

A non-predictive algal index for complex environments

Susanna Theroux, Raphael Mazor, Marcus Beck, Peter Ode, Martha Sutula, Eric Stein

Prepared for: Ecological Indicators

DRAFT

ABSTRACT

Algae are sensitive bioindicators that provide unique insights into freshwater biological condition. However, their complex species assemblages can impede the development of biological indices that require a comprehensive taxonomic dataset and a thorough catalog of species' autoecological attributes. Intensive sampling campaigns have amassed over a decade's worth of algae taxonomy records for California's wadeable streams, providing a robust algae taxonomic dataset to support the creation of a statewide biological index. Out of over 2500 algae taxonomy samples, we identified 416 reference sites and constructed a series of predictive models to use geographic location to set site-specific, reference-based expectations. We developed two separate components to the index: a measure of taxonomic completeness (the ratio of observed-to-expected taxa, O/E), and a multimetric index (MMI) to evaluate ecological structure. We developed multiple versions of each algal index, individually for diatoms, soft-bodied algae, and a hybrid approach using both algal assemblages. All indices were evaluated using either species or genus-level taxonomy data. We evaluated index performance across the state for precision, accuracy, responsiveness, and regional bias. Final index performance varied widely among all assemblages: the best performing O/E index was a diatom-only index, whereas all three algal assemblages yielded high-performing MMIs. Surprisingly, none of the best-performing MMIs incorporated modeled metrics, as predictive modeling did not improve metric performance. Combined O/E and MMIs did not perform better than MMIs alone, resulting in the selection of the MMI as the optimal indices for assessing biological health for California wadeable streams. Metrics in final MMIs relied on trait attribute data derived specifically for California taxa, highlighting the importance of a comprehensive trait attributes database for constructing a multi-metric index. Additionally, hybrid MMIs were the most sensitive to anthropogenic stressor gradients, emphasizing the value in combining multiple algal assemblages to gain a holistic assessment of biological condition.

INTRODUCTION

Algae have been used as ecological indicators for over a century (Kolkwitz and Marsson 1908; Stevenson, 2014), and in recent years there has been a renewed focus on the rich history of algal autoecological research for the development of algae-based biological indices. Algae are powerful ecological indicators due to their rapid growth rate and ability to respond to perturbations on the scale of hours instead of weeks (), their sensitivity to nutrient and chemical stressors (Hering et al., 2006; Stevenson, 2014), and their diverse phylogenetic assemblages that enable species-specific responses to a catalog of environmental conditions (Meyer et al., 2007). However, these diverse species assemblages have also acted as an obstacle to developing algae-based biological indices: many algae species lack trait attributions, therefore limiting the development of trait-based biological indices.

California presents a unique challenge for developing an algae-based biological index. California has a number of diverse ecoregions (Ode et al., 2016), ranging from temperate rainforests to deserts to grasslands (Omernik 1987), which can prove problematic in the journey to develop a universal index for all regions. Additionally, California is home to a number of endemic species (Erman 1996, Moyle and Randall 1996, Moyle et al. 1996; Sork et al., 2016), including algae (Stancheva and Sheath, 2016), with new novel species are identified annually. Novel taxa often lack trait attributes and known autoecological characteristics, not to mention the taxonomic burden to reliably identify rare or novel taxa. Luckily, California also benefits from extensive annual sampling campaigns in wadeable stream environments, efforts that have resulted in rich taxonomic and ecological datasets for both algae and benthic macroinvertebrates (BMIs). It is thanks to these sampling campaigns that California is now poised to develop a statewide algal index.

One of the primary concerns for the development of all biological indices, and especially those that are created for diverse environments, is that the index scores provide consistent meaning in variable environmental settings. To accomplish this, index development must rely on a robust and clearly-defined definition of the “reference-condition”. The reference condition approach (RCA; Reynoldson et al., 1997) uses minimally-disturbed sites (“reference” sites) to establish biological expectations at a test site. For a statewide biological index, universal reference-condition screening criteria are applied for all sites, regardless of geographic setting. In California, previous efforts to identify reference sites for ambient monitoring and biological index development (Ode et al., 2016; Mazor et al., 2016) have yielded set of screening criteria that incorporate measures of anthropogenic disturbance at a variety of spatial scales. These robust screening criteria help to evaluate a biological index’s degree of regional bias, or its ability to assess ecological condition without influence from natural environmental gradients.

Once equipped with a robust definition of the reference condition, there are multiple options for developing a biological index. A predictive modeling approach to index development allows for an opportunity to account for natural variability in biological communities by setting site-specific, reference-based expectations (Hawkins et al., 2010; Mazor et al., 2016). Pioneered by the River Invertebrate Prediction and Classification System (RIVPACS; Wright et al., 2000),

biological indices that follow this approach grade test sites as a deviation from reference-site expectations, often reported as a ratio of observed versus expected taxa (O/E) (Moss et al. 1987, Hawkins et al. 2000, Wright et al. 2000; Ode et al., 2008). Recently, studies have also focused on developing predictive multi-metric indices (MMIs) that rely on autoecological traits of biological communities instead of only species identity (Cao et al. 2007, Pont et al. 2009, Vander Laan and Hawkins 2014). In California, the use of this predictive modeling approach has been successful for the development of a BMI biological index. The California Stream Condition Index (CSCI) paired both a predictive O/E and a predictive MMI to derive a two-component index (Mazor et al., 2016). This study relied on hundreds of “minimally disturbed” reference sites (Ode et al., 2016b) to develop robust, predictive models that account for geographic setting in assessing biological condition.

Efforts to develop predictive algal indices have had varying success. Cao et al. (2007) successfully developed both an O/E and MMI for diatoms of Idaho streams. Feio et al., (2009) saw comparable performance between a predictive diatom index and non-predictive multi-metric indices, which was improved upon in Feio et al. (2014) to develop AQUAFLORE for Portuguese streams. Mazor et al. (2006) successfully developed a RIVPACS index for periphyton communities in the Fraser River. Other attempts at a predictive diatom index have met with less success, credited to greater temporal variability in diatom communities than BMI communities and high variability in reference site scores (Chessman et al., 1999). A predictive, diatom-only index has already been attempted for the Central Coast of California, but the predictive O/E index suffered from low precision and accuracy (Ritz et al., 2010).

The algal indices generated as part of this study build upon the previous regionally-specific algal indices that were developed for this region: an index of biotic integrity (IBI) for the Eastern Sierra Nevada region (Blinn and Herbst, 2003) designed using both diatoms and soft-bodied algae, and the Southern California Index of Biotic Integrity (SoCal IBI; Fetscher et al., 2014) that relied on a series of non-predictive, multi-metric indices calculated using diatoms alone, soft-bodied algae alone, or a hybrid of both assemblages. Additionally, we developed both a measure of taxonomic completeness (i.e., an O/E index) and a measure of ecological condition (i.e., an MMI) after the CSCI (Mazor et al., 2016). We tested each component of the algal index using diatoms alone, soft-bodied algae alone, or a hybrid using both assemblages after the algal IBI (Fetscher et al., 2014). We evaluated the performance of each permutation of the algal index across both environmental and disturbance gradients, with a concerted focus on selecting final indices with high precision and accuracy and low regional bias. Herein we provide the performance criteria for the various permutations of the algal indices to inform management decisions and condition assessment statewide.

METHODS

Study region: California is a species-rich biogeographic region that offers a complex history of geological change and complex modern-day geographic settings and climate fluctuations (Calsbeek et al., 2003; Sork et al., 2016). Across California’s 424,000 km², there are not only broad variations in climate, elevation, and geology, but also land-use, urban and agricultural

development (Sleeter et al., 2011). For evaluating ecoregion responses, we divided the state into 6 ecoregions (North Coast, Central Valley, Chaparral, Sierra Nevada, and South Coast) based on modified ecoregional (Omernik 1987) and hydrological boundaries (Ode et al., 2016).

Data compilation: We compiled algae taxonomic data from multiple federal, state, and regional monitoring programs in California, resulting in a dataset of 1943 sampling sites and 2588 unique sampling events in wadeable streams. All sampling events followed a standardized periphyton sampling protocol (Ode et al., 2016a). Briefly, a reach of 150 m was subdivided into eleven transects. A “multihabitat method” was employed to objectively collect subsamples of algal specimens quantitatively from a known surface area over a representative sample of stream substrata (Fetscher et al., 2009). Additionally, this quantitative fraction may be supplemented by a live, “qualitative” fraction that was used to help identify SBA species with reproductive structures. Algae samples were composited and proportioned into diatom and soft-bodied algae aliquots for laboratory analysis (Stancheva et al., 2015). Sampling events in close proximity (within 300m) were treated as repeat samples from a single site. For sites with multiple years of sampling, we focused on the most recent sample for index development and used previous sampling events for performance analyses (e.g. standard deviation of index scores within a single sampling site).

Data curation: Samples with fewer than 300 diatom valves were excluded from diatom and hybrid analyses. Due to the uneven inclusion of the qualitative fraction for soft-bodied algae taxonomy samples, only quantitative taxonomy data was included. All taxonomy results were converted from count (diatoms) or biovolume (soft bodied algae) data to presence/absence data in anticipation of future assemblage data being derived from DNA-based methods. All unpublished species names were subjected to a name-harmonization with AlgaeBase (algaebase.org) and Biodata taxa names (<http://aquatic.biodata.usgs.gov>) to remove ambiguous identifiers; harmonized name lists are provided at github.com/sccwrp.

Environmental variables: We assembled environmental data from multiple sources, including GIS-derived variables from the National Hydrography Dataset Plus (NHD, <http://www.horizon-systems.com/nhdplus>), the National Landcover Data Set (<http://www.epa.gov/mrlc/nlcd-2006.html>), the National Inventory of Dams (<http://geo.usace.army.mil>), Mineral Resource Data System (<http://tin.er.usgs.gov/mrds>), and predicted specific conductance from Olson and Hawkins (2012). These data included measures of climate, elevation, geology, land cover, land use, road density, hydrologic alteration and mining activities (**Table 2**). We used environmental variables that characterize immutable natural gradients as candidate predictors for the O/E and MMI models, whereas environmental variables influenced by anthropogenic factors were used for screening reference sites and for assessing index performance along stressor gradients (Mazor et al., 2016; Ode et al., 2016b).

Classifying sampling sites into disturbance classes: We assessed the influence of anthropogenic activity using measures of surrounding land use as well as local habitat data after Ode et al. (2016b). We followed a “least disturbed” reference concept after Stoddard et al., 2006 for the identification of “reference” sites, and identified high-activity sites after Mazor

et al., 2016 (**Table 1**). We used reference sites for calibrating our O/E and MMIs and for determining the biological composition of sites with minimal human disturbance (Hawkins et al., 2010; Mazor et al., 2016). Reference site screening thresholds closely followed those used in the construction of the CSCI (Mazor et al., 2016; Ode et al., 2016b), with the exception that conductivity was not used to eliminate a site from the reference pool (**Table 1**). Stressed, or high-activity sites, were used in scoring MMIs as well as in evaluating performance of both O/E and MMIs (see below). Any sites that did not pass the reference or stressed screening thresholds were included in the “intermediate” site pool. We further divided each dataset into a calibration (80%) and validation (20%) subset and stratified assignment by subregion for equal representation of different environmental settings.

O/E index construction: The O/E index is a measure of taxonomic completeness that quantifies the loss of expected native taxa with increasing exposure to environmental perturbation and stressors (Hawkins, 2006). The expected number of taxa (E) is derived from the environmental setting of the sampling site, whereas the observed number of taxa (O) represents the proportion of expected taxa observed. First, we clustered reference calibration sites based on their presence-absence taxonomic composition. Second, we developed a random-forest model (Cutler et al., 2007) to predict membership within a taxonomic cluster using select environmental variables that are minimally affected by human perturbation (**Table 2**). We then used this random-forest model to predict cluster membership of test sites based on their environmental setting. The capture probability (P_c), or probability of observing a taxon at a test site, was calculated as the cluster-membership probability-weighted frequencies of occurrence summed across all clusters (Mazor et al., 2016). The sum of the capture probabilities is the expected number of taxa (E) in a sample from a site.

We performed the clustering of reference site algae populations using a presence/absence transformed data matrix and excluded all taxa occurring in < 2.5% and >95% of reference calibration sites (Hawkins et al., 2000). For each of the O/E indices, we determined an optimal number of clusters through a computational permutation (clusters ranging from 3-15, capture probability of 0.4 or 0.5) and iterated the cluster numbers and minimum capture probabilities with k-means clustering using the k-means function in the *stats* R package (R Core Team, 2013). The iteration with the best performing O/E index was selected as the final cluster number for that assemblage. The rare species excluded in the clustering step were retained for subsequent steps in the O/E development.

To select environmental variables for use in a model to predict cluster group membership, we used recursive feature elimination (RFE) as implemented in the *caret* package in R (Wing et al., 2017). In brief, we used RFE to identify the optimal number of environmental predictor variables whose model accuracy was within 1% of the best model. We included environmental variables as candidate predictors that were minimally influenced by human activity (**Table 2**). We then used the *randomForest* package (Liaw and Wiener, 2002) to construct a final 500-tree O/E model using the predictor variables selected in the RFE analysis.

Evaluating O/E index performance: We optimized the models to minimize the standard deviation (SD) of reference site O/E scores at calibration reference sites. We evaluated O/E index performance by comparing the SD of the predictive O/E index to the SD of the null O/E index (i.e., all sites are in a single cluster and capture probabilities for each taxon are the same for all sites), to ensure that the predictive index had a lower SD than the null index. Additionally, we compared the predictive O/E index SD to the highest attainable precision possible based on estimates of the standard deviation among replicate samples (SDRS; Van Sickle et al., 2005). We also evaluated the O/E index for bias by regressing O against E within reference sites, with slopes close to 1 and intercepts close to 0 indicative of better performance. We evaluated the model's ability to differentiate reference from stressed sites using a t-test as implemented in the R *stats* packages. We modified scripts written by J. Van Sickle for evaluating O/E model performance (John Van Sickle, USEPA, 2005) in addition to developing custom scripts that are available online (github.com/sccwrp). As low numbers of expected taxa have been shown to impact index performance (Mazor et al., 2016; Hamalaine et al., 2018), we evaluated O/E model sensitivity to low E values by calculating precision (standard deviation of reference calibration site scores for each index), sensitivity (proportion of high activity sites not in reference condition), and accuracy (percent of reference calibration sites above 10th percentile of reference) at each maximum E value from 0 to 30 expected taxa.

MMI construction: The MMIs were constructed in the following steps: 1) calculate metric raw values; 2) develop random forest models to predict metric values at reference sites and replace raw values with differences from predicted values if appropriate; 3) score metrics; 4) select best-performing metrics and assembled proto-MMIs; 5) assemble MMI with most frequent high-performing metrics.

Metric calculation: As in the O/E model development, we performed all metric calculations on a presence/absence data matrix. Using species-level data, all metrics were calculated as both a proportion of total taxa and total count of taxa that met specific trait attributes. For the hybrid MMI, we used a combined table including both diatom and SBA taxa, for calculating metric values. Species' trait attributes were obtained from previously published algae attribute lists (Bahls, 1993; Dam et al., 1994; Porter et al., 2008; Potapova and Charles, 2007; Spaulding et al., 2010), previously published empirically-derived traits for southern California taxa (Fetscher et al., 2014), and empirically-derived traits as part of this study. Additionally, we leveraged trait attributes developed as part the French diatom index SPI (Specific Polluosensitivity Index, (Cemagref, 1982; Coste, 1986)) as well as traits derived by a panel of algal ecologists and taxonomists for the California Biological Condition Gradient (BCG; Paul et al., *in prep*). For both the SPI and BCG trait attributes, higher numbers (range 1-5) indicate species with greater tolerance for disturbed or stressed condition. For deriving a genus-level trait attributes table, when all species within a genus shared a trait attribute with 100% fidelity, that trait attribute was assigned for the genus. For identifying "sensitive" and "tolerant" California taxa as part of this study, we used an Indicator Species Analysis as implemented in the *multipatt* function in the *indicspecies* R package (Caceres and Legendre, 2009) and classified sensitive taxa as enriched at reference sites and tolerant taxa as those taxa enriched at stressed sites. We grouped metrics into thematic categories based on their autoecological, morphological,

taxonomic, or species tolerance guilds (**Table 3**). We calculated all metrics using a combination of the R language *vegan* package (Oksanen et al., 2017) and custom R scripts (github.com/sccwrp).

Prediction of metric values at reference sites: We developed random-forest models to predict values for all metrics at reference calibration sites based on the same candidate environmental variables that were used for O/E index development (**Table 2**). We again used an RFE approach to select the best performing predictive models for each metric and created a final 500-tree random-forest model for each metric based on the predictors used in the model selected by RFE. We then used the final model to predict metric values for all sites. To evaluate how well each model predicted metric values, we regressed raw observed values against predicted values for reference sites. Slopes close to 1 and intercepts close to 0 indicate better model performance. If the pseudo- R^2 of the model (calculated as $1 - \text{mean squared error [MSE]}/\text{variance}$) was > 0.2 (following Mazar et al., 2016), we adjusted metric values by subtracting predicted values from observed values. For other metrics, we used the observed metric values, presuming that the influence of natural gradients on metric values was too small to introduce bias.

Metric scoring: Metric values were scored to account for differences in scale direction of response to stress among metrics (Blocksom, 2003). This scoring transformed metrics to a 0 to 1 scale, with lower scores indicative of more stressed conditions and greater deviation from reference sites. We scored metrics after Cao et al. (Cao et al., 2007). We scored metrics that decrease with human activity as

$$(\text{Observed} - \text{Min}) / (\text{Max} - \text{Min})$$

where Min is the 5th percentile of high-activity calibration sites and Max is the 95th percentile of reference calibration sites. We scored metrics that increase with human activity as

$$(\text{Observed} - \text{Max}) / (\text{Min} - \text{Max})$$

where Min is the 5th percentile of reference calibration sites, and Max is the 95th percentile of high-activity sites. We trimmed all scores outside the range of 0 to 1.

Metric selection: We selected metrics for possible inclusion in an MMI based on a series of screening criteria (**Table 4**). First, metrics were eliminated if they had inadequate range, which we defined as containing $> 1/3$ zero or one scores (Stoddard et al., 2008) and median scores at reference and stressed sites > 0 (Stevenson and Zalack, 2013). We evaluated metric the signal-to-noise (S:N) ratio, the ratio of the variance among all sites (signal) to the variance of repeated visits to the same site (noise) (Kaufmann et al., 1999). Metrics passed with a S:N ratio > 1.5 (Stoddard et al., 2008). Additionally, we eliminated metrics with a ratio of between-site to within-site variance < 3 (Mazar et al., 2016; Stoddard et al., 2008). We evaluated metric responsiveness by t-test of the metric scores in reference and stressed sites (**Table 4**). We

assessed bias by determining whether metric values at reference sites varied among regions, using a threshold of an ANOVA F-statistic > 3 .

Assembling the MMI: All metrics that met screening criteria for each assemblage were assembled into all possible combinations (proto-MMIs). We selected the top-performing proto-MMIs using a screening criteria of ANOVA F statistic for regional bias < 3 and a precision standard deviation of reference calibration scores < 0.2 . We then calculated the most frequent metrics in these top performing proto-MMIs and grouped these metrics into thematic types, ensuring a distribution of final metrics across thematic categories (**Table 3**). The winning metrics were assembled into the final MMI. We calculated scores for the final MMI by averaging metrics scores and rescaling by the mean of reference calibration site scores. This process results in an index that, like the O/E index, has a mean of 1 at reference calibration sites.

We calculated a combined O/E + MMI score by averaging final MMI and O/E scores for each assemblage.

MMI performance evaluation: Multi-metric index performance was evaluated using a series of criteria: a) accuracy, or the performance of the index as unbiased against environmental setting or time of sampling; b) precision, or the low variability of the index score within reference sites and among samples from repeated visits within sites; c) responsiveness, or the ability to show large responses to human activities and d) sensitivity, or the ability to score a non-reference site below the impairment threshold. We compared predictive index performance to its null counterpart and also compared performance scores for all indices with both calibration and validation datasets (**Table 7**). Accuracy was assessed by comparing mean scores at reference sites. Additionally, we assessed regional bias by calculating the ANOVA F statistic for all reference site scores across Perennial Stream Assessment (PSA) ecoregions. To further assess regional bias and the influence of natural gradients, we created random forest models using all available environmental variables (**Table 2**) to predict reference site scores, with a lower variance score indicative of a smaller influence of environmental variables on index scores at reference sites. For precision, we evaluated the standard deviation of index scores at reference sites and between sites with repeat sampling events. For responsiveness, we compared t-statistic values for t-test analyses comparing reference and stressed site scores. Additionally, we created random forest models using all available stressor gradient variables (**Table S2**) to predict scores at all sites, with a higher variance score indicative of a larger response to stressor gradients. Lastly, we calculated a series of Spearman's correlations between index scores and key environmental and stressor gradients (**Table 6; Table S2**).

Reference-based thresholds and comparison to other indices: To determine reference-based thresholds for all indices, we determined the 1st, 10th, and 30th percentile of reference calibration scores for all indices using the `qnorm` function as implemented in the `stats` package in R (R Core Team, 2013). To compare algal index performance from this work to previously published algal Southern California Index of Biotic integrity (Fetscher et al., 2014), we divided the SoCal IBI by the average scores at reference calibration sites. Because the CSCI scores have

a reference calibration mean of 1, this normalization was not necessary. To compare CSCI and algal index performance, we created a series of generalized additive models as implemented in the R package *mgcv* (Wood, 2017) to model the response of the indices to stressor gradients. To compare algal dataset structure to the CSCI index developmental dataset, we calculated Bray-Curtis distances among all reference calibration sites using “vegdist” as implemented in the R package *vegan* (Oksanen et al., 2017) on a presence/absence transformed matrix for both the BMI and algae taxonomy datasets. Benthic macroinvertebrate taxonomy data was sourced from Mazor et al., (2016).

RESULTS

Development dataset statistics: Of 1943 sampling locations, we classified 28% of sites as “Reference”, 33% as “Intermediate”, and 37% of sites as “Stressed” (**Figure S1**). The greatest number of Reference sites were found in the high elevation Sierra Nevada region, and fewest in the Central Valley region (**Table 10**).

O/E model for predicting algal community composition at reference sites: Diatoms, soft-bodied algae, and hybrid assemblages exhibited minimal geographic clustering across California’s ecoregions (**Figure S2**); instead, sites with similar biological assemblages were intermingled and dispersed across the state. For diatoms, soft-bodied algae, and hybrid assemblages, we identified 13, 9, and 9 clusters, respectively. Predicted conductivity (i.e. background reference conductivity) and watershed area were selected as predictor variables for all three assemblage types (**Table 2**). The hybrid model predictor variables were shared in common with the diatom and SBA models, whereas the diatom and SBA models shared few predictor variables with each other. The genus-level diatom MMI had only two predictor variables (predicted conductivity and site elevation) and shared both of these with the diatom and hybrid O/E models.

Although all three assemblages resulted in O/E indices that performed better than indices developed from null models, the diatom O/E index had the best performance overall. For example, it had the best precision, as indicated by the lowest standard deviation among reference calibration site scores and lowest variability at sites with replicate samples (**Figure 1**). It also had the least biased prediction, as indicated with regression slope nearest to 1 and intercept closest to 0 (**Table 5**). Whereas the diatom O/E index had the best precision, the soft-bodied algae index was the most responsive to key stressors such as nitrogen and phosphorous gradients.

Predicting metric values and assembling the MMIs: For all species-level metrics, few metrics passed the preliminary screening stage for consideration as a modeled metric (pseudo- R^2 value > 0.2; **Table S3**): five metrics for diatoms, four metrics for the hybrid assemblage, and none for the soft-algae. For both the diatom and hybrid modeled metrics, the metrics were those based on trait attributes that had been assigned specifically for the California dataset and were focused on highly tolerant species: metrics that calculated numbers of BCG Level 4+5 taxa and highly tolerant taxa from the indicator species analysis. After the final species-level MMIs were

assembled, all resulting MMIs were comprised of only unmodeled metrics due to predictive metrics being excluded for other screening criteria, often the measure of regional bias (**Table S3**). Of the genus-level metrics we evaluated, only 4 diatom metrics required modeling to account for the influence of natural factors: metrics based on the presence of nitrogen or phosphorus indicators, the proportion of planktonic taxa, and the proportion of SPI Level 4+5 taxa. No genus-level metrics for the other assemblages required modeling (**Table 6**).

Metrics selected for the MMIs: Of 150 metrics evaluated for the species-level assemblages, 3, 8, and 7 metrics passed all screening thresholds for the diatom, soft-bodied algae, and hybrid species MMIs, respectively (**Table S3**). The most frequently failed screening metrics was the measure of regional bias. The remaining acceptable metrics yielded between 100,000 and 200,000 proto-MMIs for each assemblage. After screening the prototype MMIs for regional bias, precision, sensitivity, and redundancy, we assembled the winning MMIs. For diatoms, soft algae and hybrid species-level MMIs, there were a final total of 6, 7, and 8 unmodeled metrics, respectively (**Table 7**). For the genus-level diatom metrics, 3 metrics passed all screening thresholds for the diatoms and hybrid assemblages, and no metrics pass all screens for the soft algae. Both the SBA and the hybrid genus-level MMIs failed to produce any proto-MMIs that passed the screening thresholds related to responsiveness and precision. The winning genus-level diatom MMI was comprised of a single modeled metric (proportion of SPI level 4+5 taxa) and 6 unmodeled metrics (**Table 6**).

Effects of predictive modeling on metric performance: Multiple metrics across species-level assemblages showed improvements in their regional bias when these metrics were modeled compared to unmodeled metrics. For example, the diatom metric for the proportion of species from the low phosphate indicator taxa, which saw an almost 20-fold decrease in the regional bias (from 19.7 to 0.9 ANOVA *F* statistic) when modeled. Yet, despite this improvement, these metrics were never selected for inclusion in the final MMIs. One example is the diatom metric for the proportion of species from the low phosphate indicator taxa, which saw an almost 20-fold decrease in the regional bias (from 19.7 to 0.9 ANOVA *F* statistic) when modeled. Frequently, the modeled metric was much less responsive than the unmodeled version, as indicated by a lower *t*-statistic in discriminating high-activity versus reference sites (**Table S3**). For example, for the diatom metric of species count for BCG 5 taxa, the *t*-statistic dropped from 31 to 18 for the raw and predictive metrics, respectively, whereas the regional bias dropped from an ANOVA *F* statistic of 32 to 26. For 12 of the 13 modeled metrics (both species and genus level), the absolute *t*-statistic (reference vs. stressed site scores) was higher (difference in $|t| > 1$) for the raw metric than for the modeled metric. For these same metrics, 8 of the 13 modeled metrics had a better regional bias score in the modeled metric than the unmodeled metric (**Table S3**).

Performance evaluation of the O/E, MMI, and combined indices: As detailed above, the predictive O/E models outperformed their null counterparts, whereas the final species-level MMIs were comprised of non-predictive metrics, only. Across all assemblages, the MMIs outperformed the O/E models for precision and responsiveness (**Table 8**); at the same time,

regional bias was slightly higher for the MMIs than the O/E models (**Table 8**). The performance of the combined index (O/E + MMI) was variable across species assemblages. In general, all assemblages had lower regional bias and weaker responsiveness to nutrient stressors with the combined index compared to the MMI index alone. Soft-bodied algae and the hybrid combined indices also had worse precision than the MMI indices alone (**Table 8**). Of the three MMI indices, the hybrid MMI had the best performance statistics with regards to precision and responsiveness (**Table 8**), making it arguably the best performing of the three species-level MMIs. The genus-level diatom MMI had worse responsiveness than the species-level MMIs, although good accuracy and precision.

Additionally, final indices with a predictive component (i.e., the O/E indices and the genus-level MMI) had less influence by environmental gradients than their null counterparts (accuracy variance, **Table 8**). For the non-predictive indices, the soft-bodied algae MMI had the lowest natural variance scores. Notably, the diatom and hybrid non-predictive MMI's had validation data with very low natural gradient variance scores, indicating that influence from natural factors was likely very low (**Figure 2**). Within-site and between-site standard deviation for reference calibration site scores was strongly positively correlated (Pearson's $r = 0.9$) for all final indices, with the soft-bodied algae O/E having the worst precision (reference calibration within site 0.38 SD). Both the species- and genus-level MMIs had the highest responsiveness, as indicated by their ability to discriminate reference from stressed sites (t statistic, **Table 8**), whereas the O/E indices had the worst responsiveness. The same pattern was observed in looking at index response to stressor gradients (nitrogen, phosphorus, specific conductivity), with the hybrid MMI being the most responsive index (**Table 8, Figure 4**).

Effect of low numbers of expected taxa on O/E performance: Performance of the O/E indices was better at high-E than at low-E sites, with null models out-performing O/E models at E values less than 10 (**Figure 5**). For diatoms, soft-bodied algae, and hybrid O/E indices, the null indices had consistently poorer accuracy and precision when E values were greater than 10, with both the null and predictive models averaging 88% of reference calibration sites scoring above the 10th percentile of reference site scores. With the regards to sensitivity, the predictive O/E model out-performed the null models for diatoms and hybrid but not for the soft-bodied algae model. In contrast, the soft-bodied algae predictive O/E model had the most improvement over the null model with regards to precision, with the predictive model averaging 0.38 for the predictive model and 0.43 for the null model.

Distribution of biological condition classes across state: We established 4 biological condition classes based on the distribution of algal index scores at reference calibration sites (**Table 9; Table 10**). Statewide, 47% of streams were likely to be intact (hybrid MMI ≥ 0.93 [30th percentile of reference calibration sites]). Another 18% were possibly altered (hybrid MMI ≥ 0.83 [10th percentile]), 19% were likely to be altered (hybrid MMI ≥ 0.69 [1st percentile]), and 16% were very likely to be altered (< 1st percentile; **Table 10**).

Although 38% of stressed, high-activity sites were very likely to be altered, this number varied considerably by region. The South Coast and the Chaparral had the highest percent of stressed

sites that were considered very likely to be altered (52 and 27%, respectively), while only 4% of the North Coast and none of the Sierra Nevada stressed sites were classified as likely altered. Statewide, about 75% of reference sites were classified as likely to be intact, with the highest percentage in the Sierra Nevada and North Coast regions (82 and 78%, respectively), and lowest in the Central Valley where there were only 2 reference sites (**Table 10**).

Relationships to environmental and stressor gradients: All three assemblages yielded O/E and MMI indices that were unbiased with respect to environmental variables (**Figure 2; Table 8**). The O/E indices had the lowest correlation to environmental gradients, indicating that the geographic modeling helped to account for some regional variability. However, the MMI indices also had environmental gradient correlation values that were very low and often within the within site standard deviation (0.8-0.9) (**Figure 3; Table S2**). The strongest environmental gradient relationship to the MMI indices (at reference calibration sites) were the measures of elevation (elevation range and maximum elevation); multiple O/E models used measures of elevation as predictor variables, indicating elevation plays a key role in algal species distributions. MMI indices at reference calibration sites also had a slight positive correlation to the sampling day of year (**Figure 3; Table S2**). However, given the large latitudinal variation of California's sampling regions, day of year may be an artifact of latitudinal variability in sampling periods: Northern California sites tend to be sampled later in the season, whereas Southern California sites dry up earlier in the sampling season. While all three assemblages yielded species-level MMIs that responded strongly to stressor gradients, the hybrid MMI clearly was the most responsive across both nutrient and urbanization gradients (**Figure 6; Table S2**).

DISCUSSION

Our analyses have demonstrated that while the final species-level MMI indices generated as part of this study lacked any predictive metrics, we were able to derive MMIs with low regional bias and statewide applicability. The success of these indices is a credit to the large California stream algae reference dataset that allows for the development of robust and sensitive models in the face of diverse environmental gradients.

O/E predictive modeling: Although predictive modeling has become common in benthic macroinvertebrate bioassessment applications, only a handful of studies have developed predictive models for algal communities, namely diatoms (Mazor et al., 2006; Philibert et al., 2006; Feio et al., 2009, 2012; Almeida and Feio, 2012). Notably, previous studies that have attempted to model algal assemblages for the central coast region of California encountered difficulties in generating robust predictive models (Ritz et al., 2010; Rollins et al., 2012). Previous studies that have generated O/E models for diatom assemblages have resulted in models with similar precision and sensitivity to the diatom and hybrid O/E models generated in this study; for example, Cao et al. (2007), built an O/E index for diatoms in Idaho, resulting in an index with a standard deviation among reference sites of 0.17, on par with the diatom O/E model from this study. Reported measures of precision of macroinvertebrate O/E models have ranged from 0.13 (Hawkins, 2009) to 0.18 (Van Sickle et al., 2005) to 0.26 (Vander Laan and Hawkins, 2014). Linke et al. (2005) recommend O/E models with a slope ranging between 0.85-

1.15, a criterion met by all three O/E models in this study. Nonetheless, indices from these O/E models had poor precision and responsiveness.

Three key factors may be at play in the difficulty of predicting algal assemblages, as in the creation of O/E models: a) geographic parameters used as candidate predictor variables are largely not responsible for shaping algal species distribution patterns; b) environmental characteristics that are good predictors of assemblages in some settings (e.g. high elevation) may not be good predictors of algal assemblages in other settings (e.g. lowland streams); c) high isolation of freshwater habitats may result in decreased niche partitioning, leading to similar geographic areas supporting more distinct algal communities (Vander Laan and Hawkins, 2014). To address the first potential explanation, we investigated whether expanding our predictor variables to include those parameters influenced by human activity (including nitrogen and phosphorus, conductivity, *in situ* temperature) improved our ability to generate O/E models. We found that even when including these additional predictors, our O/E models saw very little improvement in performance (results not show). Although other studies have found that locally measured environmental variables (such as light availability and velocity) may improve model performance (Feio et al., 2012, Sabater et al., 2006; Veraart et al., 2008; Bornette and Pujalon, 2011; Aguiar et al., 2011, Lamb and Lowe, 1987; Soinenen, 2004), such models are less useful for bioassessment index development, as they may be unable to separate the influence of natural variation in these factors from anthropogenic disturbance (Hawkins et al., 2010; Reynoldson et al., 1997).

To investigate the role of geographic variability in the model performance, we evaluated ecoregion as a candidate predictor variable in our O/E models (**Table 2**), although it was never selected as a final predictor for any of the O/E or metric models. Additionally, we evaluated ecoregion-specific O/E models for two regions with the largest data density (and available reference sites), but their precision did not improve over the statewide model and predictive models only marginally out-performed null models (**Table S1**). The failure of the eco-region scale O/E models to out-perform the statewide models suggests that the O/E model performance is not constrained by high natural variability in the reference site pool.

Lastly, to assess the potential for isolated habitats to be supporting disparate communities of algae, we analyzed the average Bray-Curtis distances among all reference calibration sites for both algal populations as well as benthic macroinvertebrate populations (**Figure S6**). We found that on average, pairwise Bray-Curtis dissimilarity between reference sites for algae assemblages averaged 0.76, whereas dissimilarities based on BMI assemblages average 0.69, indicating that BMI taxa had greater similarity between sites of the reference pool than did algal populations. Relatedly, we identified approximately 691 species of BMI taxa at these reference sites, whereas there were over 1136 algae species, with average alpha diversity for these reference sites was 56 (SD 14) for BMI taxa and 36 (SD 16) for algae taxa. These results demonstrate that in comparison to BMIs, algal assemblages have lower species richness at reference sites, while at the same time having greater overall species richness (at reference sites) across the state and thus greater average dissimilarity between two reference sites. Taken together, these results suggest that algal populations at two sites of comparable

ecological status and similar geographic setting may have only minimal overlap in species composition. Additionally, these dissimilar reference site algal assemblages can help explain the low E values that were common across the O/E models: with a greater number of taxa spread across reference sites, and few taxa shared among sites, the O/E models were only able to confidently predict a handful of taxa at each site. These low E values resulted in poor precision and sensitivity in each of the three assemblage O/E models (**Figure 5**), consistent with previous benthic macroinvertebrate O/E model development (Mazor et al., 2016).

Trait attribute completeness dictates metric performance: The majority of all metrics that were included in final species-level MMIs were metrics that increased with stress, such as percent tolerant taxa and high nitrogen indicators. The decreasing metrics included in the species-level MMIs were the result of an indicator species analysis, meaning they were specifically tuned to the California taxa. Likewise, the best-performing species-level modeled metrics were those that resulted from an indicator species analyses (**Table 6**). This result suggests that we are better able to account for high abundance of taxa at highly degraded sites, as opposed to accounting for taxa that become increasingly rare under stressed conditions. The use of the indicator species analyses allows for the identification of these sensitive taxa, a trait attribute that may be otherwise overlooked in empirical studies. Similarly, increaser taxa are often the focus of specialized studies to develop stressor-tuned indices (Teply 2010a,b; Teply and Bahls 2005,2006) and therefore comprise the majority of trait attributes available in published literature. In contrast, the genus-level diatom MMI was split almost evenly between increasing and decreasing metrics. Aggregating trait attributes to the genus level may have helped improve the ecological signal of the decreasing metrics that otherwise have sparsely populated species attributions. Likewise, the single predictive metric from the genus-level diatom MMI was a measure of pollution tolerance (SPI value), an attribute that is assigned specifically at both the genus and species level. Taken together, these results suggest that the paucity of trait attributes available for algae species contributes to the inability to develop predictive models for individual metrics, and that trait attributions for sensitive taxa should be a priority focus of future studies.

Performance of MMI: Our analyses revealed that modeling metric scores using geographic setting did not improve metric performance. Given that algae can exhibit regionally-specific responses to environmental perturbation (Stevenson, 2014), we wanted to ensure that the statewide index was not biased by regional variation in natural gradients. We evaluated regional bias using a test of reference site scores across regions (PSA ecoregions), a screening measure that often improved with predictive modeling. Eliminating metrics with high regional bias scores often resulted in final MMIs with lower sensitivity in discriminating reference versus stressed sites (**Table S5**); however, it is worth noting, this loss in sensitivity is in fact a removal of geographic artifacts that give way to spurious results masquerading as index sensitivity. All final MMIs had low regional bias scores (**Figure 2**) in spite of their lack of predictive metrics. For this reason, we recommend future algal index development follows strict screening criteria to eliminate regional bias that can easily be conflated with index sensitivity.

Comparison to other California biological indices: The Southern California Index of Biotic Integrity (SoCal IBI, Fetscher et al., 2014) served as a precursor to the MMIs created as part of this study, and their final MMIs shared many elements. In particular, there were a number of final metrics included in the species-level MMIs that were also in the SoCal IBIs, including metrics generated using the indicator species analysis trait attributes for nitrogen, phosphorus, copper, and DOC. (Table 6). Of the three assemblages, the species-level hybrid MMI had the strongest relationship to the SoCal IBI hybrid MMI (H20, Fetscher et al., 2014) with an R^2 of 0.49 (Figure S4) for statewide scores. Like the SoCal IBI, the hybrid MMI in this study was the most responsive index. In contrast to the findings of Lavoie et al., (2004), the incorporation of SBA improved index performance over that of diatom assemblages alone. While SBA and diatoms are known to respond differently to stressor gradients, such as nutrient over-enrichment (Schneider et al., 2013), it is important to emphasize that the responsiveness of the index may reflect both an ecological characteristic of the algae as well as the sensitivity of the index itself. Our data suggests that the hybrid approach may offer an expanded suite of trait attributes that a single assemblage approach does not, therefore providing a more comprehensive trait attributes dataset to leverage for metric construction.

The final species-level MMIs showed clear concordance with the benthic macroinvertebrate CSCI index across a variety of stressor gradients (Figure 7). In particular, the CSCI demonstrated a more sensitive response to physical habitat measure of percent sands and fines than any of the algal MMIs presented here, in agreement with previous studies comparing the CSCI to algal IBI performance (Rehn, 2016). Across all sites with both CSCI and ASCI scores, the two indices both ranked a site as “very likely intact” ($> 10^{\text{th}}$ percentile of reference for both CSCI and hybrid MMI) 75% of the time. When the two indices did not agree, the CSCI ranked a site below the 10^{th} percentile of reference 15% of the time, while the hybrid MMI did 10% of the time (Figure S7). The Sierra Nevada and North Coast regions had the greatest proportion of sites in agreement, with both indices giving high scores. Notably, the South Coast and the Chaparral regions had strong agreement between the two indices, even when both indices scored a site below the 10^{th} percentile of reference. As BMI and algae are known to have different sensitivities to environmental stressors (Sonneman et al. 2001; Johnson et al 2009), the two indices provide complimentary lines of evidence on biological condition, as highlighted by both their agreement and disagreement in scoring sites across broad environmental gradients. Future work will explore the potential environmental drivers of variable performance between the two indices.

Weighing the variable performance measures for select the best combination of metrics: As discussed above, the O/E model performance fell within the range of previously published O/E studies (), in particular for the diatom and hybrid models, yet the O/E index responsiveness made them less effective at discriminating reference from stressed condition. In comparison, the species-level MMIs performed well across all screening metrics, including precision, sensitivity, accuracy, and responsiveness, despite the lack of modeling to account for the influence of natural factors. The motivation to include an O/E component in the resulting final index stems from the fact that O/E measures an inherently different component of the biological health than the MMI: the O/E index is evaluating the taxa lost to disturbance,

whereas the MMI is assessing ecological structure of the community (Mazor et al., 2016). To this end, we evaluated the inclusion of the O/E+MMI as a final index (**Table 8**). However, for all three assemblages, the combination of the O/E to the MMI only served to weaken the MMI performance into ranges outside our target performance thresholds, specifically for precision and accuracy. Therefore, the MMI models were stronger independent of the O/E models and the O/E models should be used only as a supplemental information on predicted species at test sites and not as a component of the final index score.

Choosing indices for different purposes: Since all three assemblages yielded high-performing MMIs, the critical next task for a regulatory program is selecting an appropriate index for making assessments. There are a number of factors that may go into this decision, including a) performance of index across environmental and stressor gradients; b) cost and regional capacity to perform taxonomic analyses; c) ease and consistency of field and laboratory results. Our results have demonstrated that the hybrid index is the most responsive index to nutrient stressor gradients (**Table 8**). The soft-bodied algae index had the lowest regional bias, whereas the hybrid and soft-bodied algae species level MMIs had the best precision. The genus-level diatom MMI also had low regional bias, but it was less responsive than the species-level MMIs. From a cost perspective, a hybrid index results in the largest economic burden for analysis, requiring two separate assemblages, but this expense may prove worthwhile if it is the most sensitive, informative index. Diatoms require less processing for lab analyses and are often more cost-effective () whereas soft-bodied algae are important for linking nutrient enrichment to algal biomass and determining the presence of potentially toxigenic species (Manoylov, 2014; Stancheva and Sheath, 2018). Lastly, monitoring programs rely on standardized sampling protocols and extensive quality control processes. A number of states as well as the National Water Quality Assessment program (NAWQA) have focused primarily on diatoms indices because of the difficulty in sampling soft-bodied algae in the field and standardization of lab analyses (), while California has included both assemblages in their bioassessment field protocols (Ode et al., 2016). The advent of molecular methods in bioassessment may help to eliminate many of the obstacles facing multiple-assemblage approaches, namely that sequencing approaches have the potential to decrease the expense of generating taxonomic data, enabling the analysis of multiple species assemblages. Additionally, DNA-based analytical approaches can rely on bioinformatic and standardized computational pipelines as opposed to expert taxonomists (Pawlowski et al., 2018). All of these factors should be weighed by regulatory agencies when evaluating the appropriate algal index for their unique programs.

CONCLUSIONS

This study evaluated a suite of both predictive and non-predictive indices for diatoms and soft-bodied algae in California. Designed after the benthic macroinvertebrate California Stream Condition Index (CSCI), we generated both O/E and MMI indices, as well as a combined O/E and MMI indices. Across all species assemblages, the MMI indices had superior precision and responsiveness, while the O/E indices had superior regional bias performance due to their incorporation of geographic parameters in their predictive models. However, the O/E indices were plagued by poor precision, likely due to the highly diverse algal species assemblages that

limit predictability of species distributions, resulting in low expected taxa values and unstable indices.

Our best performing O/E indices were developed using genus-level taxonomy data, in part because aggregating taxonomy data to genus level helped to increase the expected taxa values (E) and therefore help stabilize the index performance. In contrast, the best-performing MMIs used species-level taxonomy data, a consequence of the fact that trait attributes, the cornerstone of metric calculations, are often attributed at the species (or lower) level, and the majority of genera in our statewide dataset currently lack trait attributes. This difference highlights the need for a more comprehensive effort to derive trait attributes for taxa in California, given that many taxa are endemic and without known trait attributes. With an improved attributes dataset, there is an opportunity in the future to develop genus-level multi-metric indices for soft-algae and diatoms that may help alleviate the statewide taxonomic capacity burden that is currently an obstacle to many monitoring programs.

The primary goal of this study was to develop an algal index that has statewide applicability for California. The MMIs developed as part of this study build upon the work of Fetscher et al. (2014) and with the advantage of a larger, more comprehensive dataset we were able to develop a suite of robust indices that can be applied statewide. The diatom, soft-bodied algae, and hybrid species-level MMIs can be used as a secondary indicator alongside the CSCI for performing condition assessments of wadeable streams, as well as for identifying priority waterbodies for protection and performing causal assessments. With the goal of this novel index having widespread applicability to regulatory agencies and monitoring programs, we have provided all MMI calculator code (github.com/sccwrp) and encourage users to contact us with any questions on the calculation of these index scores.

ACKNOWLEDGEMENTS: The authors acknowledge funding support from the California State Water Resources Control Board. We would like to thank the Biostimulatory and Biointegrity Science Advisory Panel, especially Jan Stevenson and Charles Hawkins, for their helpful discussion; the SWAMP Bioassessment workgroup, especially Betty Fetscher, Chad Loflen, Kevin Lunde, and Andy Rehn for helpful discussion and feedback; Rosalina Stancheva for her input and algae expertise; and Aaron Jones and Kenny McCune for analytical and database support.

Table 1. Stressor and human-activity gradients used to identify reference sites and high-activity sites. Sites that did not exceed the listed reference thresholds were used as reference sites. Sites that exceeded at least one high-activity threshold (str threshold) were used as high-activity (“stressed”) sites. Sources A = National Landcover Data Set; B = custom roads layer; C = National Hydrography Dataset Plus; D = National Inventory of Dams; E = Mineral Resource Data System; F = field-measured variables. WS = watershed; 5 km = watershed clipped to a 5-km buffer of the sampling point; 1 km = watershed clipped to a 1-km buffer of the sampling point; W1_HALL = proximity-weighted human activity index (Kaufmann et al., 1999), Code 21 = land-use category that corresponds to highly managed vegetation, such as roadsides, lawns, cemeteries, and golf courses.

Variable	Scale	Ref threshold	Str threshold	Unit	Source
% agricultural	1 km, 5 km, WS	< 3	> 50	%	A
% urban	1 km, 5 km, WS	< 3	> 50	%	A
% agricultural + % urban	1 km, 5 km, WS	< 5		%	A
% Code 21	1 km, 5 km	< 7	> 50	%	A
	WS	< 10	> 50	%	A
Road density	1 km, 5 km, WS	< 2	> 5	km/km ²	B
Road crossings	1 km	< 5		crossings	B, C
	5 km	< 10		crossings	B, C
	WS	< 50		crossings	B, C
Dam distance	WS	< 10		km	D
% canals and pipelines	WS	< 10		%	C
Instream gravel mines	5 km	< 0.1		mines/km	C, E
Producer mines	5 km	0		mines	E
W1_HALL	Reach	< 1.5	> 5	NA	F
% sands and fines	Reach			%	F
Slope	Reach			%	F

Table 2. Natural gradients evaluated as predictor variables for inclusion in both the O/E and genus-level diatom MMI Random Forest models. Genus-level diatom MMI predicted metric = proportion SPI species 4+5. A = PRISM climate database ([\(Stancheva and Sheath, 2016\)](http://climate.geoscorp.gov/)); C = Olson and Hawkins (2012); D = National Atmospheric Deposition Program National Trends Network (<http://nadp.slh.wisc.edu/ntn/>); M = National Elevation Dataset (<http://ned.usgs.gov/>).

Variable	Description	Diatom O/E	SBA O/E	Hybrid O/E	Diatom MMI	Ref
		<i>Gini</i>	<i>Gini</i>	<i>Gini</i>	<i>MSE</i>	
AREA_SQKM	Area		45.27			D
AtmCa	Catchment mean of mean 1994–2006 annual precipitation-weighted mean Ca+ concentration		45.54			D
AtmMg	Catchment mean of mean 1994–2006 annual precipitation-weighted mean Mg+ concentration					
BDH_AVE	Catchment mean bulk density					C
CaO_Mean	Calcite mineral content					C
CondQR50	Median predicted conductivity (reference)	43.47	43.98	41.99		C
DayOfYear	Day of year		49.02	40.91		
ELEV_RANGE	Elevation range of catchment					M
KFCT_AVE	Catchment mean soil erodibility (K) factor					C
LogWSA	Area of the unit of analysis	46.40	33.16	45.22	15.88	
LPREM_mean	Catchment mean log geometric mean hydraulic conductivity					C
LST32AVE	Average of mean 1961–1990 first and last day of freeze	39.97		40.74		D
MAX_ELEV	Maximum elevation in catchment	42.14		39.70		M
MAXWD_WS	Catchment mean of 1961–1990 annual maximum number of wet days	40.87				D
MEANP_WS	Catchment mean of mean 1971–2000 annual precipitation			39.28		D
MgO_Mean	Magnesium oxide mineral content					C
MIINP_WS	Catchment mean of mean 1971–2000 minimum monthly precipitation					D
Month	Month		18.67			
New_Lat	Latitude		47.85	38.60		
New_Long	Longitude	42.07		37.01		
PCT_CENOZ	Percent Cenozoic sediments					
PCT_NOSED	Percent non-sedimentary or volcanic geology					
PCT_QUART	Percent Quarternary geology					
PCT_SEDIM	Percent sedimentary geology					
PCT_VOLCNC	Percent volcanic geology					
PPT_00_09	10-y (2000–2009) average annual precipitation				19.44	A
PRMH_AVE	Catchment mean soil permeability					C
PSA6c	PSA region					
S_Mean	S-bearing geology					
SITE_ELEV	Site elevation	40.45		39.62		M
TEMP_00_09	10-y (2000–2009) average monthly temperature	37.96				A
TMAX_WS	Catchment mean of mean 1971–2000 maximum temperature		48.99			D
UCS_Mean	Catchment mean unconfined compressive strength					F
XWD_WS	Catchment mean of mean 1961–1990 annual number of wet days					D

Table 3. Metric classes for multi-metric index. BCG taxa = taxa identified as indicative of a Biological Condition Gradient (BCG) Level indicators from the California BCG effort (Paul et al., *in prep.*); SPI taxa = taxa with a pollution tolerance indicator level designation (Coste, 1986).

Category	Example metrics
Tolerance/sensitivity	Association with specific water-quality constituents (nitrogen, phosphorus, organic carbon, metals)
	Low/High oxygen preference
	Salinity tolerance
	SPI index level
Autecological guild	Nitrogen-fixers
	Saprobic/heterotrophic taxa
Morphological guild	Sedimentation indicators
Relationship to reference	Taxa associated with reference vs. non-reference sites
	BCG taxa level
Taxonomic groups	Chlorophyta, Rhodophyta, Zygnemataceae, heterocystous cyanobacteria, <i>Suriella</i> , <i>Cyclotella</i>
Diversity	Richness

DRAFT

Table 4. Metric screening criteria for evaluating inclusion in the MMI. Description of criteria, statistical test, and threshold for passing.

Description	Test	Threshold	Reference
Regional bias	ANOVA of metric values at reference sites by ecoregion (PSA)	F statistic < 3	Mazor et al., 2016
Sensitivity	t-test comparing reference/stressed site scores (t statistic > 10)	F statistic > 2	Mazor et al., 2016
Frequency of Zero	Frequency of score = 0	< 33% of scores	Stoddard et al., 2008
Frequency of One	Frequency of score = 1	< 33% of scores	Stoddard et al., 2008
Range of Ref scores	Median score at reference sites	> 0	Stevenson and Zalack, 2013
Range of Stress scores	Median score at stressed sites	> 0	Stevenson and Zalack, 2013
Signal to Noise	Variance across all sites / variance at repeat site visits	> 1	Stoddard et al., 2008
Repeat visit variation	ANOVA on repeat samplings of station codes	F statistic < 3	Mazor et al., 2016

DRAFT

Table 5. Performance results for the O/E predictive models for each assemblage at calibration (cal) and validation (val) reference sites.

	Variation explained O		Regression slopes		Intercepts	
	<i>Cal</i>	<i>Val</i>	<i>Cal</i>	<i>Val</i>	<i>Cal</i>	<i>Val</i>
Diatom	0.49	0.34	1.015	0.871	-0.079	1.226
Soft-bodied algae	0.52	0.49	0.977	1.037	0.117	-0.314
Hybrid	0.60	0.55	1.037	0.962	-0.305	0.180

DRAFT

Table 6. Metrics selected in final diatom, soft-bodied algae, and hybrid MMIs. Asterisk denotes modeled metrics. BCG 3 taxa = taxa identified as indicative of a Biological Condition Gradient (BCG) Level 3 indicator status from the California BCG effort (Paul et al., *in prep.*); SPI 2 taxa = Taxa with a pollution tolerance indicator level 2 designation; NHHONF = N-heterotrophic-high organic N (facultative); SPI 4+5 taxa = Taxa with a pollution tolerance indicator level 4 or 5 designation; NAHON = N-autotrophic-high organic N; least tolerant taxa = sensitive taxa identified through indicator species analysis.

Description	Species-level			Genus-level	Response to stress
	Diatom	SBA	Hybrid	Diatom	
Count species: BCG 3 taxa	x	x	x		Increase
Count species: high copper indicators from Fetscher et al., 2014		x			Increase
Count species: high DOC indicators from Fetscher et al., 2014		x			Increase
Count species: low total phosphorous indicators from Fetscher et al., 2014		x			Decrease
Count species: SPI 2 taxa				x	Decrease
Proportion individuals: most tolerant taxa		x			Increase
Proportion species: Cyclotella taxa	x		x	x	Increase
Proportion species: Green algae		x			Increase
Proportion species: high copper indicators from Fetscher et al., 2014			x		Increase
Proportion species: high DOC indicators from Fetscher et al., 2014			x		Increase
Proportion species: low total nitrogen indicators from Fetscher et al., 2014			x		Decrease
Proportion species: low total phosphorous indicators from Fetscher et al., 2014	x				Decrease
Proportion species: NHHONF taxa	x		x		
Proportion species: non-ref indicators from Fetscher et al., 2014		x			Increase
Proportion species: SPI 4+5 taxa				x*	Increase
Proportion species: Suriella taxa	x		x	x	Increase
Proportion species: taxa requiring at least 10% oxygen	x		x		Increase
Richness: NAHON taxa				x	Increase
Proportion species: Gomphonema taxa				x	Decrease
Proportion species: least tolerant taxa				x	Decrease

Table 7. Screening metric results for chosen metrics in each MMI. Rsq=pseudo R-squared from RandomForest modeling; PSA(F) = F statistic for ANOVA analysis of metric scores at reference sites across PSA regions (Central Valley excluded); Ref/Int/Str (F) = F statistic for ANOVA analysis of metric scores across Reference, Intermediate, and Stressed sites; Ref/Str t = t statistic for t-test comparison of metric scores at Reference versus stressed sites; Freq = frequency of zero and one metric scores; Repeat visits (F) = F statistic for ANOVA analysis of variation in site scores across repeat samplings; Range = range in scores at reference (ref) and stressed (str) sites; Signal:Noise = the ratio of the variance among all sites (signal) to the variance of repeated visits to the same site (noise). BCG= Biological Condition Gradient; ISA = Indicator Species Analysis; SPI = Specific Polluosensitivity Index; NALON = N-autotrophic-low organic N; CRUS = Cladophora glomerata + Rhizoclonium hieroglyphicum + Ulva flexuosa + Stigeoclonium spp. (Stancheva and Sheath; 2016); ZHR = Zygnemataceae + heterocystous cyanobacteria + Rhodophyta (Stancheva and Sheath; 2016).

	rsq	PSA (F)	Ref/Int/Str (F)	Ref/Str t	Freq Zero	Freq One	Repeat visits (F)	Range (ref)	Range (str)	Signal:Noise	Assem.
Proportion species: NHHONF taxa	NA	3.55	110.87	12.52	0.54	0.00	2.73	1.07	0.91	2.09	diatom
Proportion species: taxa requiring at least 10% oxygen	NA	5.06	155.78	14.25	1.08	0.00	2.54	1.19	0.79	1.86	diatom
Proportion species: Cyclotella taxa	NA	1.21	112.25	13.53	1.08	0.00	2.29	1.06	0.93	2.16	diatom
Proportion species: Suriella taxa	NA	1.11	99.30	12.55	1.08	0.00	1.51	1.03	1.03	1.68	diatom
Count species: BCG 3 taxa	NA	5.50	113.57	10.43	1.36	0.00	3.09	0.95	0.60	3.77	diatom
Proportion species: low total phosphorous indicators from Fetscher et al., 2014	NA	0.71	67.86	8.89	0.27	0.00	2.09	0.62	0.31	1.96	diatom
Proportion species: most tolerant taxa	NA	4.71	134.37	14.55	0.48	0.00	1.21	1.05	0.91	2.33	sba
Count species: low total phosphorous indicators from Fetscher et al., 2014	NA	2.17	129.32	14.13	1.21	0.00	2.30	1.11	0.86	2.20	sba
Count species: high copper indicators from Fetscher et al., 2014	NA	1.87	188.79	16.43	0.24	0.00	2.20	1.03	0.85	3.13	sba
Count species: high DOC indicators from Fetscher et al., 2014	NA	5.05	190.97	17.23	0.00	0.00	2.15	1.06	0.86	2.56	sba
Proportion species: non-ref indicators from Fetscher et al., 2014	NA	1.63	31.79	7.21	2.90	0.00	2.69	1.08	0.99	2.23	sba
Proportion species: Green algae	NA	2.60	40.48	7.65	1.93	0.00	2.34	1.04	0.59	2.11	sba
Count species: BCG 3 taxa	NA	1.33	54.68	8.23	13.77	0.00	1.97	0.80	0.27	2.26	sba
Proportion species: NHHONF taxa	NA	3.18	157.97	15.51	1.20	0.00	1.96	1.04	0.84	1.75	hybrid
Proportion species: taxa requiring at least 10% oxygen	NA	4.65	170.55	14.99	0.72	0.00	1.89	1.19	0.76	1.59	hybrid
Proportion species: Cyclotella taxa	NA	1.66	175.66	17.61	1.20	0.00	1.87	1.06	0.84	1.96	hybrid
Proportion species: Suriella taxa	NA	1.32	95.71	12.69	0.96	0.00	1.33	1.03	1.03	1.90	hybrid
Proportion species: high copper indicators from Fetscher et al., 2014	NA	2.33	134.57	13.44	0.48	0.00	1.28	1.06	0.93	2.54	hybrid
Proportion species: high DOC indicators from Fetscher et al., 2014	NA	1.29	111.00	11.88	0.72	0.00	2.09	1.07	0.92	1.93	hybrid
Count species: BCG 3 taxa	NA	4.07	120.99	11.05	2.15	0.00	2.19	0.92	0.50	2.76	hybrid

Proportion species: low total nitrogen indicators from Fetscher et al., 2014	NA	1.12	128.00	11.51	3.35	0.00	1.85	0.84	0.38	1.86	hybrid
Count species: SPI 2 taxa	NA	4.58	253.05	21.39	0.00	0.00	2.95	1.03	0.69	4.57	diatom-genus
Richness: NAHON taxa	NA	0.54	13.39	5.29	2.44	0.00	1.73	1.03	1.03	1.58	diatom-genus
Proportion species: Cyclotella taxa	NA	3.37	271.52	20.91	0.27	0.00	1.82	1.12	0.56	2.88	diatom-genus
Proportion species: Gomphonema taxa	NA	5.32	25.95	5.81	8.13	0.00	2.86	1.01	0.86	2.18	diatom-genus
Proportion species: least tolerant taxa	NA	5.32	25.95	5.81	8.13	0.00	2.86	1.01	0.86	2.18	diatom-genus
Proportion species: SPI 4+5 taxa	0.2153	0.84	47.40	10.08	2.17	0.00	1.23	0.88	0.66	1.57	diatom-genus
Proportion species: Suriella taxa	NA	1.66	117.71	14.05	0.27	0.00	1.92	1.08	1.08	1.99	diatom-genus

DRAFT

Table 8. Performance measures to evaluate all final indices. MMI = multi-metric index and observed (O)/ expected (E) taxa index at calibration (Cal) and validation (Val) sites, level = taxonomic level, Assem= assemblage, Type = null or predictive indices. For accuracy tests, only reference sites were used. Accuracy: mean score (ref) = mean score of reference sites (* indicates value is mathematically fixed at 1); F = F-statistic for differences in scores at reference calibration sites among 5 PSA regions (Central Valley); Var = variance in index scores explained by natural gradients at reference sites. Precision: among sites = standard deviation of scores at reference sites; within sites = standard deviation of within-site residuals for reference calibration and validation sites with multiple samples. Responsiveness: t = t-statistic for difference between mean scores at reference and stressed sites, var = variance in index scores explained by human-activity gradients at all sites. Spearman's correlation Rho values for key stressor gradients; for a complete list of Spearman's correlations, see **Table S2**.

Index	Level	Assem	Type	Accuracy				Precision				Responsiveness				Spearman's (Rho)				
				Mean score	F	Var		Among sites (SD)		Within sites (SD)		t		Var		TN	TP	SpCond		
				Cal	Val	Cal	Val	Cal	Val	Cal	Val	Cal	Val	Cal	Val	Cal	Val			
O/E+MMI	both	diatoms	both	1*	0.99	0.34	1.75	0.13	0.13	0.14	0.15	0.07	0.07	18.68	10.04	0.50	0.53	-0.44	-0.37	-0.48
O/E+MMI	both	hybrid	both	1*	0.97	2.60	1.42	0.05	0.10	0.17	0.18	0.09	0.11	17.70	8.22	0.35	0.38	-0.40	-0.36	-0.40
O/E+MMI	both	sba	both	1*	0.99	1.74	1.37	0.07	0.18	0.24	0.28	0.13	0.10	20.56	9.86	0.39	0.40	-0.40	-0.43	-0.32
O/E	genus	diatoms	null	1*	0.99	4.96	1.93	0.15	0.17	0.20	0.23	0.08	0.08	7.21	3.14	0.31	0.33	-0.26	-0.16	-0.24
O/E	genus	diatoms	pred	1.01	1.00	0.49	0.75	-0.13	0.01	0.18	0.19	0.11	0.07	9.5	4.9	0.30	0.36	-0.31	-0.18	-0.31
O/E	genus	hybrid	null	1*	0.94	6.17	1.37	0.02	0.10	0.27	0.30	0.15	0.17	2.95	-0.52	0.10	0.12	-0.18	-0.14	-0.16
O/E	genus	hybrid	pred	1.01	0.98	0.48	1.04	-0.18	-0.01	0.25	0.24	0.16	0.14	8.0	3.4	0.20	0.23	-0.29	-0.20	-0.27
O/E	genus	sba	null	1*	0.96	1.77	1.59	0.14	0.16	0.44	0.48	0.21	0.17	14.00	6.58	0.33	0.35	-0.28	-0.36	-0.20
O/E	genus	sba	pred	1.01	0.96	0.66	1.84	-0.11	0.03	0.38	0.42	0.29	0.17	15.7	7.0	0.27	0.30	-0.32	-0.36	-0.23
MMI	species	diatoms	null	1*	0.99	3.31	1.86	0.16	-0.04	0.17	0.15	0.09	0.09	22.30	13.78	0.52	0.52	-0.49	-0.49	-0.59
MMI	genus	diatoms	null	1*	1.01	15.78	5.89	0.14	0.16	0.29	0.30	0.16	0.15	15.47	7.87	0.46	0.45	-0.52	-0.51	-0.53
MMI	genus	diatoms	pred	1*	0.99	1.91	2.84	-0.17	0.15	0.17	0.22	0.13	0.12	22.65	9.75	0.32	0.32	-0.42	-0.41	-0.40
MMI	species	hybrid	null	1*	1.01	2.28	0.95	0.14	0.02	0.13	0.13	0.08	0.08	27.20	18.61	0.59	0.60	-0.55	-0.51	-0.55
MMI	species	sba	null	1*	1.02	1.34	0.96	-0.08	0.08	0.14	0.15	0.09	0.08	21.86	15.19	0.40	0.41	-0.45	-0.33	-0.41

Table 9. O/E, MMI, and O/E + MMI index performance statistics.

	Diatom O/E	SBA O/E	Hybrid O/E
10th percentile of reference	0.78	0.52	0.69
% reference calibration above 10th percentile of reference	89	89	89
% reference validation above 10th percentile of reference	88	82	87
% stressed below 10th percentile of reference	26	45	16
	Diatom MMI	SBA MMI	Hybrid MMI
10th percentile of reference	0.78	0.82	0.83
% reference calibration above 10th percentile of reference	93	89	91
% reference validation above 10th percentile of reference	94	93	92
% stressed below 10th percentile of reference	62	58	73
	Diatom ASCI	SBA ASCI	Hybrid ASCI
10th percentile of reference	0.83	0.73	0.80
% reference calibration above 10th percentile of reference	89	89	89
% reference validation above 10th percentile of reference	89	85	90
% stressed below 10th percentile of reference	53	63	45
	Diatom MMI genus	SBA MMI genus	Hybrid MMI genus
10th percentile of reference	0.79	0.83	0.92
% reference calibration above 10th percentile of reference	90	90	89
% reference validation above 10th percentile of reference	88	78	71
% stressed below 10th percentile of reference	63	49	76

Table 10. Percentage of sites in different condition classes by region and site status based on hybrid MMI scores. Percentiles refer to the distribution of scores at reference calibration (Cal) sites. For reference, intermediate, and stressed sites, numbers in the last 6 columns are percentage of sites.

		Sites	Likely to be intact (i.e., ≥30th percentile of reference sites, 0.93)	Possibly altered (i.e., 30th– 10th percentile of reference sites, 0.83)	Likely to be altered (i.e., 1st–10th percentile of reference sites, 0.69)	Very likely to be altered (i.e., <1 st percentile of reference sites)
Statewide	Reference	737	74	18	7	1
	Intermediate	856	54	23	16	7
	Stressed	841	16	13	33	38
	Total		47	18	19	16
North Coast	Reference	100	78	13	9	0
	Intermediate	105	73	22	5	0
	Stressed	46	74	9	13	4
	Total		75	16	8	1
Desert/Modoc	Reference	49	67	22	8	2
	Intermediate	58	60	26	10	3
	Stressed	8	50	13	25	13
	Total		63	23	10	3
Chaparral	Reference	169	68	22	8	2
	Intermediate	192	51	27	18	4
	Stressed	214	18	17	37	27
	Total		44	22	22	12
Central Valley	Reference	2	50	50	0	0
	Intermediate	14	50	29	21	0
	Stressed	80	30	26	30	14
	Total		33	27	28	11
South Coast	Reference	216	70	20	8	1
	Intermediate	366	39	25	23	13
	Stressed	481	5	9	34	52
	Total		30	17	25	29
Sierra Nevada	Reference	199	82	15	3	0
	Intermediate	121	85	11	4	0
	Stressed	12	100	0	0	0
	Total		84	13	3	0

Figure 1. Index scores by reference, intermediate, and high activity ("stressed") sites for the O/E, MMI, and combined O/E+MMI indices.

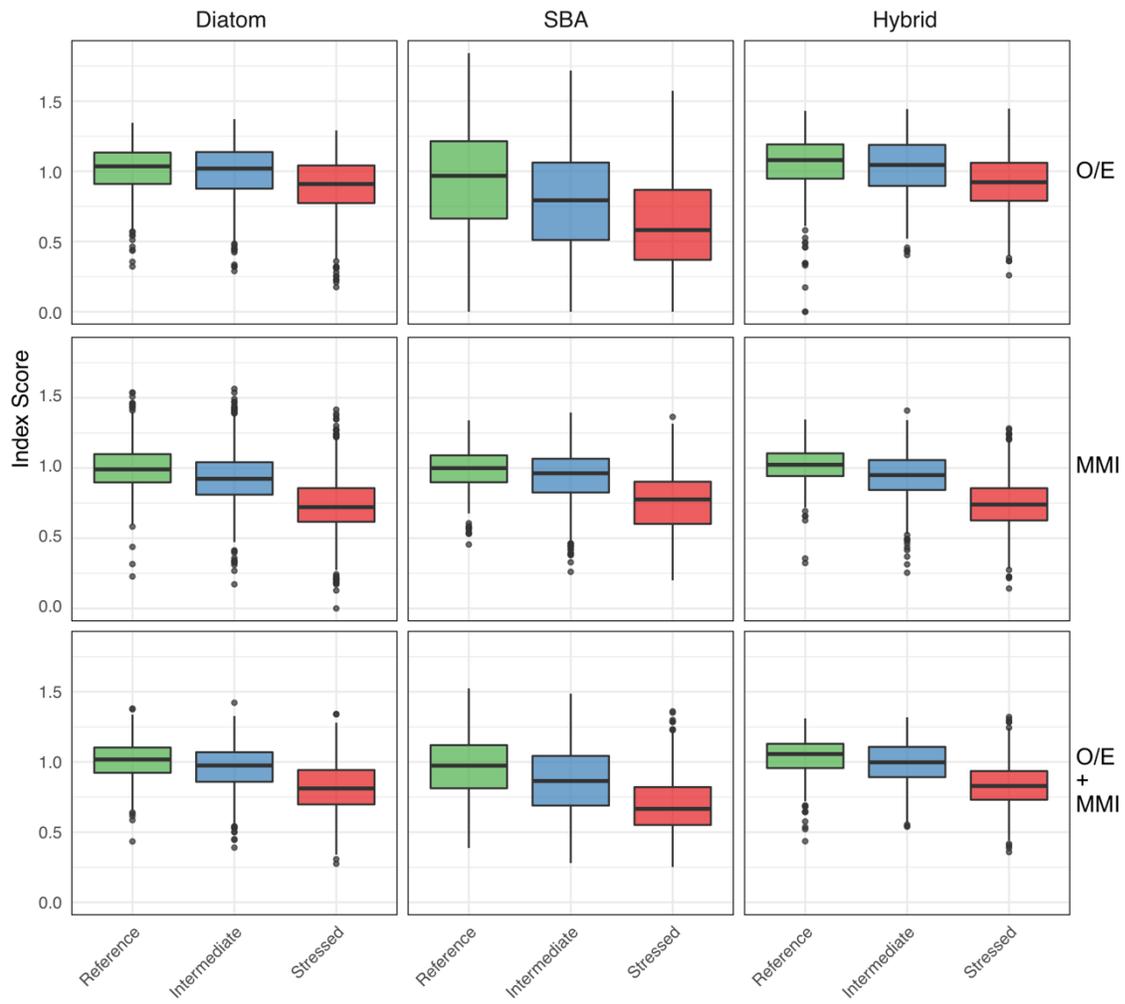


Figure 2. Algal index scores at reference calibration sites across PSA regions. CV= Central Valley, SC= South Coast, CH=Chaparral, DM=Desert/Modoc, NC=North Coast, SN=Sierra Nevada.

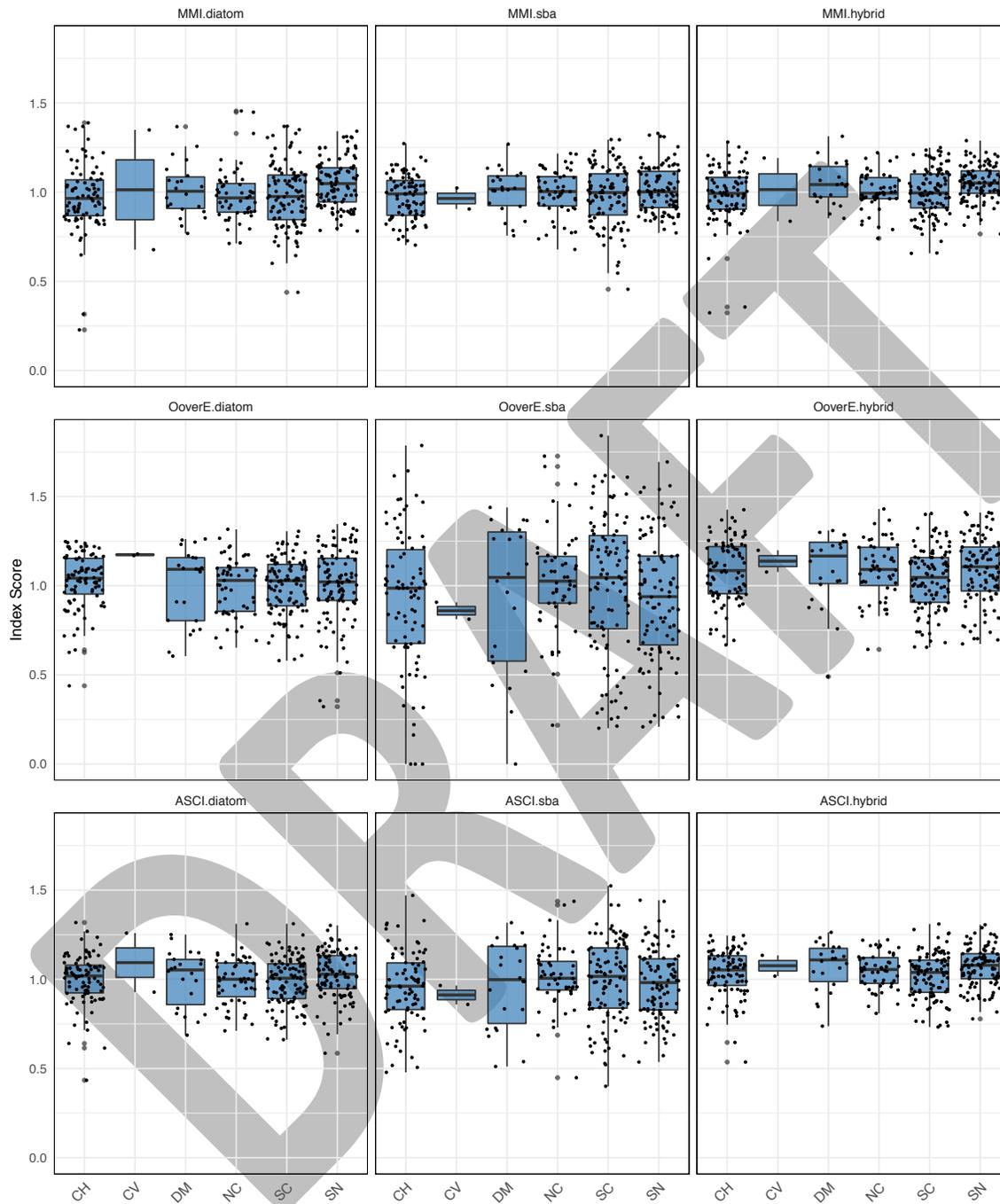


Figure 3. Diatom, SBA, and hybrid MMI scores across environmental gradients for reference calibration sites. R^2 values for linear regression of relationship as shown ($p < 0.001$).

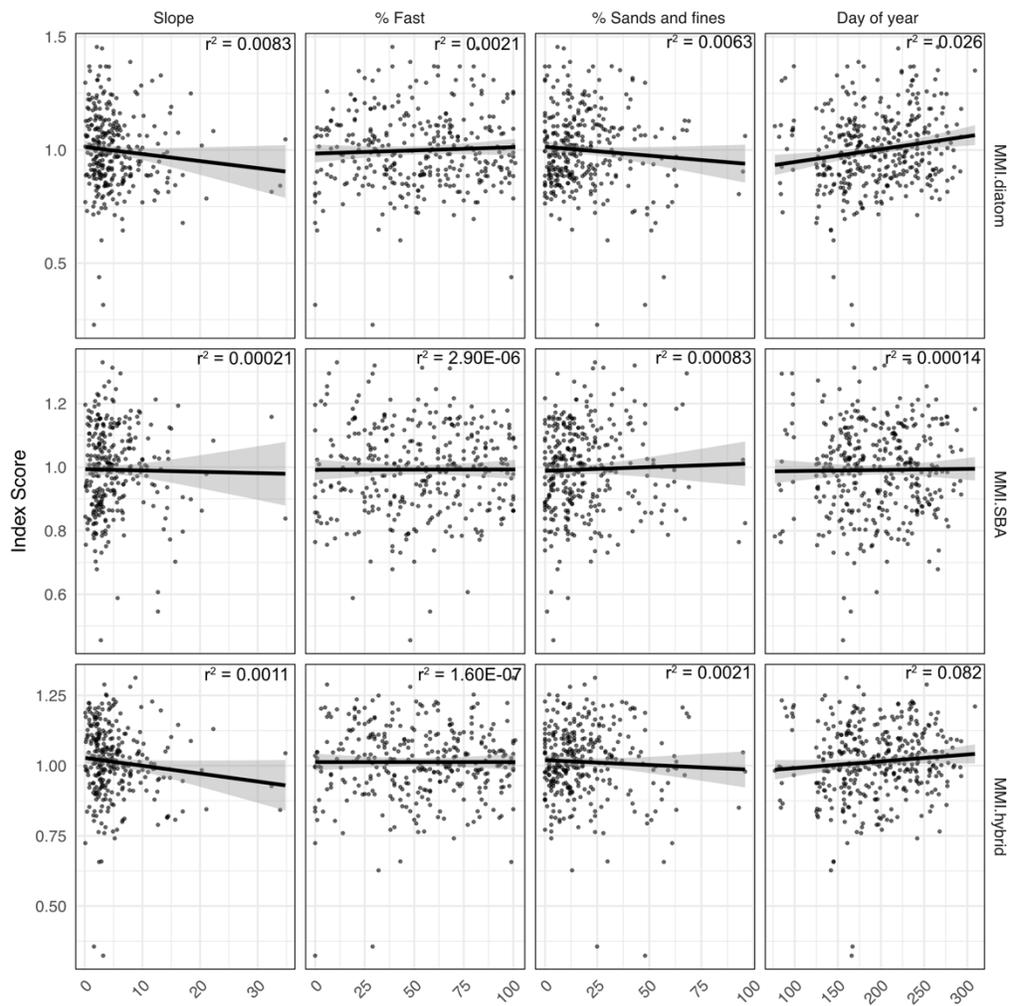


Figure 4. MMI scores across stressor gradients (log transformed) of percent urban development at the watershed scale, riparian activity (W1 Hall), percent sands and fines, total nitrogen, and total phosphorus. R^2 values for linear regression of relationship as shown ($p < 0.001$).

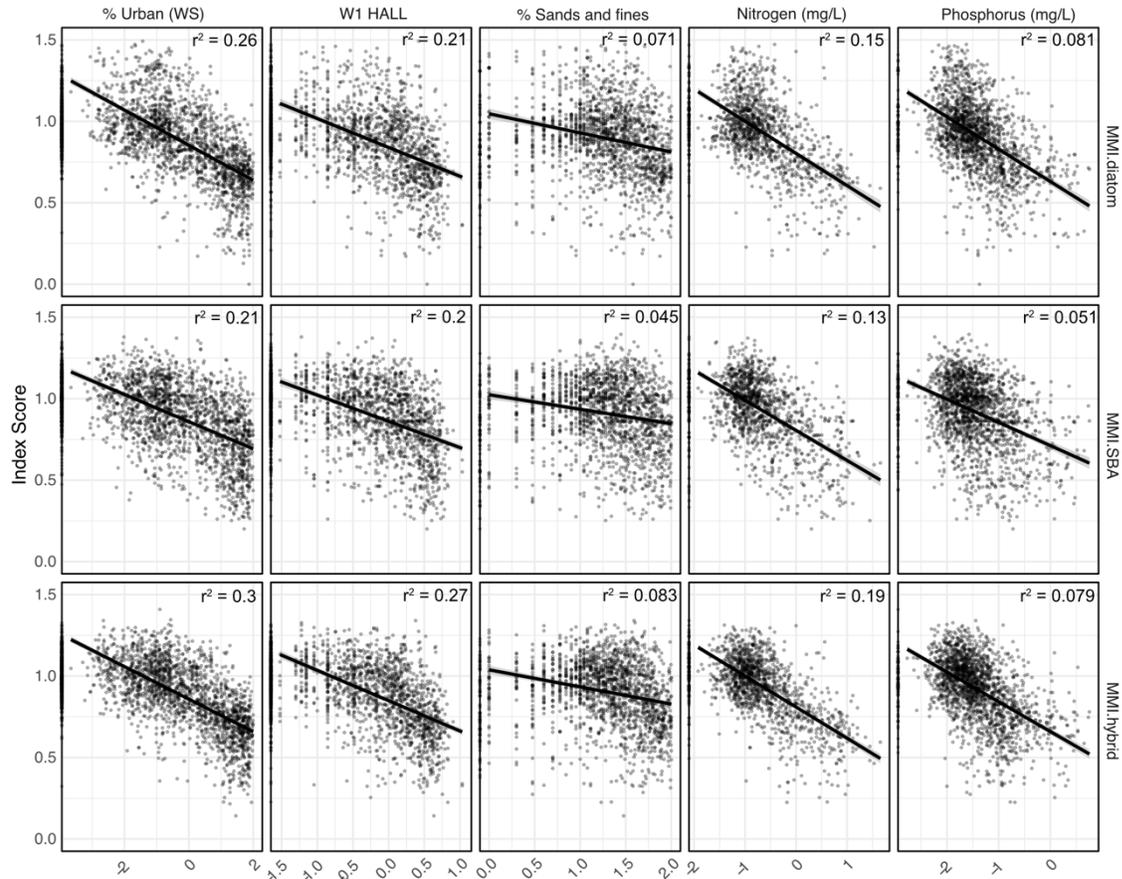


Figure 5. Effect of expected number of taxa (E) on accuracy, precision, and sensitivity of the O/E index performance for both the predictive and null O/E models. Accuracy = proportion of reference calibration sites in reference condition (i.e., score >10th percentile of reference calibration sites) for each index. Precision = standard deviation of reference calibration sites for each index. Sensitivity = proportion of high-activity sites not in reference condition.

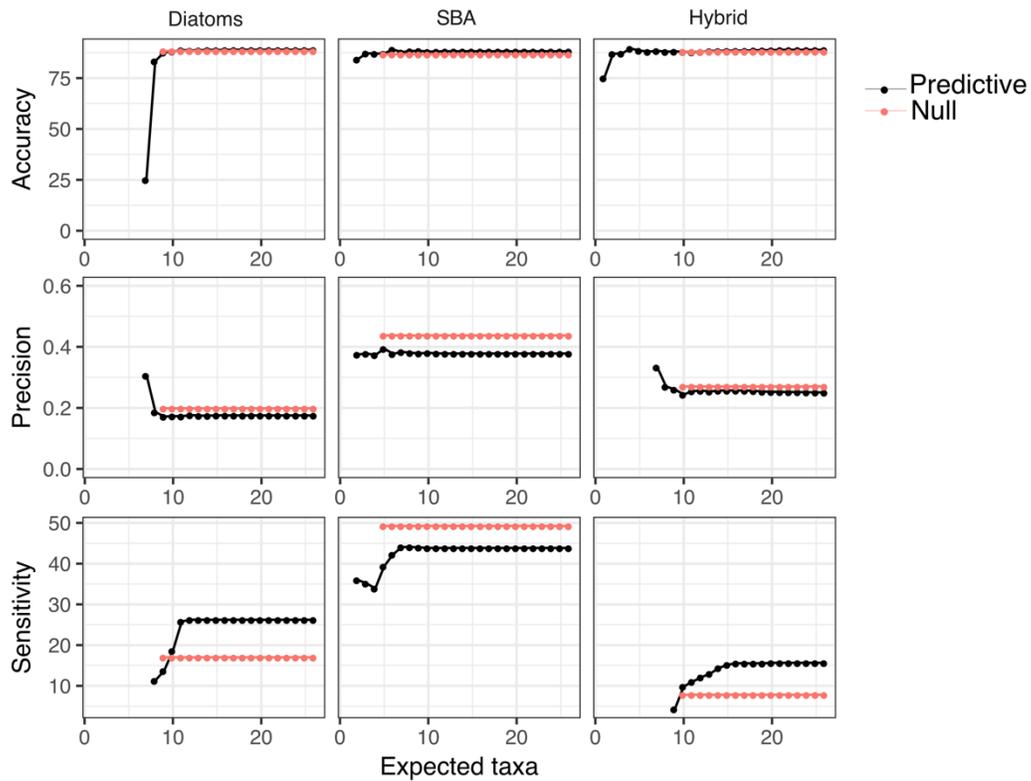


Figure 6. Hybrid MMI results across California. Inset: detail view of hybrid MMI scores for the Bay Area (top) and Los Angeles (bottom).

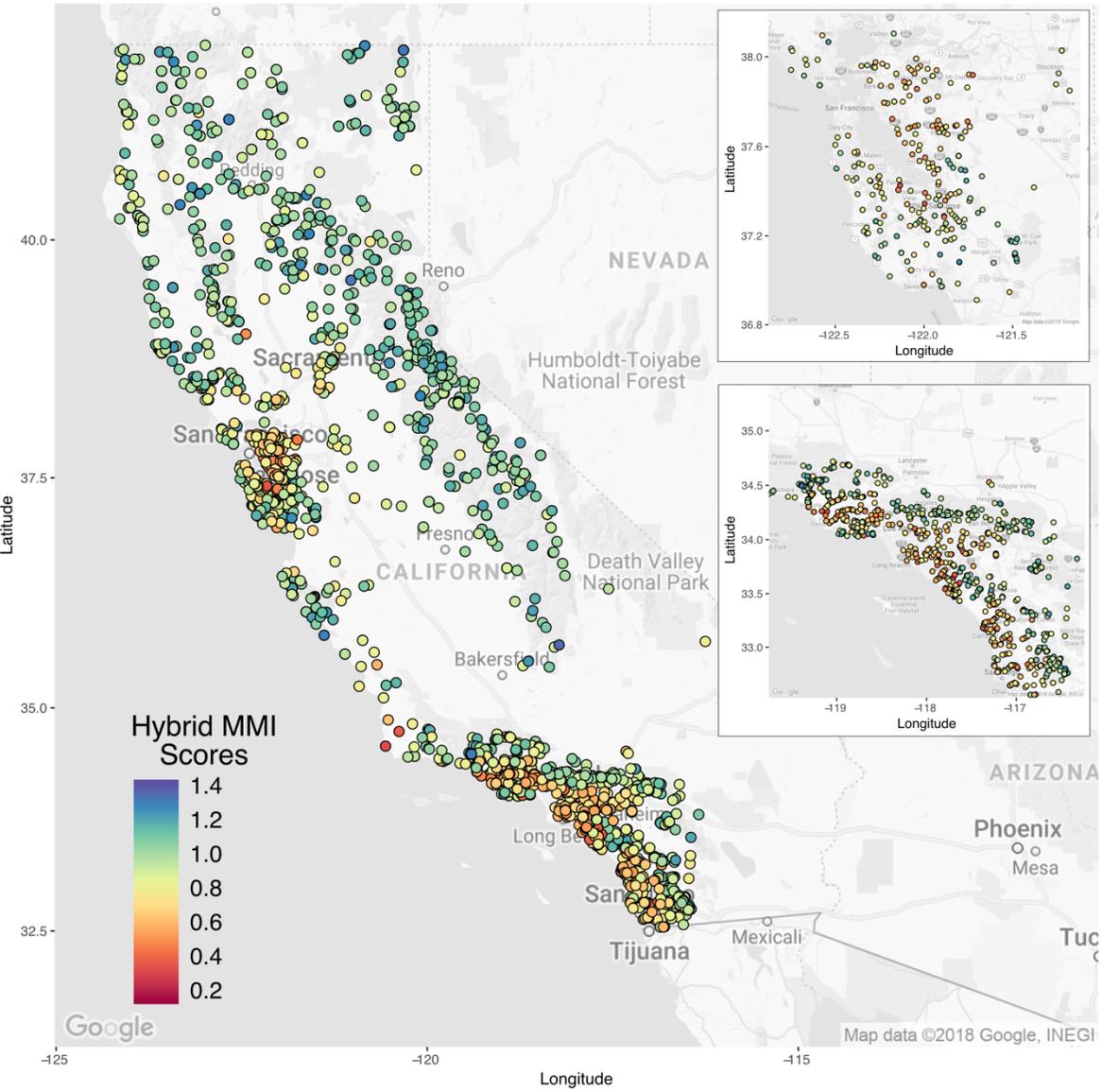


Figure 7. GAM model results of algal index scores and CSCI scores versus stressor gradients.

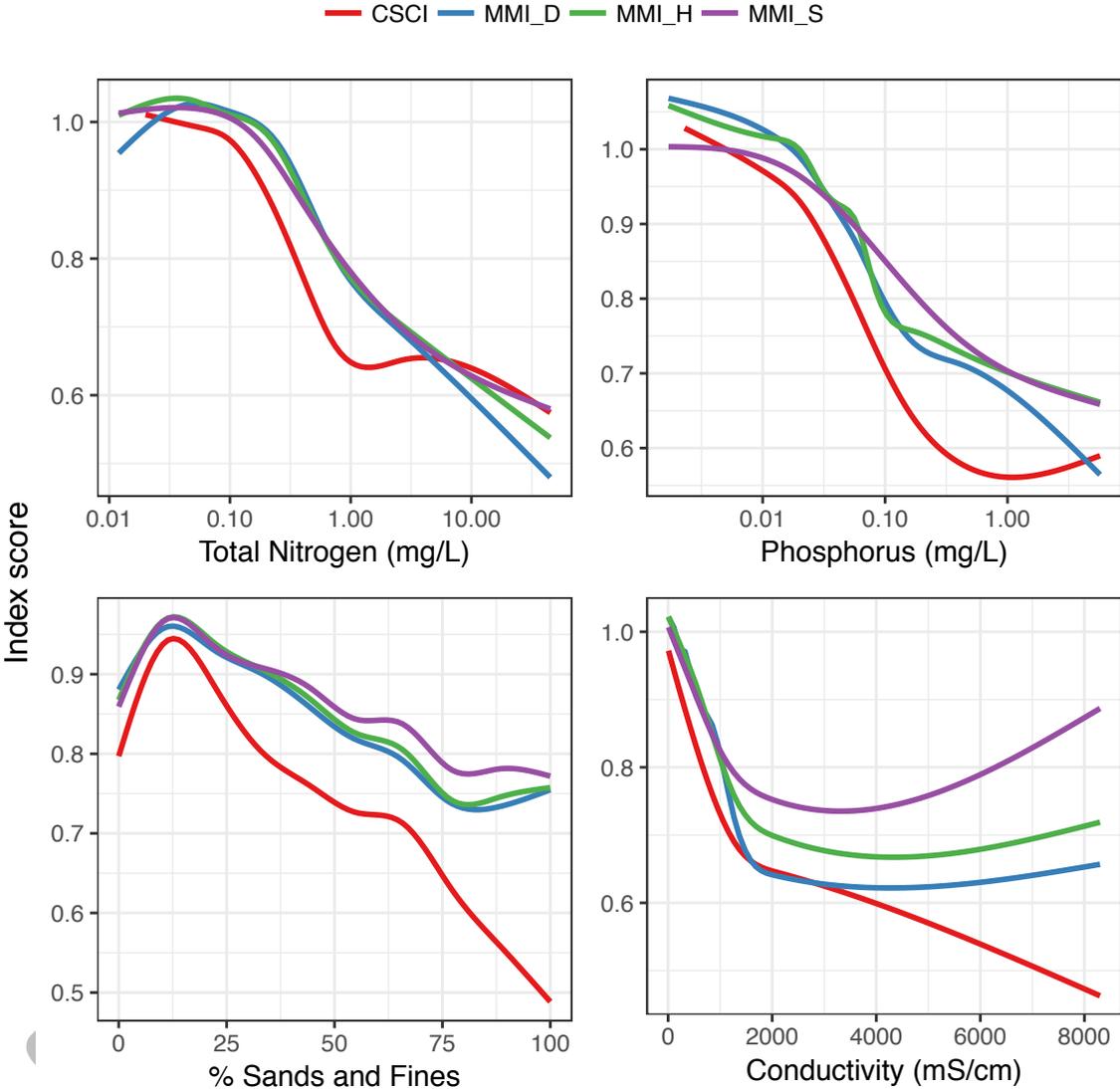


Table S1. Ecoregion O/E model performance metrics. Standard deviation (SD) of the predictive model and the null model (SD null) for each O/E index for the reference calibration dataset.

Assemblage	EcoRegion	SD	SD null
Diatoms	Sierra Nevada	0.18	0.25
Soft-bodied algae	Sierra Nevada	0.32	0.33
Hybrid	Sierra Nevada	0.27	0.28
Diatoms	South Coast	0.18	0.21
Soft-bodied algae	South Coast	0.34	0.40
Hybrid	South Coast	0.21	0.23

DRAFT

Table S2. Spearman's correlation Rho values for final index scores versus stressor gradients (top) and reference calibration scores versus stressor gradients (bottom). Abbreviations as in Table 1 and Table 2. Values < -0.4 and > 0.4 highlighted red for emphasis.

Stressor gradients	Diatom			Hybrid			SBA		
	O/E + MMI	MMI	O/E	O/E + MMI	MMI	O/E	O/E + MMI	MMI	O/E
Chloride (mg/L)	-0.52	-0.61	-0.28	-0.47	-0.59	-0.27	-0.37	-0.42	-0.24
DOC (mg/L)	-0.39	-0.43	-0.22	-0.34	-0.41	-0.20	-0.32	-0.31	-0.25
TN (mg/L)	-0.46	-0.49	-0.31	-0.47	-0.55	-0.29	-0.44	-0.45	-0.32
% sands and fines	-0.17	-0.23	-0.06	-0.21	-0.25	-0.13	-0.33	-0.18	-0.33
pH	-0.04	-0.02	-0.05	-0.01	-0.03	0.00	-0.01	-0.10	0.04
Total Phosphorus (mg/L)	-0.39	-0.49	-0.18	-0.39	-0.51	-0.20	-0.42	-0.33	-0.36
Road Density (5K)	-0.46	-0.55	-0.23	-0.43	-0.57	-0.20	-0.41	-0.43	-0.28
Specific Cond (uS/cm)	-0.52	-0.59	-0.31	-0.46	-0.55	-0.27	-0.35	-0.41	-0.23
TEMP_00_09	-0.43	-0.44	-0.28	-0.41	-0.45	-0.28	-0.36	-0.40	-0.24
URBAN_2011_5K	-0.53	-0.58	-0.32	-0.50	-0.62	-0.28	-0.47	-0.50	-0.32
W1_HALL_SWAMP	-0.38	-0.44	-0.20	-0.38	-0.48	-0.20	-0.35	-0.40	-0.23
Environmental gradients	Diatom			Hybrid			SBA		
	O/E + MMI	MMI	O/E	O/E + MMI	MMI	O/E	O/E + MMI	MMI	O/E
Day of Year	0.08	0.16	-0.04	0.10	0.11	0.07	-0.04	-0.02	-0.05
Month	0.07	0.14	-0.03	0.09	0.09	0.07	-0.05	-0.01	-0.07
AREA (km ²)	0.18	0.22	0.07	0.10	0.17	0.03	0.02	0.00	0.03
AtmCa	0.00	0.02	0.00	-0.03	0.01	-0.05	-0.04	-0.04	-0.04
AtmMg	-0.17	-0.10	-0.18	-0.05	-0.08	-0.01	-0.10	-0.13	-0.06
BDH_AVE	-0.02	-0.04	0.02	-0.04	-0.03	-0.04	-0.01	-0.03	0.01
CaO_Mean	0.04	-0.01	0.05	0.05	0.04	0.03	-0.10	-0.01	-0.11
CondQR50	-0.10	-0.14	-0.06	-0.05	-0.06	-0.04	-0.06	-0.09	-0.03
ELEV_RANGE	0.15	0.19	0.05	-0.01	0.06	-0.05	-0.07	-0.01	-0.08
KFCT_AVE	-0.03	-0.08	0.02	0.05	0.01	0.05	0.00	-0.01	0.02
LogWSA	0.18	0.22	0.07	0.10	0.17	0.03	0.02	0.00	0.03
LPREM_mean	0.04	0.00	0.05	0.08	0.02	0.10	-0.08	0.00	-0.09
LST32AVE	0.17	0.21	0.08	0.09	0.13	0.05	0.04	0.11	0.00
MAX_ELEV	0.19	0.27	0.05	0.04	0.13	-0.01	-0.03	0.03	-0.05
MAXWD_WS	0.05	0.06	0.01	0.04	-0.01	0.06	0.01	0.02	-0.01
MEANP_WS	0.02	0.08	-0.03	-0.02	-0.01	-0.01	0.04	0.05	0.02
MgO_Mean	0.06	-0.02	0.09	0.09	0.03	0.10	-0.05	-0.01	-0.06
MINP_WS	0.10	0.13	0.06	0.05	0.05	0.05	0.04	0.07	0.00
Latitude	0.04	0.03	0.02	0.06	0.00	0.07	0.05	0.05	0.03
Longitude	0.02	0.04	0.00	-0.02	0.04	-0.05	-0.03	-0.02	-0.02
PCT_CENOZ	-0.03	0.02	-0.10	-0.04	0.01	-0.06	-0.02	-0.06	-0.01
PCT_NOSED	-0.01	0.02	-0.01	-0.07	0.01	-0.09	0.03	0.02	0.03
PCT_QUART	-0.08	-0.06	-0.05	-0.06	-0.06	-0.03	-0.06	0.00	-0.07
PCT_SEDIM	0.00	0.00	-0.03	0.01	0.01	0.00	-0.03	-0.02	-0.03
PCT_VOLCNC	0.05	0.03	0.07	0.09	0.01	0.11	0.03	0.03	0.02
PPT_00_09	-0.01	0.04	-0.04	-0.03	-0.03	-0.02	0.04	-0.01	0.03
PRMH_AVE	0.05	0.07	0.04	-0.02	0.00	-0.03	0.03	0.01	0.02
PSA6cnum	0.06	0.17	-0.04	0.02	0.13	-0.03	0.05	0.07	0.03
S_Mean	-0.06	-0.03	-0.08	-0.06	-0.01	-0.08	-0.04	-0.06	-0.02
SITE_ELEV	0.13	0.19	0.04	0.07	0.13	0.02	0.02	0.08	-0.01
TEMP_00_09	-0.09	-0.14	-0.03	-0.06	-0.09	-0.03	-0.01	-0.08	0.03
TMAX_WS	-0.11	-0.14	-0.07	-0.04	-0.05	-0.03	0.01	-0.05	0.05
UCS_Mean	-0.03	-0.03	0.01	-0.07	-0.05	-0.06	-0.01	-0.04	0.00
XWD_WS	0.06	0.07	0.03	0.05	-0.01	0.07	0.03	0.05	0.00

Table S3. Metric screening results for both species- and genus-level metrics for inclusion in the MMI. Assem = assemblage. RF (rsq) = RandomForest pseudo-R-squared. PSA ANOVA = PSA ecoregion ANOVA test for regional bias. T-test = t test (t statistic and p value) comparing average reference calibration and high activity (stressed) site scores. S:N = signal to noise test. SD = standard deviation of reference calibration site scores. For complete table, click [HERE](#).

	Assem.	Level	RF (rsq)	PSA ANOVA (p)	PSA ANOVA (F)	T-test (t)	T-test (p)	Freq Zero	Freq One	S:N (p)	S:N (F)	Range (ref)	Range (stress)	SD
Achnanthes.richness_raw	diatom	spp	NA	0.51	0.83	3.01	0.00	1.90	0.00	0.00	1.90	1.02	1.02	0.14
Amphora.richness_raw	diatom	spp	NA	0.00	7.36	8.68	0.00	1.63	0.00	0.78	0.91	1.01	1.01	0.15
cnt.ind.least.tol_raw	diatom	spp	NA	0.02	3.02	8.75	0.00	0.00	0.00	0.00	2.25	1.65	0.08	0.80
cnt.spp.BCG12_raw	diatom	spp	NA	0.00	17.47	22.07	0.00	2.71	0.00	0.00	3.07	1.00	0.20	0.52
cnt.spp.BCG2_raw	diatom	spp	NA	0.00	17.80	22.05	0.00	2.71	0.00	0.00	3.14	1.00	0.20	0.52
cnt.spp.BCG3_raw	diatom	spp	NA	0.00	5.50	10.43	0.00	1.36	0.00	0.00	3.09	0.95	0.60	0.51
cnt.spp.BCG4_raw	diatom	spp	NA	0.00	25.28	15.46	0.00	0.54	0.00	0.00	3.45	1.03	0.68	0.31
cnt.spp.Halo_raw	diatom	spp	NA	0.00	21.58	17.50	0.00	0.00	0.00	0.00	3.28	1.00	0.67	0.30
cnt.spp.HighMotility_raw	diatom	spp	NA	0.00	9.38	5.57	0.00	0.00	0.00	0.00	2.39	1.37	1.15	0.57
cnt.spp.IndicatorClass_TN_high_raw	diatom	spp	NA	0.00	13.53	11.05	0.00	0.27	0.00	0.00	3.18	1.08	0.77	0.40
cnt.spp.IndicatorClass_TN_low_raw	diatom	spp	NA	0.00	12.09	18.31	0.00	0.27	0.00	0.00	4.11	0.94	0.40	0.46
cnt.spp.IndicatorClass_TP_high_raw	diatom	spp	NA	0.00	16.78	12.16	0.00	0.54	0.00	0.00	3.23	1.05	0.72	0.39
cnt.spp.IndicatorClass_TP_low_raw	diatom	spp	NA	0.00	6.95	17.09	0.00	0.27	0.00	0.00	3.69	1.05	0.44	0.47
cnt.spp.least.tol_raw	diatom	spp	NA	0.02	3.02	8.75	0.00	0.00	0.00	0.00	2.25	1.65	0.08	0.80

Table S4. All O/E, MMI, and combined O/E+MMI index scores for all sites. For complete table, click [HERE](#).

	SampleDate	SiteSetSample2	O.diatom	E.diatom	OoverE.diatom	OoverE.null.diatom	MMI.diatom	OE+MMI.diatom
000CAT148_8.10.10_1	8/10/10	Reference	7	8.01	0.87	0.92	1.27	1.07
000CAT228_8.10.10_1	8/10/10	Reference	8	8.79	0.91	0.92	1.16	1.04
102PS0139_8.9.10_1	8/9/10	Intermediate	8	8.18	0.98	0.92	0.92	0.95
102PS0177_8.28.12_1	8/28/12	Reference	7	8.00	0.87	0.92	1.33	1.10
102PS0177_8.28.12_2	8/28/12	Reference	7	8.08	0.87	0.81	1.33	1.10
103CDCHHR_9.14.10_1	9/14/10	Reference	8	7.55	1.06	0.92	0.84	0.95
103CDCHHR_9.14.10_2	9/14/10	Reference	7	7.62	0.92	1.04	0.91	0.92
103FC1106_7.15.14_1	7/15/14	NA	8	8.10	0.99	0.81	1.10	1.04
103FCA168_7.24.13_1	7/24/13	Intermediate	9	7.96	1.13	0.92	0.97	1.05
103KLCMSR_9.14.10_1	9/14/10	Reference	9	8.97	1.00	1.04	0.89	0.94
103PS0217_9.10.13_1	9/10/13	Intermediate	10	7.78	1.29	1.15	0.80	1.04
103RDCBCC_7.21.11_1	7/21/11	Intermediate	8	8.06	0.99	0.81	0.80	0.89
103STCNSR_8.9.11_1	8/9/11	Reference	4	8.64	0.46	0.46	0.85	0.66
103WER026_8.15.12_1	8/15/12	Reference	6	7.81	0.77	0.92	1.00	0.88
103WER026_8.15.12_2	8/15/12	Reference	7	6.92	1.01	0.92	0.99	1.00
103WER029_7.16.14_1	7/16/14	Reference	NA	NA	NA	NA	NA	NA
103WER030_8.10.11_1	8/10/11	Reference	5	7.93	0.63	0.69	0.89	0.76
105BVCAGC_9.21.10_1	9/21/10	Reference	6	7.97	0.75	0.58	0.82	0.79
105CE0329_7.23.14_1	7/23/14	Reference	NA	NA	NA	NA	NA	NA
105CLCATC_8.10.10_1	8/10/10	Reference	8	7.77	1.03	1.04	0.94	0.99
105CR0901_7.23.14_1	7/23/14	NA	9	9.75	0.92	1.04	1.25	1.09
105CR0901_8.11.14_1	8/11/14	NA	13	9.75	1.33	1.27	1.00	1.17
105CR0906_8.12.14_1	8/12/14	NA	12	9.75	1.23	1.27	1.27	1.25
105CR0908_7.22.14_1	7/22/14	NA	11	9.52	1.16	1.15	0.99	1.07

Figure S1. Development dataset distribution of sites deemed representative of Reference, Intermediate, or high activity (Stressed) conditions after reference screening.

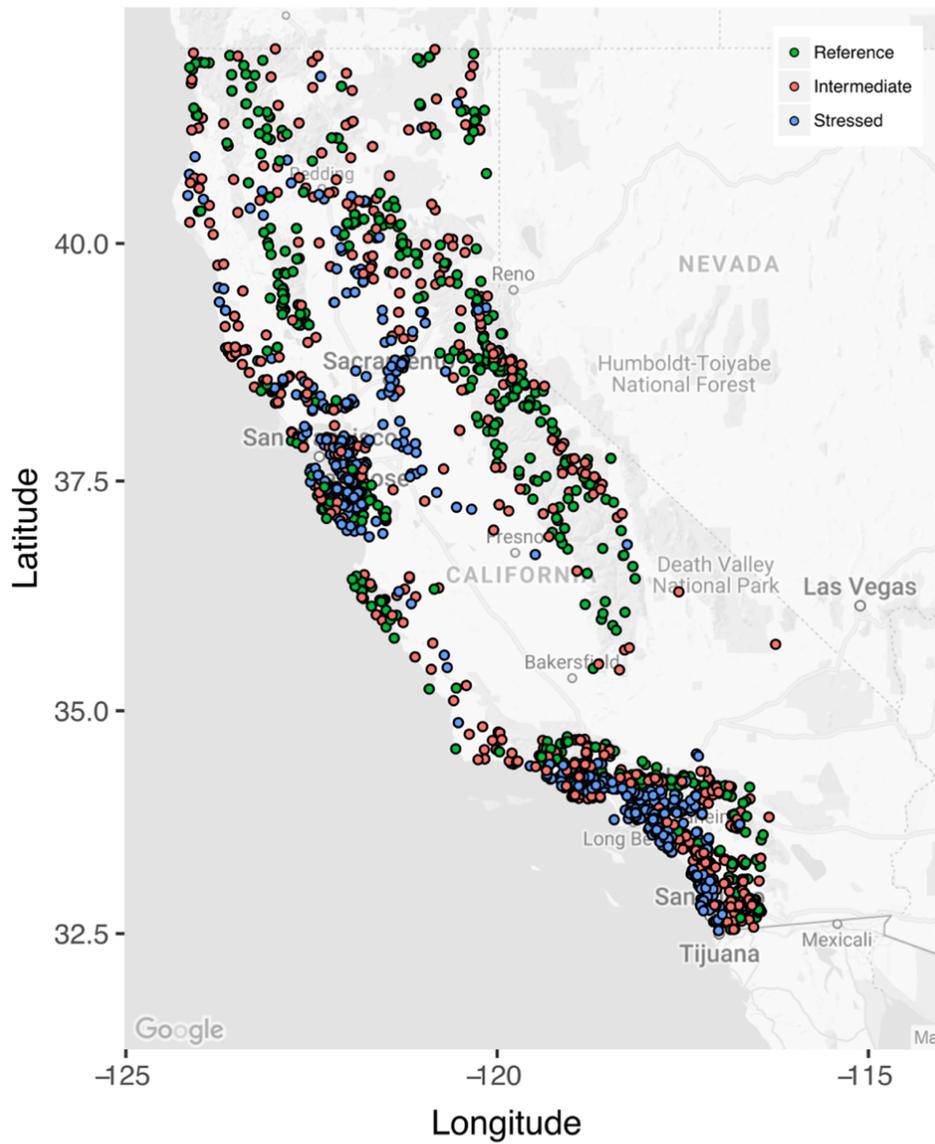


Figure S2. O/E model reference calibration site cluster groupings for each algal assemblage.

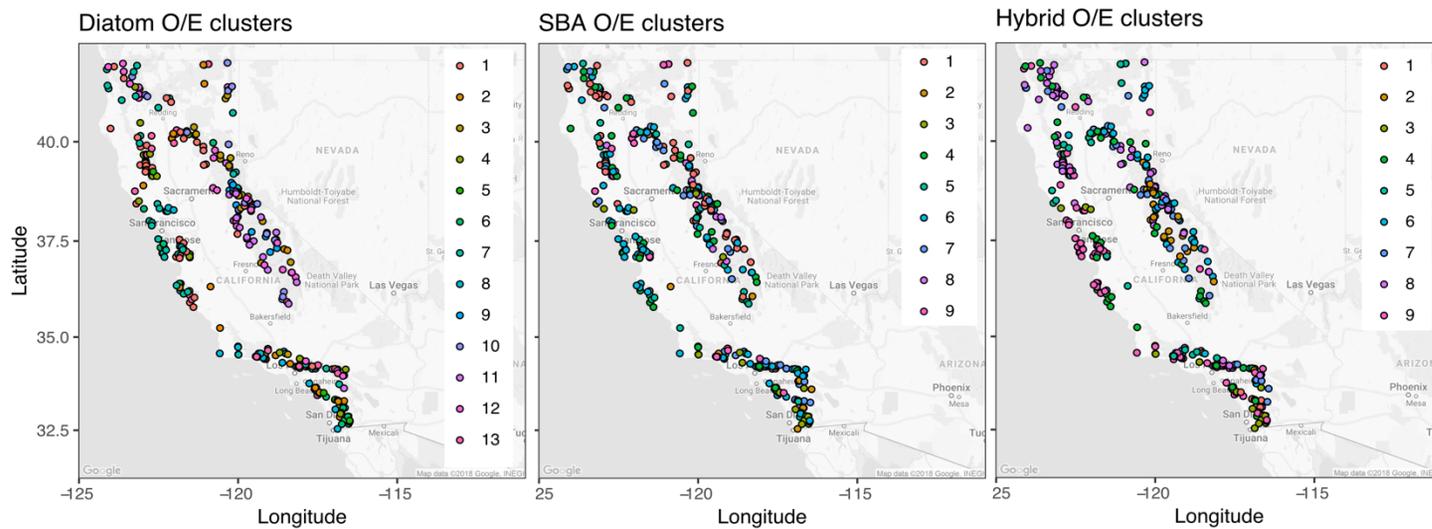


Figure S3. Index performance at Reference and Stressed sites across PSA regions. CV= Central Valley, SC= South Coast, CH=Chaparral, DM=Desert/Modoc, NC=North Coast, SN=Sierra Nevada.

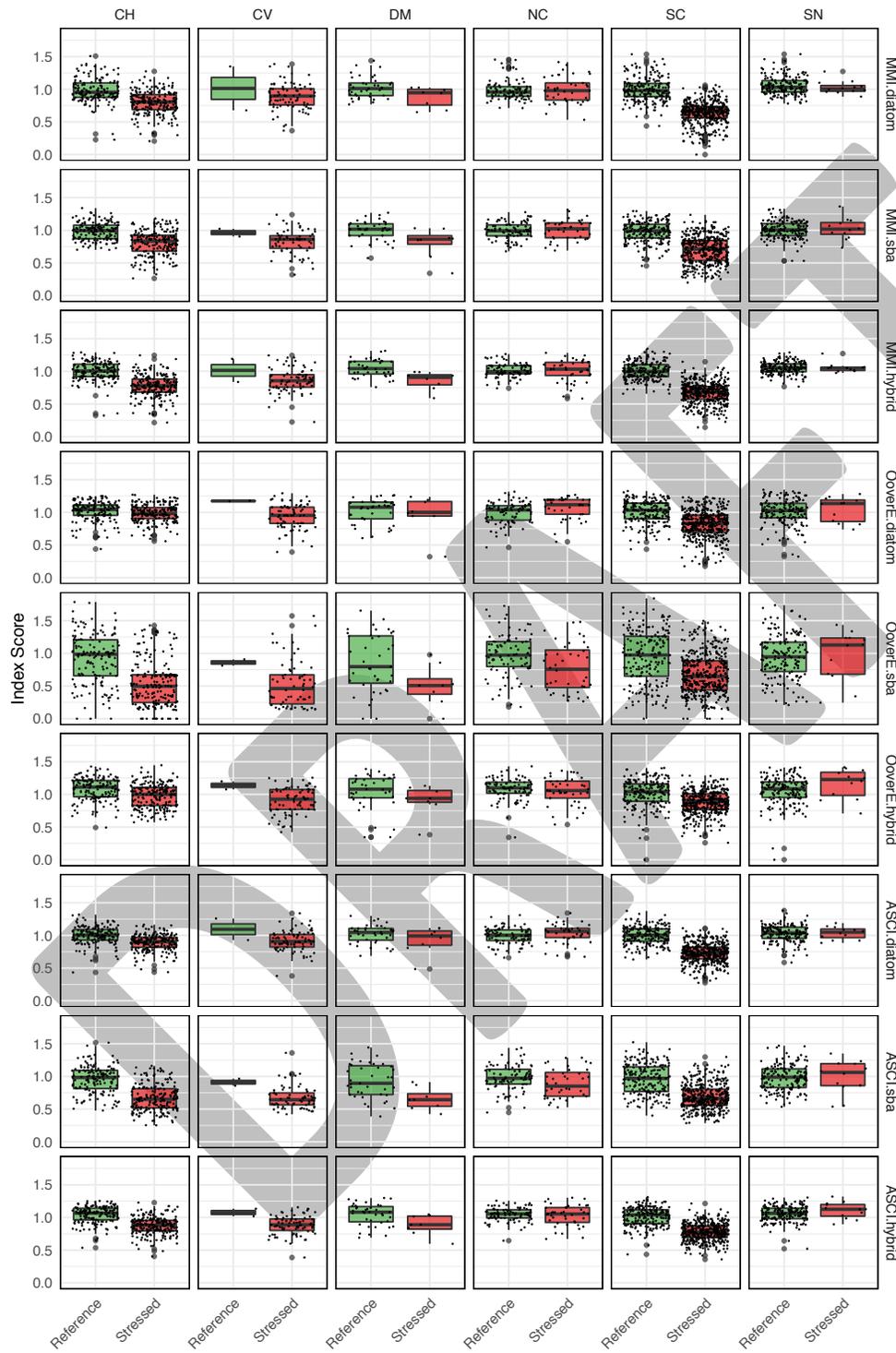


Figure S4. California Stream Condition Index (CSCI) and Southern California Index of Biotic Integrity (IBI) scores versus algal index scores (this paper). The CSCI index has O/E, MMI and a combined CSCI index. The IBI has a diatom-only (D18), soft-bodied algae only (S2), and hybrid (H20) MMI indices. The dashed line indicates a 1:1 ratio; the blue line represents a linear model best fit relationship with 95% confidence interval as shown. R^2 values for the linear model as shown.

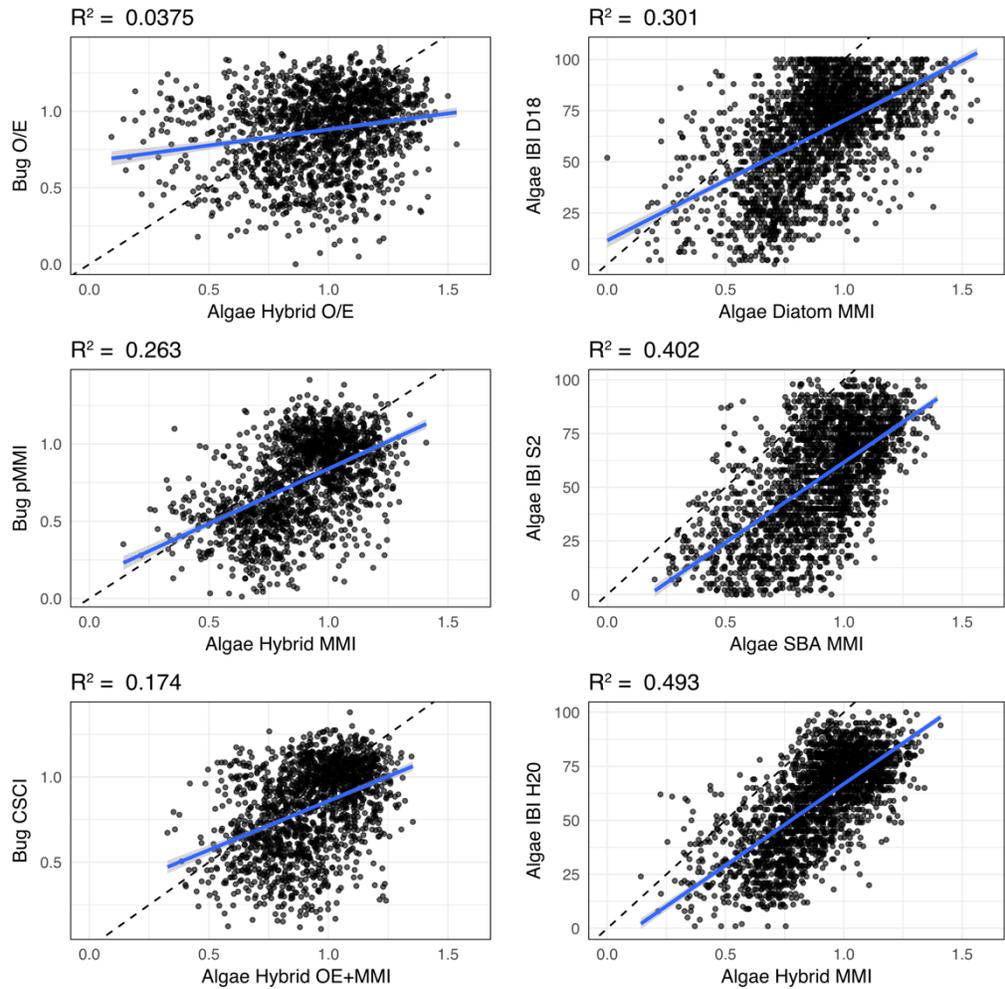


Figure S5. Comparison between stressor and environmental gradients and the difference between CSCI and hybrid MMI scores. the blue line represents a linear model best fit relationship with 95% confidence interval as shown.

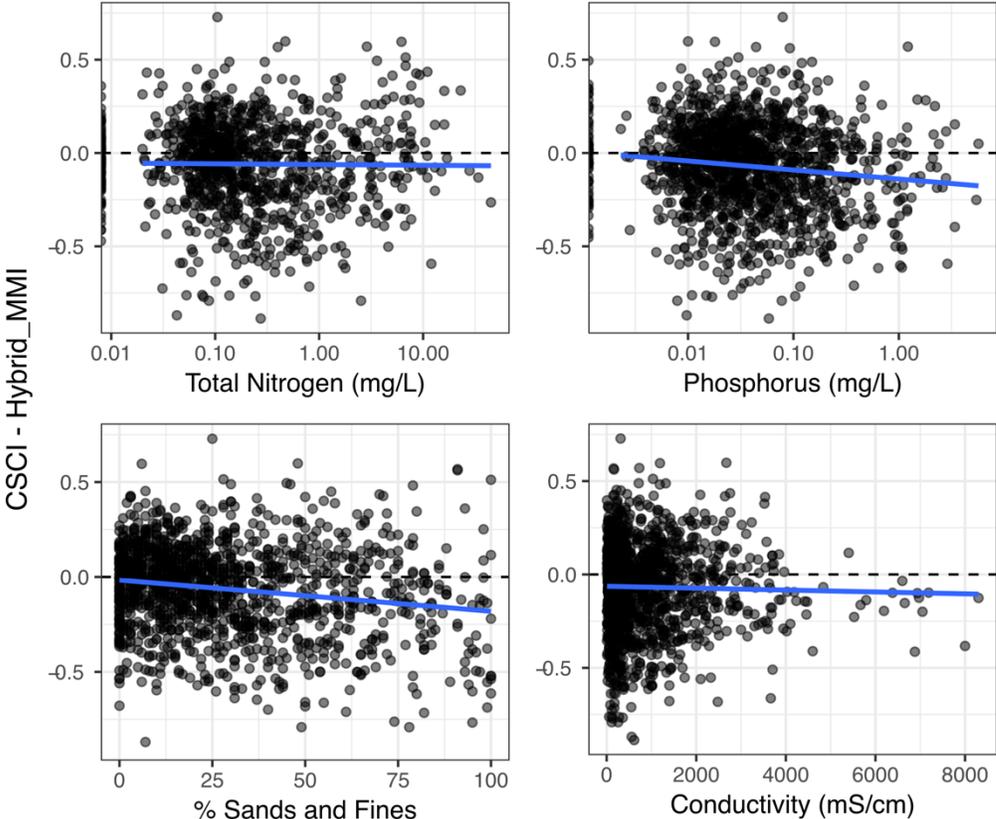


Figure S6. Comparison of algae and benthic macroinvertebrate (BMI) Bray-Curtis distances for 200 reference calibration sites. Graphs show distribution of Bray-Curtis distances for comparison of all site x site comparisons.

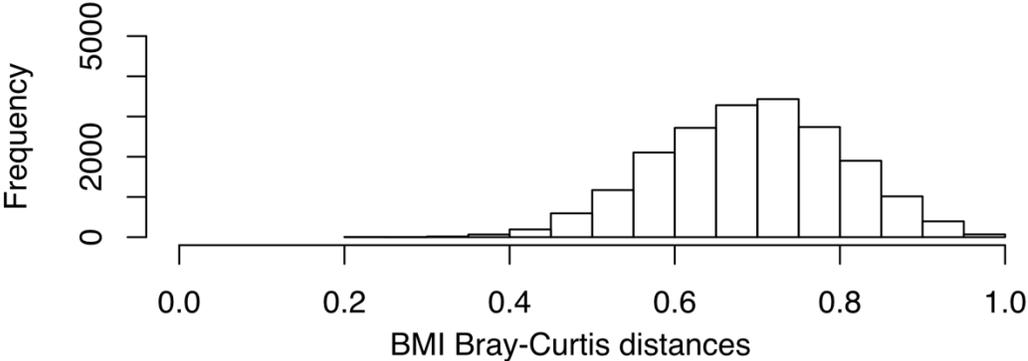
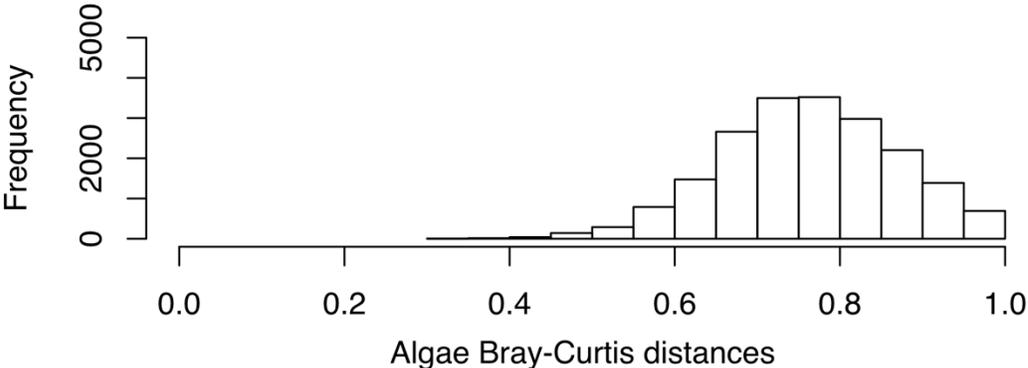
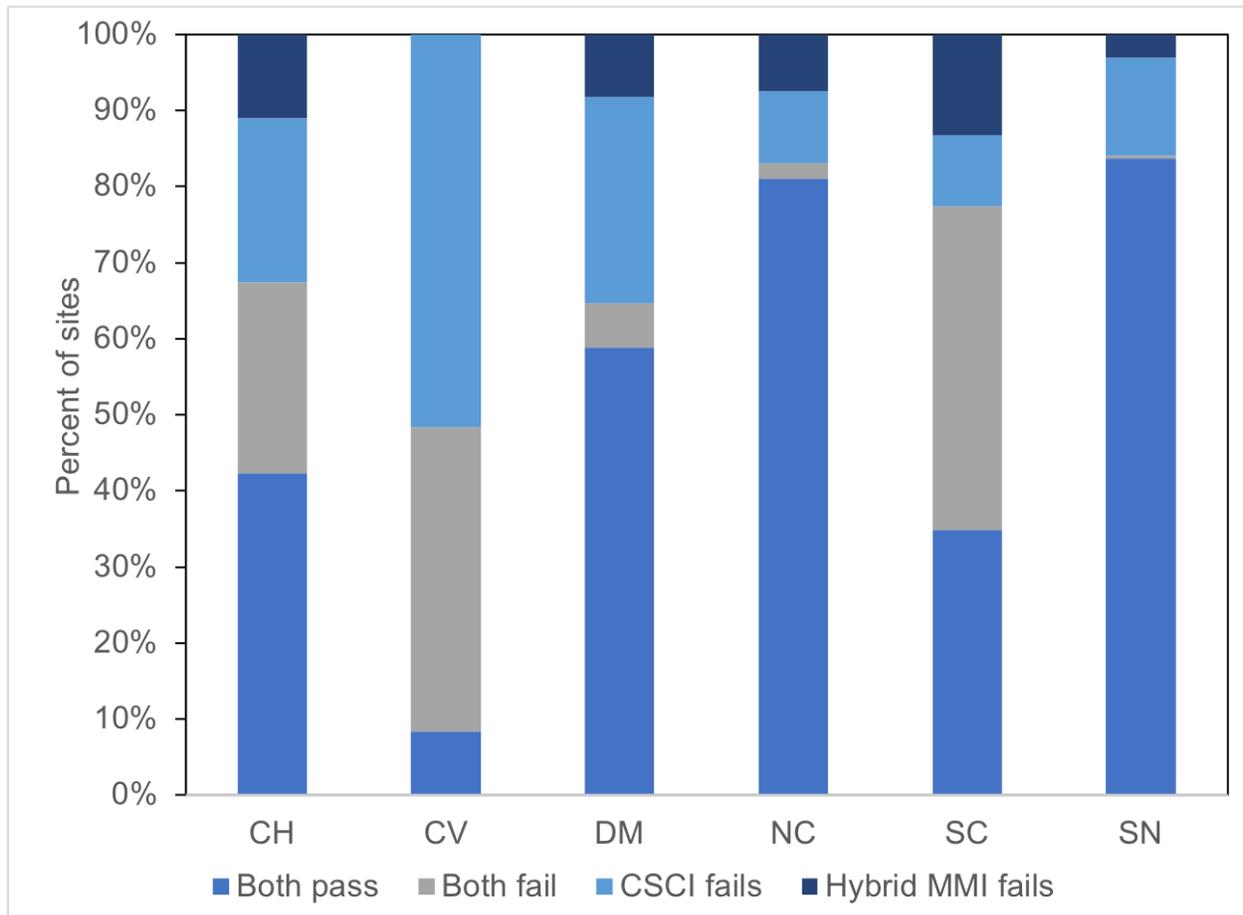


Figure S7. Distribution of CSCI and algal hybrid MMI scores agreement across PSA regions. Sites were evaluated based on score exceeding the 10th percentile of reference site scores, 0.79 for the CSCI and 0.83 for the algal hybrid MMI. CV= Central Valley, SC= South Coast, CH=Chaparral, DM=Desert/Modoc, NC=North Coast, SN=Sierra Nevada.



REFERENCES

- Bahls, L., 1993. Periphyton bioassessment methods for Montana streams. Montana Water Quality Bureau, Department of Health and Environmental Science, Helena, USA.
- Blinn, D., Herbst, D., 2003. Use of diatoms and soft algae as indicators of stream abiotic determinants in the Lahontan Basin, USA (Final report to the California regional water quality control board, Lahontan region).
- Blocksom, K.A., 2003. A Performance Comparison of Metric Scoring Methods for a Multimetric Index for Mid-Atlantic Highlands Streams. *Environ. Manage.* 31, 0670–0682. <https://doi.org/10.1007/s00267-002-2949-3>
- Caceres, M.D., Legendre, P., 2009. Associations between species and groups of sites: indices and statistical inference.
- Cao, Y., Hawkins, C.P., Olson, J., Kosterman, M.A., 2007. Modeling natural environmental gradients improves the accuracy and precision of diatom-based indicators. *J. North Am. Benthol. Soc.* 26, 566–585. <https://doi.org/10.1899/06-078.1>
- Cemagref, 1982. Étude des méthodes biologiques quantitative d’appréciation de la qualité des eaux. Bassin Rhône-Méditerranée-Corse. Centre National du Machinisme Agricole, du Génie rural, des Eaux et des Forêts, Lyon, France.
- Coste, M., 1986. Les méthodes microfloristiques d’évaluation de la qualité des eaux. Cemagref, Bordeaux.
- Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. RANDOM FORESTS FOR CLASSIFICATION IN ECOLOGY. *Ecology* 88, 2783–2792. <https://doi.org/10.1890/07-0539.1>
- Dam, H.V., Mertens, A., Sinkeldam, J., 1994. A coded checklist and ecological indicator values of freshwater diatoms from The Netherlands. *Netherland J. Aquat. Ecol.* 28, 117–133. <https://doi.org/10.1007/BF02334251>
- 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. <https://doi.org/10.1007/s10811-013-0088-2>
- Hawkins, C.P., Cao, Y., Roper, B., 2010. Method of predicting reference condition biota affects the performance and interpretation of ecological indices. *Freshw. Biol.* 55, 1066–1085. <https://doi.org/10.1111/j.1365-2427.2009.02357.x>
- Hering, D., Johnson, R.K., Kramm, S., Schmutz, S., Szoszkiewicz, K., Verdonschot, P.F.M., 2006. Assessment of European streams with diatoms, macrophytes, macroinvertebrates and fish: a comparative metric-based analysis of organism response to stress. *Freshw. Biol.* 51, 1757–1785. <https://doi.org/10.1111/j.1365-2427.2006.01610.x>
- Kaufmann, P.R., Levine, P., Peck, D.V., Robison, E.G., Seeliger, C., 1999. Quantifying physical habitat in wadeable streams. USEPA [National Health and Environmental Effects Research Laboratory, Western Ecology Division].
- Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. *R News* 2, 18–22.
- 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>
- Meyer, J.L., Strayer, D.L., Wallace, J.B., Eggert, S.L., Helfman, G.S., Leonard, N.E., 2007. The contribution of headwater streams to biodiversity in river networks. *J. Am. Water Resour. Assoc.* Vol 43 No 1.
- Ode, P.R., Fetscher, A.E., Busse, L.B., 2016a. Standard Operating Procedures (SOP) for the Collection of Field Data for Bioassessments of California Wadeable Streams: Benthic Macroinvertebrates, Algae, and Physical Habitat (No. Technical Report 835). State Water Resources Control Board Surface Water Ambient Monitoring Program. Sacramento, CA.
- 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., 2016b. 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>
- Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2017. *vegan: Community Ecology Package*.
- Olson, J.R., Hawkins, C.P., 2012. Predicting natural base-flow stream water chemistry in the western United States. *Water Resour. Res.* 48. <https://doi.org/10.1029/2011WR011088>
- Porter, S.D., Mueller, D.K., Spahr, N.E., Munn, M.D., Dubrovsky, N.M., 2008. Efficacy of algal metrics for assessing nutrient and organic enrichment in flowing waters. *Freshw. Biol.* 53, 1036–1054. <https://doi.org/10.1111/j.1365-2427.2007.01951.x>
- Potapova, M., Charles, D.F., 2007. Diatom metrics for monitoring eutrophication in rivers of the United States. *Ecol. Indic.* 7, 48–70. <https://doi.org/10.1016/j.ecolind.2005.10.001>
- R Core Team, 2013. *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. North Am. Benthol. Soc.* 16, 833–852. <https://doi.org/10.2307/1468175>
- Sork, V.L., Gugger, P.F., Chen, J.-M., Werth, S., 2016. Evolutionary lessons from California plant phylogeography. *Proc. Natl. Acad. Sci.* 113, 8064–8071.
<https://doi.org/10.1073/pnas.1602675113>
- Spaulding, S., Lubinski, D., Potapova, M., 2010. Diatoms of the United States [WWW Document]. Diatoms U. S. URL <http://westerndiatoms.colorado.edu> (accessed 10.10.17).
- Stancheva, R., Busse, L., Kociolek, J.P., Sheath, R., 2015. Standard Operating Procedures for Laboratory Processing and Identification of Stream Algae in California (No. SWAMP-SOP-2015-0003). California State Water Resources Control Board Surface Water Ambient Monitoring Program (SWAMP) Bioassessment SOP 0003.
- Stancheva, R., Sheath, R.G., 2016. Benthic soft-bodied algae as bioindicators of stream water quality. *Knowl. Manag. Aquat. Ecosyst.* 15. <https://doi.org/10.1051/kmae/2016002>
- Stevenson, J., 2014. Ecological assessments with algae: a review and synthesis. *J. Phycol.* 50, 437–461. <https://doi.org/10.1111/jpy.12189>

- Stoddard, J.L., Herlihy, A.T., Peck, D.V., Hughes, R.M., Whittier, T.R., Tarquinio, E., 2008. A process for creating multimetric indices for large-scale aquatic surveys. *J. North Am. Benthol. Soc.* 27, 878–891. <https://doi.org/10.1899/08-053.1>
- Van Sickle, J., Hawkins, C.P., Larsen, D.P., Herlihy, A.T., 2005. A null model for the expected macroinvertebrate assemblage in streams. *J. North Am. Benthol. Soc.* 24, 178–191. [https://doi.org/10.1899/0887-3593\(2005\)024<0178:ANMFTE>2.0.CO;2](https://doi.org/10.1899/0887-3593(2005)024<0178:ANMFTE>2.0.CO;2)
- Wing, M.K.C. from J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Team, the R.C., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt, T., 2017. caret: Classification and Regression Training.
- Wood, S.N., 2017. Generalized Additive Models: An Introduction with R (2nd edition). Chapman and Hall/CRC.

DRAFT