The California Stream Condition Index (CSCI): A New Statewide Biological Scoring Tool for Assessing the Health of Freshwater Streams

Prepared by: Andrew C. Rehn[1], Raphael D. Mazor[2,1], Peter R. Ode[1]

[1]*California Department of Fish and Wildlife*
[2]*Southern California Coastal Water Research Project*

SWAMP Technical Memorandum
SWAMP-TM-2015-0002
September 2015

TABLE OF CONTENTS

OBJECTIVE

The objective of this technical memo is to summarize the development, features and use of SWAMP's next-generation index for monitoring stream health in California.

OVERVIEW

California's dramatic environmental diversity supports a broad array of natural stream types throughout the state. Bioassessment of freshwater stream and rivers is especially challenging in such a region because the reference condition, or the benchmark of biological condition expected when human disturbance in the environment is absent or minimal, varies greatly among natural stream types. Previous indices used by monitoring programs were developed on a regional basis to help partition the state's environmental diversity, but statewide assessments were confounded by different criteria used in different regions. The CSCI, which translates complex data about individual benthic macroinvertebrates (BMIs) found living in a stream into an overall measure of stream health, was developed specifically to address some of the shortcomings of earlier indices. First, the CSCI was developed with a much larger, more representative data set that makes it applicable statewide and that covers the broad range of environmental variability among natural stream types. Second, the CSCI sets biological benchmarks for a site based on its site-specific environmental setting. Finally, the CSCI combines two separate types of indices, each of which provides unique information about the biological condition of a stream: a multi-metric index (MMI) that measures ecological structure and function, and an observed-to-expected (O/E) index that measures taxonomic completeness. Together they provide multiple lines of evidence about the condition of a stream, providing greater confidence in results than a single index.

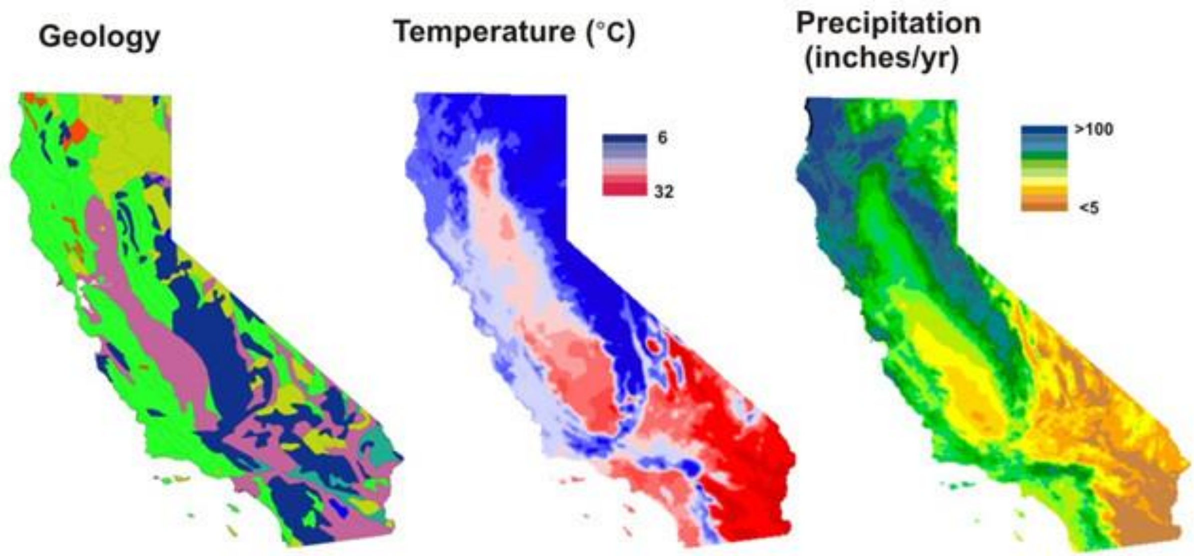

Figure 1. Extreme natural gradients in California result in a high degree of natural variation in biological expectations among stream types.

Introduction

California contains continental-scale environmental diversity within its borders, encompassing some of the most extreme gradients in elevation, climate and geology found in the United States (Figure 1). It supports temperate rainforests in the North Coast, alpine forests and meadows in the mountains, deserts in the east, and chaparral, oak woodlands, and grasslands with a Mediterranean climate in most remaining parts of the state. Such great physiographic complexity correspondingly supports a broad array of natural stream types, which in turn hosts a rich diversity of aquatic organisms. Bioassessment, which is the science of using aquatic organisms as indicators of stream health and function, is greatly complicated in such regions because the reference condition varies greatly among natural stream types (Figure 2).

Previous indices used by stream monitoring programs to rank or "score" biological condition at sampling sites relative to reference conditions were developed for specific subregions of California as a means of partitioning the state's environmental variability (e.g., Ode et al. 2005, Rehn 2009). While this approach allowed the establishment of defensible impairment thresholds within regions, comparison among regions was confounded for two closely related reasons: 1) the criteria used to define reference conditions varied among regions, and 2) each index was composed of different metrics so that deviation from the reference benchmark was not equivalently measured in all settings and did not have the same ecological meaning across the entire state.

Moreover, some portions of the state and certain stream types were unrepresented. To support the ongoing development of California's statewide Biological Integrity Plan, the State Water Board funded the development of a new index that was applicable statewide, encompassed as much natural environmental variability as possible, and allowed consistent and equivalent scoring thresholds in all regions of the state.

Figure 2. Bioassessment is complicated in regions with natural environmental complexity because the reference condition varies greatly among natural stream types.

The California Stream Condition Index (CSCI) is a new statewide biological scoring tool that translates complex data about benthic macroinvertebrates (BMIs) found living in a stream into an overall measure of stream health. Finalized in 2013 and recently accepted for publication (Mazor et al. in press), the CSCI represents the latest generation of biological indicators for assessing stream health in California. The CSCI combines two separate types of indices, each of which provides unique information about the biological condition at a stream: a multi-metric index (MMI) that measures ecological structure and function, and an observed-to-expected (O/E) index that measures taxonomic completeness. Unlike previous MMI or O/E indices that were applicable only on a regional basis or under-represented large portions of the state, the CSCI was built with a statewide dataset that represents the broad range of environmental conditions across California. The CSCI provides consistency and accuracy in the interpretation of biological data collected by both statewide and regional monitoring programs and will be the basis of the new statewide Biological Integrity Plan. This memo summarizes the development and key features of the CSCI, including its performance characteristics, recommended scoring thresholds, and data requirements for its use. Full details of CSCI development can be found in Mazor et al. (in press).

Compilation of Data Sets

Benthic datasets for CSCI development were compiled from more than 20 federal, state, and regional monitoring programs that sampled streams sites between 1999 and 2011. Standardization of BMI data was necessary because the level of taxonomic effort used to identify organisms and the number of specimens identified per sample varied among programs. Somewhat different data standardization approaches were used for the MMI and the O/E, but to accommodate data reduction that occurs during standardization, 600-count BMI samples identified to "Level 2a" as defined by the Southwest Association of Freshwater Invertebrate Taxonomists (SAFIT, Richards and Rogers 2011) were required[1]. BMI samples with insufficient numbers of organisms or taxonomic resolution were excluded from analyses. A final data set from 1,985 sites met all requirements and was used for development and evaluation of both the O/E and MMI indices (Figure 3).

[1] SAFIT Level 2a identifies most taxa to species and Chironomidae to subfamily.

Quantifying Natural and Anthropogenic Gradients Across Sites

Environmental data were gathered from multiple sources to characterize natural and anthropogenic factors known to affect benthic communities such as climate, elevation, geology, land cover, road density, hydrologic alteration, and mining. GIS variables that characterized natural and relatively stable environmental factors (e.g., topography, geology, climate) were used as predictors for O/E and MMI models, whereas variables related to human activity (e.g., land use, road density, etc.) were used to classify sites as reference and to evaluate responsiveness of O/E and MMI indices to human activity gradients. Most variables related to human activity were calculated at three spatial scales: within the entire upstream drainage area (watershed), within the contributing area 5 km upstream of a site, and within the contributing area 1 km upstream of a site (Appendix 1). Quantifying human activity at multiple spatial scales allowed sites to be screened for both local and catchment-scale impacts. By contrast, variables used as predictors for O/E and MMI indices were calculated at either the site (i.e., "point") scale or the watershed scale, but not at the local (1k and 5k) scales (Appendix 2).

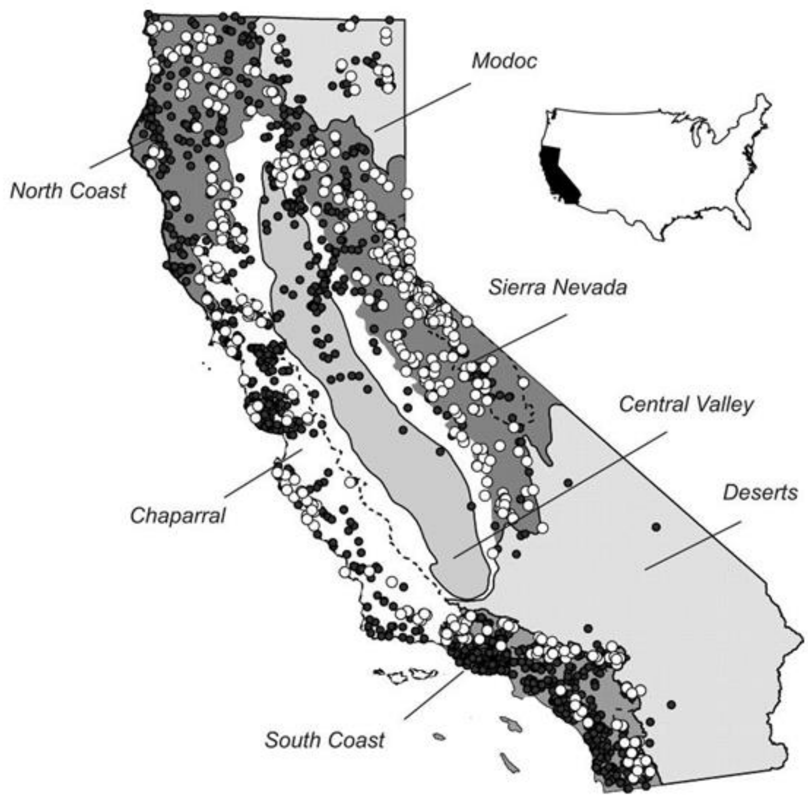

Figure 3. Distribution of 1,985 sampling sites used in development and validation of the CSCI. White circles are sites that passed reference screens (n = 590; see text) and black circles are sites that failed one or more screening criteria. Major ecological regions are those used as reporting units for the Perennial Streams Assessment (PSA).

Sites were divided into three sets for development and evaluation of indices: reference (i.e., low-activity), moderate-activity, and high-activity sites. Uniform statewide criteria for defining reference sites were recently established by Ode et al. (in press; also see Appendix 1) with an emphasis on achieving a balance between thorough environmental representativeness while still maintaining a pool of "minimally-disturbed" sites *sensu*

Stoddard et al. (2006). Nearly 600 of the 1,985 sites included in the data set for CSCI development passed reference screening criteria (Figure 3), a fairly high success rate due to an emphasis being placed on data sets likely to contain high-quality reference sites during data compilation. In addition to good geographic coverage, the final reference pool also represented several biologically important natural gradients (Figure 4). Identification of high-activity sites was necessary for MMI calibration (described below) and for performance evaluation of both MMI and O/E. High-activity sites were defined as meeting any of the following criteria: ≥50% developed land (i.e., % agricultural + % urban) at any spatial scale; ≥ 5 km/km² road density at any spatial scale; or riparian disturbance index (W1_HALL of Kaufmann et al. 1999) ≥ 5. Sites not identified as either reference or high-activity were designated as moderate-activity.

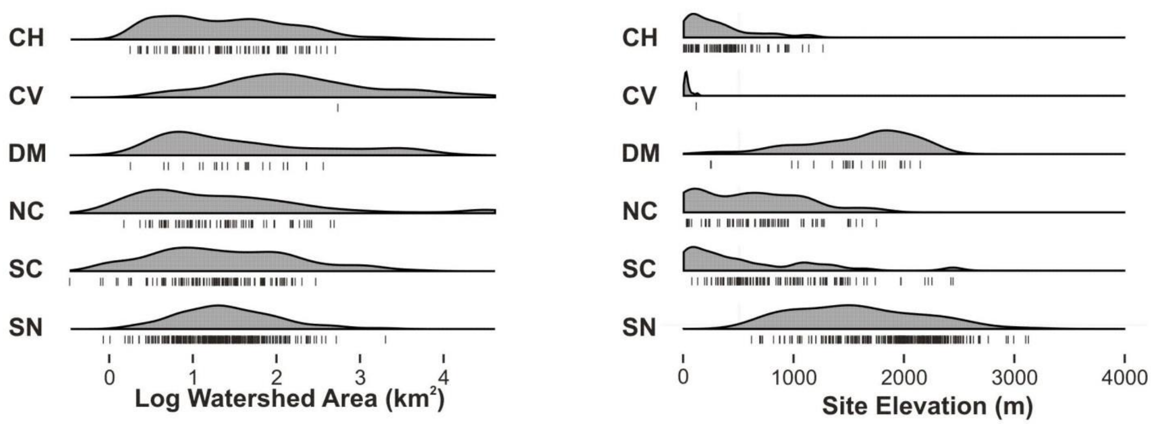

Figure 4. Examples of natural gradients that are important drivers of biological variability and are well-represented by the reference site pool. Unbiased estimates of natural gradient distributions in California were derived from probabilistic surveys conducted between 2000 and 2011 and are shown as kernel density estimates. Values of individual reference sites are shown as small vertical lines. Regions (see Figure 2) are abbreviated as follows: SN = Sierra Nevada, SC = South Coast, NC = North Coast, DM = Deserts + Modoc, CV = Central Valley, CH = Chaparral.

Building Predictive Models for O/E and MMI

The CSCI combines two different types of indices that have traditionally been used separately in stream assessments and provide unique information about the biological condition of a stream; an observed-to-expected (O/E) index that measures taxonomic completeness, and a multi-metric index (MMI) that measures ecological structure and function. Predictive modeling has been used in the development of O/E indices since their inception (Moss et al. 1987), but its use in the development of MMIs is relatively new (e.g., Pont et al. 2009). In each case, modeling improves index performance, but the process through which modeling helps achieve better performance differs somewhat between the approaches.

O/E indices assess the taxonomic completeness of a site by comparing observed and expected taxa lists. The taxa expected at a new assessment site, or a "test" site, are predicted by statistical modeling of relationships between taxonomic composition and natural environmental gradients at reference sites. Biological condition at a test site is then measured as the number of expected taxa (E) that are actually observed (O), and degradation of biological condition is quantified as loss of expected native taxa. Modeling relationships between taxonomic composition and natural environmental gradients produces indices that are more precise compared to null models where all taxa are assumed to have an equal probability of occurrence at all sites. In addition, the statistical modeling process in development of an O/E index produces site-specific expectations for each assessment site.

A multi-metric index aggregates several measures of BMI attributes, or metrics (e.g., percent predators, number of pollution tolerant species, etc.), into a single measure of biological condition. Metrics include measures of assemblage richness, composition, and diversity, and are chosen based on their responsiveness to human disturbance gradients and/or their ability to discriminate between reference and degraded condition. The challenge is that expected metric values at reference sites vary greatly among natural stream types, and natural gradients often co-vary with human disturbance gradients, thereby confounding metric response to disturbance. Previous MMIs developed for use in various subregions of California utilized regionalization approaches to control for the effects of natural variation in biological expectations, where "one size fits all" expectations were developed within large, mostly geographically defined areas (e.g., chaparral vs. mountains). Regionalization approaches have often been shown to poorly account for natural variation among sites (Hawkins et al. 2010). Therefore, models were developed to predict expected metric values at reference sites based on multiple natural environmental gradients (Appendix 2). Metric residuals (the difference between observed and expected values) were used as new metric values instead of raw metrics because they measure the range of metric variation after removing the effect of natural environmental variables. The models developed for reference sites were then used to predict expected metric values and calculate residuals at moderate- and high-activity sites. The advantages of this approach are twofold: 1) the expected metric values for any given assessment site are site-specific; and 2) metric residuals provide a more accurate evaluation of metric response to human disturbance gradients because they model out the effects of variation across natural environmental gradients. Use of modeled metrics produced an MMI with much better performance characteristics than un-modeled (null) metrics.

O/E indices do not require scoring because, as a simple ratio of observed-to-expected taxa, they are already scaled so that the mean score at reference sites is 1. Scoring is required for MMIs because individual metrics have different scales and different responses to stress, i.e., as human activity increases, some metrics decrease while others increase (Blocksom 2003). Scoring transforms metrics to a standard scale ranging from 0 (i.e., most stressed) to 1 (i.e., similar to reference sites). After scoring, final metrics[2] were chosen based on their ability to discriminate between reference and high-activity sites and by lack of bias among PSA regions (Figure 3). Scores for the final MMI at each site were then calculated by averaging the scores of the final selected metrics and rescaling (dividing) by the mean of reference calibration sites. Rescaling of MMI scores ensures that MMI and O/E are expressed in similar scales (i.e., as a ratio of observed to reference expectations), improving comparability of the two indices. A combined index (the California Stream Condition Index, CSCI) was calculated by averaging final MMI and O/E scores.

Setting Scoring Thresholds for the CSCI

The CSCI was calibrated during its development so that the mean score of reference sites is 1. Scores that approach 0 indicate great departure from reference condition and degradation of biological condition.

[2] Six metrics representing different aspects of assemblage composition (richness, trophic structure, tolerance, etc.) were chosen for inclusion in the final MMI: Taxonomic Richness, Shredder Taxa Richness, Percent Clinger Taxa, Percent Coleoptera Taxa, Percent EPT Taxa, and Percent Intolerant Individuals.

Scores > 1 can be interpreted to indicate greater taxonomic richness and more complex ecological function than predicted for a site given its natural environmental setting. In practice, CSCI scores observed from nearly 2000 study reaches sampled across California range from about 0.1 to 1.4. For the purposes of making statewide assessments, three thresholds were established based on the 30th; 10th; and 1st percentiles of CSCI scores at reference sites[3]. These three thresholds divide the CSCI scoring range into 4 categories of biological condition as follows: ≥0.92 = likely intact condition; 0.91 to 0.80 = possibly altered condition; 0.79 to 0.63 = likely altered condition; ≤0.62 = very likely altered condition.

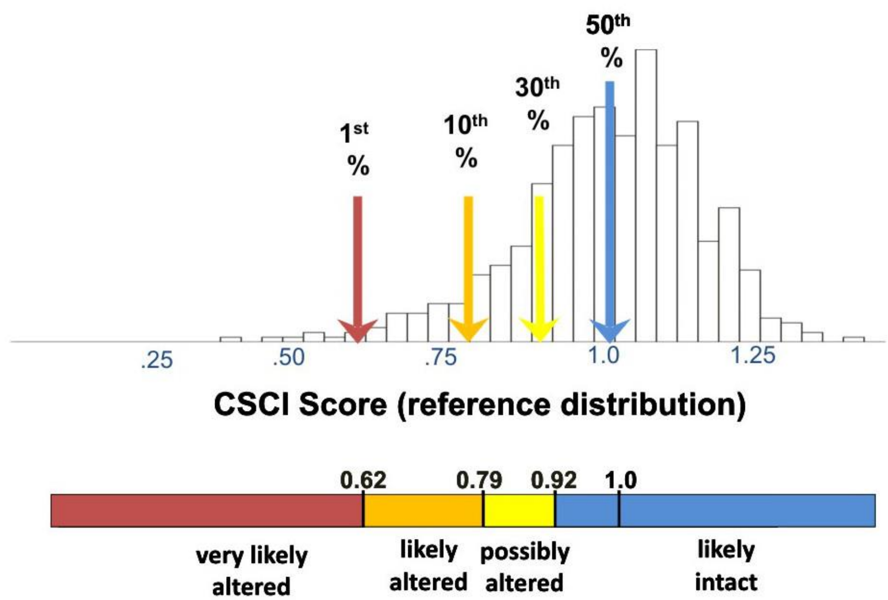

Figure 5. Distribution of CSCI scores at reference sites with thresholds and condition categories.

CSCI Performance

The CSCI had better performance than its null (un-modeled) counterpart in terms of accuracy and bias, precision, responsiveness, and sensitivity (see Appendix 3 for definitions of performance criteria). For example, mean regional differences in null CSCI scores at reference sites were large and significant, but were mostly absent in predictive CSCI scores (Figure 6a). The CSCI also was strongly responsive to human disturbance gradients, and the response was not confounded by the effects of natural gradients because those effects were modeled out by the use of metric residuals (Figure 6b).

[3] The rationale for these thresholds was to balance Type 1 errors (inferring degradation when it does not exist) and Type II errors (inferring a site is in reference condition when it is degraded). Similar thresholds have a precedent in bioassessment literature, but other methods for setting thresholds are possible, and if applied, might be equally valid.

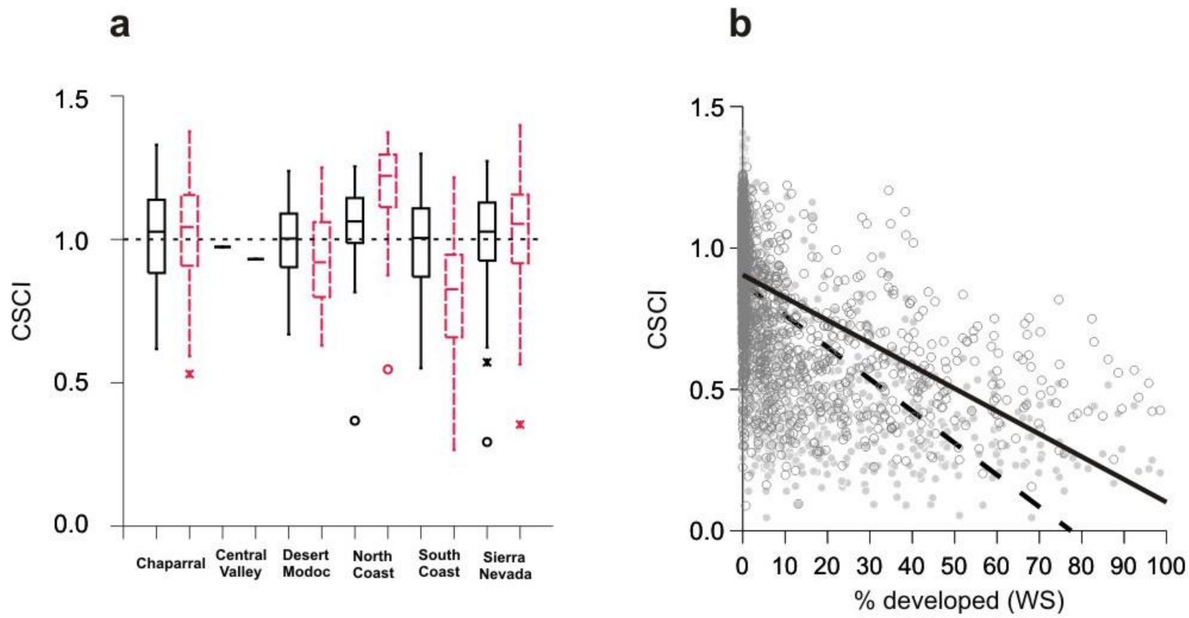

Figure 6. Examples of CSCI performance: a) Distribution of scores for predictive (black boxes) and null models (red dashed boxes) for the CSCI by geographic region. The horizontal dashed line indicates the expected value at reference sites (i.e., 1). Boxes represent the median, first, and third quartiles. Whiskers represent 1.5 times the interquartile range. Circles and X's represent outliers.
b) Relationship between predictive CSCI score (open circles and solid line) and null CSCI score (gray symbols and dashed line) and percent development in the watershed (= % urban + % ag). Note that the null CSCI has a steeper slope than the predictive CSCI because un-modeled metrics partially respond to natural gradients. By contrast, the predictive CSCI provides a more accurate response to disturbance gradients because the effects of metric variation across natural gradients have been modeled out.

Calculating the CSCI

Two types of data are required to calculate the CSCI: biological data and environmental data. Biological data are generated from benthic macroinvertebrate samples collected in accordance with standard SWAMP protocols (Ode 2007) and identified to the required taxonomic level of effort. SWAMP currently recommends a 600-count sample identified by a qualified taxonomist to *at least* SAFIT level 2a (Richards and Rogers 2015), with most taxa identified to species and Chironomidae identified to subfamily. Environmental data (e.g., watershed area, geology, precipitation) are generated by a specialist following standard geographic information system (GIS) protocols. Interim instructions (Mazor et al. 2015) that describe all steps in calculating the CSCI can be found at the SWAMP Bioassessment Program website. The first section describes the process for using GIS to delineate catchment polygons and then calculate environmental predictors (see Appendix 2 for required predictors). The second section describes the process for using the environmental predictors in conjunction with taxonomic data to calculate CSCI scores using custom libraries and scripts in the R statistical programming language. SWAMP is currently developing online tools to generate CSCI scores from user-supplied biological data and site coordinates, requiring minimal technical expertise.

More information about the SWAMP Bioassessment Program can be found at: http://www.waterboards.ca.gov/water_issues/programs/swamp/bioassessment. Those wishing to arrange training in CSCI calculation should contact Calvin Yang: calvin.yang@waterboards.ca.gov

REFERENCES

Blocksom, K. A. 2003. A performance comparison of metric scoring methods for a multimetric index for Mid-Atlantic highland streams. Environmental Management 42: 954-965.

Hawkins, C. P., Y. Cao, and B. Roper. 2010b. Methods of predicting reference condition biota affects the performance and interpretation of ecological indices. Freshwater Biology. 55: 1066-1085.

Kaufmann, P. R., P. Levine, E. G. Robinson, C. Seeliger, and D. V. Peck. 1999. Surface waters: Quantifying physical habitat in wadeable streams. EPA/620/R-99/003. US EPA. Office of Research and Development. Washington, DC.

Mazor, R.D. 2015. Bioassessment of streams in southern California: A report on the first 5 years of the Stormwater Monitoring Coalition's regional stream survey. Southern California Coastal Water Research Project Technical Report 844.

R. D. Mazor, P. R. Ode, A. C. Rehn, M. Engeln, T. Boyle, E. Fintel, and S. Verbrugge. 2015. The California Stream Condition Index (CSCI): Guidance for calculating scores using GIS and R. SCCWRP Technical Report #883. SWAMP-SOP-2015-0004.

Mazor, R.D., A.C. Rehn, P.R. Ode, M. Engeln, K.C. Schiff, E.D. Stein, D. Gillett and C.P. Hawkins. In press. Bioassessment in complex environments: designing an index for consistent meaning in different settings. Freshwater Science.

Moss, D., T. Furse, J. F. Wright, and P. D. Armitage. 1987. The prediction of macro-invertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. Freshwater Biology 17: 41-52.

Ode, P.R. 2007 Standard operating procedures for collecting benthic macroinvertebrate samples and associated chemical and physical data for ambient bioassessments in California. California State Water Resources Control Board Surface Water Ambient Monitoring Program (SWAMP) Bioassessment SOP 001.

Ode, P. R., A. C. Rehn, and J. T. May. 2005. A quantitative tool for assessing the integrity of southern coastal California streams. Environmental Management 35: 493-504.

Ode, P.R., T.M. Kincaid, T. Fleming and A.C. Rehn. 2011. Ecological Condition Assessments of California's Perennial Wadeable Streams: Highlights from the Surface Water Ambient Monitoring Program's Perennial Streams Assessment (PSA) (2000-2007). A collaboration between the State Water Resources Control Board's Non-Point Source Pollution Control Program (NPS Program), Surface Water Ambient Monitoring Program (SWAMP), California Department of Fish and Game Aquatic Bioassessment Laboratory, and the U.S. Environmental Protection Agency.

Ode, P.R., A.C. Rehn, R.D. Mazor, K.C. Schiff, E.D. Stein, J.T. May, L.R. Brown, D.B. Herbst, D. Gillett, K. Lunde and C.P. Hawkins. In press. Evaluating the adequacy of a reference site pool for the ecological assessment of streams in environmentally complex regions. Freshwater Science.

Olson, J. R., and C. P. Hawkins. 2012. Predicting natural base-flow stream water chemistry in the western United States. Water Resources Research 48. W02504. doi: 10.1029/2011WR011088

Pont, D., R. M. Hughes, T. R. Whittier, and S. Schmutz. 2009. A predictive index of biotic integrity model for aquatic-vertebrate assemblages of Western U.S. streams. Transactions of the American Fisheries Society. 138: 292-305.

Rehn, A. C. 2009. Benthic macroinvertebrates as indicators of biological condition below hydropower dams on West Slope Sierra Nevada streams, California, USA. River Research and Applications 25: 208-228.

Richards, A. B., and D. C. Rogers. 2011. List of freshwater macroinvertebrate taxa from California and adjacent states including standard taxonomic effort levels. Southwest Association of Freshwater Invertebrate Taxonomists. Chico, CA. Available from www.safit.org.

Stoddard, J. L., D. V. Peck, S. G. Paulsen, J. Van Sickle, C. P. Hawkins, A. T. Herlihy, R. M. Hughes, P.R. Kaufmann, D. P. Larsen, G. Lomnicky, A. R. Olsen, S. A. Peterson, P. L. Ringold, and T. R. Whittier. 2005. An Ecological Assessment of Western Streams and Rivers. EPA 620/R-05/005, U.S. Environmental Protection Agency, Washington, DC.

SUGGESTED CITATION

Rehn, A.C., R.D. Mazor and P.R. Ode. 2015. The California Stream Condition Index (CSCI): A New Statewide Biological Scoring Tool for Assessing the Health of Freshwater Streams. Swamp Technical Memorandum SWAMP-TM-2015-0002.

APPENDIX 1. STRESSOR AND HUMAN ACTIVITY GRADIENTS USED TO IDENTIFY REFERENCE SITES AND EVALUATE INDEX PERFORMANCE.

Sites that did not exceed the listed thresholds were used as reference sites. WS: Watershed. 5 km: Watershed clipped to a 5-km buffer of the sample point. 1 km: Watershed clipped to a 1-km buffer of the sample point. Variables marked with an asterisk (*) indicate those used in the random forest evaluation of index responsiveness. W1_HALL: proximity-weighted human activity index (Kaufmann et al. 1999). Sources are as follows: 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: Predicted specific conductance (Olson and Hawkins 2012). G: Field-measured variables. Code 21 is a land use category that corresponds to managed vegetation, such as roadsides, lawns, cemeteries, and golf courses.

	Variable	Scale	Threshold	Unit	Data source
*	% Agriculture	1 km, 5 km, WS	<3	%	A
*	% Urban	1 km, 5 km, WS	<3	%	A
*	% Ag + % Urban	1 km, 5 km, WS	<5	%	A
*	% Code 21	1 km and 5 km	<7	%	A
*		WS	<10	%	A
*	Road density	1 km, 5 km, WS	<2	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
	Specific conductance	Site	99/1**	prediction interval	F
	W1_HALL	Reach	<1.5	NA	G
	% Sands and fines	Reach		%	G
	Slope	Reach		%	G

** The 99th and 1st percentiles of predictions were used to generate site-specific thresholds for specific conductance. Because the model was observed to under-predict at higher levels of specific conductance (data not shown), a threshold of 2000 µS/cm was used as an upper bound if the prediction interval included 1000 µS/cm.

APPENDIX 2. NATURAL GRADIENTS USED AS PREDICTORS FOR DEVELOPMENT OF O/E AND MMI INDICES.

Variable	Data Source
Site (i.e., "point")	
Latitude	
Longitude	
Elevation	A
Catchment Morphology	
Log watershed area	A
Elevation Range	A
Climate	
10-year (2000-2009) average precipitation at the sample point	B
10-year (2000-2009) average air temperature at the sample point	B
Mean June to September 1971-2000 monthly precipitation, averaged across the catchment	B
Geology	
Average bulk soil density	C
Average soil erodibility factor (k)	C
Log % phosphorus-bearing geology	C

Sources:

A. National Elevation Dataset (http://ned.usgs.gov/)

B. PRISM climate mapping system (http://www.prism.oregonstate.edu)

C: Generalized geology, mineralogy, and climate data derived for a conductivity prediction model (Olson and Hawkins 2012)

Predictors that were evaluated but not selected for any model include percent sedimentary geology, nitrogenous geology, soil hydraulic conductivity, soil permeability, sulfur-bearing geology, calcite-bearing geology, and magnesium oxide-bearing geology.

APPENDIX 3. SUMMARY OF PERFORMANCE EVALUATIONS FROM MAZOR ET AL. (IN PRESS)

Aspect	Description	Indication of good performance
Accuracy and Bias	Scores are minimally influenced by natural gradients	- Approximately 90% of validation reference sites have scores above the 10th percentile of calibration reference sites. - Landscape-scale natural gradients explain little variability in scores at reference sites, as indicated by a low pseudo-R^2 for a 500-tree random forest model. - No visual relationship evident in plots of scores at reference sites against field measurements of natural gradients.
Precision	Scores are similar when measured under similar settings	- Low standard deviation of scores among reference sites (one sample per site) - Low pooled standard deviation of scores among samples at reference sites with multiple sampling events.
Responsiveness	Scores change in response to human activity gradients	- Large t-statistic in comparison of mean scores at reference and high-activity sites. - Landscape-scale human activity gradients explain variability in scores, as indicated by a high pseudo-R^2 for a 500-tree random forest model.
Sensitivity	Scores indicate poor condition at high-activity sites	- High percentage of high-activity sites have scores below the 10th percentile of calibration reference sites.