10-2020

Characterizing benthic macroinvertebrate and algal biological condition gradient models for California wadeable Streams, USA

Michael J. Paul

Ben Jessup

Larry R. Brown

James L. Carter

Marco Cantonati

See next page for additional authors

Authors

Michael J. Paul, Ben Jessup, Larry R. Brown, James L. Carter, Marco Cantonati, Donald F. Charles, Jeroen Gerritsen, David B. Herbst, Rosalina Stancheva, Jeanette Howard, Bill Isham, Rex Lowe, Raphael D. Mazor, Patina K. Mendez, Peter R. Ode, Alison O'Dowd, John Olson, Yangdong Pan, Andrew C. Rehn, Sarah Spaulding, Martha Sutula, and Susanna Theroux

Ecological Indicators 117 (2020) 106618

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Characterizing benthic macroinvertebrate and algal biological condition gradient models for California wadeable Streams, USA

Michael J. Paul[a,*], Ben Jessup[b], Larry R. Brown[c], James L. Carter[d], Marco Cantonati[e,f], Donald F. Charles[e], Jeroen Gerritsen[a], David B. Herbst[g], Rosalina Stancheva[h], Jeanette Howard[i], Bill Isham[j], Rex Lowe[k], Raphael D. Mazor[l], Patina K. Mendez[m], Peter R. Ode[n], Alison O'Dowd[o], John Olson[p], Yangdong Pan[q], Andrew C. Rehn[n], Sarah Spaulding[r], Martha Sutula[k], Susanna Theroux[k]

[a] Tetra Tech, Inc., Ecological Sciences, Research Triangle Park, NC, United States
[b] Tetra Tech, Inc., Ecological Sciences, Montpelier, VT, United States
[c] United States Geological Survey, Sacramento, CA, United States
[d] Emeritus, United States Geological Survey, Menlo Park, CA, United States
[e] Academy of Natural Sciences of Drexel University, Patrick Center for Environmental Research, Philadelphia, PA, United States
[f] MUSE – Museo delle Scienze, Limnology & Phycology Section, Trento, Italy
[g] University of California, Santa Cruz, Institute of Marine Sciences, CA, United States
[h] California State University San Marcos, San Marcos, CA, United States
[i] The Nature Conservancy, San Francisco, CA, United States
[j] Wood Environment and Infrastructure, Inc., San Diego, CA, United States
[k] Bowling Green University, Bowling Green, OH, United States
[l] Southern California Coastal Water Research Project, Costa Mesa, CA, United States
[m] University of California, Berkeley, Department of Environmental, Science, Policy & Management, Berkeley, CA, United States
[n] California Department of Fish and Wildlife, Sacramento, CA, United States
[o] Humboldt State University, Department of Environmental Science & Management, CA, United States
[p] California State University Monterey Bay, Monterey, CA, United States
[q] Portland State University, Portland, OR, United States
[r] United States Geological Survey, Boulder, CO, United States

ARTICLE INFO

Keywords:
Wadeable streams
Biological criteria
Aquatic life use criteria

ABSTRACT

The Biological Condition Gradient (BCG) is a conceptual model that describes changes in aquatic communities under increasing levels of anthropogenic stress. The BCG helps decision-makers connect narrative water quality goals (e.g., maintenance of natural structure and function) to quantitative measures of ecological condition by linking index thresholds based on statistical distributions (e.g., percentiles of reference distributions) to expert descriptions of changes in biological condition along disturbance gradients. As a result, the BCG may be more meaningful to managers and the public than indices alone. To develop a BCG model, biological response to stress is divided into 6 levels of condition, represented as changes in biological structure (abundance and diversity of pollution sensitive versus tolerant taxa) and function. We developed benthic macroinvertebrate (BMI) and algal BCG models for California perennial wadeable streams to support interpretation of percentiles of reference-based thresholds for bioassessment indices (i.e., the California Stream Condition Index [CSCI] for BMI and the Algal Stream Condition Index [ASCI] for diatoms and soft-bodied algae). Two panels (one of BMI ecologists and the other of algal ecologists) each calibrated a general BCG model to California wadeable streams by first assigning taxa to specific tolerance and sensitivity attributes, and then independently assigning test samples (264 BMI and 248 algae samples) to BCG Levels 1–6. Consensus on the assignments was developed within each assemblage panel using a modified Delphi method. Panels then developed detailed narratives of changes in BMI and algal taxa that correspond to the 6 BCG levels. Consensus among experts was high, with 81% and 82% expert agreement within 0.5 units of assigned BCG level for BMIs and algae, respectively. According to both BCG models, the 10th percentiles index scores at reference sites corresponded to a BCG Level 3, suggesting that this type of threshold would protect against moderate changes in structure and function while allowing loss of some sensitive taxa. The BCG provides a framework to interpret changes in aquatic biological condition along a

---

[*] Corresponding author at: Tetra Tech Inc., P.O. Box 14409, 1 Park Drive, Suite 200, Research Triangle Park, NC 27709, United States.
E-mail address: Michael.Paul@tetratech.com (M.J. Paul).

https://doi.org/10.1016/j.ecolind.2020.106618
Received 27 January 2020; Received in revised form 2 June 2020; Accepted 9 June 2020

gradient of stress. The resulting relationship between index scores and BCG levels and narratives can help decision-makers select thresholds and communicate how these values protect aquatic life use goals.

1. Introduction

Quantitative water quality goals to protect and restore biological integrity are cornerstones of federal and state water quality protection programs (e.g., California Porter Cologne Act 1969, Federal Water Pollution Control Act 1972, EU Water Framework Directive 2000). However, setting quantitative goals is challenging because of complexities in translating species composition and abundance (biological condition) into targets that represent narrative goals like "integrity" or "balanced". Many states in the USA have developed quantitative bioassessment indices to measure and assess attainment of biological integrity goals (Davis and Simon, 1995; USEPA, 2002; Yoder and Barbour, 2009). These indices are typically based on assemblage structure assessed relative to that observed in reference sites with comparable environmental settings (Reynoldson et al., 1997; Hawkins et al., 2010). These indices rely on empirical data to identify present-day reference conditions, quantified from least disturbed landscapes and used to construct the index. Biological integrity goals or numeric targets are frequently characterized by deviation from the reference site population, calculated as a statistical characteristic (e.g., 30th, 10th, or 1st percentile) of reference site bioassessment index score distributions (e.g., Barbour et al., 1999; Mazor et al., 2016). However, because reference sites are typically based on a "best available" or "least disturbed" definition (Stoddard et al., 2006), their condition may include historic degradation, unmeasured anthropogenic stress (e.g., chemicals for which there are no measurements), or changes over time associated with factors such as climate change (Poff et al., 2010). As a result, moderately disturbed reference sites may influence the reference distribution and reference-based indices would then potentially assess sites against an already degraded benchmark. More importantly, although these indices and percentile thresholds provide statistically reproducible and unbiased assessments of biological integrity, communicating to policy makers, managers, and the public the appropriateness of a selected statistical value as a representation of biological integrity goals is challenging. However, communicating what different values of the indices mean can be enhanced when linked to a clear narrative of the biological structural or functional changes that are protected or lost along the numeric range.

The Biological Condition Gradient (BCG) is a conceptual model that describes structural and functional changes in stream systems as they degrade in response to human disturbance (Fig. 1) (Davies and Jackson, 2006; United States Environmental Protection Agency (USEPA), 2016). The BCG is a standard biological response gradient intended to have a universal meaning, so that interpretations do not vary across regions, and the theoretical levels can be applied to any waterbody. It can be used for interpreting biological indices and for comparing and reconciling regional differences in reference condition, types of indices, or even indices for different assemblages. It was developed, in part, to supplement biological threshold interpretations based on statistical index properties with additional interpretations based on ecological properties (i.e., taxa richness, species composition, tolerance and functional organization) and help managers identify values along a numeric biological index where protection of aquatic life use lies.

The process of calibrating BCG models to a new region begins with identifying experts familiar with the ecology of degraded and natural ecosystems in the study area and training them in the BCG concept and approach. They are then asked to assign BCG taxa attributes (sensitivity/tolerance, endemic, invasive, etc.) to regional taxa and to place sites into one of the 6 BCG levels based on taxonomic composition and abundance, accompanied by geographic information on site location to inform expectation (Davies and Jackson, 2006; United States Environmental Protection Agency (USEPA), 2016; Gerritsen et al., 2017). Although experts are provided limited environmental information about the sites (e.g., elevation, climate, geology), they are given no

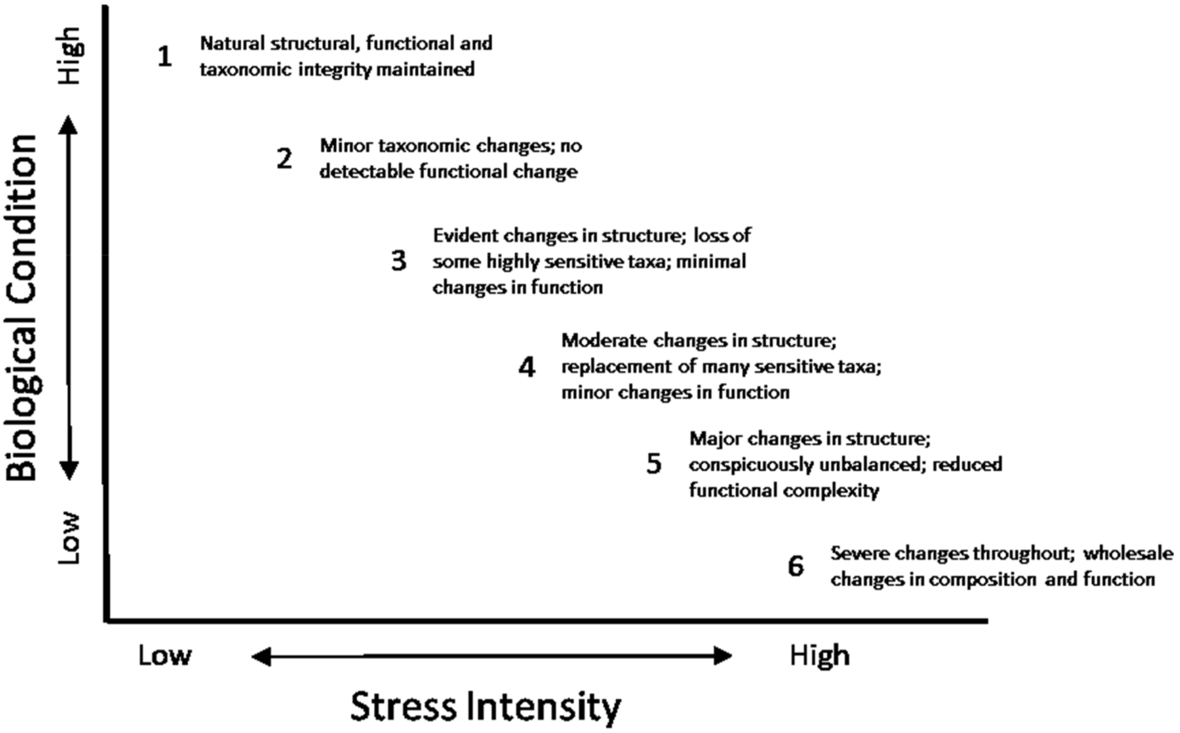

Fig. 1. The Biological Condition Gradient (BCG) conceptual model (adapted from Davies and Jackson, 2006 and USEPA 2016).

information about disturbance or human activities in the watershed or about stressors measured at the site (e.g., water chemistry or habitat data). Therefore, their judgments about biological condition are based purely on biological data (i.e., taxonomic data and metrics calculated from them). Experts are also asked to record their decision-making process in ecological terms. The BCG calibration efforts, thus, capture the breadth and depth of expert ecological interpretation regarding sample composition along disturbance gradients. The final BCG level assignments can be plotted and modeled against bioassessment index scores calculated for the same samples. BCG scores and their consensus narrative ecological descriptions can then be linked to any specific biological index score.

BCG narratives, which contain language similar to state aquatic life use narratives, can support the interpretation of bioassessment indices, provide additional rationale for the selection of numeric biological integrity goals and communicate to stakeholders the ecological changes associated with these goals (Davies and Jackson, 2006; Gerritsen et al., 2017). California is a prime example of where such decision support is helpful. The state has a strong wadeable stream bioassessment program, supported by well-established protocols, training and quality assurance, and a broad network of least disturbed reference sites (Ode et al., 2016). A robust statewide bioassessment dataset exists, representing both BMI and algal assemblages as well as a comprehensive set of data on stressors (e.g., chemical, physical habitat). The California State Water Resources Control Board (Water Board) staff is proposing a policy to protect biological integrity in wadeable streams using thresholds of numeric indices to protect narrative aquatic life uses. The Water Board adopted the California Stream Condition Index (CSCI; Mazor et al., 2016) which provides a numeric score of biological condition based on benthic macroinvertebrate (BMI) bioassessment data and supported the development of an algal stream condition index (ASCI; S. Theroux, pers. comm.), which is based on both diatom and soft-bodied algal assemblages. Both indices are expressed as ratios. Values close to 1 indicate a sample similar to reference expectation; lower scores indicate biological assemblages that differ from the reference expectation. Mazor et al. (2016) proposed narratives of “likely intact,” “possibly altered,” “likely altered,” and “very likely altered” associated with the greater than 30th, 30-10th, 10-1st, and < 1st percentiles of reference as interpretation of the CSCI. Similar thresholds will be developed for the ASCI. However, communication of what these percentiles of reference represent in terms of loss of ecosystem structure and function would be helpful for supporting policy decisions on numeric biological integrity goals.

The purpose of this work was to: 1) assign BMI and algal BCG taxon attribute scores to taxa; 2) assign BCG levels to 250 + California wadeable stream samples along human disturbance gradients; 3) develop narratives of structural and functional changes for BCG levels specifically associated with degradation of California wadeable streams; and 4) compare BCG levels (Fig. 1, levels 1–6) with CSCI and ASCI numeric scores to relate percentile of reference thresholds (30th, 10th, 1st) to narratives associated with the loss of benthic invertebrate and algal community structure and function.

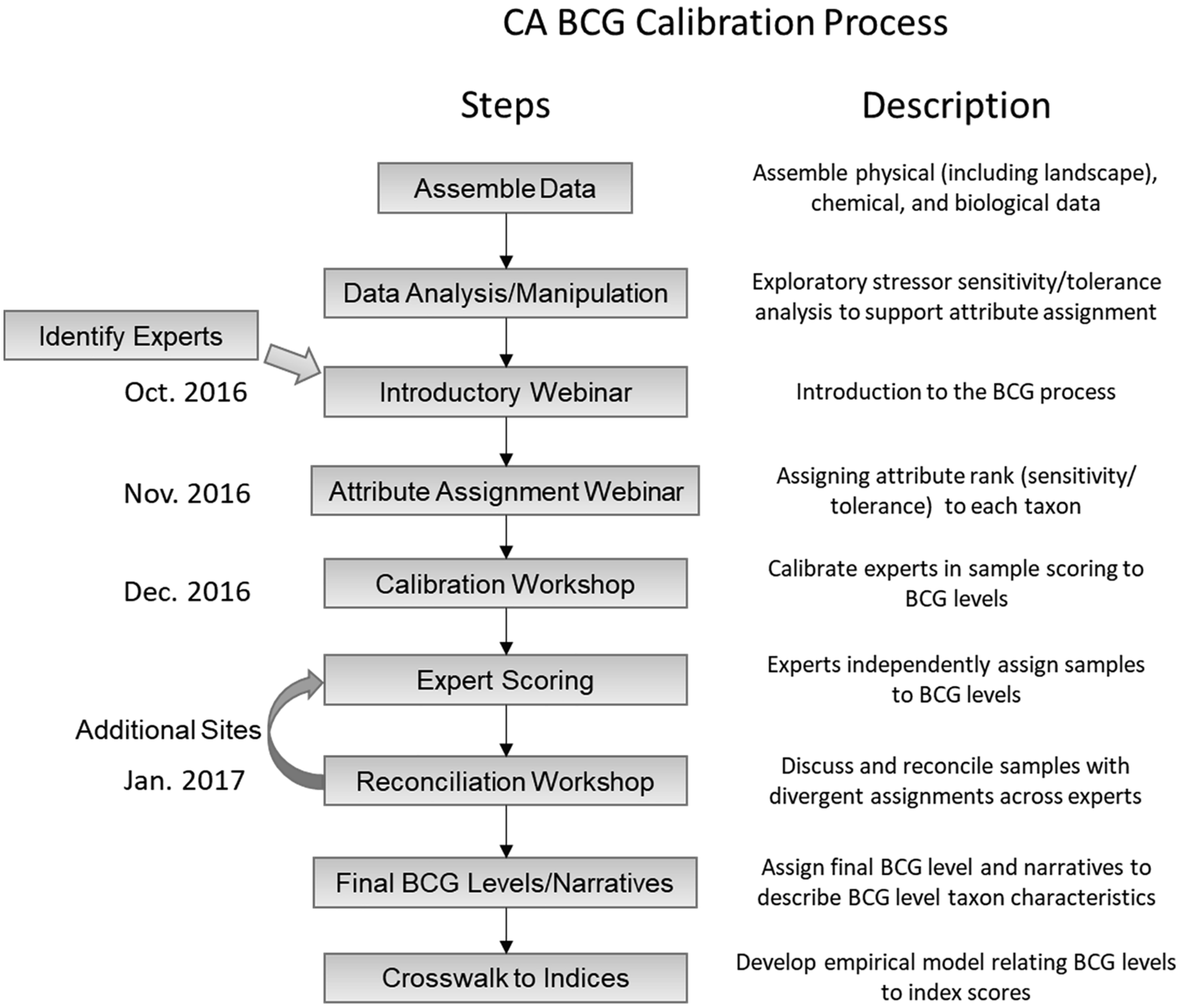

Fig. 2. Process diagram for calibration of the BCG in California.

2. Methods

2.1. The BCG process

A multistep process was followed to calibrate a BCG to California wadeable stream conditions (Fig. 2). The process included assembling data, orienting experts to the taxa attribute and level assignment processes and using an expert BCG level assignment framework to describe the BCG in terms of observed assemblage response to anthropogenic stressors. This calibration process is like those used in BCG development in other regions (Gerritsen et al., 2017; United States Environmental Protection Agency (USEPA), 2016). Assessing condition of biological assemblages (e.g., by interpreting bioassessment indices) involves professional judgment (e.g., selection of a percentile) even when such judgment may be embedded within objective, quantitative approaches (e.g., Steedman, 1994; Borja et al., 2004; Weisberg et al., 2008). This California BCG calibration relies explicitly on professional judgment and development of consensus supported by data and uses both individual and group interpretations.

BCG calibration began with the assembly and analysis of biological monitoring data. Samples were selected to represent a full gradient of natural and stressor conditions based on existing classification schemes, stressor information, and biological indices. Data were organized to support analyses and review by experts. Nine experts in BMI and five experts in algal ecology in California were identified. All 9 BMI experts were from California, with specific expertise including southern to northern as well as Sierra Nevada to coastal stream assemblages. Algal experts included a California expert, and 4 experts from outside the state but with experience working with California taxa from national surveys. Experts were given an orientation on the theoretical basis of the BCG, the BCG level assignment process, and an introduction to taxa attributes (taxa characteristics; SI Table 1). After training, the experts gathered for the first workshop, in which they agreed on and completed taxa attribute assignments and received training in assigning samples to BCG levels. The training demonstrated how experts were expected to interpret sample data in the context of BCG level definitions, assign BCG levels to samples independently, compose rationale for their ratings, and reconcile multiple ratings per sample. Between the first and second workshop, experts independently assigned sites to BCG levels (1–6, Fig. 1). The second workshop was held to reconcile ratings and eventually agree on a consensus BCG level assignment for each sample. The results of the expert consensus process included ratings for 250 + samples for each assemblage and narrative statements describing the biological characteristics of each BCG level. The narratives were based on the rationale of independent experts when deciding on their ratings and from group discussions. The final steps in the calibration process included relating the BCG level assignments to CSCI and ASCI scores to provide managers a narrative description of ecological conditions associated with percentile of reference.

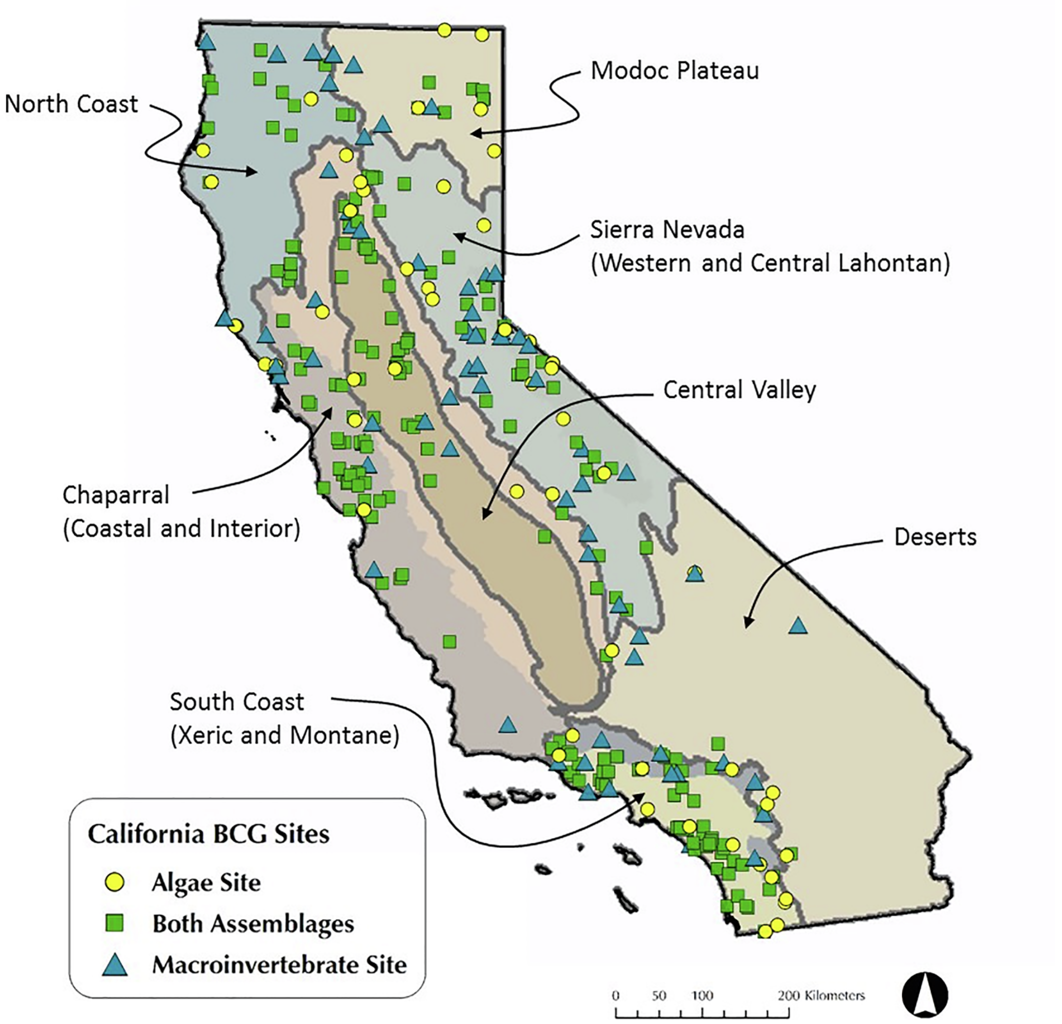

Fig. 3. Sample site locations throughout California, showing the assemblages sampled and the Perennial Stream Assessment (PSA) region.

Table 1
Descriptive statistics for natural and stressor variables in California samples reviewed by BMI and algal expert panels.

Variable	Min	Mean	Max
Collection Year	1998	2011	2015
Catchment Area (km^2)	1	315	8812
Elevation (m)	3	651	3130
Average Air Temperature (°C)	6.4	20.9	29.7
Precipitation (m/y)	0.09	0.66	2.0
Agricultural and Urban Cover (%)	0.0	10.0	88.8
Total Phosphorus (mg/L)	0.0	0.18	5.1
Total Nitrogen (mg/L)	0.0	1.02	34
Specific Conductance (μS/cm)	8	540	6381

2.2. BCG taxon attribute development

In the BCG conceptual model, biological attributes of aquatic ecosystems change along a gradient of increasing anthropogenic stress including factors like nutrient enrichment, physical habitat degradation, and flow modification. All taxa attributes were relevant in the California BCG calibration (SI Table 1). Detailed attribute definitions were thoroughly reviewed by the expert panels before assigning attributes to any taxa (United States Environmental Protection Agency (USEPA), 2016). Because the number of taxa to which experts were required to assign attributes was large (> 1200 combined), initial taxa assignments were estimated by BCG facilitators using information from existing tolerance values and California specific tolerance metric calculations consistent with other BCG calibration efforts (Hausmann et al., 2016; Gerritsen et al., 2017). Experts used these tolerance metrics and supplemental analyses (e.g., taxon specific stress-response curves) to refine the initial attribute assignments.

Tolerance metrics (17) calculated as central tendencies, environmental limits and optima were calculated for taxa that occurred in at least 10 samples from environmental stress-response curves and made available to experts for the purpose of assigning attribute values to each taxon (see Yuan, 2006). These analyses examined the response of taxa to stressor variables (e.g., conductivity, total nitrogen, and total phosphorus). Tolerance metrics, expressed as central tendencies, describe the average environmental conditions under which a taxon is likely to occur. Central tendencies were estimated by computing the mean of the product of taxon abundance and the environmental stressor variable, calculated using both abundance and presence/absence, assuming a normal distribution across the environmental gradient. The width of the bell shape is defined as tolerance and should not be confused with the tolerance scale used to describe general taxon sensitivity (Hilsenhoff, 1987). Weighted cumulative distribution functions (CDFs) were used to estimate tolerance in non-uniform sample distributions. Environmental limits attempt to capture the maximum or the minimum level of an environmental variable under which a taxon is likely to persist, while optima define the environmental conditions that are most preferred by a given taxon. Both limits and optima can be derived from observational data or regression relationships. The area under the curve of 95th percentile cumulative percentiles (CPs) was used to represent the environmental limits a taxon can tolerate. Taxa optima (i.e., the central tendency) were estimated using the median values of the CDF and the CP of the regression models. Regression estimates of taxon-specific stress response relationships were developed using linear regression models (LRM), quadratic logistic regression models (QLRM), and generalized additive models (GAM). Tolerance metrics developed from these statistical methods were generally correlated, so variations due to statistical approaches were minimized by taking an average of results from all methods. Tolerance metrics from both abundance and presence/absence based models, were ranked and taxa scores translated to assign the initial taxa attribute levels from II to V. These initial analyses cannot infer attribute level I taxa, which were assigned by experts in the first workshop. The stressor response analysis was applied for 769 invertebrate taxa for which there was enough data, while stressor response analysis was applied for 318 diatom genera and species and for 58 soft bodied algae (SBA) genera. At the species level, there were generally few SBA occurrences that could be used to model stressor-response patterns.

During the first workshop, experts refined the initial taxa attribute assignments, including identifying level I taxa, based largely on their knowledge of the taxa and the conditions under which they occur, informed by calculations of the taxon-specific tolerance metrics or literature derived tolerance values (e.g., Rott et al., 1999; Potapova and Charles, 2007; Stevenson et al., 2008; Lange-Bertalot et al., 2017). Open panel discussions focused on taxa for which there was disagreement which arose largely from continent-scale geographic differences in taxonomic expertise. Unlike site scoring, attribute assignment was not anonymous, so a potential for biased influence in attribute assignments may have existed. Given the number of taxa and the limited disagreements, we feel this bias was likely minimal. Final taxa attribute assignments were included as information for site scoring and used by experts to rate sites along the BCG scale.

2.3. Sample selection for expert scoring

Sites for expert review were selected from the California Surface Water Ambient Monitoring Program (SWAMP), Perennial Streams Assessment (PSA), Stormwater Monitoring Coalition (SMC) program databases and other sources. A total of 264 sites were scored for macroinvertebrate BCG and 248 for algae BCG (Fig. 3). They cover a wide range of characteristics representing the diverse environmental settings encountered throughout California (Table 1). Selection criteria were established to distribute site types among physiographic regions, biological condition based on biological metrics and reference status, and stressor types based on land cover and water quality. That is, sites were selected to represent diverse natural settings found across the state, as well as to represent gradients of stress from a range of representative human activities, including: forestry, agriculture, urbanization, channelization, and hydropower. Efforts were made to include some sites representing unusual or challenging circumstances.

Approximately 200 samples with paired BMI and algal data were reviewed by experts during the first round. After initial scoring, additional samples were requested by experts to expand the distribution among California PSA regions (Fig. 3); in contrast to the initial set, these additional BMI and algal samples were not paired.

2.4. Data preparation and presentation for BCG scoring

Data provided to experts for sample scoring included raw taxonomic data, metrics of taxa attributes upon which experts had previously agreed (SI Fig. 1), and data descriptive of natural biogeographic gradients in California wadeable streams (e.g., elevation, mean annual precipitation, dominant geology). Raw taxonomic data were provided to experts in the form of taxa lists with enumeration per taxon. Macroinvertebrate samples had approximately 600 organisms collected using standardized methods known to the experts (Ode et al., 2016) and identified to a range of taxonomic resolutions from species to class, with the majority identified to genus. Similarly, algal samples were also collected using standard methods (Fetscher et al., 2009). Sample data included 600 valve target diatom counts, lists of soft algae taxa observed, collection type (qualitative, macroalgae, microalgae, or epiphyte), and calculated biovolume (Fetscher et al., 2014).

In addition to the raw data for each assemblage, sample data included metrics commonly used in assessment (e.g., tolerance, functional feeding group) and metrics based on taxa attributes (e.g., percent BCG level 2 individuals). BMI experts had some information from the CSCI model. The CSCI is a combination of a predictive multimetric index (MMI) and an observed/expected (O/E) taxonomic completeness model (Mazor et al., 2016). For BMI experts, basic metrics were immediately displayed with each sample, including those that compose the MMI, whereas expected taxa richness (from the O/E model of the CSCI) was hidden, giving experts an opportunity to interpret the sample composition without revealing expected richness. CSCI scores were never displayed. BMI experts could have calculated an O/E score from expected taxon capture probabilities but calculating a CSCI score would have been difficult. For algal experts, the ASCI, which consists of three multimetric indices for diatoms (dASCI), soft bodied algae (sbaASCI), and a hybrid (hASCI, respectively; S. Theroux, pers. comm.) had not been completed during site scoring by algal experts. We selected the hybrid ASCI (hASCI) as the index to compare with algal BCG level assignment results because of its performance relative to the other two algal indices (S. Theroux, pers. comm.).

Biogeographic data based on GIS analysis was also provided and included biological regions, climate, watershed area, geology, predicted background conductivity and other predicted background water quality and geological characteristics (Olson and Hawkins, 2012). These were displayed during the BCG level assignment process to help experts set their expectations for organismal presence and abundance. The format for the BCG level assignment exercises was standardized to show all biological and physical data together for each sample (SI Fig. 1).

Variables related to stress or human activity were compiled and used for site selection, stressor-response analysis, and *post hoc* evaluation, but were not displayed to experts during the sample BCG level assignment process. They included land use, road density, mines, dams, and field data, including chemistry and physical habitat. Water chemistry included specific conductivity, chloride, total nitrogen (TN), total phosphorus (TP), and pH in addition to several less common measures. Physical habitat was not available for all sites, but included dominant substrate, habitat complexity, riparian vegetation, shading, and channel morphology (SWAMP, 2018, Accessed on January 22, 2020).

2.5. Assignment of samples to BCG levels

Under typical BCG model development (e.g., Hausmann et al., 2016; Gerritsen et al., 2017), open discussions occur among experts during the process of sample BCG level assignment. This open discussion might allow for limited bias to influence ratings if some experts are perceived to be more qualified or are more persuasive than others. To reduce this bias, BCG development for California wadeable streams was based on a "Delphi" approach, which uses independent expert interpretation of sample information followed by reconciliation and discussion to arrive at consensus on assignment of a BCG level (Nair et al., 2011).

During the first round of BCG level assignments, experts worked independently and anonymously to evaluate sites, using BCG level definitions (Fig. 1), taxa lists, taxa attribute summaries, and site characteristics not subject to human disturbance. The latter included ecoregional information, collection date, catchment area, elevation, water temperature, precipitation and other predicted background water quality and geological characteristics important in understanding expectations for taxa (Olson and Hawkins, 2012) (e.g., SI Fig. 1). Excluded from site information reviewed by experts were site locations, stressor information (land use, pollutant concentration, and habitat assessments), and existing biological assessment scores.

After this first round, sample ratings were reviewed; ratings were considered in agreement if no more than three of nine macroinvertebrate experts and one of five algal experts rated a sample at one level above or below that of the other experts. For samples in agreement, the majority expert BCG level assignment was used as the consensus for the sample. For all samples with divergent scores (i.e., greater than 1 level difference) among experts, a second-round reconciliation process was applied.

For reconciliation, samples were anonymously presented to the experts with the independently written scoring rationale of each expert. After considering their colleagues rationale, experts publicly shared their second-round BCG level assignments, which may have changed towards the initial median, changed towards an extreme BCG level assignment as a result of convincing or previously unrecognized data interpretation, or not changed at all.

After the second-round voting (the re-vote), a set of rules were approved by experts to make a final consensus BCG level assignment for a sample from potentially disparate assignments. The final level for these samples was the most common one assigned (i.e., the mode) and, if there was a tie, the mid-value was selected and rounded down (i.e., to the lower BCG level). Because there were 9 macroinvertebrate experts, the median level assignment was selected. Algal experts allowed themselves to assign BCG levels for each sample in "core" or "qualified" levels, where core referred to BCG levels 1–6 and qualified indicated conditions somewhat better (+) or worse (−) than the core. Therefore, each algal BCG level had three possible ratings (e.g., 3−, 3, or 3+). Algal samples were ultimately given a core value BCG level assignment based on the decision rules above.

2.6. Relating BCG levels to numeric index values

The BCG levels assigned to the approximately 250 samples throughout California were related to existing CSCI and hASCI bioassessment indices for the same samples graphically and statistically. Box plots of index values by BCG level were made to visually compare the relationship between index scores and BCG level. Similar comparisons classified by natural variables (elevation, watershed area, temperature and precipitation) and PSA regions were made to review the extent to which relationships between responses were either universal or might be context dependent.

Proportional-odds logistic regression modeling was used to estimate ranges of index values that are likely to fall within each BCG level (polr package in R, R Core and Team., 2016). Proportional-odds modeling is an ordinal logistic regression model that allows illustration of the points at which an index is more likely to be associated with one BCG level in comparison to all other levels (Agresti, 2002; Venables, 2002).

Box plots of BCG scores with TN, TP, specific conductance, and percent agricultural/urban landuse were examined to verify BCG response to stressor gradients.

2.7. BCG narrative development

A narrative description for each BCG level was created to communicate the ecological characteristics recognized by the experts, and generally paralleled the original descriptive definitions for the levels (United States Environmental Protection Agency (USEPA), 2016). The rationale for assigning samples to BCG levels was explored at several points in the BCG scoring process. In training discussions, the experts were asked to conceptually characterize the best biological conditions possible or observable in California so that all experts had an agreed upon benchmark. These conceptual characterizations were generally narrative and qualitative statements of taxon richness, biomass or abundance of certain types, and occurrence of indicator taxa for BCG levels 1 and 2. Effects of naturally occurring stressors on biological conditions were also discussed, recognizing that expectations are dependent on environmental setting due to variability in background natural stressors, the timing of sample collection, or the confounding (or compounding) of natural and anthropogenic stressors. During the sample scoring process, experts again wrote and discussed their rationale for assigning each sample to a BCG level. The rationale included general qualitative comparisons, qualitative and quantitative expectations based on attribute and taxonomic trait metrics, and expectations for indicator taxa. As the evidence built for assignments at each level, the group came to an agreement regarding general narrative descriptions for each BCG level. This agreement was captured in narrative statements compiled through expert review and consensus.

3. Results

3.1. Taxa attribute assignments

There were 769 invertebrate taxa, 546 diatom taxa, and 419 soft algal taxa considered by the expert panels. Of these, 92%, 77%, and 37% of taxa, respectively, were assigned to taxa attribute levels. This process took most of the first workshop (2 days), in addition to the time required by facilitators to prepare the supporting analyses (approximately 1 week). Attributes were not assigned (NA) for taxa that occurred in < 10 samples, were unfamiliar to the experts, or were of ambiguous taxonomic resolution. Most taxa (77%, 68%, and 69% for BMI, diatoms, and soft-bodied algae, respectively) were assigned to taxa attributes III and IV, the sensitive and moderately tolerant categories (Table 2). There were relatively few taxa in attributes I and VI (the historically documented, sensitive, long-lived or regionally endemic and the non-native or intentionally introduced taxa attributes, respectively).

3.2. Sample ratings

In the first round of ratings, assignments were generally more consistent among experts for BMI than algae but both expert groups achieved good consensus during reconciliation. For BMI, most individual BCG assignments (81%) were within 0.5 BCG level of the median or majority assignment and 95% were within 1 BCG level. After the second-round voting of all macroinvertebrate samples, 78% were within 0.5 level and 100% within 1 level (SI Fig. 1). For the algae, 51% of independent ratings agreed with the median BCG level assignment for each sample during the first round of review. In the consensus dataset, 57% of ratings were the same as the median. A total of 82% of ratings were within 1/3 of the median (no more than a + or – difference), and 94% of ratings were within 2/3 of the median. Among the five algal experts, only 6% of the individual ratings differed from the median by a whole BCG level or more. In the samples with some discrepancy, 38% of ratings were the same as the median. Following the second-round voting, expert agreement with the median increased to 50% of samples. Of the 85 samples that required reconciliation to arrive at consensus, 11 samples showed a change in median ratings of more than 1/3 level (e.g. the difference between 3 and 3+) between rounds.

For both BMI and algae, assignments were roughly evenly distributed among the intermediate BCG levels, with few assigned to Level 1 or 6; this pattern was more pronounced for algae than for BMI, especially sites where both assemblages were sampled (Table 3). Of the 264 macroinvertebrate samples rated, 239 samples were assigned to BCG levels 2 through 5, in almost equal proportions (Fig. 4). Few sites were assigned to BCG levels 1 and 6, < 20% of the average of those identified to levels 2 through 5. Of the 236 algal samples with hASCI scores that were assigned a BCG level, most were assigned to BCG levels 3 and 4, in equal proportions (Fig. 4). Almost equal numbers of sites were assigned to levels 2 and 5, though less than half of those in levels 3 and 4. Only 1 sample was assigned to level 6 and no level 1 samples were identified.

Although assignments to BCG levels were roughly even at the statewide level, different patterns were evident in certain regions (e.g., Fig. 5). For example, most samples in the Central Valley (a heavily agricultural region with few undeveloped areas) were assigned to Level 4 or 5 by BMI experts (30 of 39) and by algae experts (26 of 37), and only 4 algal and no BMI samples were assigned lower than 3. In contrast, experts assigned more samples to Level 3 or lower (58 of 75 for BMI and 58 of 66 for algae) than to Level 4 or higher in the largely forested Sierra Nevada and North Coast regions.

3.3. Relating BCG levels to the CSCI and hASCI

Bioassessment index scores (i.e., CSCI and hASCI) declined with

Table 2
The number of taxa and representative taxa assigned to each taxa attribute level. The taxa shown represent the most common and abundant for each level. BMI = benthic macroinvertebrate. SBA = soft-bodied algae.

Attribute	Number of BMI Taxa	Representative BMI Taxa	Number of Diatom Taxa	Representative Diatom Taxa	Number of SBA Taxa	Representative SBA Taxa
I. Specialists: Historically documented, sensitive, long-lived, or regionally endemic taxa	9	*Vorticifex effusa* *Nerophilus californicus* *Sierraperla cora*	3	*Encyonema latum* *Navicula aurora* *Gomphoneis mamilla*	1	*Zygnema aplanosporum*
II. Highly SENSITIVE: Highly sensitive (typically uncommon) taxa	95	*Drunella doddsii* *Yoraperla nigrisoma* *Ameletus*	96	*Reimeria sinuata* *Epithemia sorex* *Rhopalodia gibba*	33	*Chamaesiphon polymorphus* *Homoeothrix varians* *Calothrix parietina*
III. Sensitive: Intermediate sensitive taxa	245	*Lepidostoma* *Micrasema* *Epeorus*	129	*Achnanthidium minutissimum* *Rhoicosphenia abbreviata* *Nitzschia dissipata*	62	*Nostoc verrucosum* *Calothrix epiphytica* *Aphanothece minutissima*
IV. Indiscriminate: Taxa of intermediate tolerance	301	*Oligochaeta* *Baetis tricaudatus* *Orthocladius complex*	134	*Planothidium lanceolatum* *Nitzschia inconspicua* *Ulnaria ulna*	46	*Heteroleibleinia* sp. 1 *Leptolyngbya foveolarum* *Aphanocapsa delicatissima*
V. Tolerant: Tolerant taxa	56	*Ostracoda* *Hyalella* *Dicrotendipes*	57	*Achnanthidium exiguum* *Nitzschia palea* *Navicula gregaria*	14	*Scenedesmus ellipticus* *Rhizoclonium hieroglyphicum* *Desmodesmus abundans*
VI. Exotic: Nonnative or intentionally introduced species	6	*Potamopyrgus antipodarum* *Corbicula* *Cambaridae*				

Table 3
Correspondence matrix of BMI and algal BCG expert site scores for the 194 samples in common. Shaded cells are one to one correspondence.

Algal BCG Level Assignments	BMI BCG Level Assignments						
	Levels	1	2	3	4	5	6
	1						
	2		17	9	1		
	3	3	24	21	12	6	
	4		3	12	20	33	7
	5				3	13	10

BCG level (Spearman rank order correlations $p < 0.05$, Fig. 4). There was less difference in the distribution of CSCI scores between levels 1 and 2 than between other levels, even though most adjacent level interquartile ranges overlapped somewhat. There was a gap between interquartile ranges from levels 3 to 4 and between 5 and 6. For macroinvertebrate samples, the relationship between the BCG levels and the CSCI appears to be robust across PSA regions, except for the South Coast, where CSCI scores were consistently higher in BCG levels than in other regions (Fig. 5). Within BCG levels 3 and 4, CSCI scores among Desert Modoc samples were also generally higher than those from other PSA regions. Comparison of CSCI index scores to BCG level using the proportional odds model showed that CSCI values greater than 1.0 were more likely to be BCG level 2 or 1 than any other level. Furthermore, scores below 0.3 were more likely to be within BCG level 6. It is worth noting that the CSCI and hASCI indices are ratios, with an average reference sample population mean of 1. Samples with greater species richness and more sensitive taxa than expected on average can receive a score of 1 or greater. The CSCI treats these scores as part of the natural variability, yet BCG experts treated this excess richness and proportion sensitive taxa as meaningful differences in BCG levels. According to the proportional odds model, the highest probability that samples would be assigned BCG levels 3, 4, and 5 occurred when sample CSCI scores were approximately 0.90, 0.65, and 0.40, respectively.

hASCI scores similarly declined with increasing BCG levels, but not as systematically as CSCI scores (Spearman rank order correlations $p < 0.05$, Fig. 4). Compared to macroinvertebrates, the algal BCG levels appear to be somewhat compressed. BCG levels 2 and 3 overlapped substantially. At BCG levels greater than 3, scores declined, except in level 6 which had only 1 sample assigned to it. Interquartile ranges were generally similar across BCG levels. Across regions, index score distributions declined with BCG scores in Chaparral, Central Valley, Desert-Modoc, South Coast, and Sierra Nevada across BCG levels 2–5 but were relatively similar for the North Coast (Fig. 5). For hASCI, values above 0.90 are more likely to be BCG level 3 or 2, whereas values below 0.10 are likely to be 6 (Fig. 6). Again, according to the proportional odds model, the highest probability that samples would be assigned BCG levels 4 and 5 occurred at hASCI scores of approximately 0.70 and 0.30, respectively.

Except for algal BCG and watershed area, both assemblage BCG levels were significantly (Spearman rank correlation $p < 0.05$) correlated with environmental settings (Fig. 7) and stressor/human activity gradients (Fig. 8). BMI BCG scores increased in larger watersheds and BCG level assignments for both assemblages decreased with elevation and precipitation and increased with temperature. These associations with environmental factors might be driven by underlying stressors, which are likely greater at lower elevations in California, reflecting greater agricultural and urban development. Nutrient concentrations, conductivity, and land cover disturbance are higher in sites with BCG levels 4, 5, and 6 scores when compared to conditions in sites with lower BCG scores (levels 1–3) (Fig. 8).

3.4. *Narrative BCG level descriptions for the biological assemblages*

The narratives of sensitivity and tolerance, embodied in respective decisions for taxa attribute assignments, and the information on taxon presence, absence and abundance were recorded and refined by each group of experts. These represented revised BCG descriptions for California streams, attempting as much as possible to indicate region specific changes (SI Table 2). The different expert groups also attempted to record some of the specific taxonomic distinctions between BCG levels as reflected in presence, absence, and abundance.

4. Discussion

The BCG model for California's wadeable streams integrates bioassessment tools into management programs by providing an easily understood way to interpret and communicate statistically complex numeric representations of stream biological condition, namely the bioassessment index scores. This addresses a major challenge in raising appreciation for bioassessment efforts: putting data into a meaningful context for anyone interested in ecological response or in explaining the benefits of (often expensive) remediation efforts to non-technical audiences (Poikane et al., 2016). By linking index scores to a BCG model, we create new opportunities to engage with audiences (both the general

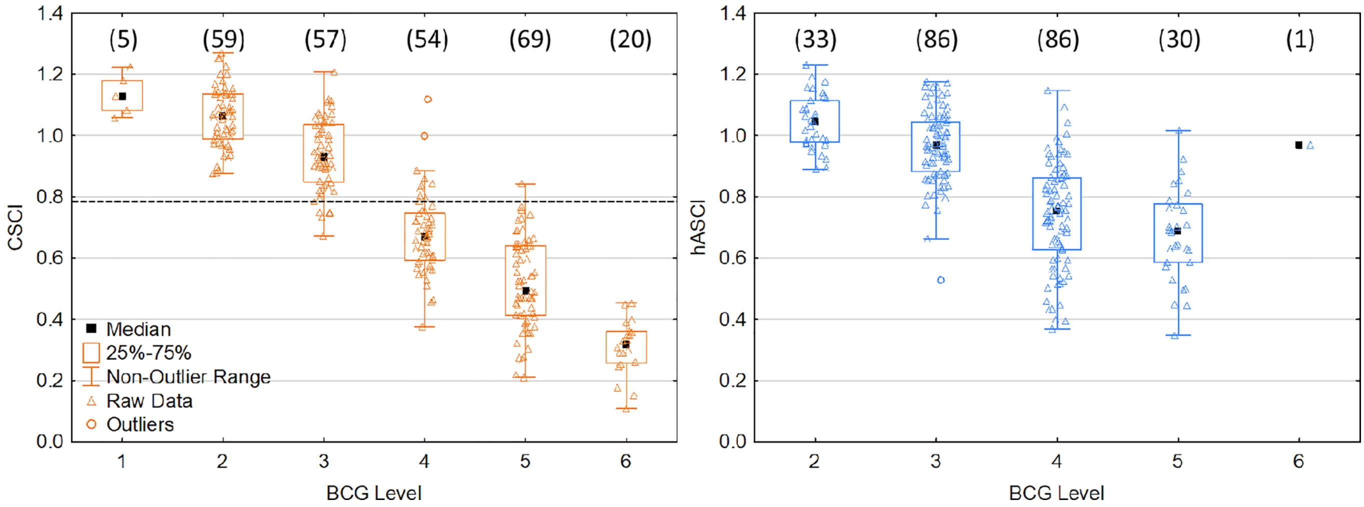

Fig. 4. Distributions of CSCI and hASCI values by BCG levels. Boxes show medians (black squares), quartiles (boxes), and non-outlier extremes (whiskers). Individual sample values are shown and sample sizes are given in parenthese above the boxes. The horizontal hatched line indicates a CSCI value of 0.79, the 10th percentile of reference site scores.

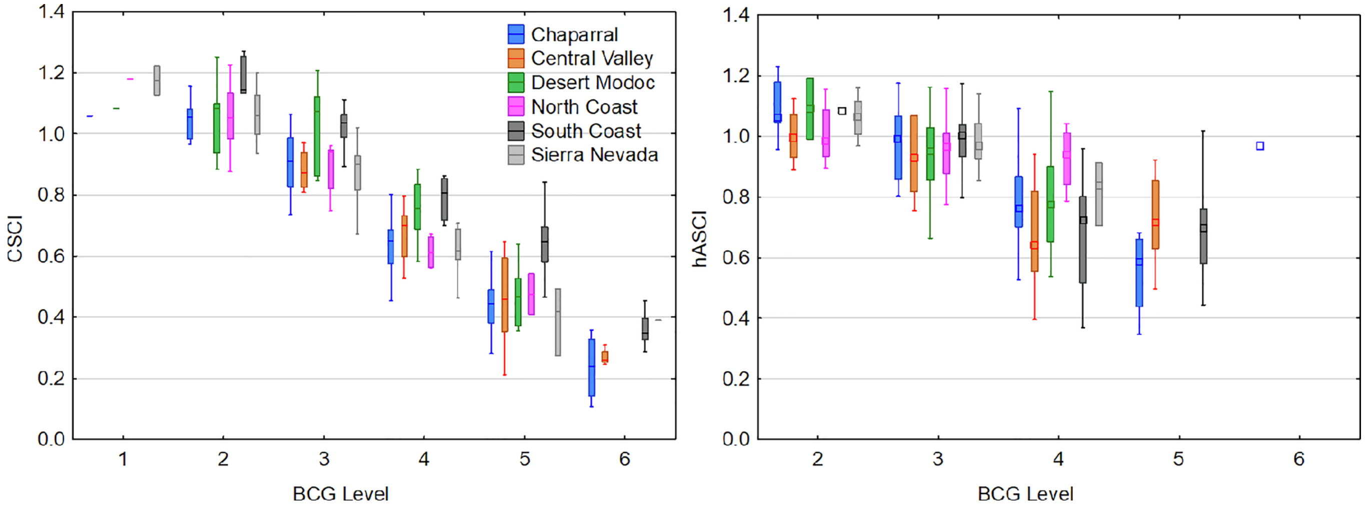

Fig. 5. CSCI and hASCI distributions in relation to final BCG levels by PSA region. Boxes show medians, quartiles (boxes), and non-outlier extremes (whiskers).

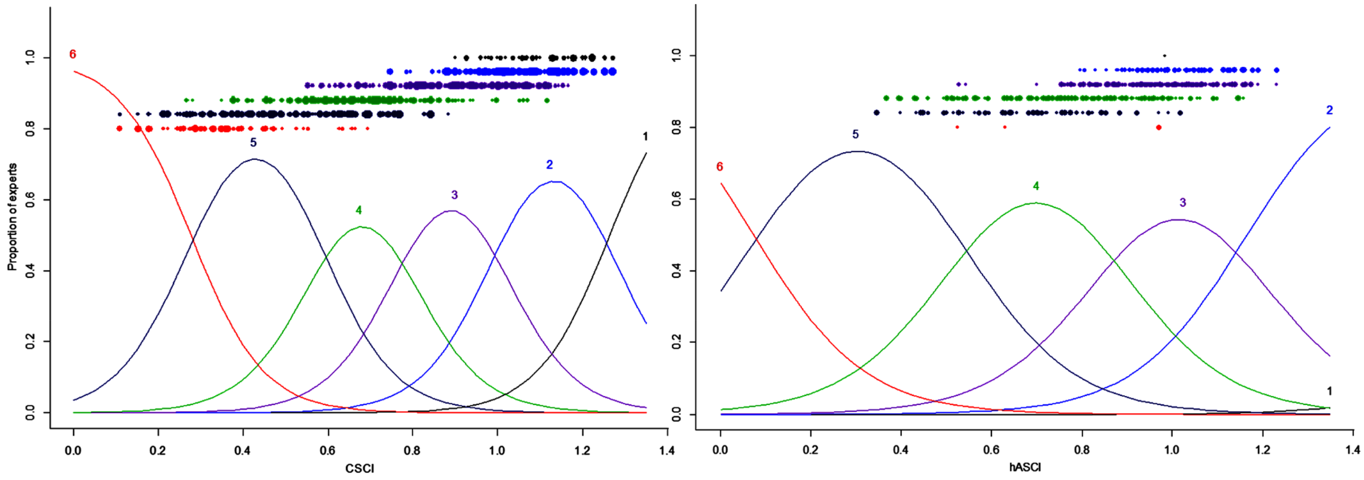

Fig. 6. Proportional odds diagram relating BCG levels to the CSCI (left) and hASCI (right). Curves represent the modeled proportion of expert ratings that would be expected for a sample with a given CSCI/hASCI score. The points above the curve reflect actual ratings, with marker size indicating relative frequency and color indicating BCG level corresponding to the curves.

public as well as agency staff) that may lack familiarity with many of the concepts underpinning numeric bioassessment indices and find the technical and conceptual complexity of the tools a potential barrier to their adoption or use.

California wadeable stream BCG models showed close agreement among experts within each assemblage. High agreement among experts has also been seen in BCG exercises conducted elsewhere in North America (Hausmann et al., 2016; Gerritsen et al., 2017) and with other expert processes, including: the evaluation of marine benthic invertebrates (Weisberg et al., 2008; Teixeira et al., 2010), marine sediment quality (Bay et al., 2007; Bay and Weisberg, 2012), and fecal contamination (Cao et al., 2013). BCG levels determined by the experts differentiated BMI and algal condition along gradients of TN and TP, specific conductivity, and agricultural and urban land use, all of which reflect common stressors associated with the ambient human disturbance gradients of these regions (California Surface Water Ambient Monitoring Program (SWAMP), 2015). Our modification of the typical BCG model development included incorporation of a modified Delphi approach in which first round BCG scoring and revision was kept anonymous. This helped reduce bias from strong individual expert panelist opinion (Nair et al., 2011) compared to previous applications of BCG model development.

Several states in the USA have developed expert-driven BCG models to support biological threshold selection based on percentiles of reference population index values; e.g., Minnesota, where BCG models have supplemented state-supported bioassessment indices for threshold development (Gerritsen et al., 2017). California has developed sound monitoring tools to measure biological condition in streams, has a peer-reviewed index for macroinvertebrates (CSCI, Mazor et al., 2016) and recently developed one for algae (ASCI, S. Theroux, pers. comm.). The CSCI has been widely implemented for reporting waterbody status, evaluating restoration, and in wastewater and stormwater permitting. In California, BMI and Algae BCG level 1–6 narratives can facilitate communication among the Water Board and public stakeholders by describing the implications of statistically based thresholds and providing alternatives in terms of specific loss of structure and function (Davies and Jackson, 2006). For example, the 10th percentile reference of CSCI (0.79), the preferred target to protect biological integrity in the proposed San Diego Regional Water Board bio-objectives policy (San Diego Regional Board, 2020), corresponds to BMI BCG level 3 (Fig. 4), in which anticipated changes include: “some changes in structure due to loss of some rare native taxa; shifts in relative abundance of taxa, although sensitive–ubiquitous taxa are common and abundant; ecosystem functions are fully maintained through redundant attributes of the system” (SI Table 2). This narrative provides the public a description of the biological condition protected at this reference percentile.

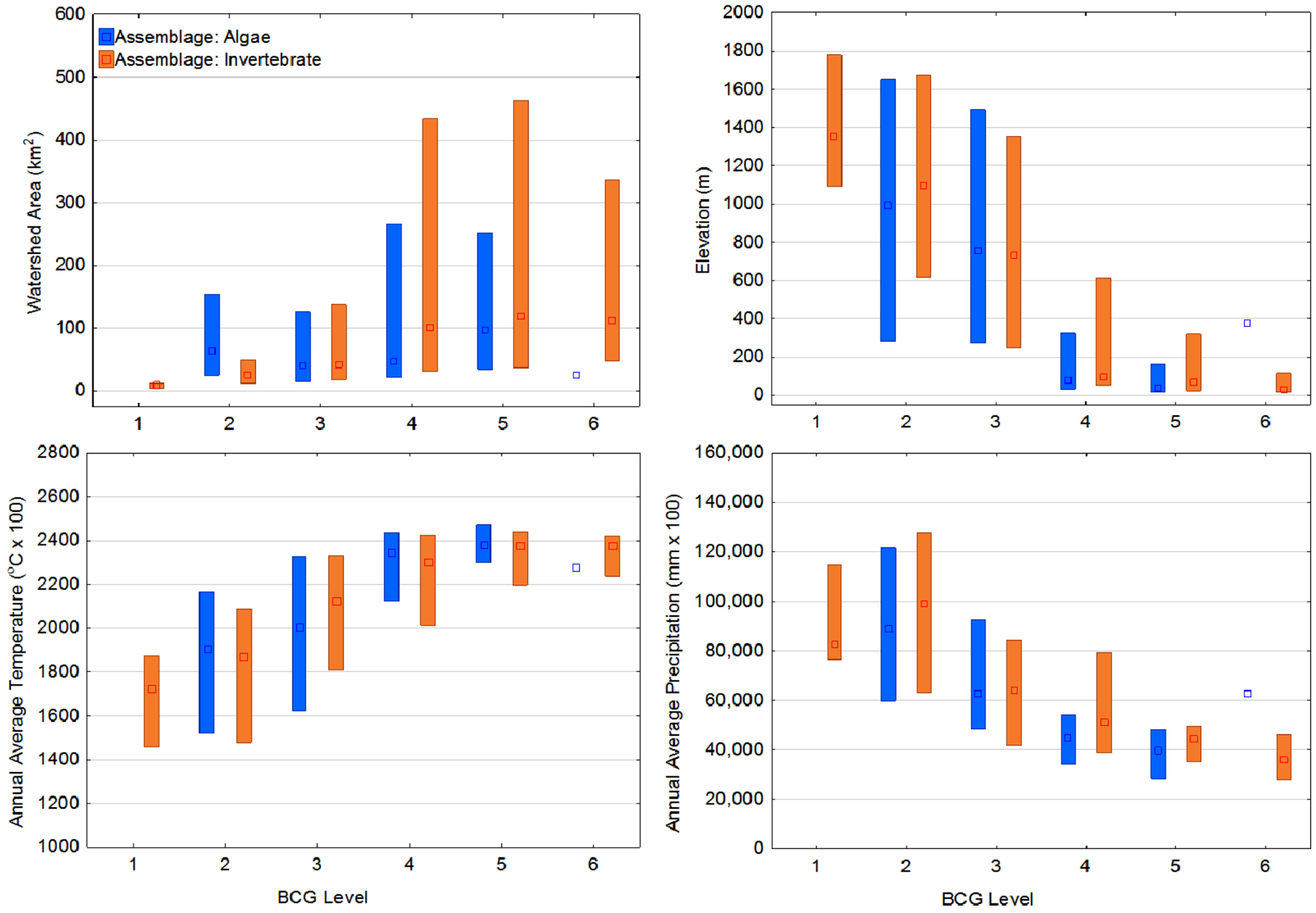

Fig. 7. Distributions of selected environmental variables in relation to macroinvertebrate and algal final BCG levels. Variables include watershed area, site elevation, modeled annual average site temperature, and modeled annual average site precipitation. Boxes show medians (squares) and intra-quartile ranges.

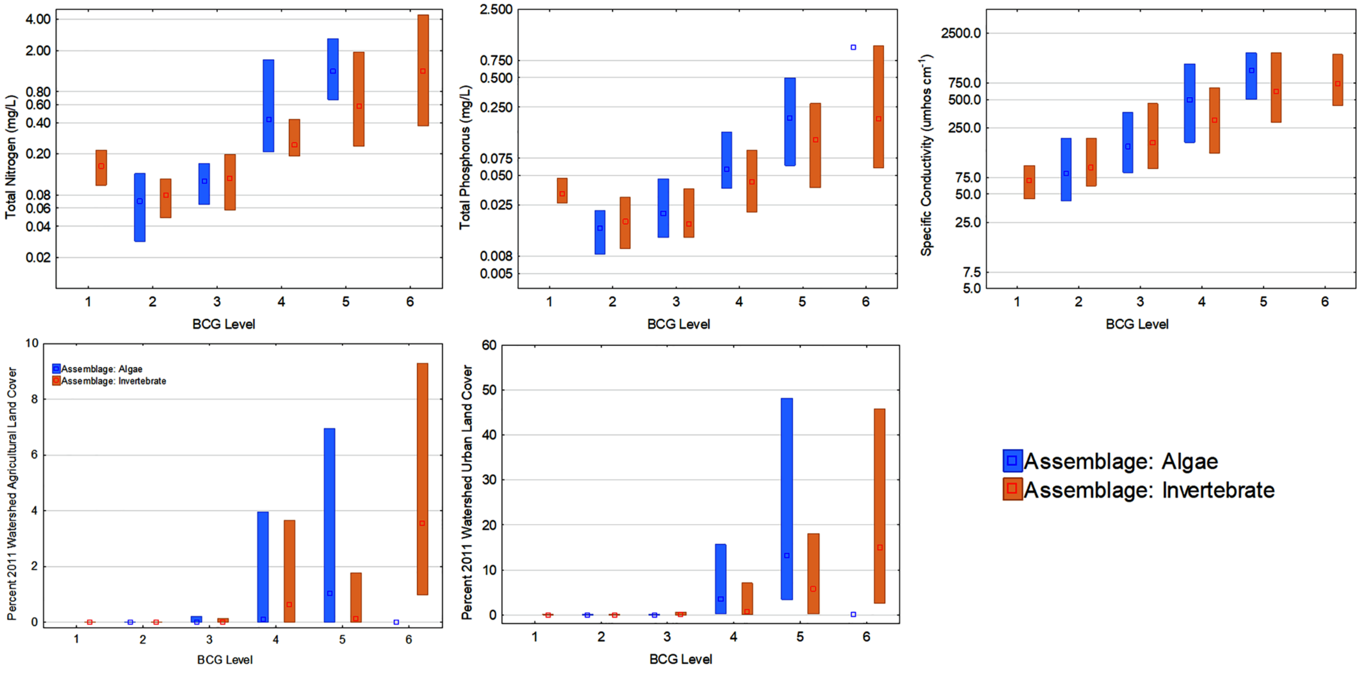

Fig. 8. Distributions of total nitrogen, total phosphorus, specific conductivity, percent agricultural land cover and percent urban land cover (2011) in site catchments by final BCG levels. Boxes show medians and intra-quartile ranges.

The consensus BCG levels assigned by the expert panels were in concordance with the CA bioassessment index scores (strongly with the CSCI, but also with the hASCI). This is not surprising given that the BCG level assignments are derived from a subset of the same monitoring program dataset, developed using a uniform set of protocols, and representing a similar range of conditions (Davies and Jackson, 2006; Mazor et al., 2016). The primary difference being that the BCG uses expert knowledge rather than statistical predictions to infer deviation from the range of natural variability (United States Environmental Protection Agency (USEPA), 2016). Each approach has its advantages and together they can be considered complementary for assessment and threshold setting. Bioassessment indices, particularly those based on O/E approaches, can account for some site-specific variability in natural gradients and thus minimize potential biases in scores among ecoregions (Hawkins et al., 2010). The level of natural variability inherent in states as large and as topographically diverse as California may have exceeded the experts' ability to account for such high variability during the scoring process. Statistically-based bioassessment indices have been criticized for being ambiguous at the index level and for arbitrarily combining metrics (Suter, 1993), problems that are not inherent in expert derived models (Gerritsen et al., 2017). Here, the correspondence between BMI and algal BCG and their corresponding indices may reduce these concerns. BCG levels may be better able to resolve meaningful differences at the extremes of the disturbance gradient, which indices sometimes poorly discern. They can certainly be used to describe the ecological conditions and consequences of sites in those settings. This may be particularly useful for states considering management options such as tiered aquatic life uses (Yoder and Rankin, 1995a,b; Davies and Jackson, 2006). This clearer resolution of extremes was evident in the comparison of BCG scoring to the CSCI more than for the hASCI.

BCG levels based on BMI showed broad distribution across the entire range of CSCI scores, indicating greater discrimination, whereas algal BCG levels were more compressed within a subset of the hASCI's scoring range. This occurred because few algae taxa were designated as taxa attribute I and none as attribute VI indicators, suggesting a limitation in information about the algal taxa to reliably characterize specialist or exotic taxa (Table 2). Moreover, no algal experts assigned sites to BCG level 1 and only 1 to level 6. This trend has been observed in previous algal BCG efforts as well, including the New Jersey diatom BCG (Charles et al., 2019) where no sites were assigned to levels 1 and 6, and a combined Mid-Atlantic region algal BCG effort in which only a handful of algal taxa received a BCG Level 1 or Level 6 assignment (Hausmann et al., 2016). The under representation of sites classified as BCG Level 1 and 6 using algae may be due to a variety of factors including: differential responses of algae to stress, more tolerance in general among algae, incomplete autecological understanding of algal taxa, incompletely describing the range of least to most disturbed conditions for the algae, or bias associated with the BCG process in scoring algal assemblages. The latter includes thinking that endemic or non-native taxa be required to be present to score in levels 1 or 6 (i.e., confounding BCG taxa attributes with scoring levels). There was some feeling that this also influenced BMI scoring and is an area for greater exploration within the BCG development community. Historical impacts and natural variability may constrain assigning level 1 scores, but some experts felt there should have been more sites meeting the spirit of level 1 than were assigned.

Although the BCG framework developed in this study complements bioassessment indices as a communication and decision support tool for policy development, we recommend some guidelines for its use in supporting management decisions in California. First, unlike in other states where a quantitative logic model was developed from expert decisions to provide an assessment tool to replicate expert scoring for new samples (e.g. Gerritsen et al., 2017), we did not develop our CA BCG framework as a substitute or alternative biological scoring tool to standardized indices already incorporated into management programs, like the CSCI or ASCI. This was primarily because of considerable opposition within the management community to adopting a BCG model for California over fears that it would undermine efforts to standardize assessments and create conflict over inconsistent or biased selection of indices. Emphasizing that the main use of the BCG framework is as a decision support tool for communication is essential in assuaging this opposition.

Second, the BCG should not be used to define the highest ecological potential for any site. For example, if a site biological index score falls within BCG level 4, its potential BCG level could be significantly higher (i.e., 1–4), but the site may also be constrained by factors that are difficult to control such as urban development. So, the restorability of a site (i.e., the feasibility of attaining a higher BCG level) cannot be inferred from the score alone.

Third, for many stakeholders, the concept of a BCG model is linked to that of tiered uses, as this has been a common application (Davies and Jackson, 2006; Gerritsen et al., 2017). However, tiered biological objectives were not a considered application in California, which led to confusion as to the intended use of the model. However, as we demonstrate, a BCG model has potential value to stream biological assessment outside of setting tiered objectives. Communicating this use of the BCG model to stakeholders was important.

The BCG supports establishment of biological integrity water quality goals in California wadeable streams by providing an understanding of the ecological implications of different index thresholds. The BCG results can also help communicate to stakeholders the appropriateness of specific state biological index values as valid representations of biological integrity goals. Both needs are aided by the BCG descriptions of biological structure and function protected or lost for different values of frequently used indices.

CRediT authorship contribution statement

Michael J. Paul: Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Resources, Supervision, Writing - review & editing, Visualization, Funding acquisition. **Ben Jessup:** Conceptualization, Methodology, Formal analysis, Resources, Writing - original draft, Writing - review & editing, Visualization. **Larry R. Brown:** Investigation, Writing - original draft, Writing - review & editing. **James L. Carter:** Investigation, Writing - original draft, Writing - review & editing. **Marco Cantonati:** Investigation, Writing - original draft, Writing - review & editing. **Donald F. Charles:** Investigation, Writing - original draft, Writing - review & editing. **Jeroen Gerritsen:** Investigation, Writing - original draft, Writing - review & editing. **David B. Herbst:** Investigation, Writing - original draft, Writing - review & editing. **Rosalina Stancheva:** Investigation, Writing - original draft, Writing - review & editing. **Jeannette Howard:** Investigation, Writing - original draft. **Bill Isham:** Investigation, Writing - original draft, Writing - review & editing. **Rex Lowe:** Investigation, Writing - original draft, Writing - review & editing. **Raphael Mazor:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **Patina K. Mendez:** Investigation, Writing - original draft, Writing - review & editing. **Peter Ode:** Investigation, Writing - original draft, Writing - review & editing. **Alison O'Dowd:** Investigation, Writing - original draft, Writing - review & editing. **John Olson:** Investigation, Writing - original draft, Writing - review & editing. **Yangdong Pan:** Investigation, Writing - original draft, Writing - review & editing. **Andrew C. Rehn:** Investigation, Writing - original draft, Writing - review & editing. **Sarah Spaulding:** Investigation, Writing - original draft, Writing - review & editing. **Martha Sutula:** Conceptualization, Methodology, Resources, Supervision, Writing - original draft, Supervision, Writing - review & editing, Funding acquisition, Project administration. **Susanna Theroux:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors wish to thank Cajun James and Sierra Pacific Industries for sharing data. We would also like to thank Kenneth McCune and Erik Leppo for assisting with data compilation, management, and preparation. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. This work was supported by the California State Water Resources Control Board and the United States Environmental Protection Agency.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2020.106618.

References

Agresti, A., 2002. Categorical Data Analysis, 2nd ed., Wiley series in probability and statistics. Wiley-Interscience, New York.

Barbour, M.T., Gerritsen, J., Snyder, B.D., & Stribling, J.B., 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish. EPA 841-B-99-002. US Environmental Protection Agency, Office of Water, Washington, DC. 1999.

Bay, S., Berry, W., Chapman, P.M., Fairey, R., Gries, T., Long, E., MacDonald, D., Weisberg, S.B., 2007. Evaluating consistency of best professional judgment in the application of a multiple lines of evidence sediment quality triad. Integr. Environ. Assess. Manage. 3, 491–497.

Bay, S.M., Weisberg, S.B., 2012. Framework for interpreting sediment quality triad data. Integr. Environ. Assess. Manage. 8, 589–596. https://doi.org/10.1002/ieam.118.

Borja, A., Franco, J., Muxika, I., 2004. The Biotic Indices and the Water Framework Directive: the required consensus in the new benthic monitoring tools. Mar. Pollut. Bull. 48, 405–408. https://doi.org/10.1016/j.marpolbul.2003.10.024.

California Surface Water Ambient Monitoring Program (SWAMP). 2015. The Perennial Streams Assessment (PSA): An Assessment of Biological Condition Using the New California Stream Condition Index. SWAMP-MM-2015-0001. Surface Water Ambient Program, California Department of Fish and Wildlife, Sacramento, CA. (www.waterboards.ca.gov/water_issues/programs/swamp/bioassessment/docs/psa_memo_121015.pdf, Accessed 1/23/2020).

SWAMP. 2018. An Index to Measure the Quality of Physical Habitat in California Wadeable Streams. SWAMP-TM-2018-0005. Surface Water Ambient Program, California Department of Fish and Wildlife, Sacramento, CA.

Cao, Y., Hagedorn, C., Shanks, O.C., Wang, D., Ervin, J., Griffith, J., Layton, B., McGee, C.D., Riedel, T., Weisberg, S., 2013. Towards establishing a human fecal contamination index in microbial source tracking. Int. J. Environ. Eng. Syst. 4, 46–58.

Charles, D.F., Tuccillo, A.P., Belton, T.J., 2019. Developing nutrient criteria for rivers and streams: diatoms and the biological condition gradient. Ecol. Ind. 96, 258–269.

Davies, S.P., Jackson, S.K., 2006. The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. Ecol. Appl. 16, 1251–1266.

Davis, W.S., & Simon, T.P., 1995. Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making. CRC Press.

Fetscher, A.E., Busse, L., & Ode, P., 2009. Standard Operating Procedures for Collecting Stream Algae Samples and Associated Physical Habitat and Chemical Data for Ambient Bioassessments in California (No. Bioassessment SOP 002). California State Water Resources Control Board Surface Water Ambient Monitoring Program (SWAMP).

Fetscher, A.E., Stancheva, R., Kociolek, J.P., Sheath, R.G., Stein, E.D., Mazor, R.D., Ode, P.R., Busse, L.B., 2014. Development and comparison of stream indices of biotic integrity using diatoms vs. non-diatom algae vs. a combination. J. Appl. Phycol. 26, 433–450.

Gerritsen, J., Bouchard, R.W., Zheng, L., Leppo, E.W., Yoder, C.O., 2017. Calibration of the biological condition gradient in Minnesota streams: a quantitative expert-based decision system. Freshw. Sci. 36, 427–451. https://doi.org/10.1086/691712.

Hausmann, S., Charles, D.F., Gerritsen, J., Belton, T.J., 2016. A diatom-based biological condition gradient (BCG) approach for assessing impairment and developing nutrient criteria for streams. Sci. Total Environ. 562, 914–927. https://doi.org/10.1016/j.scitotenv.2016.03.173.

Hawkins, C.P., Olson, J.R., Hill, R.A., 2010. The reference condition: predicting benchmarks for ecological and water-quality assessments. J. N. Am. Benthol. Soc. 29, 312–343. https://doi.org/10.1899/09-092.1.

Hilsenhoff, W.L., 1987. An improved biotic index of organic stream pollution. Great Lakes Entomol. 20, 31–39.

Lange-Bertalot H., Hofmann G., Werum M. & Cantonati M. 2017. Freshwater Benthic Diatoms of Central Europe: Over 800 Common Species Used in Ecological Assessment. M. Cantonati, M.G. Kelly & H. Lange-Bertalot (Eds.): 942 pp. Koeltz Botanical Books (ISBN 978-3-946583-06-6). English edition with updated taxonomy and added species of Hofmann et al., 2013.

Mazor, R.D., Rehn, A.C., Ode, P.R., Engeln, M., Schiff, K.C., Stein, E.D., Gillett, D.J., Herbst, D.B., Hawkins, C.P., 2016. Bioassessment in complex environments: designing an index for consistent meaning in different settings. Freshw. Sci. 35, 249–271. https://doi.org/10.1086/684130.

Nair, R., Aggarwal, R., Khanna, D., 2011. Methods of formal consensus in classification/diagnostic criteria and guideline development. Semin. Arthritis Rheum. 41, 95–105. https://doi.org/10.1016/j.semarthrit.2010.12.001.

Ode, P.R., Rehn, A.C., Mazor, R.D., Schiff, K.C., Stein, E.D., May, J.T., Brown, L.R., Herbst, D.B., Gillett, D., Lunde, K., Hawkins, C.P., 2016. Evaluating the adequacy of a reference-site pool for ecological assessments in environmentally complex regions. Freshw. Sci. 35, 237–248. https://doi.org/10.1086/684003.

Olson, J.R., Hawkins, C.P., 2012. Predicting natural base-flow stream water chemistry in the western United States: predicting water chemistry. Water Resour. Res. 48, W02504. https://doi.org/10.1029/2011WR011088.

Poff, N.L., Pyne, M.I., Bledsoe, B.P., Cuhaciyan, C.C., Carlisle, D.M., 2010. Developing linkages between species traits and multiscaled environmental variation to explore vulnerability of stream benthic communities to climate change. J. N. Am. Benthol. Soc. 29, 1441–1458.

Poikane, S., Kelly, M.G., Cantonati, M., 2016. Benthic algal assessment of ecological status in European lakes and rivers: challenges and opportunities. Sci. Total Environ. 568, 603–613. https://doi.org/10.1016/j.scitotenv.2016.02.027.

Potapova, M., Charles, D.F., 2007. Diatom metrics for monitoring eutrophication in rivers of the United States. Ecol. Ind. 7, 48–70. https://doi.org/10.1016/j.ecolind.2005.10.001.

R Core Team. 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Reynoldson, T.B., Norris, R.H., Resh, V.H., Day, K.E., Rosenberg, D.M., 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. J. N. Am. Benthol. Soc. 16, 833–852. https://doi.org/10.2307/1468175.

Rott, E., Van Dam, H., Pfister, P., Pipp, E., Pall, K., Binder, N., & Ortler K. 1999. Indikationslisten für Aufwuchsalgen. Teil 2: Trophieindikation, geochemische Reaktion, toxikologische und taxonomische Anmerkungen. Publ. Wasserwirtschaftskataster, BMLF, 1-248.

San Diego Regional Board. 2020. Biological Objectives Basin Plan Amendment. San Diego Regional Water Board, San Diego, CA, (https://www.waterboards.ca.gov/sandiego/water_issues/programs/basin_plan/bio_objectives/, Accessed 1/23/2020).

Steedman, R.J., 1994. Ecosystem health as a management goal. J. N. Am. Benthol. Soc. 13, 605–610. https://doi.org/10.2307/1467856.

Stevenson, R.J., Pan, Y., Manoylov, K., Parker, C.A., Larsen, D.P., Herlihy, A.T., 2008. Development of diatom indicators of ecological conditions for streams of the western United States. J. N. Am. Benthol. Soc. 27, 1000–1016.

Stoddard, J.L., Larsen, P., Hawkins, C.P., Johnson, R.K., Norris, R.H., 2006. Setting expectations for the ecological condition of running waters: the concept of reference condition. Ecol. Appl. 16, 11.

Suter, G.W., 1993. A critique of ecosystem health concepts and indexes. Environ. Toxicol. Chem. 12, 1533–1539. https://doi.org/10.1002/etc.5620120903.

Teixeira, H., Borja, Á., Weisberg, S.B., Ananda Ranasinghe, J., Cadien, D.B., Dauer, D.M., Dauvin, J.-C., Degraer, S., Diaz, R.J., Grémare, A., Karakassis, I., Llansó, R.J., Lovell, L.L., Marques, J.C., Montagne, D.E., Occhipinti-Ambrogi, A., Rosenberg, R., Sardá, R., Schaffner, L.C., Velarde, R.G., 2010. Assessing coastal benthic macrofauna community condition using best professional judgement – Developing consensus across North America and Europe. Mar. Pollut. Bull. 60, 589–600. https://doi.org/10.1016/j.marpolbul.2009.11.005.

United States Environmental Protection Agency (USEPA), 2016. A Practitioner's Guide to the Biological Condition Gradient: A Framework to Describe Incremental Change in Aquatic Ecosystems (No. EPA-842-R-16-001). U.S. Environmental Protection Agency, Washington, D.C.

USEPA, 2002. Summary of Biological Assessment Programs and Biocriteria Development for States, Tribes, Territories, and Interstate Commissions: Streams and Wadeable Rivers (No. EPA-822-R-02-048). U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

Venables, W.N., & Ripley, B.D., 2002. Modern Applied Statistics with S, Fourth Edition. ed. Springer.

Weisberg, S., Thompson, B., Ranasinghe, A., E. Montagne, D., Cadien, D., Dauer, D., Diener, D., Oliver, J., Reish, D., Velarde, R., & Word, J., 2008. The level of agreement among experts applying best professional judgment to assess the condition of benthic infaunal communities. Ecological Indicators 8, 389–394. https://doi.org/10.1016/j.ecolind.2007.04.001.

Yoder, C.O., & Rankin, E.T., 1995a. The role of biological criteria in water quality monitoring, assessment and regulation. Ohio EPA Technical Report MAS/1995-1-3.

Yoder, C.O., & Rankin, E.T., 1995b. Biological criteria program development and implementation in Ohio, pp. 109-144 (Chapter 9). In W.S. Davis and T. Simon (eds.). Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making. Lewis Publishers, Boca Raton, FL.

Yoder, C.O., Barbour, M.T., 2009. Critical technical elements of state bioassessment programs: a process to evaluate program rigor and comparability. Environ. Monit. Assess. 150, 31–42. https://doi.org/10.1007/s10661-008-0671-1.

Yuan, L.L., 2006. Estimation and Application of Macroinvertebrate Tolerance Values (No. EPA/600/P-04/116F). U.S. Environmental Protection Agency, ORD, National Center for Environmental Assessment, Washington, D.C.