (and hence more sensitive in detecting biological alteration), because effects of natural environmental gradients on biological composition tends to be better or more easily controlled. However, regional indices potentially suffer from comparability issues if either the quality of reference sites vary across regions or the specific metrics selected for use in indices differ across regions. A related issue is that the range of natural variation (i.e., the naturally occurring ecological dynamics of assemblages) can vary across regions. The technical team needs to carefully consider if different thresholds (that imply 'impairment') should be established across regions. The need for separate thresholds can be evaluated by examining variability in reference site index scores across regions.

Lester: Analysis should be possible to explicitly describe these trade-offs.

Statistical Models to Link CSCI and ASCI assessment endpoints to numeric targets for nutrients and organic matter abundance.

3. **The Science Plan details a work element intended to support Water Board staff decisions on numeric targets for nutrients and organic matter by developing statistical stress-response models of the relationship between bioassessment indices (CSCI, ASCI) and concentrations of nutrients and organic matter abundance (benthic chl-a, ash-free dry mass).**

Cliff: *Some of my thoughts and recommendation for charge question one might be considered here. Specifically, might there be indices from some of your condition measurements that might work better as an index than simply correlating CSCI and ASCI against a single variable like TN or TP collected at a single moment in time. For example, you would like to know if the wadeable stream under study shows the likelihood for nitrogen limitation, phosphorus limitation, joint limitation, or no nutrient limitation. Combining information on the concentrations of both nutrients and the elemental ratio (TN/TP) might provide a "nutrient" index that is more informative than just the single variables alone. Similarly, if there are time series of data for some of the condition measurements you might explore some type of a composite value rather than relying on a single value.*

   a. **Given the exploratory analyses showing relationships between CSCI, ASCI and the eutrophication indicators of interest, what concerns should the technical team keep in mind when developing statistical stress-response models, and how might they be addressed?**

Paul: As discussed above, BIs such as CSCI and ASCI are probably not the best application for setting numeric nutrient and organic matter targets. The statistical analyses show some of the complications and uncertainties inherent in the approach. An analysis of nutrient concentration data could be used in a traditional "reference" approach. Linking nutrients to productivity metrics such as organic matter could also be framed in a productivity target, and a reference approach set for measurable limits on productivity. This is a more traditional and direct relationship between nutrients as a stressor and uncomplicated responses of receptors that are most sensitive to nutrient inputs and reflective in the index. Standing crop is often used as a surrogate for productivity, but has associated problems with natural and confounding stressors. In any case, even the most effective or stringent management of nutrients may not provide a remedy for ALUS impairment because it does not incorporate the full suite of stressors that are reflected in BIs, but it can be a more direct, practical and perhaps effective way to set targets for nutrient stressors in a single-pollutant management scenario.

The team needs to recognize that much of the scatter in plots of index values versus nutrient/OM values is likely associated with effects of extraneous environmental variables and the fact that the nutrient/OM may not represent the actual stressor gradient appropriate to different regions. Also, consider using O-E as the x-axis rather than raw stressor values/concentrations and partial dependence plots to better isolate effects of the stressor of interest (e.g., Fig. 4, see *Vander Laan, J. J., Hawkins, C. P., Olson, J. R., & Hill, R. A. 2013. Linking land use, in-stream stressors, and biological condition to infer causes of regional ecological impairment in streams. Freshwater Science, 32:801-820*).

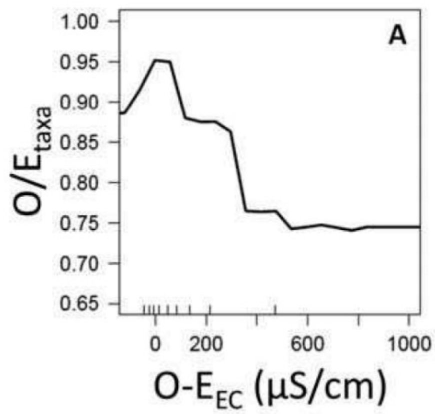

Figure 4. Partial dependence plot showing how invertebrate O/E values changed in response to how strongly salinity (as electrical conductivity) differed from natural, background levels.

Lester: Additional work to improve accuracy of estimated relationships

1. Use conceptual model (Element 1.3.1) to define statistical analyses
   a. Consider multistage model, linking TN and TP to chl-a (and other eutrophication metrics), which are then linked to bug metrics (see Qian and Miltner 2016)
   b. Examine effects of covariates identified in conceptual model with analysis when data are available, or with literature, when lacking data. At the minimum, consider the effects of known stressors that covary with nutrients (habitat degradation and conductivity).
2. Use models to partition variability
   a. Hierarchical models can be used to estimate the contribution of temporal and sampling variability to observed response metric values. Models can be used to estimate mean summer TN and TP from grab samples.
      ii. "Binned ranges": probably more effective to develop models that link directly to CSCI and ASCI values, rather than using binned ranges.
      iii. Work with state board to define regulatory use of statewide defaults
         1. Do statewide defaults apply when local data are lacking?
         2. Are statewide defaults used only as a screening assessment, to be followed by watershed plan that lead to effluent limits? Or,
         3. Are statewide defaults to be viewed as numeric nutrient criteria that are to be adopted by regional water boards and used directly in permitting?
         4. When specifying statewide defaults consider their ultimate use.
         a. Might be best to specify a statewide model that can be adapted by regional water boards…would require a modeling structure that is amenable to incorporation of local data, or…

Perhaps define ranges of criterion values based on the range of possible outcomes a given nutrient concentration.

Jan: The exploratory analyses showing relationships between CSCI, ASCI and the eutrophication are useful. Yes, there are issues, but they can be addressed.

You need to parse out natural effects on nutrient effects with sensitivity plots derived from partial dependence plots. For example, what is the sensitivity to N along a P gradient. Do BRT or RF sensitivity analysis with partial dependence plots to determine how nutrient effects vary with other stressors and naturally varying factors.

Work on communication, per Ken's comments. For example, I don't put a line in a graph if relationship is. Also spread the spread scale so we can see the detail that cluster of data.

Think about deviation from natural somehow in stressor response relationships to account for natural variability. However, there are issues because many stressor-response relationships are not linear. So when accounting for natural conditions, the same magnitude in deviation in a stressor from natural can produce very different responses depending upon where natural is along that non-linear response gradient.

Look for thresholds (valuable for stakeholder consensus) in individual metrics of biological condition, versus MMIs for reasons stated above. Relationships showing some assimilative capacity are particularly valuable for selecting NNEs with stakeholder support.

Use stressor indicators based on species to evaluate whether NNEs are met or not, because biological indicators for stressors can be more accurate than one time sampling water chemistry.

Finding ecological responses in high nutrient concentrations is important for establishing management targets in agriculturally dominated watersheds. These may emerge at the metric versus MMI level.

Remember CART threshold ideas and ability to establish the probability (risk) of threshold response occurring.

Consider using structure equation modeling for causal analysis, as well as other approaches for causal analysis.

How do you plan to incorporate downstream impacts?