An Index to Measure the Quality of Physical Habitat in California Wadeable Streams

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

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

SWAMP Technical Memorandum
SWAMP-TM-2018-0005
August 2018

SWAMP
Surface Water Ambient Monitoring Program

TABLE OF CONTENTS

SUMMARY

- Assessing the condition of physical habitat in streams is vital in supporting Clean Water Act goals to protect and restore the chemical, physical and biological integrity of the nation's waters.
- The condition of physical habitat is a fundamental driver of stream ecosystem health; high quality (natural) habitat is critical for maintaining beneficial uses. Physical habitat components such as streambed substrate, channel morphology, flow-microhabitat complexity, in-stream cover-type

complexity, and riparian vegetation cover contribute to the overall physical and biological integrity of a stream.

- Physical characteristics of a site vary due to both natural factors and human disturbance. Statistical models based on a large statewide reference data set can help distinguish natural variability from anthropogenic stress. These models work across the diverse stream types found in California.
- Based on these models, a multimetric index was developed that characterizes physical habitat condition for streams in California. Index scores near 1 indicate physical habitat conditions similar to reference, whereas lower scores indicate degradation.
- This index may be used in a variety of stream management applications, including assessing potential causes of poor biological condition, setting targets for restoration, and prioritizing sites for protection or intervention.

INTRODUCTION

Physical characteristics of stream and river channels and their proximate riparian zones, including stream size, substrate type and diversity, diversity of current velocities, canopy shading and channel gradient have long been recognized as being among the primary factors that influence the structure and composition of biological assemblages at a given site (Hynes 1970; Karr et al. 1986; Parsons et al. 2004). However, these habitat attributes can also be directly or indirectly altered by anthropogenic activities, with habitat degradation and loss related to human land use often found to co-occur with poor biological condition (USEPA 2018). Quantifying the condition of physical habitat is an important component in the management and recovery of stream biota (Honea et al. 2009). As such, field-based physical habitat assessments (either qualitative or quantitative) are typically conducted as part of stream quality monitoring surveys to help characterize the overall ecological condition of study sites, to support inference about causation when biological impairment is observed, and to determine whether changes in management or restoration efforts improve physical habitat over time. Physical habitat assessments typically exclude water chemistry and other water column attributes like water clarity, temperature and light intensity, but typically include measures of aquatic macrophytes, riparian vegetation, and woody debris because of their roles in modifying habitat structure, light input, and nutrient input, although the latter could be considered biological measures.

Quantitative physical habitat protocols (Figure 1) were introduced to stream bioassessment in California in 2000-2003 as part of U.S. EPA's Environmental Monitoring and Assessment Program (EMAP; Peck et al. 2006), and subsequently were adopted by state bioassessment programs through 2007 (Ode et al. 2011). Ode (2007) modified certain elements of EMAP protocols to reduce the time required by 2-person field crews from approximately 3 hours for physical habitat assessment alone (Hughes et al. 2010) down to 2 hours for physical habitat assessment ***and*** benthic macroinvertebrate sampling, allowing for more intensive sampling of algae assemblages and completion of 2 sites per day if within close enough driving or hiking distance (EMAP protocols usually took too long to allow completion of 2 sites in a day). However, many elements of the EMAP protocol were kept unchanged or were only slightly modified, making many endpoint variables calculated from 2007 onward identical or directly comparable to endpoints derived from earlier data sets.

Endpoint variables, or metrics, are typically reach-scale averages of various physical habitat measures derived from transect-based measurements and observations. As with the protocols themselves, many physical habitat metrics used by SWAMP were also adopted directly from EMAP, namely the manual "*Quantifying Physical Habitat in Wadeable Streams*" by Kaufmann et al. (1999). These metrics have been used mostly as independent explanatory variables to evaluate how biota respond to individual PHAB elements (e.g., Ode et al. 2005), although exceedance thresholds have been established for a few to include in relative risk calculations for statewide stream condition assessments (Rehn 2015). Despite the importance of physical conditions for instream biological assemblages, no attempt has been made in California to combine metrics into an index that reflects an overall measure of physical habitat condition at a site, and performance of most physical habitat metrics has never been evaluated, especially whether metrics discriminate between reference sites and high-activity (or "stressed") sites, or how strongly they vary across natural gradients, such that statistical modeling may be required to factor out natural variance before combining metrics into an index.

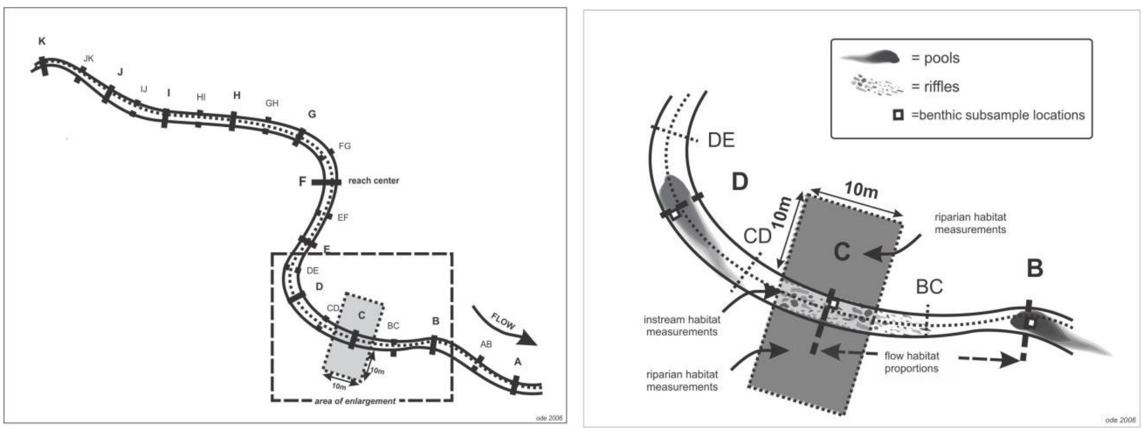

Figure 1. Physical habitat measurements are made at eleven equally-spaced spaced transects (A-K) and inter-transects from within a 150-meter sampling reach (or 250 meters for streams > 10 meters wide). Modified from Ode (2007).

Objectives

Evaluating the condition of individual attributes provides insight into specific differences between reference and non-reference streams, but overall assessments that incorporate multiple attributes into a single measure of the status of stream habitat are especially valuable (Rheinhardt et al. 2007) because the overall condition of streams may be more relevant for use in decision making by managers (Al-Chokhachy et al. 2010; San Diego Water Board 2016; K. Worcester pers. comm.). An easily interpretable index that integrates a number of individual habitat attributes into an overall measure of physical condition could address several management objectives and, together with biological indices, would provide much needed support for the State's 303(d) Listing Policy that recommends managers "*Evaluate physical habitat data and other water quality data, when available, to support conclusions about the status of [a] water segment*". The purpose of this report is to describe the development of a physical habitat index, analogous to the recently developed California Stream Condition Index (CSCI; Mazor et al. 2016) for benthic macroinvertebrates and the Algal Stream Condition Index (ASCI; Theroux et al. in prep) for benthic algae, that synthesizes individual metrics into an overall measure of physical habitat condition at a site. The same approach was used as in development of CSCI and ASCI: 1) expected metric values were modeled (predicted) at reference sites based on natural gradients; 2) for metrics with large amounts of variance explained by natural gradients, metric residuals (i.e., the observed value minus the predicted value) were used to factor out that response; 3) metrics were selected that were non-redundant statistically and that best responded to stressor gradients; 4) multiple types of metrics were included (e.g., substrate, in-channel cover, riparian complexity, etc.) to ensure different aspects of physical habitat were represented.

METHODS

Data Sets

Physical habitat data were compiled from >20 federal, state and regional bioassessment programs conducted in California from 2000-2016. Data were acquired from the SWAMP data warehouse and the Stormwater Monitoring Coalition (SMC), SWAMP's partner program in southern coastal California. Projects were included in index development if they had at least the "full" level of effort recommended by Ode (2007). In all, data were available from 1,797 unique sampling stations (Figure 2) and 2,205 unique sampling events (some sites had repeat visits over time). Unique stations were subset into reference, intermediate and high-activity sites using the same criteria as for CSCI development (Mazor et al. 2016). Stations were further subset (randomly) into 75% calibration sites (*n* = 1,355) and 25% validation sites (*n* = 442 sites) with assignments stratified by land use class and PSA region to ensure representation of all environmental settings in both data subsets. Calibration data were used for random forest statistical modeling of the extent to which metrics vary across natural gradients and to establish metric scoring criteria. Validation data were used to assess applicability of the final index to novel data, i.e., to test whether the index has similar performance properties for new, independent data as it did for calibration data. For sites with repeat visits, a single visit was randomly selected for calibration or validation; remaining visits were used to evaluate precision of the final index.

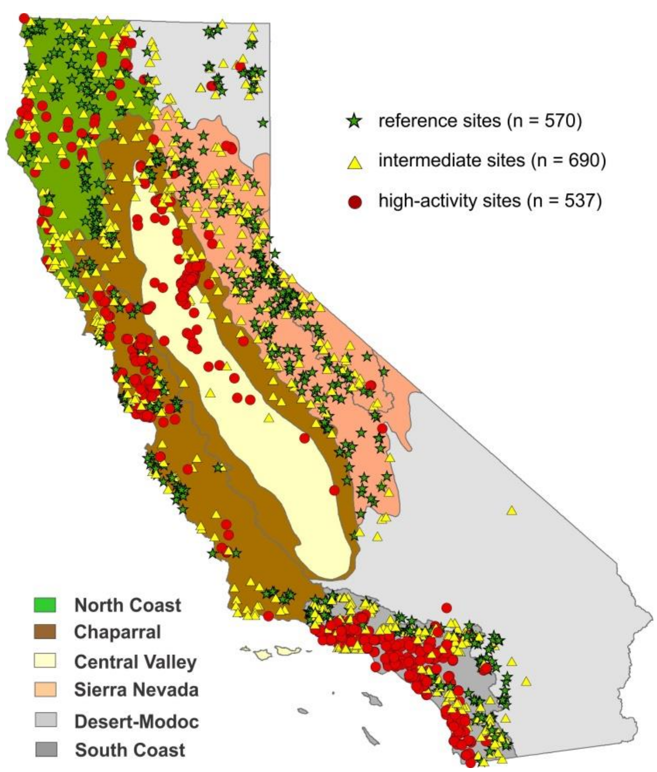

Figure 2. Geographic distribution of 1,797 sampling sites where data were collected for use in development and validation of the index of physical habitat condition. Criteria for defining reference, intermediate and high-activity site groups were the same as those used by Mazor et al. (2016) in CSCI development.

Metrics and Modeling

Fifty-seven candidate physical habitat metrics were chosen for statistical modeling and potential inclusion in the final index (Appendix 1) based on best professional judgement whether they quantify important physical attributes, whether each metric has a plausible influence on biological assemblages in natural (i.e., reference) settings and whether each has a plausible response to human disturbance gradients. Metrics based on components of the field protocol that quantify human disturbance at the scale of the sampling reach were not considered as candidates for the index , as the index was intended to measure response to disturbance, not directly incorporate disturbance measures. Metrics were calculated using the SWAMP reporting module, a custom Microsoft Access™ application that allows flexible management and reporting of several bioassessment data types. Candidate metrics were assigned to five thematic groups representing different physical attributes: substrate, riparian vegetation (including structure and shading), flow habitat variability, in-channel cover and channel morphology. Thematic groups were used to help ensure that metrics included in the final index were distributed among different habitat elements.

Like biological attributes, physical habitat characteristics vary naturally according to their physiographic and environmental setting. The challenge for characterizing these relationships is that 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. Developers of multi-metric indices for biological assemblages and physical habitat have increasingly begun to use statistical modeling to predict expected metric values at reference sites based on multiple natural environmental gradients (e.g., Pont et al. 2009, Vander Laan and Hawkins 2014, and Mazor et al. 2016 for biological assemblages, and Al-Chokhachy et al. 2010 for physical habitat). Metric residuals (i.e., the difference between observed and expected values) are then used as new metric values instead of raw metrics because they measure the range of metric variation after removing (or at least reducing) the effect of natural environmental variables. Models developed for reference sites are then used to predict expected metric values and calculate residuals at moderate- and high-activity sites. The advantages of this approach are twofold: 1) 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 factor out the effects of variation across natural environmental gradients included in the model.

Because the rationale and approach to development of a predictive multi-metric index has been well-documented for the CSCI (Mazor et al. 2016), only a short synopsis of the steps is presented here:

- 500-tree random forest models (using the *randomForest* package in R, Liaw and Wiener 2002) were used to predict values for all 57 metrics at reference calibration sites based on similar GIS-derived candidate variables that were used for CSCI development (Appendix 2).

- Recursive feature elimination (RFE) was used to select the simplest model (i.e., the model with the fewest predictors) whose root mean square error (RMSE) was no more than 2% greater than the RMSE of the optimal model (i.e., the one with the lowest RMSE). Only models with 10 or fewer predictors were considered. Limiting the complexity of models typically reduces overfitting and improves model validation (Strobl et al. 2007), but also reduces the number of GIS variables required for users to calculate. RFE was implemented with the *caret* package in R using the default settings for random forest models (Kuhn et al. 2012).

- For metrics with ≥ 20% of their variance explained by models, metric residuals (observed minus predicted) were used; otherwise, raw metric values were used for scoring.

- Metrics were scored following Cao et al. (2007). Metrics that decrease with human activity were scored as follows:

  (Observed value - Min)/(Max - Min)

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

  (Observed value - Max)/(Min - Max)

  where Min is the 5th percentile of reference calibration sites and Max is the 95th percentile of high-activity sites. Scores outside the range of 0 to 1 were then trimmed to 0 or 1.

Metric Performance and Selection

The performance of scored metrics was evaluated based on their discrimination and bias. For discrimination, t-tests of the differences in means between reference and high-activity site groups in the calibration data set were used, with higher t-values indicating better discrimination. No "hard" threshold was established for what constituted good discrimination as was done for the CSCI where only metrics with t >10 were considered. Rather, metrics with the highest and most statistically significant t-value within a given thematic group were considered. For bias, ANOVA was used to check for differences in mean score at calibration reference sites among PSA regions; an F-statistic >2 and with statistical significance of $p < 0.05$ was considered to represent high bias, or unequal means among regions. Also, metrics with low range (i.e., ≥ half of observations at reference calibration sites = 0) were not considered for inclusion in the index.

The best performing metric (or two in the case of substrate metrics, see below) was then chosen from within each thematic group and it was confirmed that the final chosen metrics were not strongly correlated (i.e., Pearson correlations among final metrics was <0.7). Selection of the final subset of metrics included in the index was more subjective than in previous development of biological indices used in California (e.g., Ode et al. 2005, Mazor et al. 2016), but several metrics within each thematic group that showed good discrimination also showed strong bias even after modeling, thus reducing the number of metrics to choose from; for those that performed well by both measures, emphasis was placed on more integrative and complex metrics rather than on more simple metrics (e.g., riparian cover sum of three layers was chosen over mean woody shrubs ground cover). The final metrics were assembled into an index, hereafter referred to as the "IPI" (Index of Physical Habitat Integrity), by dividing the mean metric score at each site by the mean score at reference calibration sites.

RESULTS

Selected Metrics

Five metrics were selected for inclusion in the IPI representing four of the five thematic groups (the channel morphology group was not represented; see Appendix 3 for descriptions of how each metric characterizes elements of physical habitat integrity):

Ev_FlowHab = evenness of flow habitat types (an unmodeled decreaser)
H_SubNat = Shannon diversity of natural substrate types (an unmodeled decreaser)
PCT_SAFN = percent sand and fine substrate (a modeled increaser)
XCMG = riparian cover, sum of 3 layers (canopy, mid-layer and ground; a modeled decreaser)
H_AqHab = Shannon diversity of natural in-channel cover types (a modeled decreaser)

The strongest Pearson correlation among the final five metrics was 0.47, well below the *a prior* threshold of <0.7 established above. In an early version of the index, mean wetted width/depth ratio was included as a modeled, bidirectional metric (i.e., negative residuals were treated as a decreasing metric and positive residuals were treated as an increasing metric) to represent the channel morphology thematic group, but after an evaluation of reference and high-activity sites that scored either very poorly or very well for that metric, it often didn't make intuitive sense why they scored the way they did, and it was decided not to include the metric in the final index. The main issues were that: 1) the mean wetted width/depth ratio on a given day is highly dependent on flow stage (which varies seasonally, and even diurnally in well-vegetated sreams with pronounced evapotransporation, or snow-melt driven streams that experience hot days and cool nights), making the metric relatively imprecise (Kaufmann et al. 1999); and 2) associations between geomorphic variables and biotic composition are highly variable and scale-dependent, and further investigations are required to concisely establish baseline ecological-geomorphological relationships, especially if those relationships are to be applied in management decision making or setting targets for stream restoration.

Also, it is well-known that percent sand and fines can show a bidirectional response to different types of human disturbance (e.g., channel armoring downstream of dams or fining of sediment in response to erosion), but that bi-directionality was not apparent in raw (i.e., unmodeled and unscored) data: high-activity sites had significantly more sand and fines compared to reference sites, and when PCT_SAFN *was* scored as a modeled, bidirectional metric, it showed high bias among ecoregions ($F = 12$!). Therefore, PCT_SAFN was treated as a unidirectional, increasing metric. By scoring PCT_SAFN as an increaser, the index could potentially overestimate the condition of streams where sands and fines have been depleted. Within the development data set, this situation most commonly occurred in concrete channels maintained for flood control. Scoring PCT_SAFN as an increaser meant that having less sand and fines than predicted is "better", such that concrete-bottomed channels sometimes scored better than soft-bottomed channels (Figure 3). This result is contrary to the intent of measuring deviation from the natural (reference) state, so at the recommendation of SMC stakeholders who reviewed an early version of the index, a *post-hoc* "concrete correction" was made whereby percent concrete was added to PCT_SAFN prior to scoring, thus eliminating cases where completely concrete-lined channels scored better than soft-bottomed channels that still retain some small vestige of "naturalness". The rationale for this correction is that excess fine sediment and concrete both represent the lack of suitable substrate to support intact benthic assemblages.

Figure 3. Two highly urbanized sites in the south coast, on the left with a natural soft bottom and on the right with a concrete bottom. Originally, the site on the left had an index score of 0.44 and the site on the right a score of 0.62, neither of which is an especially good score, but it was counter-intuitive that a site with a natural bottom would have a lower score than a completely concrete channel. After the "concrete correction" described in the text, the score at the site on the right was reduced to 0.35.

Index Performance

The IPI showed good discrimination between reference and high-activity sites in both calibration and validation data sets and did not show bias among PSA regions (Figure 4). The IPI also showed highly significant responses to continuous stressor gradients measured at both local and watershed scales (Figure 5). By contrast, IPI scores at reference sites showed no relationship with select natural gradients that co-vary with human disturbance gradients (Figure 6), indicating that the index is not influenced by natural environmental setting. Finally, IPI precision was evaluated based on 281 sites with at least 2, but with as many as 6, repeat visits over time (total $n = 689$), with the mean standard deviation of IPI scores among site revisits = 0.08, which is similar to that reported for the CSCI (Mazor et al. 2016).

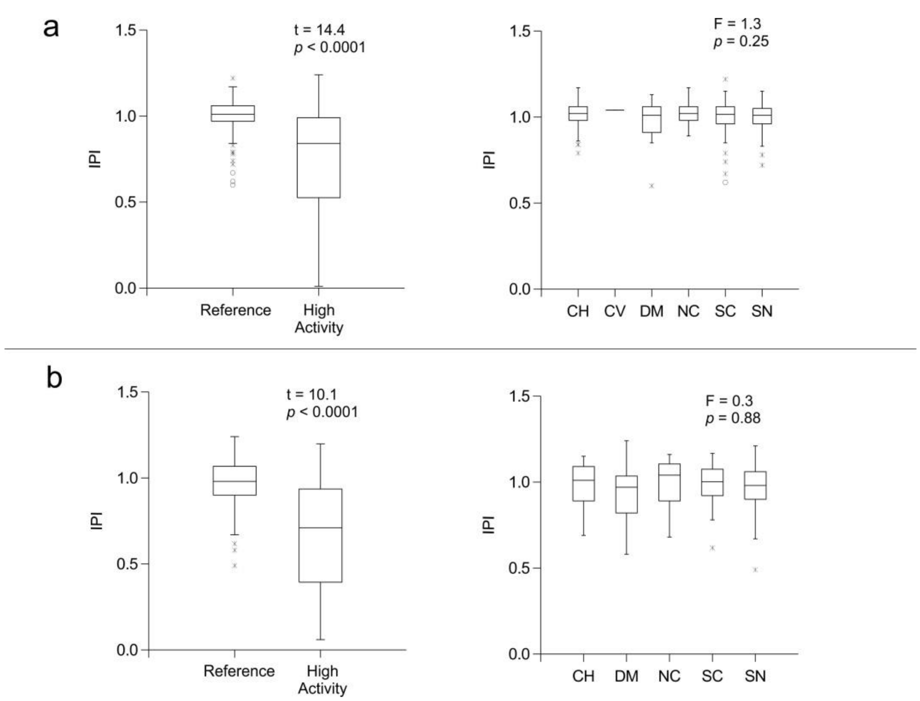

Figure 4. Evaluations of IPI discrimination (left plots) and bias (right plots) in a) calibration data, and b) validation data. CH= chaparral; CV = Central Valley; DM = desert-Modoc; NC= north coast; SC = south coast; SN = Sierra Nevada.

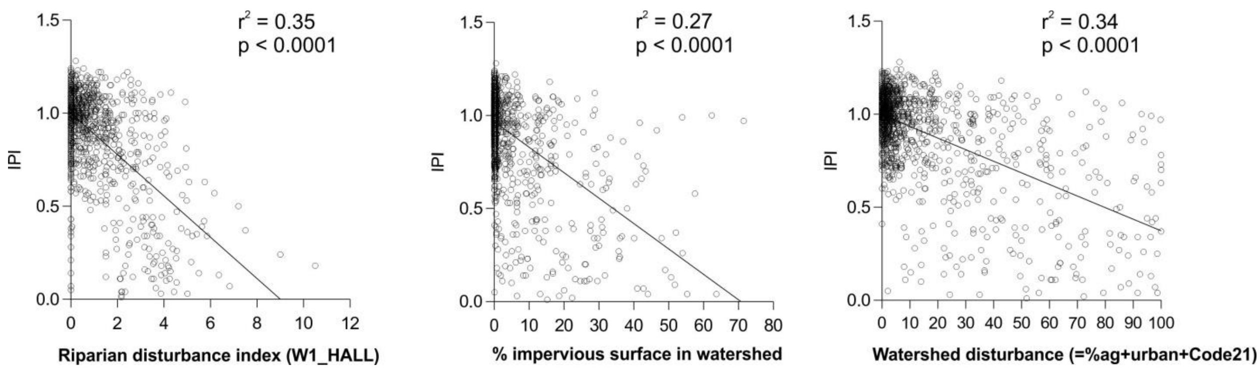

Figure 5. Scatterplots of IPI response to continuous stressor gradients at local scale (riparian disturbance index) and watershed scale (% impervious and %ag + %urban + Code 21). Code 21 is an NLCD land use category that includes developed open space like golf courses, city parks and road side vegetation.

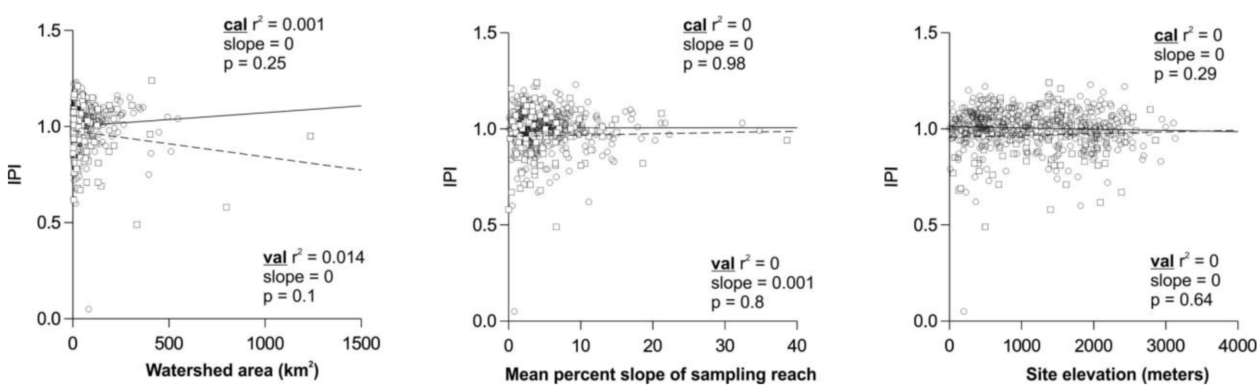

Figure 6. Scatterplots of relationships between IPI and select natural gradients; zero slopes indicate index scores are not dependent on natural environmental setting. Cal and val indicate least-squares regression statistics for calibration (open circles and solid regression lines) and validation (open squares and dotted regression lines) data subsets, respectively.

Relationships Between IPI and Other Indices

Although the IPI is intended as a stand-alone, independent line of evidence in stream condition assessments, the importance of high-quality physical habitat to healthy biological assemblages and overall ecological integrity, and the potential of the IPI to explain biological degradation, invites an evaluation of relationships between IPI and other commonly used indices. The IPI was significantly and positively correlated with other biological and physical indices (Figure 7), being more strongly correlated with CSCI than with the H20 hybrid algae index of Fetscher et al. (2014) based on diatoms and soft algae, a finding consistent with other analyses (e.g., Rehn 2016) that suggest benthic macroinvertebrates respond more strongly to physical habitat while algae responds more strongly to water chemistry. Least-squares linear regressions were used as a simple way to characterize the direction and relative strength of relationships, but relationships between biological indices and IPI are non-linear, showing more of a threshold response at IPI values around 0.6. It is perhaps unsurprising that IPI was most strongly (and linearly) correlated with CRAM, another index based partially on physical measures of habitat that are largely complementary to physical habitat protocols used in bioassessment.

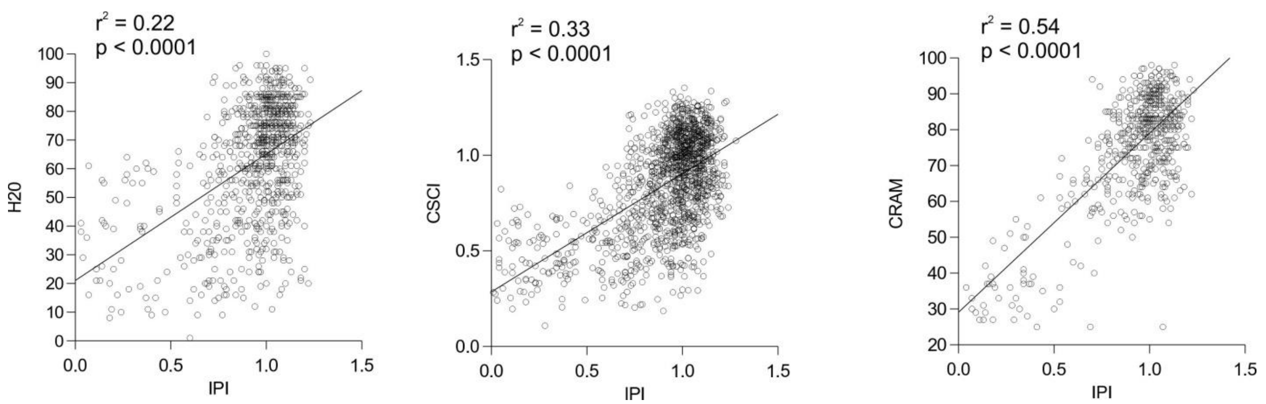

Figure 7. Least-squares linear regressions of H20 (based on diatoms and soft algae), CSCI (based on benthic macroinvertebrates), and CRAM (partially based on physical habitat measures complementary to bioassessment physical habitat protocols) on IPI.

Provisional Thresholds for the IPI

The IPI 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 physical condition. Scores > 1 can be interpreted to indicate greater physical complexity than predicted for a site given its natural environmental setting. In practice, IPI scores observed from nearly 1,800 study reaches sampled across California range from nearly zero at highly-urbanized, concrete-lined box channels to almost 1.3 at sites that have greater physical complexity than expected. More important than the overall scoring range of the index, however, is that sites in different ecoregions can have good physical conditions (and similar index scores) but look very different from one another, i.e., not all sites with good physical conditions are expected to look the same (Figure 8). For the purposes of making statewide assessments, three thresholds analogous to those used for the CSCI were established based on the 30th; 10th; and 1st percentiles of IPI scores at reference sites. These three thresholds divide the IPI scoring range into 4 categories of physical condition as follows: ≥0.94 = likely intact condition; 0.93 to 0.84 = possibly altered condition; 0.83 to 0.71 = likely altered condition; ≤0.70 = very likely altered condition.

Novato Creek (site 206PS0073 sampled 2010); IPI = 0.28 | Napa River (site 206PS0165 sampled 2012); IPI = 1.12

Lone Tree Creek (site 531PS0429 sampled 2011); IPI = 0.14 | Tuolumne River (site 535PS0265 sampled 2009); IPI = 1.15

Figure 8. Examples of sites with highly altered physical conditions (left photos) and better-than-expected physical conditions (right photos) in the San Francisco Bay Area (top photos) and the Central Valley (bottom photos), illustrating that the best physical conditions do not necessarily look identical in different ecoregions. Arrows represent a continuum of condition within regions.

DISCUSSION AND CONCLUSIONS

This study represents the first standardized, comprehensive assessment of the performance of quantitative physical habitat metrics calculated by stream bioassessment programs in California. Understanding whether (and which) metrics respond to human disturbance and/or need to be modeled to account for variation across natural gradients is an important step in determining how best to use and interpret them in statewide and regional stream condition assessments. Also, the fact that the best performing metrics were successfully combined into an index is an important step toward fully integrated assessments of biological, chemical and physical integrity in support of Clean Water Act goals.

The index is mostly intended as a support tool for evaluating biological data, rather than as a surrogate measure of aquatic life use support. For example, the San Diego Regional Water Quality Control Board could use the IPI as a secondary line of evidence in its proposed biological objectives, where the primary evidence for listing status is derived from CSCI scores. The index could also support causal assessments, e.g., a poor IPI score in association with degraded biology could be used as a coarse screen that indicates investigators should look more closely at individual physical habitat metrics to help diagnose causation, and may help to distinguish between water quality and habitat impacts to biological condition. Similarly, identifying differences in physical integrity between impacted and unimpacted sites could provide supporting lines of evidence in enforcement cases dealing with streambed or riparian alteration, or in less-developed regions where timber harvest and grazing are more prevalent land uses than agriculture or urbanization. Several regions and watershed groups have developed "report cards" that either integrate multiple index scores into a single score or display multiple index scores side-by-side. Having a physical habitat index that concisely pulls together multiple measures of habitat quality into a single measure allows physical condition to be expressed alongside other indices, providing a more complete assessment of overall ecological condition at a site.

In addition to the discussions of mean wetted width/depth ratio and percent sand and fines presented above, observations concerning a few other individual metrics are worth noting. For example, the metric mean aerial cover of natural shelter types (or "XFC_NAT") has been used in both national assessments (USEPA 2018) and statewide assessments (Rehn 2015) because of its integrative characterization of in-stream cover, but was rejected in this study because of bias among ecoregions, which modeling was unable to remove. The metric does have value in terms of showing strong discrimination between reference and high-activity sites, but users should be aware of its regional variability and the consequent need for regionally-specific thresholds. The same can be said of any metric with good discrimination but high bias (e.g., mean cobble embededdness, median substrate particle size, etc.), and the goal of this memo is not to recommend discontinued use of such metrics, as they may still be valuable in specific applications. Percent pool habitat has been shown by other authors to decrease in response to human disturbance (Al-Chokhachy et al. 2010), but in our data was a slight increaser that showed poor enough discrimination that scoring was deemed not worthwhile. Finally, mean barren ground cover performed much better as a bi-directional metric than as an increaser, perhaps because in more natural settings human disturbance tends to increase barren cover relative to reference conditions, but in more developed settings disturbance tends to increase cover via invasives like Himalayan blackberry, arundo, tamarisk, etc.

Next Steps

Finally, although the index shows good performance attributes and is based on a large development data set, the use of an integrated physical habitat index is fairly new to stream bioassessment in California relative to indices based on benthic macroinvertebrates or algae, which have received a fair amount of acceptance over the course of several years. Additional validation of the IPI, including ground-truthing or closer investigation of sites with unusual scores (e.g., high- activity sites with good scores or reference sites with poor scores) and application of the index in case studies like stream restorations or causal assessments where biological degradation has been observed, will be important steps towards its full adoption in statewide and regional bioassessment programs. To that end, tools to assist the bioassessment community (in and out of SWAMP) in calculation of the index are being developed for broader distribution.

REFERENCES

Al-Chokhachy, R., B.B. Roper and E.K. Archer. 2010. Evaluating the status and trends of physical stream habitat in headwater streams within the interior Columbia River and Upper Missouri River basins using an index approach. Transactions of the American Fisheries Society 139: 1041-1059.

Cao, Y., C.P. Hawkins, J. Olson and M.A. Kosterman. 2007. Modeling natural environmental gradients improves the accuracy and precision of diatom-based indicators. Journal of the North American Benthological Society 26: 566-585.

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

Hughes, R.M., A.T. Herlihy and P.R. Kaufmann. 2010. An evaluation of qualitative indexes of physical habitat applied to agricultural streams in ten U.S. states. Journal of the American Water Resources Association 46: 792-806.

Honea, J.M., J.C. Jorgensen, M.M. McClure, T.D. Cooney, K. Engie, D.M. Holzer and R. Hilborn. 2009. Evaluating habitat effects on population status: influence of habitat restoration on spring-run Chinook salmon. Freshwater Ecology 54: 1576-1592.

Hynes, H.B.N. 1970. *The Ecology of Running Waters*. University of Toronto Press, Toronto.

Karr, J.R., K.D. Fausch, P.L. Angermeier, P.R. Yant and I. J. Schlosser. 1986. Assessing biological integrity in running waters: a method and its rationale. Illinois Natural History Survey Special Publication 5.

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.

Kuhn, M., J. Wing, S. Weston, A. Williams, C. Keefer and A. Engelhardt. 2012. caret: Classification and Regression Training. R package version 5.15-045.

Liaw, A., and M. Wiener. 2002. Classification and Regression by random Forest. R News 2: 18-22.

Mazor, R.D., A.C. Rehn, P.R. Ode, M. Engeln, K.C. Schiff, E.D. Stein, D. Gillett, D.B. Herbst and C.P. Hawkins. 2016. Bioassessment in complex environments: designing an index for consistent meaning in different settings. Freshwater Science 35: 249-271.

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. 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., 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. Available from: https://www.waterboards.ca.gov/water_issues/programs/swamp/docs/reports/psa_smmry_rpt.pdf

Parsons, M., M.C. Thoms and R.H. Norris. 2004.Development of a standardized approach to river habitat assessment in Australia. Environmental Monitoring and Assessment 98: 109-130.

Peck, D.V., A.T. Herlihy, B.H. Hill, R.M. Hughes, P.R. Kaufmann, D.J. Klemm, J.M. Lazorchak, F.H. McCormick, S.A. Peterson, P.A. Ringold, T. Magee and M. Cappaert. 2006. Environmental Monitoring and Assessment Program—Surface Waters Western Pilot study: field operations manual for wadeable streams. EPA/620/R-06/003. Office of Research and Development. US Environmental Protection Agency. Corvallis, OR.

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. 2015. The Perennial Streams Assessment (PSA): An assessment of biological condition using the new California Stream Condition Index (CSCI). Available from: https://www.waterboards.ca.gov/water_issues/programs/swamp/bioassessment/docs/psa_memo_121015.pdf

Rehn, A.C. 2016. Using multiple biological and habitat condition indices for bioassessment of California stream. SWAMP Technical Memorandum: SWAMP-TM-SB-2016-0003. Available from: https://www.waterboards.ca.gov/water_issues/programs/swamp/bioassessment/docs/multiple_indices_tech_memo.pdf

Rheinhardt, R.D., M.M. Brinson, R.R. Christian, K.H. Miller and G.F. Meyer. 2007. A reference-based framework for evaluating the ecological condition of stream networks in small watersheds. Wetlands 27: 524-542.

San Diego Water Board. 2016. Clean Water Act Sections 305(b) and 303(d) Integrated Report for the San Diego Region. October 12, 2016. https://www.waterboards.ca.gov/water_issues/programs/tmdl/integrated2014_2016.shtml

Strobl, C., A.L. Boulesteix, A. Zeileis, and T. Hothorn. 2007. Bias in random forest variable importance measures: Illustrations, sources, and a solution. BMC Bioinformatics 8:25.

USEPA. 2018. DRAFT National Rivers and Streams Assessment 2012-2014: A Collaborative Survey. EPA XXX-X-XX-XXX. U.S. Environmental Protection Agency, Washington, DC. Accessed March 3, 2018. https://nationalriversandstreamsassessment.epa.gov/

Vander Laan, J.J., and C.P. Hawkins. 2014. Enhancing the performance and interpretation of freshwater biological indices: An application in arid zone streams. Ecological Indicators 36: 470-482.

SUGGESTED CITATION

Rehn, A.C., R.D. Mazor and P.R. Ode. 2018. An index to measure the quality of physical habitat in California wadeable streams. SWAMP Technical Memorandum SWAMP-TM-2018-0005.

APPENDIX I. FIFTY-SEVEN METRICS EVALUATED FOR INCLUSION IN THE IPI.

Min, med, max and sd = minimum, median, maximum and standard deviation of metric values observed in calibration data, respectively. rsq = pseudo r-squared from random forest models (a measure of how much variance models explain for each metric). For Response: D = decreaser; I = increaser; B = bidirectional. Metrics for which no response or associated performance statistics are listed are those that failed range test. t = t-statistic from 2-way t-tests, F = F-statistic from ANOVA. raw = values from raw metrics; scored = values from scored metrics. *** = p-value <0.0001; ** = p value < 0.01; * = p value < 0.05.

							Calibration			Validation		
Variable Class/Name	min	med	max	sd	rsq	Response	t raw	t scored	F scored	t raw	t scored	F scored
Channel morphology												
Mean bankfull width	1.1	7.4	56	5.9	0.45	B	na	4.1***	1.7	na	1.1	0.13
Mean wetted width/depth ratio	4.3	37.7	282.8	25.2	0.24	B	na	7.2***	0.54	na	4.5***	0.59
Mean cross-sectional area	0	0.4	11.2	1.1	0.51	B	na	9.2***	2.9*	na	3.1**	1.3
Coefficient of variation of depth	0.7	1.2	2.5	0.3	0.13	D	6.7***	7.2***	4.8**	3.1**	3**	2.5
Coefficient of variation of wetted width	0.1	0.4	1.3	0.2	0.12	D	3.9***	5.2***	3.8**	2.5*	3.7**	1.1
Flow habitat												
Evenness of flow habitat types	0	0.72	0.98	0.2	0.06	D	8.1***	8.1***	3.2	3.2**	3.2**	0.95
Shannon diversity (H) of flow habitat	0	0.99	1.7	0.4	0.13	D	10.8***	10.7***	5.2***	5.3***	5.3***	2.1
Percent cascade/falls	0	0	86	9.3	0.36							
Percent glide	0	33	97	26.4	0.24	I	10.7***	5***	2.9*	5.5***	1.3	2.1
Percent pool	0	7	68	13.1	-0.02	I?	2.2*			3**		
Percent rapid	0	0	70	9.8	0.12							
Percent riffle	0	33	100	26.4	0.13	D	9.5***	9.6***	3.2*	6.3***	6.4***	1.9
Percent run	0	0	63	11.5	0.24							
Percent fast water	0	52	100	28.5	0.30	D	11.1***	5.2***	0.4	6.6***	1.7	2.5
Percent slow water	0	48	100	28.0	0.30	I	12***	5.9***	0.37	7.5***	2.6**	2.4
Instream Cover												
Evenness of instream cover types	0.1	0.8	0.99	0.2	0.26	D	2.8**	5.4***	2.8*	0.07	2.2*	3.1*
Shannon diversity (H) of instream cover	0.07	1.4	1.99	0.4	0.22	D	5.5***	7.5***	0.68	0.36	2.9**	4.3**

Appendix 1 continued.

							Calibration			Validation		
Variable Class/Name	min	med	max	sd	rsq	Response	t raw	t scored	F scored	t raw	t scored	F scored
Instream Cover (continued)												
Mean filamentous algae cover	0	0.5	57.5	9.0	0.2	I	7.5***	3.8***	0.57	4.3***	1.9	1.3
Mean aquatic macrophytes cover	0	1.4	76.6	9.2	0.11	I	4.6***	4.7***	10.1***	3.1**	3.3**	4.4**
Mean woody debris <0.3m cover	0	5	87.5	7.7	0.04	D	2.5*	4.3***	1.1	0.81	0.52	1.6
Mean woody debris >0.3m cover	0	1.4	38.4	5.6	0.06	D	3.2**	5***	5.1**	2.6*	2.4*	2.3
Mean live tree roots cover	0	1.8	29.4	3.9	0.03	D	1.4	0.67	3.7**	2.5*	2.1*	1.02
Mean overhanging vegetation cover	0	5	60.5	9.7	0.09	D	1.4	0.14	2.5*	0.98	1.1	2.7*
Mean boulders cover	0	20.6	87.5	19.4	0.36	D	13.8***	12.3***	0.49	8.9***	5.8***	0.91
Mean undercut banks cover	0	0.9	54.1	4.4	0.06	D	0.02	1.1	1.02	0.96	0.48	2.1
Natural shelter cover - EMAP	4.6	41	148.3	26.6	0.16	D	8.7***	10.3***	4.6**	5.1***	5.4***	4**
Natural shelter cover - SWAMP	5.5	48.5	170.4	29.6	0.12	D	6.4***	7***	5.9***	2.4*	3.2**	4.2**
Riparian cover												
Mean upper canopy cover	0	28	97	19.5	0.37	D	4.1***	8.4***	0.06	0.1	0.5	0.35
Mean mid-layer canopy cover	0	33.5	83	13.9	0.11	D	5.8***	6.3***	1.1	1.8	1.8	1.6
Mean mid and upper canopy cover	0	62	148	26.1	0.26	D	5.8***	9.3***	0.48	0.9	2.3*	1.3
Mean woody shrubs ground cover	0	26	68	13.9	0.10	D	7***	7.6***	1.7	2.3*	2.4*	1.1
Riparian cover sum of 3 layers	14	118.5	231	35.0	0.26	D	5.7***	9.2***	0.39	0.4	2.1*	0.98
Riparian woody cover sum of 3 layers	0	89.5	201	32.6	0.17	D	7.3***	8.1***	3.6**	1.9	2.2*	0.93
Mean mid-channel canopy density	0	66	100	30.3	0.47	D	5***	4.9***	0.62	1.4	0.39	0.47
Mean barren soil/duff ground cover	0	38.5	88	22.1	0.25	I	0.86	2.1*	0.22	2.6*	0.35	0.76
Mean barren soil/duff ground cover	0	38.5	88	22.1	0.25	B	na	9***	0.58	na	1.2	0.89
Mid and upper canopy presence	0	0.91	1	0.3	0.21	D	7.9***	7.8***	0.11	1.2	1.7	0.43
Riparian vegetation all 3 layers present	0	0.91	1	0.3	0.22	D	8.2***	7.5***	0.13	1.2	1.5	0.31
Substrate												
Evenness of natural substrate types	0.2	0.8	0.98	0.1	-0.05	D	11.8***	11.9***	0.31	4.5***	4.6***	0.69

Appendix 1 continued.

							Calibration			Validation		
Variable Class/Name	min	med	max	sd	rsq	Response	t raw	t scored	F scored	t raw	t scored	F scored
Substrate (continued)												
Shannon diversity (H) of natural substrate	0.4	1.6	2.03	0.2	-0.03	D	14.4***	14.5***	1.8	7.2***	7.2***	0.52
Percent cobble	0	23	62	11.8	0.13	D	16.5***	17***	9.1***	8.7***	9.2***	9.3***
Percent CPOM presence	1	31	96	22.8	0.2	I	1.4	1.9	1.64	3**	2.2*	2.3
Percent fines	0	1	72	7.9	0.16	I	7.5***	6.9***	0.89	3.1**	3.8***	1.6
Percent coarse gravel	0	18	61	12.2	0.28	D	5.9***	12.8***	0.18	3.4**	8.1***	1.4
Percent fine gravel	0	8	50	7.8	0.11	D	1.3	1.6	2.63*	0.89	0.58	1.49
Percent rough bedrock	0	0	64	6.6	-0.05							
Percent smooth bedrock	0	0	78	10.2	0.06							
Percent bedrock (rough + smooth)	0	3	78	11.8	0.14	D	7.5***	9.1***	2.7*	3.6***	3.9***	0.58
Percent sand	0	10	92	13.1	0.29	I	8.2***	3.5***	0.37	5.6***	3.1**	1.9
Percent small boulders	0	11	46	10.1	0.31	D	16***	13.3***	0.66	8.6***	6.1***	0.3
Percent large boulders	0	3	58	9.8	0.10	D	9.1***	9.8***	1.2	6.8***	7.7***	3.3*
Percent Boulders (large + small)	0	16.5	67	15.0	0.42	D	16.5***	14.4***	0.47	9.8***	7.2***	1.8
Percent wood	0	1	17	2.3	-0.04	D	0.89	1.9	0.77	2.1*	1.6	3.9**
10th percentile of particle size (d10)	0.03	1.03	157	17.6	0.12	D	6.4***	7.8***	10.6***	2.7**	4.8***	3.9**
Median particle size (d50)	0.03	73	5660	787.2	0.07	D	9.8***	13.8***	9.9***	6.8***	8.8***	8.1***
90th percentile of particle size (d90)	1.03	625	5660	2252.2	0.24	B	na	3.6***	0.77	na	1.1	1.8
90th percentile of particle size (d90)	1.03	625	5660	2252.2	0.24	D	10.6***	7.4***	1.1	7.4***	3.9***	3*
Mean cobble embeddedness	0	28	92	14.3	0.18	I	6.8***	6.9***	12.1***	4.7***	4.9***	7.7***
Percent sand and fines (<2 mm)	0	12	96	16.8	0.40	B	na	13.4***	12***	na	6.7***	4.9**
Percent sand and fines (<2 mm)	0	12	96	16.8	0.40	I	11.3***	5.6***	0.5	6.6***	4.3***	1.5

Note: Mean barren ground cover and percent sand and fine sediment were scored and evaluated as both increaser and bi-directional metrics; d90 of sediment particle size was scored and evaluated as both a decreaser and bi-directional metric.

APPENDIX 2. NATURAL GRADIENTS (PREDICTOR VARIABLES) USED IN MODELING OF PHYSICAL HABITAT METRICS.

Predictors with an asterisk next to the abbreviation are required for IPI calculation. P = point variables; WS = variables calculated at the whole catchment scale.

	Abbreviation	Definition	Scale
Location			
	New_Lat*	Latitude	P
	New_Long*	Longitude	P
	SITE_ELEV*	Site elevation (m)	P
Climate			
	PPT_00_09*	10-y (2000-2009) average precipitation at the sample point	P
	TEMP_00_09	10-y (2000-2009) average air temperature at the sample point	P
	SumAve_P	Catchment mean of mean June-Sep 1971-2000 monthly ppt	WS
	MINP_WS*	Catchment mean of mean 1971-2000 min monthly ppt	WS
	MEANP_WS*	Catchment mean of mean 1971-2000 annual ppt	WS
	TMAX_WS	Catchment mean of mean 1971-2000 max temperature	WS
	XWD_WS	Catchment mean of mean 1961-1990 annual number of wet days	WS
	MAXWD_WS	Catchment mean of 1961-1990 annual max number of wet-days	WS
Catchment morphology			
	ELEV_RANGE*	Range in elevation from sample point to maximum elevation in catchment	WS
	MAX_ELEV*	Maximum elevation (m) in catchment	WS
	AREA_SQKM*	Catchment area in square kilometers	WS
Field measures	XSLOPE*	Mean slope of sampling reach	P
	SINU	Sinuosity of sampling reach	P
	XBEARING	Mean bearing of sampling reach	P
	XBKF_W*	Mean bankfull width (m) of sampling reach	P
Geology			
	BDH_AVE	Catchment mean bulk soil density	WS
	LPREM_mean	Catchment mean log geometric mean hydraulic conductivity	WS
	KFCT_AVE*	Catchment mean soil erodibility (K) factor	WS
	N_Mean	Catchment mean whole rock N	WS

Appendix 2 continued.

	Abbreviation	Definition	Scale
Geology continued			
	P_Mean	Catchment mean whole rock P	WS
	PRMH_AVE	Catchment mean soil permeability	WS
	MgO_Mean	Catchment mean whole rock MgO	WS
	S_Mean	Catchment mean whole rock S	WS
	CaO_Mean	Catchment mean whole rock CaO	WS
	UCS_Mean	Catchment mean unconfined Compressive Strength	WS
	PCT_CENOZ	Percent Cenozoic geology	WS
	PCT_NOSED	Percent non-sedimentary geology	WS
	PCT_QUART	Percent quaternary geology	WS
	PCT_SEDIM	Percent sedimentary geology	WS
	PCT_VOLCNC	Percent igneous geology	WS

APPENDIX 3. DESCRIPTIONS OF HOW IPI METRICS CHARACTERIZE ELEMENTS OF PHYSICAL HABITAT INTEGRITY.

Evenness of flow habitat types
Ev_FlowHab measures the evenness of riffles, pools, and other flow microhabitat types. Optimal physical conditions include a relatively even mix of velocity/depth regimes, with regular alternation between riffles (fast-shallow), runs (fast-deep), glides (slow-shallow) and pools (slow-deep). Poor conditions occur when a single microhabitat dominates (usually glides, with pools and riffles absent). A stream that has a uniform flow regime will typically support far fewer types of organisms than a stream that has a variety of alternating flow regimes. Riffles in particular are a source of high-quality habitat and diverse fauna, and their regular occurrence along the length of a stream greatly enhances the diversity of the stream community. Pools are essential for many fish and amphibians.

Shannon diversity of natural substrate types
H_SubNat measures the diversity of natural substrate types, assessing how well multiple size classes (e.g., gravel, cobble and boulder particles) are represented. In a stream with high habitat quality for benthic macroinvertebrates, layers of cobble and gravel provide diversity of niche space. Occasional patches of fine sediment, root mats and bedrock also provide important habitat for burrowers or clingers, but do not dominate the streambed. Lack of substrate diversity, e.g., where >75% of the channel bottom is dominated by one particle size or hard-pan, or with highly compacted particles with no interstitial space, represents poor physical conditions. Riffles and runs with a diversity of particle sizes often provide the most stable habitat in many small, high-gradient streams.

Percent sand and fine substrate
PCT_SAFN measures the amount of small-grained sediment particles (i.e., <2 mm) that have accumulated in the stream bottom as a result of deposition. Deposition may result from soil disturbance in the catchment, landslides, and bank erosion. Sediment deposition may cause the formation of islands or point bars, filling of runs and pools, and embeddedness of gravel, cobble, boulders and snags, with larger substrate particles covered or sunken into the silt, sand, or mud of the stream bottom. As habitat provided by cobbles or woody debris becomes embedded, and as interstitial spaces become inundated by sand or silt, the surface area available to macroinvertebrates and fish is decreased. High levels of sediment deposition are symptoms of an unstable and continually changing environment that becomes unsuitable for many organisms. Although human activity may sometimes deplete sands and fines (e.g., by upstream dam operations), and this depletion may harm aquatic life, the IPI treats only increases in this metric as a negative impact on habitat quality, although a *post-hoc* "concrete correction" was made whereby the metric percent concrete is added to PCT_SAFN before scoring (see main text above for explanation).

Riparian vegetation cover, sum of 3 layers
XCMG measures the amount of vegetative protection afforded to the stream bank and the near-stream portion of the riparian zone. The root systems of plants growing on stream banks help hold soil in place, thereby reducing the amount of erosion likely to occur. The riparian zone also serves as a buffer to pollutants entering a stream from runoff and provides shading and habitat and nutrient input into the stream. Banks that have full, multi-layered, natural plant growth are better for fish and macroinvertebrates than are banks without vegetative protection or those shored up with concrete or riprap. Vegetative removal and reduced riparian zones occur when roads, parking lots, fields, lawns, bare soil, riprap, or buildings are near the stream bank. Residential developments, urban centers, golf courses, and high grazing pressure from livestock are the common causes of anthropogenic degradation of the riparian zone. Even in undeveloped areas, upstream hydromodification and invasion by non-native species can reduce the cover and quality of riparian zone vegetation.

Appendix 3 continued.

Shannon diversity of natural instream cover

H_AqHab measures the relative quantity and variety of natural structures in the stream, such as cobble, large and small boulders, fallen trees, logs and branches, and undercut banks available as refugia or as sites for feeding or spawning and nursery functions of aquatic macrofauna. A wide variety and/or abundance of submerged structures in the stream provides macroinvertebrates and fish with a large number of niches, thus increasing habitat diversity. When variety and abundance of cover decreases (e.g., due to hydromodification, increased sedimentation, or active stream clearing), habitat structure becomes monotonous, diversity decreases, and the potential for recovery following disturbance decreases. Snags and submerged logs—especially old logs that have remained in-place for several years--are among the most productive habitat structure for macroinvertebrate colonization and fish refugia in low-gradient streams.