Prioritizing management goals for stream biological integrity within the developed landscape context

Marcus W. Beck ([marcusb@sccwrp.org](mailto:marcusb@sccwrp.org)), Raphael D. Mazor ([raphaelm@sccwrp.org](mailto:raphaelm@sccwrp.org)), Scott Johnson ([scott@aquaticbioassay.com](mailto:scott@aquaticbioassay.com)), Karin Wisenbaker ([karin@aquaticbioassay.com](mailto:karin@aquaticbioassay.com)), Joshua Westfall ([jwestfall@lacsd.org](mailto:jwestfall@lacsd.org)), Peter R. Ode ([peter.ode@wildlife.ca.gov](mailto:peter.ode@wildlife.ca.gov)), Ryan Hill ([hill.ryan@epa.gov](mailto:hill.ryan@epa.gov)), Chad Loflen ([Chad.Loflen@waterboards.ca.gov](mailto:Chad.Loflen@waterboards.ca.gov)), Martha Sutula ([marthas@sccwrp.org](mailto:marthas@sccwrp.org)), Eric D. Stein ([erics@sccwrp.org](mailto:erics@sccwrp.org))

Version Date: Sun Aug 11 22:39:49 2019 -0700

# Abstract

Stream management goals for biological integrity may be difficult to achieve in developed landscapes where channel modification and other factors constrain in-stream conditions. To evaluate potential constraints on biological integrity, we developed a statewide landscape model for California that estimates ranges of likely scores for a macroinvertebrate-based index that are typical at a site with the observed level of landscape alteration. This context can support prioritization decisions for stream management, like identifying reaches for restoration or enhanced protection based on how observed scores relate to the model expectations. Median scores were accurately predicted by the model for all sites in California with bioassessment data (Pearson correlation r = 0.75 between observed and predicted for calibration data, r = 0.72 for validation). The model also predicted that 15% of streams statewide are constrained for biological integrity within their present developed landscape, particularly for urban and agricultural areas in the South Coast, Central Valley, and Bay Area regions. We worked with a local stakeholder group from the San Gabriel River watershed (Los Angeles County, California) to evaluate how the statewide model could support local management decisions. To achieve this purpose, we created an interactive application, the Stream Classification and Priority Explorer (SCAPE), that compares observed scores with expectations from the landscape model to assign priorities. We observed model predictions consistent with the land use gradient from the upper to lower watershed, where potential limits to achieving biological integrity were more common in the heavily urbanized lower watershed. However, most of the sites in the lower watershed scored within their expected ranges, and were therefore given a low priority for restoration. In contrast, two low-scoring sites in the undeveloped upper watershed were prioritized for causal assessment and possible future restoration, whereas three high-scoring sites were prioritized for protection. The availability of geospatial and bioassessment data at the national level suggests that these tools can easily be applied to inform management decisions at other locations where altered landscapes may limit biological integrity.

Key words: Bioassessment, biotic integrity, streams, urbanization, modified channels, landscape stressors, random forests, prioritization, data visualization, stakeholder group

# Introduction

The widespread use of bioassessment data to assess ecological condition of aquatic environments is a significant advance over chemical or physical methods of assessment, yet managers and stakeholders require contextual information for synthesizing and interpreting biological information. The reference condition concept that is built into many biological indices provides a broad context for observed condition relative to unaltered habitats for a particular region (Reynoldson et al. [1997](#ref-Reynoldson97), Stoddard et al. [2006](#ref-Stoddard06)). However, achieving a reference condition of biological integrity (i.e., having structure and function comparable to natural habitat for the same region, Karr et al. ([1986](#ref-Karr86))) may be challenging if landscape conditions (e.g., watershed imperviousness) limit the spatial and temporal scales that can be effectively managed (Chessman and Royal [2004](#ref-Chessman04), Chessman [2014](#ref-Chessman14)). Resource management decisions could be improved if information is available that describes these limitations. A landscape context is required that describes how likely a site is to achieve biological integrity, which can inform how bioassessment data supports decisions or be used to identify priorities.

Prioritizing among sites that are affected by landscape alteration is a critical challenge for managers in urban and agricultural settings (Walsh et al. [2005](#ref-Walsh05), Beechie et al. [2007](#ref-Beechi07), Paul et al. [2008](#ref-Paul08)). In developed landscapes, the majority of stream miles are in poor biotic condition (USGS (US Geological Survey) [1999](#ref-USGS99), Finkenbine et al. [2000](#ref-Finkenbine00), Morgan and Cushman [2005](#ref-Morgan05)). Restoring streams in urban or agricultural settings can be costly, success is not universally defined, and achieving regional reference-like conditions may be difficult (Bernhardt et al. [2007](#ref-Bernhardt07), Kenney et al. [2012](#ref-Kenney12), Shoredits and Clayton [2013](#ref-Shoredits13)). Conventional approaches to protect and restore biological integrity have commonly focused on direct improvements at the site level to mitigate instream stressors (Carline and Walsh [2007](#ref-Carline07), Lester and Boulton [2008](#ref-Lester08), Roni and Beechi [2012](#ref-Roni12), Loflen et al. [2016](#ref-Loflen16)), in addition to upstream preventive measures that may be incentivized or enforced through regulation. Although these approaches can lead to improvements in ecological condition(e.g., Bernhardt et al. ([2007](#ref-Bernhardt07))), there is no universal remedy for achieving biological integrity in streams. In urban areas, protective thresholds for biological integrity have been debated (Cuffney et al. [2011](#ref-Cuffney11)). For biological integrity, several states have implemented a tiered aquatic life use or alternative use designation as potential approaches to account for baseline shifts in ecosystem condition from channel modification (e.g., Florida Department of Environmental Protection ([2011](#ref-FLDEP11)), US Environmental Protection Agency, Region 10 ([2013](#ref-USEPA13)), Midwest Biodiversity Institute ([2016](#ref-MBI16)), permitted under section 303(c)(2) of the Clean Water Act). Other approaches may include site-specific criteria or alternative thresholds with implementation clarifications (e.g., Board ([2016](#ref-SDWB16))).

Herein, we define constrained streams as those where present landscapes are likely to limit management options for restoring biological integrity. This definition describes a biological expectation and is distinct from the classical definition used in the general stream ecology literature (e.g., a physically constrained channel in the morphological sense). By describing an expected range of biological conditions due to factors that constrain biointegrity and may be difficult to manage, efforts to improve or protect condition could be prioritized at less constrained sites where alternative or more easily managed factors are influencing condition. For example, a monitoring site with an observed biological index score that is above an expected range of attainability could be assigned a higher management priority (i.e., for protection) relative to a site that is scoring within the expected range based on landscape development. A predictive model of bioassessment scores that is based on landscape metrics (e.g., imperviousness) could describe constraints on biological integrity, particularly for factors that are difficult to manage and are often associated with instream stressors. Analysis methods that characterize biotic and abiotic factors that limit assemblage composition have been explored by others (i.e., limiting factor theory, Chessman et al. ([2008](#ref-Chessman08)), Chessman ([2014](#ref-Chessman14))). Similar concepts have been applied in a landscape context to describe variation in biological communities and metrics at different spatial scales (Waite [2013](#ref-Waite13), Waite et al. [2014](#ref-Waite14)), although they have not been developed to describe constraints as defined above.

Consistent and empirical links between land use thresholds and poor biotic integrity have been identified in many cases (Allan et al. [1997](#ref-Allan97), Wang et al. [1997](#ref-Wang97), Clapcott et al. [2011](#ref-Clapcott11)) and previous modelling efforts have successfully used geospatial data to predict stream condition at regional or national scales (Vølstad et al. [2004](#ref-Volstad04), Carlisle et al. [2009](#ref-Carlisle09), Brown et al. [2012](#ref-Brown12), Hill et al. [2017](#ref-Hill17)). Many of these models are based on the understanding of relationships between stream condition and watershed characteristics (Hynes [1975](#ref-Hynes75), Johnson et al. [1997](#ref-Johnson97), Richards et al. [1997](#ref-Richards97)), which can be broadly conceptualized within the Driver-Pressure-Stress-Impact-Response (DPSIR) framework that describes relationships between the origins and consequences of environmental problems (Smeets and Weterings [1999](#ref-Smeets99)). However, past efforts have primarily focused on characterizing condition at unsampled locations, often predicting the most likely condition by estimating averages. Alternative modelling approaches, such as quantile-based methods (e.g., Cade and Noon ([2003](#ref-Cade03))), could be used to predict a range of expectations for biotic integrity from geospatial data. This approach differs fundamentally from previous efforts of estimating average condition by providing an estimate of the minimum and maximum scores that are likely given the landscape context. Once the responses of macroinvertebrate communities to landscape changes at large spatial scales are understood, expectations can be compared to field samples and sites can be prioritized by local managers based on deviation from the expectation.

The goal of this study was to present the development and application of a landscape model to predict a lower and upper bioassessment score that would be expected at a stream reach based on land use. Our specific objectives were to 1) develop and validate the landscape model, 2) apply the model results to categorize all stream segments in California into constraint classes, and 3) provide a case study within a single watershed to demonstrate how model predictions and classifications can be used to prioritize management actions at a local scale. The model was developed and applied to all streams and rivers in California, specifically focusing on the potential of urban and agricultural land use to constrain biological condition. The case study demonstrated how the statewide model could be used to classify and prioritize at the regional scale using guidance from a local stakeholder group from a heavily urbanized watershed. Managers currently have no prioritization tools for evaluating the context of biological integrity scores in their watershed. An interactive software application, the Stream Classification and Priority Explorer (SCAPE), was developed for our case study to help stakeholders choose regional management priorities from the statewide landscape model.

# Methods

## Study area and data sources

California covers 424,000 km of land with extreme diversity in several environmental gradients, such as elevation, geology, and climate (Figure 1a, Ode et al. ([2016b](#ref-Ode16))). Temperate rainforests occur in the north (North Coast region, NC), deserts and plateaus in the northeast and southeast (Deserts and Modoc Plateau region, DM), and Mediterranean climates in coastal regions (Chaparral and South Coast regions, CH and SC). The Central Valley region (CV) is largely agricultural and drains a large mountainous area in the east-central region of the state (Sierra Nevada region, SN). Urban development is concentrated in coastal areas in the central (San Francisco Bay Area, Chapparal region) and southern (Los Angeles, San Diego metropolitan area, South Coast) regions of the state. Landscape alteration has been relatively recent, with one estimate that developed lands increased in California by 38% from 1973 to 2000 (Sleeter et al. [2011](#ref-Sleeter11)). Silviculture and logging activities have also occurred in forested regions (SN, NC). For analysis, the state was evaluated as a whole and by the major regions described above (Ode et al. [2011](#ref-Ode11)).

The landscape model was developed using land use data, national stream hydrography layers, and biological assessments. A general assumption was that water quality issues could be conceptually linked to societal and economic drivers, reflected through the link between land use and stream biotic integrity (e.g., under the DPSIR framework, Figure 2, Smeets and Weterings ([1999](#ref-Smeets99))). Stream data from the National Hydrography Dataset Plus (NHD-plus) (McKay et al. [2012](#ref-McKay12)) were used to identify stream segments in California for modelling biological integrity. The NHD-plus is a surface water framework that maps drainage networks and associated features (e.g., streams, lakes, canals, etc.) in the United States. Stream segments designated in the NHD-plus were used as the discrete spatial unit for modelling biological integrity. Here and throughout, “segment” is defined based on NHD-Plus flowlines. Hydrography data were combined with landscape metrics available from the StreamCat Dataset (Hill et al. [2016](#ref-Hill16)) that provided estimates of land use at the riparian zone (i.e., a 100-m buffer on each side of the stream segment), the catchment (i.e., nearby landscape flowing directly into the immediate stream segment, excluding upstream segments), and the entire upstream watershed for each NHD-Plus segment. Many of the metrics in StreamCat were derived from the 2006 National Land Cover Database (Fry et al. [2011](#ref-Fry11)).

The California Stream Condition Index (CSCI) (Mazor et al. [2016](#ref-Mazor16)) was used as a measure of biological condition in California streams. The CSCI is a predictive index that compares the observed taxa and metrics at a site to those expected under least disturbed reference conditions (Stoddard et al. [2006](#ref-Stoddard06)). Expected values at a site are based on models that estimate the likely macroinvertebrate community in relation to factors that naturally influence biology, e.g., watershed size, elevation, climate, etc. (Moss et al. [1987](#ref-Moss87), Cao et al. [2007](#ref-Cao07)). The index score at a site can vary from 0 to ~ 1.4, with values near 1 indicating less deviation from reference state. Because the index was developed to minimize the influence of natural gradients, the index scores have consistent meaning across the state (Mazor et al. [2016](#ref-Mazor16)). A CSCI threshold of 0.79, based on the tenth percentile of scores at all reference calibration sites for the original index, has been proposed as a threshold below which a site does not meet designated biological uses (Board [2016](#ref-SDWB16)). As described below, the expected CSCI scores obtained from the landscape model were compared to this threshold to identify different constraint classes.

Benthic macroinvertebrate data were used to calculate 6270 individual CSCI scores at nearly 3400 unique sites between 2000 and 2016 (Figure 1b). We aggregated data collected under more than 20 federal, state, and regional bioassessment programs. Some of these programs employed a spatially balanced probabilistic design (e.g., the statewide Perennial Stream Assessment (Rehn [2015](#ref-Rehn15)), or the Stormwater Monitoring Coalition’s survey of southern California streams, Mazor ([2015](#ref-Mazor15))), while other programs used different designs for project-specific purposes (such as the statewide Reference Condition Monitoring Program, Ode et al. ([2016b](#ref-Ode16))). Most of these programs targeted perennial streams, although an unknown number of intermittent streams with flows lasting into the normal sampling period were included (Mazor et al. [2014](#ref-Mazor14)), particularly in more arid southern California. Because these programs are extensive, most regions and stream-types where perennial wadeable streams are located were represented in the calibration data set.

Field samples were collected during base flow conditions typically between May and July following methods in Ode et al. ([2016a](#ref-Ode16b)). Bioassessment sites were snapped to the closest NHD-plus stream segment in ArcGIS (ESRI (Environmental Systems Research Institute) [2016](#ref-ESRI16)). In cases where multiple sites were located on the same segment, the most downstream site was selected for model calibration under the assumption that the landscape data in StreamCat were most relevant to this site. One sample date was chosen randomly for sites with multiple dates so that one CSCI score was matched to a site. This created a final dataset of 2620 unique field observations used to calibrate and validate the landscape model.

## Building and validating the landscape model

Expected CSCI scores were modelled using estimates of canal/ditch density, imperviousness, road density/crossings, and urban and agricultural land use for each stream segment (Table 1, Figure S1). StreamCat was used as the only source for predictor variables because of consistent methods and linkage to NHD-Plus flowlines (Hill et al. [2016](#ref-Hill16)). Preliminary analyses indicated that these variables produced a predictive model with comparable performance relative to a larger model with additional predictors. These variables were chosen specifically as indicators of land-management activities that were most likely to limit the attainability of biological integrity. Landscape variables were preferred over in-stream data because landscape stressors can be more challenging to manage, and we wanted to quantify biological impacts relative to these challenges. Further, presence or absence of channel modification was not used to quantify limits on biological integrity because landscape predictors were more broadly inclusive of the problem (e.g., modified channels are often but not always constrained, constrained channels are not always modified). Overall, the model was associative by design and was intended as a predictive tool that does not describe specific mechanisms of biological alteration. We assumed that deviation of observed scores from the model predictions (i.e., residuals) could be used to describe in-stream factors associated with condition for follow-up analysis.

The model was developed using quantile regression forests to estimate ranges of likely CSCI scores in different landscapes (Meinshausen [2006](#ref-Meinshausen06), [2017](#ref-Meinshausen17)). Random forests are an ensemble learning approach to predictive modelling that aggregates information from a large number of regression trees and have been used extensively in bioassessment applications (Carlisle et al. [2009](#ref-Carlisle09), Chen et al. [2014](#ref-Chen14), Mazor et al. [2016](#ref-Mazor16), Fox et al. [2017](#ref-Fox17)). Random forest models can quantify complex, non-linear relationships and interactions between variables and can be more effective with large datasets relative to more commonly-used approaches, such as multiple regression (Breiman [2001](#ref-Breiman01), Hastie et al. [2009](#ref-Hastie09)). Quantile models, such as quantile regression forests, evaluate the conditional response across the range of values that are expected, in contrast to conventional models that provide only an estimate of the mean response (Cade and Noon [2003](#ref-Cade03)). This modelling approach can estimate a lower and an upper limit for the conditional distribution of likely scores that might be expected at a site given land use, which can be used to identify sites where that range includes management targets. Quantile regression forests were used to predict CSCI scores in each stream segment at five percent increments (i.e., 5th, 10th, etc.) from the 5th to 95th percentile of expectations. The statewide validated model (described below) was used to predict percentile expectations of CSCI scores at all stream segments where predictors were available. For example, the 50th percentile prediction was the most likely score for a stream segment given observed values for landscape variables, whereas a lower (e.g., 5th percentile) and upper (95th percentile) conditional quantile (and points in between) were also predicted. The quantregForest package (Meinshausen [2017](#ref-Meinshausen17)) for the R Statistical Programming Language (RDCT (R Development Core Team) [2018](#ref-RDCT18)) was used to develop the landscape model using the default settings.

We stratified sample data to ensure sufficient representation of landscape gradients across major regions in the state (Figure 1). Calibration data for the landscape model were obtained from a random selection of 75% of segments with observed CSCI scores, where the selection was based on a random draw from strata of quartiles defined by increasing watershed imperviousness relative to each region (n = 1965 segments). This ensured that the model was calibrated with data that covered the variation of landscape development among regions (i.e., regions with low development were not under-represented and those with high development were not over-represented). The remaining 25% of sites were used for model validation (n = 655). Where multiple samples were available at a single site, one sample was selected at random for both calibration and validation purposes. All predictions for the calibration dataset were obtained using out-of-bag estimates from the random forest model to prevent bias and over-fitting. Out-of-bag predictions are based on the subset of trees in the random forest model in which calibration data were excluded during training (Mazor et al. [2016](#ref-Mazor16), Meinshausen [2017](#ref-Meinshausen17)).

Model performance metrics were chosen to evaluate both predictive ability of the landscape model and potential for bias which may vary depending on different land use gradients across the state. Performance was assessed for the statewide dataset and within each major region by comparing differences between observed CSCI scores and median predictions at the same locations. Differences were evaluated using Pearson correlations and root mean squared errors (RMSE); high correlation coefficients and low RMSE values indicated good predictive ability. Regression analysis between observed and predicted scores was used to assess potential bias based on intercept and slope values differing from 0 and 1, respectively.

## Statewide application of the landscape model

We applied the landscape model to 138716 stream segments statewide to estimate the extent of streams in one of four different constraint classes: likely unconstrained, possibly unconstrained, possibly constrained, and likely constrained (Table 2). Ranges in predictor variables between the statewide and calibration datasets were similar, such that over-extrapolation of the model domain to the statewide data was unlikely. The classification process is described in Figure 3a through c. Classifications were based on the comparison of a CSCI threshold representing a management goal and the predicted range or predicted median score at a segment. These two decision points (i.e., the threshold and the size of the predicted range) were critical in defining segment classifications. We used a CSCI threshold of 0.79 following previous examples (Board [2016](#ref-SDWB16), Mazor et al. [2016](#ref-Mazor16)) and a lower and upper bound from the 10th to the 90th percentiles of expected CSCI scores. Stream segments where the predicted 90th quantile score was below the threshold were considered likely constrained, whereas those where the predicted 10th percentile was above the threshold were considered likely unconstrained (Figure 3c). The remaining sites were classified as possibly unconstrained or possibly constrained, based on whether the median expectation was above or below the threshold respectively (Table 2).

The influence of the key decision points on the extent of segment classifications created by the landscape model was evaluated. Stream segment classifications depend on the percentile range of score expectations (or certainty) from the landscape model (Figure 3b) and the CSCI threshold for evaluating the overlap extent (Figure 3c). For the certainty range, these bounds do not describe statistical certainty in the traditional sense (e.g., confidence interval), but rather a desired range that is defined as a potentially acceptable lower and upper limit around the median prediction for a CSCI score given landscape development. Eight different ranges of values for the score expectations from wide to narrow were evaluated at five percent intervals, i.e., 5th-95th, 10th-90th, …, 45th-55th. Different CSCI thresholds were also evaluated using values of 0.63, 0.79, and 0.92, corresponding to the 1st, 10th, and 30th percentile of scores at reference calibration sites used to develop the CSCI (Figure 1b) (Mazor et al. [2016](#ref-Mazor16)). The percentage of stream segments in each class statewide and by major regions were estimated for each of the twenty-four scenarios (width by threshold combinations). Although some of the results can be assumed (e.g., increasing CSCI thresholds causes more sites to be classified as constrained), we expected differences among regions based on differences in land use.

A categorization scheme was developed for sites where biomonitoring data were available to compare observed CSCI scores to the range of expected scores from the model (Figure 3d). This post-hoc classification was necessary to determine if observed CSCI scores were under- or over-scoring relative to landscape expectations, which can help prioritize management actions. For example, managers may choose to prioritize sites with index scores above or below the landscape model predictions differently than those that are within the expected range. Sites with observed scores above the upper limit of the segment expectation (e.g., above the 90th percentile of expected scores) were considered “over-scoring” and sites below the lower limit (e.g., 10th percentile) were considered “under-scoring”. If neither “over-scoring” nor “under-scoring”, the site was considered as “expected” within the context of the landscape model.

# Results

## Model performance

There was generally good agreement between observed and predicted CSCI scores statewide (Table 3, Figure S2). For the calibration dataset, observed and predicted values were correlated (r = 0.75, RMSE = 0.17), with an intercept (0.04) and slope (0.93) that indicated minimal median prediction bias. Performance was similar with the validation dataset (r = 0.72, RMSE = 0.18, intercept = 0.07, slope = 0.90).

Overall, the model performed well in regions with a mix of urban, agricultural, and open land (e.g., South Coast and Chaparral regions), whereas performance was weakest in regions without strong development gradients (e.g., Sierra Nevada region) (Table 3, Figure S2, S3). Performance for the Chaparral and South Coast regions were comparable or slightly improved compared to the statewide dataset for both the calibration (r = 0.71, 0.75, respectively) and validation (r = 0.74, 0.72) datasets. Model predictions for the Central Valley, Desert/Modoc, and North Coast regions had worse performance compared to the statewide results, with correlations of approximately 0.66, 0.50, and 0.55 with observed values in the calibration dataset and 0.49, 0.55, and 0.55 in the validation dataset. Model performance was weakest for the Sierra Nevada region (calibration r = 0.45, validation r = 0.21), where timber harvesting, rather than urban or agricultural development, is the most widespread stressor. A slight bias in model predictions was observed for the Central Valley and North Coast, where the former was over-predicted and the latter was under-predicted (Figure S2).

## Statewide patterns in stream constraints

Statewide, spatial patterns in the predicted limits of biological integrity were similar to patterns in land use (Figure 4). A majority of stream segments statewide were classified as possibly constrained (11% of all stream length) or possibly unconstrained (46%), whereas a minority were likely constrained (4%) or likely unconstrained (39%) (Table 4). Likely unconstrained streams were common in the Sierra Nevada (50%), North Coast (46%), and Desert/Modoc (46%) regions, whereas likely constrained were relatively abundant in the Central Valley (22%) and South Coast (15%) regions. However, constrained and unconstrained streams were both found in every region (Figure 4)

Observed CSCI scores were within the predicted decile range as often as expected (i.e., 80% statewide, based on the 10th and 90th conditional quantiles), and over-scoring sites were roughly as common (9%) as under-scoring sites (10%) (Table 5). Similar patterns were observed within regions, although a slightly larger percentage of sites in the Central Valley were under-scoring compared to the other regions, which may have been evidence of a slight bias of over-predicting in this region. Over-scoring sites were slightly more common in the South Coast and Sierra Nevada regions, whereas under-scoring sites were more common in others (i.e., the Chaparral, Central Valley, and Desert/Modoc regions).

Changing key decision points of the landscape model affected the estimates of the extent of streams in each class (Figure 5). Unsurprisingly, decreasing the certainty of predictions from the landscape model by narrowing the quantile range (5th-95th to 45th-55th) shifted a number of streams from the possible to likely category in both constrained and unconstrained segments. Similarly, changing the CSCI threshold from relaxed to more conservative (0.63 to 0.92) increased the number of streams classified as possibly or likely constrained and decreased the number of streams as possibly or likely unconstrained. However, the effects of these decision points varied greatly by region. For example, over 80% of segments in the Central Valley were classified as likely constrained using a high CSCI threshold with the narrowest range of predictions, whereas less than 1% of segments were in this category using a low CSCI threshold with the widest range of predictions. Opposite trends were observed in regions with reduced land use pressures. For example, almost all stream segments in the North Coast and Sierra Nevada regions were classified as likely unconstrained using a low CSCI threshold and narrow range of predictions.

# Discussion

Managing for biological integrity requires the use of 1) assessment tools that can accurately evaluate condition, and 2) tools that provide an estimate of the range of attainable conditions relative to the landscape. The landscape model was developed with these needs in mind to better inform application of the CSCI for decision-making relative to landscape constraints on biological condition. Statewide application of the model demonstrated where streams are likely constrained on a regional basis, whereas application in a case study (described below) demonstrated how the model can be used by local stakeholders to prioritize management actions that are informed by landscape context. The landscape model can inform the interpretation of biotic condition and is a decision-making tool that can help identify where management goals could be focused.

Model performance was comparable to similar studies that focused on developing predictions of biological condition from geospatial data. Hill et al. ([2017](#ref-Hill17)) developed a national model to predict stream site condition that correctly classified sites at about 75% of locations, depending on region. Importantly, regionally specific models were more accurate than a single national model. For continuous predictions of biointegrity index scores, Carlisle et al. ([2009](#ref-Carlisle09)) developed a model for a large area of the eastern United States. Models for continuous data were able to correctly identify class membership from an a posteriori prediction at about 85% of sites, which was similar in precision to models that were developed solely for categorical responses. For the landscape model herein, a comparison of the percentage of correctly classified sites that were above the 10th percentile of reference site scores (0.79) for observed data compared to predicted data showed that our model had comparable performance to other studies. The landscape model had 83% predictive accuracy for classifying sites as altered (<0.79) or unaltered (>0.79) for the statewide results. However, the goal of our model was distinct from previous studies, such that our intent was not to predict bioassessment scores at unsampled locations, but rather to describe variation in scores as a function of land use to identify constraints. Interpretation of predictive accuracies between models should consider the differences in the goals for each model.

## Case study: Application of the landscape model to the San Gabriel River watershed

Results from the statewide model were applied in a regional context through local application with a stakeholder group from the San Gabriel River watershed (Los Angeles County, California, Figure S4). The statewide model provides only a range of expected scores for a stream segment. Comparison of observed index scores from an actual biological sample with the results from the model can establish a basis for how managers prioritize sites. For example, managers may prioritize sites with observed scores that are above the modelled expectation differently than those that are scoring within the ranges predicted by the model. Alternatively, a site scoring as expected in an unconstrained segment could be a higher priority for managers than a site scoring as expected in a constrained segment; the latter may require more resources for comparable changes in biotic condition. As such, the lower San Gabriel watershed is heavily urbanized with many modified channels (Figure S4b) and managers require prioritization tools to identify where efforts should be focused among many sites in the context of landscape development. Information from the landscape model allowed the stakeholder group to develop management priorities based on how actual CSCI scores compared to biological expectations from the model (Figure S5).

Management priorities for individual sites that were important for the stakeholder group included the following actions (Table S1, Figure S6):

* Investigate: Conduct additional monitoring at a site or review of supplementary data (e.g., field visits, review aerial imagery);
* Protect: Recommend additional scrutiny of any proposed development and/or projects that could affect a site;
* Restore: Pursue targeted action for causal assessment and/or restoration activity at a site.

These priority actions were first identified independent from the landscape model and then assigned to each site by the stakeholder group based on a comparison of observed CSCI scores and the expected range of scores from the landscape model. In general, stakeholders assigned higher priority for all three actions to sites in likely unconstrained segments where CSCI scores were over- or under-scoring or at sites that were possibly unconstrained but the observed CSCI scores were below the biological threshold (dotted line in Figure S6, Table S1). Constrained sites were given lower priority overall or restoration actions were recommended as a lower priority despite low CSCI scores. Continuing current practices (e.g., routine monitoring, neither of the above actions) was also identified by the stakeholders as necessary for these low priority sites. Recommended actions to investigate were applied to both over-scoring and under-scoring sites, protect was given a high priority exclusively at over-scoring sites, and restore was more common at under-scoring sites.

The landscape model is primarily an exploratory tool to help identify patterns among monitoring sites where more intensive analyses may be appropriate. This application was tested through engagement of our local stakeholder group. Rather than identifying individual sites in need of specific management actions, the group used the landscape model to characterize patterns on the landscape that were consistent with the recommended management priorities. In doing so, the group explored and discussed potential management actions relative to the landscape characteristics of the watershed. The final decision by the group to prioritize management actions for the different sites in broad categories of protect, restore, and investigate was based on group discussions to reach agreement on how outcomes from the model could be applied. Facilitated discussions that directly engage stakeholders have been suggested by others as effective mechanisms that allow recommendations provided by these tools to be adopted in formal decision-making (Stein et al. [2017](#ref-Stein17)). However, the recommended actions have relevance only in the interests of the San Gabriel Regional Monitoring Program. Localized applications of the statewide model must engage stakeholders in a similar process to develop recommendations that are specific to regional needs at the watershed scale (Brody [2003](#ref-Brody03), Reed [2008](#ref-Reed08)).

Engagement with the stakeholder group was facilitated through creation of an interactive and online application, the Stream Classification and Priority Explorer (SCAPE, Figure S7, <http://shiny.sccwrp.org/scape/>)(Beck [2018a](#ref-Beck18c)). The SCAPE application can be used to select and visualize management priorities for all monitoring sites in the San Gabriel watershed (Figure S8) and was also critical for demonstrating how results from the statewide model could be used at a regional scale. Because of this application, the stakeholder group was able to explore the potential impacts of biointegrity policies currently under review in California, such as the effect of changing a potential threshold for defining biological use attainment and how the assigned priorities shift accordingly. Additionally, the SCAPE application correctly identified sites where discrepancies between CSCI scores and other measures of stream condition had been previously observed. Without the landscape context provided by the model (i.e., Figure S5, right side), stakeholders had limited information to prioritize among sites (i.e., no context for scores, Figure S5, left side).

The SCAPE application also demonstrated core concepts of the model and allowed stakeholders to explore the key decision points that affected the model output. Specifically, drop-down menus and sliders allowed users to change certainties in the CSCI score predictions (e.g., 10th and 90th percentile predictions) and explore alternative thresholds for biological objectives (e.g., 10th percentile of reference scores that defined constraint classes). This functionality allowed the stakeholders to develop recommendations that were completely independent of the model, i.e., decisions were not hard-wired into the model nor SCAPE. Results in Figure 5 also demonstrate the broader implications of how the key decision points affected model results at regional and statewide scales. These results and the functionality provided by SCAPE demonstrate flexibility of the landscape model and the considerations that should be made for regional applications. For example, constraint classifications and the decision points that define them may have little relevance in regions without development gradients that are not captured well by the model (e.g., Sierra Nevada, North Coast). Conversely, the chosen range for the lower and upper expectation of biological integrity is a tradeoff between which constraint classes are most appropriate for a region. Wider ranges force more stream segments into the “possible” constraint classes, whereas smaller ranges provide more separation of segments into the likely constrained or likely unconstrained classes. The specific choice is a management decision and we provide the ability to evaluate tradeoffs both in SCAPE and with our results herein.

## Alternative applications of the landscape model

Results from our analysis could be used for managing the biological integrity of streams under state or federal water quality mandates (e.g. “biological criteria” under the Clean Water Act). Management activities for biological integrity could involve the protection of sites meeting or exceeding biological objectives or the restoration of sites that have the potential to meet or exceed biological objectives. The selection of appropriate management actions for streams requires the consideration of the physical and chemical condition of streams concurrent with biological monitoring results. The landscape model can place observed scores in an appropriate context relative to their expected condition for the landscape. This information could provide flexibility in the selection of regulatory or management actions at specific sites or within larger regions (e.g., hydrologic subareas), and to further prioritize where and when actions should take place based on the resources needed for protection or restoration actions. For example, for sites that meet biological objectives but where the models predict some degree of constraint, regulatory actions may be associated with protecting that condition and could be implemented in the short-term to prevent degradation. Moreover, additional actions could be recommended to determine why these sites score above the constrained expectations, such as causal assessments to identify site-specific characteristics contributing to biointegrity (e.g., intact physical habitat independent of landscape development). This flexibility is not intended to exclude sites from consideration that are less likely to achieve biological objectives, but rather to facilitate the decision-making process through a more transparent application of the model in a regulatory application. The landscape model could also help identify where tiered aquatic life uses (TALU, Davies and Jackson ([2006](#ref-Davies06))) may be needed. However, the model is not intended, nor is it is sufficient, as a standalone tool for this purpose because it lacks specificity as to what uses may apply under different landscape conditions.

Non-regulatory applications of the landscape model are also possible by identifying where additional restoration, monitoring, or protection may have the most benefit. For example, landscape models could be used to support conservation planning, particularly at the watershed scale where land use practices can be a critical factor for decision-making. Ongoing work in California has focused on setting priorities for managing biodiversity that focus on watersheds within a conservation network (Howard et al. [2018](#ref-Howard18)). Results from the landscape model could be used to enhance this network by providing supporting information on constraints in an assessment framework. More generally, these applications could represent a novel use of bioassessment data beyond the pass/fail paradigm in the regulatory sense, for example, as tools for land use planning (Bailey et al. [2007](#ref-Bailey07)). In many cases, including California, bioassessment indices have been sufficiently developed to allow large-scale condition assessment across regions, yet they are rarely used as planning tools to guide decisions on where resources should be focused (Nel et al. [2009](#ref-Nel09)). Our landscape model makes bioassessment data in California more accessible and identifies an appropriate expectation for the information, enabling the potential for both regulatory and non-regulatory applications.

Several states have implemented alternative use designations for applying bioassessment criteria in modified channels (Florida Department of Environmental Protection [2011](#ref-FLDEP11), US Environmental Protection Agency, Region 10 [2013](#ref-USEPA13), Midwest Biodiversity Institute [2016](#ref-MBI16)). Although our results generally support the link between impacted biology and channel modification, a regulatory framework based on direct channel modification may be insufficient because constraints are more accurately defined relative to landscape development. As defined for the model, a constrained channel may or may not be engineered (see supplement for Tecolote Creek example, Figure S9), but an engineered channel in a developed landscape will typically be constrained. Furthermore, channel modification does not always result in biological degradation, particularly if the contributing watershed is largely undeveloped. For example, Stein et al. ([2013](#ref-Stein13)) observed reference-like bioassessment index scores in armored reaches within national forest lands in southern California. A classification framework for biological constraints using only channel modification would provide incomplete and potentially misleading information on streams. Ideally, context for evaluating biological condition from a landscape model, in conjunction with reach-specific data on channel modification, should be used.

Our approach to assessing constrained streams is readily transferable outside of California. The landscape model could be applied to other bioassessment methods, such as a multi-metric index (the most common bioassessment approach within the US; Buss et al. ([2014](#ref-Buss14))), O/E assessments (Moss et al. [1987](#ref-Moss87)), biological condition gradients (Davies and Jackson [2006](#ref-Davies06)), or with other biological endpoints (e.g., fish or diatoms). More importantly, our use of national geospatial datasets (i.e., NHDPlus, McKay et al. ([2012](#ref-McKay12)); StreamCat, Hill et al. ([2016](#ref-Hill16))) means that these methods could be applied across the United States. National bioassessment indices have been developed and the landscape model could be developed as a national-scale product of constraints on biological condition to complement recent work that predicted probable biological conditions with the National Rivers and Streams Assessment (Hill et al. [2017](#ref-Hill17)). Global geospatial datasets of freshwater-specific environmental variables are also available and could be used to develop similar models outside of the United States (Domisch et al. [2015](#ref-Domi15)).

Extension of the landscape models beyond California should also consider landscape stressors that are predictive of biotic condition in other regions. For example, urban and agricultural gradients were sufficient to characterize constraints in many regions of California, whereas Hill et al. ([2017](#ref-Hill17)) found that the volume of water stored by dams was an important predictor of biological condition in the Northern Appalachian and Northern Plains regions of the US. In their paper, Hill et al. ([2017](#ref-Hill17)) provided an example of how predictive models could be used to identify potential sites for restoration or conservation, however, their illustration did not explicitly identify sites that were over- or under-scoring relative to a biological endpoint. Our case study provided an example of how our model helped establish priorities at the local-scale and a similar process could be used for applying different landscape models in other states.

## Model assumptions and limitations

There are several characteristics of the landscape model that could affect its performance when applied outside of urban and agricultural settings. First, the model was developed with a focus on the needs of managers that apply bioassessment tools in developed landscapes where conditions are presumably constrained. As such, landscape variables were chosen to capture the effects of development on CSCI scores in these areas (Table 1). Application of the model in regions where different stressors have strong impacts on stream condition should consider the relevance of urban and agricultural stressors and if an alternative model that better captures other stressor gradients is needed. For example, our results suggest that streams in the North Coast and Sierra Nevada regions are largely unconstrained, but the landscape model was a poor predictor of CSCI scores in these areas. The dominant stressors likely to affect stream condition in these regions originate from sources that are less common in developed landscapes, such as silviculture and cannabis cultivation. The current landscape model does not adequately capture these impacts outside of urban and agricultural environments. Moreover, poor model predictions are compounded by low sensitivity of the CSCI to relevant stressor gradients in these regions (Mazor et al. [2016](#ref-Mazor16)). Accurate data for quantifying these potential stressors are not explicitly available in StreamCat, but surrogates could be explored in future models (e.g., coverage of introduced vegetation classes as a proxy for silviculture). Regardless, investments in improving spatial data could yield significant improvements in further development of bioassessment indices and tools for their interpretation.

An additional assumption is that the landscape model can adequately discriminate between intractable constraints on biology that are spatially and temporally pervasive relative to more manageable constraints. That is, we assumed that the impacts of stressors included in the model, such as urbanization, require long-term extensive mitigation planning, whereas stressors associated with deviations from model predictions can be mitigated in the short-term using focused actions. These assumptions are not unique to our model and have been used in other applications that have evaluated biological potential (Paul et al. [2008](#ref-Paul08), Chessman [2014](#ref-Chessman14), Waite et al. [2014](#ref-Waite14)). However, many stressors excluded from the model can have long-lasting impacts, leading to management scenarios where long-term recovery may only be possible with sustained and costly application of resources. For example, logging activities can impact benthic macroinvertebrate communities for a decade or more after harvesting activities have stopped (Stone and Wallace [1998](#ref-Stone98), Quinn and Wright-Stow [2008](#ref-Quinn08)). In urban areas, pervasive and profound alteration to groundwater and hydrology is common and stream communities in groundwater fed systems may require substantial time and resources for restoration. The potential legacy impacts of large-scale alterations of the natural environment are not well-captured by the current model, neither from a spatial nor temporal perspective. A more refined application of the landscape model would be necessary to evaluate different scales of impact, which could include developing separate models for each region, as well as more careful selection of model inputs to capture scales of interest for potential impacts on stream condition.

The landscape model describes constraints at scales larger than instream characteristics as a necessary approach to accurately predict bioassessment scores. Additional analyses that evaluate how different predictors influence model performance at different quantiles could provide insight into how landscape factors relate to constraints (e.g., Koenker and Machado [1999](#ref-Koenker99)). Further, a distinction between constraints on biological condition and channel modification is implicit such that indication of the former by the model does not explicitly indicate presence of the latter. As noted above, our results consistently indicated that engineered channels are biologically constrained, but the model is based on an a priori selection of land use variables to predict biotic integrity. A correspondence between habitat limitations and channel modification is likely in many cases, but data are insufficient to evaluate biological effects statewide relative to land use constraints. Moreover, bioassessment scores can be similar in modified channels compared to natural streams independent of watershed land use, i.e., concordance between degraded stream condition and channel modification may not always be observed (Stein et al. [2013](#ref-Stein13)). More comprehensive assessments at individual sites may be needed to diagnose the immediate causes of degraded condition.

An additional consideration in using the landscape model is the meaning of biologically constrained relative to whole stream communities. Biologically constrained sites were considered those where present landscapes were likely to limit CSCI scores that describe macroinvertebrate condition. In many cases, poor biotic condition of the macroinvertebrate community translates to poor stream condition. However, a constrained macroinvertebrate community does not always mean other biological attributes of stream condition (e.g., fish assemblages) are also constrained. Urban streams sometimes support diverse algal assemblages such that algal-based measures of biotic condition may alternatively suggest good biotic condition relative to macroinvertebrate-based indices (Brown et al. [2009](#ref-Brown09), Mazor et al. [2018](#ref-Mazor18)). Broadening the landscape model to include multiple taxonomic assemblages or endpoints would allow a more complete assessment of how condition relates to landscape alteration.

Finally, there are a few concerns applying a landscape modelling approach for bioassessment using the NHD-Plus flowlines as a base layer. We applied our model to the entire network of the NHD-Plus represented in StreamCat, which included a large number of intermittent or ephemeral streams, as well as non-wadeable rivers. Therefore, the application of model results in these stream-types is open to question, valid only to the degree that the CSCI and its response to landscape disturbance can represent more relevant measures of biological integrity. In regions where ephemeral streams are particularly common (e.g., the inland deserts or the South Coast region), estimates of the extent of constrained or unconstrained streams may be inaccurate.

## Summary

We demonstrated the use of quantile regression forests to successfully predict a lower and upper range of expected biological index scores that could be observed at a stream segment as a result of landscape development. Although random forest models have been increasingly used in bioassessment applications, our approach is the first to use quantile models to develop biological expectations. As such, additional work could build on this initial approach to apply these models in different locations, to alternative biological response endpoints, or to explore different predictors that capture regionally-specific stressor gradients. The predictive performance of quantile regression forests in bioassessment applications have also not been fully explored, such as understanding the accuracy of predictions or the relative importance of predictors at different quantiles. Our approach suggests these models are promising and future work could focus on any of the above suggestions to better understand the utility of these tools in applied contexts.

The landscape model can be used to characterize the extent of biologically constrained channels in developed landscapes and provides a tool to determine how managers can best prioritize resources for stream management by understanding landscape factors that might constrain each segment. Our application to the San Gabriel watershed demonstrated how the statewide results can be used at a spatial scale where many management decisions are implemented through close interaction with a regional stakeholder group with direct interests in the local resources. The approach leverages information from multiple sources to develop a context for biological assessment that provides an expectation of what is likely to be achieved based on current land use development. This can facilitate more targeted management actions that vary depending on the landscape context and can also inform decisions on extent and effort for future monitoring locations.

# Supplement

Geospatial data of model results mapped to stream reaches in California is provided at Beck ([2018b](#ref-Beck18d)). The SCAPE model application website is available at <http://shiny.sccwrp.org/scape/>, full source code accessible at Beck ([2018a](#ref-Beck18c)). Additional content for the case study, figures, and tables are available in the supplement.

# Author contributions

MB, RM, SJ, KW, JW, PO, RH, CL, MS, and ES performed the research and analyzed the data. MB, RM, SJ, JW, PO, RH, and CL wrote the paper. RM, SJ, KW, and PO provided data. All authors discussed the methods and results and contributed to the development of the manuscript.

# Acknowledgments

The authors acknowledge support from the San Gabriel River Regional Monitoring Program and the California State Water Resources Control Board. We thank Phil Markle and Lester Yuan for reviewing an earlier draft of the manuscript. The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency. Any mention of trade names, products, or services does not imply an endorsement by the U.S. Government or the U.S. Environmental Protection Agency. The EPA does not endorse any commercial products, services, or enterprises.

# Figures

Figure 1: Urban and agricultural land use (a) and distribution of observed stream CSCI scores (b) in California. Cover of urban and agricultural land use in stream watersheds was used to develop a landscape model for stream segment expectations of bioassessment scores. Breakpoints for CSCI scores are the 1st, 10th, and 30th percentile of scores at least-disturbed, reference sites throughout the state. Altered and intact refers to biological condition (Mazor et al. 2016). Grey lines are major environmental regions in California defined by ecoregional and watershed boundaries, CV: Central Valley, CH: Chaparral, DM: Deserts and Modoc Plateau, NC: North Coast, SN: Sierra Nevada, SC: South Coast.

Figure 1: Urban and agricultural land use (a) and distribution of observed stream CSCI scores (b) in California. Cover of urban and agricultural land use in stream watersheds was used to develop a landscape model for stream segment expectations of bioassessment scores. Breakpoints for CSCI scores are the 1st, 10th, and 30th percentile of scores at least-disturbed, reference sites throughout the state. Altered and intact refers to biological condition (Mazor et al. [2016](#ref-Mazor16)). Grey lines are major environmental regions in California defined by ecoregional and watershed boundaries, CV: Central Valley, CH: Chaparral, DM: Deserts and Modoc Plateau, NC: North Coast, SN: Sierra Nevada, SC: South Coast.

Figure 2: Conceptualized response and management pathways captured by the landscape model under the Driver-Pressure-Stress-Impact-Response (DPSIR) framework (Smeets and Weterings 1999). Landscape predictors provided in StreamCat (Hill et al. 2017) were used to describe pressures from urban and agricultural development that could impact the macroinvertebrate community in streams by altering physical and chemical habitat. Biological response was measured using the CSCI (Mazor et al. 2016) as an impact indicator and then evaluated relative to ranges of CSCI scores that were expected at each site provided by the landscape model. Observed CSCI scores and context from the landscape model provide a basis for informing management actions that could address environmental impacts at different points in the response pathway, where the management pathway could address causes at different scales and efficiencies.

Figure 2: Conceptualized response and management pathways captured by the landscape model under the Driver-Pressure-Stress-Impact-Response (DPSIR) framework (Smeets and Weterings [1999](#ref-Smeets99)). Landscape predictors provided in StreamCat (Hill et al. [2017](#ref-Hill17)) were used to describe pressures from urban and agricultural development that could impact the macroinvertebrate community in streams by altering physical and chemical habitat. Biological response was measured using the CSCI (Mazor et al. [2016](#ref-Mazor16)) as an impact indicator and then evaluated relative to ranges of CSCI scores that were expected at each site provided by the landscape model. Observed CSCI scores and context from the landscape model provide a basis for informing management actions that could address environmental impacts at different points in the response pathway, where the management pathway could address causes at different scales and efficiencies.

Figure 3: Application of the landscape model to identify site expectations and bioassessment performance for sixteen example stream segments. A range of CSCI scores is predicted from the model (a) and the lower and upper limits of the expectations are cut to define a certainty range for the predictions (b). Overlap of the certainty range at each segment with a chosen CSCI threshold (c) defines the stream segment classification as likely unconstrained, possibly unconstrained, possibly constrained, and likely constrained. The observed bioassessment scores are described relative to the classification as over scoring (above the certainty threshold), expected (within), and under scoring (below) for each of four stream classes (d).

Figure 3: Application of the landscape model to identify site expectations and bioassessment performance for sixteen example stream segments. A range of CSCI scores is predicted from the model (a) and the lower and upper limits of the expectations are cut to define a certainty range for the predictions (b). Overlap of the certainty range at each segment with a chosen CSCI threshold (c) defines the stream segment classification as likely unconstrained, possibly unconstrained, possibly constrained, and likely constrained. The observed bioassessment scores are described relative to the classification as over scoring (above the certainty threshold), expected (within), and under scoring (below) for each of four stream classes (d).

Figure 4: Statewide application of the landscape model showing the stream segment classifications. Major regional boundaries are also shown (see Figure 1).

Figure 4: Statewide application of the landscape model showing the stream segment classifications. Major regional boundaries are also shown (see Figure 1).

Figure 5: Changes in stream segment classes by region and statewide for different scenarios used to define biological constraints. Twenty-seven scenarios were tested that evaluated different combinations of certainty in the CSCI predictions (nine scenarios from wide to narrow prediction ranges as identified by the tail cutoff for the expected quantiles) and potential CSCI thresholds (three scenarios from low to high). The percentage of total stream length for likely unconstrained and likely constrained is shown for each scenario. Stream classifications as possibly unconstrained or possibly constrained are not shown but can be inferred form the area of white space above or below each bar. The solid black line indicates the percentage division between unconstrained and constrained classifications. CV: Central Valley, CH: Chaparral, DM: Deserts and Modoc Plateau, NC: North Coast, SN: Sierra Nevada, SC: South Coast.

Figure 5: Changes in stream segment classes by region and statewide for different scenarios used to define biological constraints. Twenty-seven scenarios were tested that evaluated different combinations of certainty in the CSCI predictions (nine scenarios from wide to narrow prediction ranges as identified by the tail cutoff for the expected quantiles) and potential CSCI thresholds (three scenarios from low to high). The percentage of total stream length for likely unconstrained and likely constrained is shown for each scenario. Stream classifications as possibly unconstrained or possibly constrained are not shown but can be inferred form the area of white space above or below each bar. The solid black line indicates the percentage division between unconstrained and constrained classifications. CV: Central Valley, CH: Chaparral, DM: Deserts and Modoc Plateau, NC: North Coast, SN: Sierra Nevada, SC: South Coast.

# Tables

Table 1: Land use variables used to develop the landscape model of stream bioassessment scores. All variables were obtained from StreamCat (Hill et al. [2016](#ref-Hill16)) and applied to stream segments in the National Hydrography Dataset Plus (NHD-plus) (McKay et al. [2012](#ref-McKay12)). The measurement scales for each variable are at the riparian (100 m buffer), catchment, and/or watershed, scale relative to a stream segment. Combined scales for riparian measurements (e.g., riparian + catchment, riparian + watershed) are riparian estimates for the entire catchment or watershed area upstream, as compared to only the individual segment. Total urban and agriculture land use variables were based on sums of individual variables in StreamCat as noted in the description. Rp100: riparian, Cat: catchment, Ws: watershed

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Scale | Description | Unit |
| CanalDens | Cat, Ws | Density of NHDPlus line features classified as canal, ditch, or pipeline | km/sq km |
| PctImp2006 | Cat, Ws, Cat + Rp100, Ws + Rp100 | Mean imperviousness of anthropogenic surfaces (NLCD 2006) | % |
| TotUrb2011 | Cat, Ws, Cat + Rp100, Ws + Rp100 | Total urban land use as sum of developed open, low, medium, and high intensity (NLCD 2011) | % |
| TotAg2011 | Cat, Ws, Cat + Rp100, Ws + Rp100 | Total agricultural land use as sum of hay and crops (NLCD 2011) | % |
| RdDens | Cat, Ws, Cat + Rp100, Ws + Rp100 | Density of roads (2010 Census Tiger Lines) | km/sq km |
| RdCrs | Cat, Ws | Density of roads-stream intersections (2010 Census Tiger Lines-NHD stream lines) | crossings/sq km |

Table 2: Stream class definitions describing potential biological constraints. Classes are based on the overlap of the range of likely bioassessment scores with a potential threshold for a biological objective. Identifying stream classes requires selecting the cutoff range of likely scores from the landscape model and a chosen threshold for the objective.

|  |  |  |
| --- | --- | --- |
| Class | Definition | Example |
| Likely unconstrained | Lower bound of prediction range is above threshold | 10th percentile > 0.79 |
| Possibly unconstrained | Lower bound of prediction range is below threshold, but median prediction is above | 50th percentile > 0.79 |
| Possibly constrained | Upper bound of prediction range is above threshold, but median prediction is below | 50th percentile < 0.79 |
| Likely constrained | Upper bound of prediction range is below threshold | 90th percentile < 0.79 |

Table 3: Performance of the landscape model by calibration (Cal) and validation (Val) datasets in predicting CSCI scores. The statewide dataset (Figure 4) and individual regions of California (Figure 1) are evaluated. Averages and standard deviations (in parentheses) for observed and predicted CSCI values of each dataset are shown. Pearson correlations (r), root mean squared errors (RMSE), intercept, and slopes are for comparisons of predicted and observed values to evaluate model performance. All correlations, intercepts, and slopes are significant at alpha = 0.05. CV: Central Valley, CH: Chaparral, DM: Deserts and Modoc Plateau, NC: North Coast, SN: Sierra Nevada, SC: South Coast.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset | Location | n | Observed | Predicted | r | RMSE | Intercept | Slope |
| Cal | Statewide | 1965 | 0.82 (0.26) | 0.83 (0.20) | 0.75 | 0.17 | 0.04 | 0.93 |
|  | CH | 512 | 0.76 (0.27) | 0.79 (0.21) | 0.71 | 0.19 | 0.03 | 0.92 |
|  | CV | 116 | 0.51 (0.18) | 0.57 (0.15) | 0.66 | 0.15 | 0.05 | 0.81 |
|  | DM | 86 | 0.87 (0.22) | 0.91 (0.14) | 0.50 | 0.20 | 0.15 | 0.79 |
|  | NC | 208 | 0.92 (0.20) | 0.94 (0.13) | 0.55 | 0.17 | 0.12 | 0.86 |
|  | SC | 631 | 0.79 (0.24) | 0.78 (0.21) | 0.75 | 0.16 | 0.11 | 0.87 |
|  | SN | 412 | 0.98 (0.18) | 0.98 (0.09) | 0.45 | 0.16 | 0.12 | 0.88 |
| Val | Statewide | 655 | 0.82 (0.25) | 0.84 (0.20) | 0.72 | 0.18 | 0.07 | 0.90 |
|  | CH | 172 | 0.76 (0.27) | 0.81 (0.21) | 0.74 | 0.19 | -0.04 | 0.98 |
|  | CV | 40 | 0.52 (0.19) | 0.59 (0.16) | 0.49 | 0.19 | 0.16 | 0.60 |
|  | DM | 28 | 0.84 (0.17) | 0.93 (0.11) | 0.55 | 0.17 | 0.07 | 0.83 |
|  | NC | 71 | 0.94 (0.19) | 0.96 (0.11) | 0.55 | 0.16 | 0.00 | 0.98 |
|  | SC | 208 | 0.80 (0.24) | 0.78 (0.21) | 0.72 | 0.17 | 0.17 | 0.81 |
|  | SN | 136 | 0.97 (0.17) | 0.98 (0.09) | 0.21 | 0.17 | 0.57 | 0.41 |

*Table 4: (#tab:clstot) Summary of stream length for each stream class statewide and major regions of California (Figures 1, 4). Lengths are in kilometers with the percentage of the total length in a region in parentheses. All lengths are based on a CSCI threshold of 0.79 and the 10th to 90th percentile of expected scores from the landscape model. CV: Central Valley, CH: Chaparral, DM: Deserts and Modoc Plateau, NC: North Coast, SN: Sierra Nevada, SC: South Coast.*

|  | constrained | | unconstrained | |
| --- | --- | --- | --- | --- |
| Region | likely | possibly | possibly | likely |
| Statewide | 8150 (4) | 24735 (11) | 101591 (46) | 85317 (39) |
| CV | 3356 (22) | 8010 (52) | 3202 (21) | 951 (6) |
| CH | 1642 (3) | 7840 (13) | 30693 (50) | 21206 (35) |
| DM | 255 (0) | 3395 (6) | 27194 (47) | 26479 (46) |
| NC | 108 (0) | 1442 (5) | 14152 (49) | 13286 (46) |
| SN | 20 (0) | 1067 (3) | 18228 (48) | 19032 (50) |
| SC | 2770 (15) | 2981 (16) | 8122 (45) | 4363 (24) |

*Table 5: (#tab:reltot) Summary of CSCI scores by relative expectations for each stream class statewide and in each major region of California (Figures 1, 4). Average CSCI scores (standard deviation) and counts (percent) of the number of monitoring stations in each relative score category and region are shown. Sites are over-scoring if the observed scores are above the range of expectations at a segment, expected if within the range, or under-scoring if below the range. CV: Central Valley, CH: Chaparral, DM: Deserts and Modoc Plateau, NC: North Coast, SN: Sierra Nevada, SC: South Coast.*

|  | under-scoring | | expected | | over-scoring | |
| --- | --- | --- | --- | --- | --- | --- |
| Region | CSCI | n (%) | CSCI | n (%) | CSCI | n (%) |
| Statewide | 0.54 (0.21) | 267 (10) | 0.83 (0.23) | 2041 (80) | 1.08 (0.17) | 242 (9) |
| CH | 0.47 (0.18) | 89 (13) | 0.79 (0.24) | 535 (80) | 1.08 (0.17) | 45 (7) |
| CV | 0.34 (0.12) | 25 (17) | 0.54 (0.17) | 118 (81) | 0.63 (0.25) | 2 (1) |
| DM | 0.6 (0.17) | 15 (14) | 0.9 (0.17) | 89 (80) | 1.15 (0.08) | 7 (6) |
| NC | 0.66 (0.17) | 28 (10) | 0.93 (0.16) | 228 (82) | 1.15 (0.08) | 22 (8) |
| SC | 0.54 (0.22) | 56 (7) | 0.78 (0.22) | 656 (81) | 1.02 (0.2) | 97 (12) |
| SN | 0.67 (0.16) | 54 (10) | 0.99 (0.11) | 415 (77) | 1.16 (0.06) | 69 (13) |

# References

Allan, D., D. Erickson, and J. Fay. 1997. The influence of catchment land use on stream integrity across multiple spatial scales. Freshwater Biology 37:149–161.

Bailey, R. C., T. B. Reynoldson, A. G. Yates, J. Bailey, and S. Linke. 2007. Integrating stream bioassessment and landscape ecology as a tool for land use planning. Freshwater Biology 52:908–917.

Beck, M. W. 2018a. SCCWRP/SCAPE: v1.0 (Version 1.0). Zenodo, <http://doi.org/10.5281/zenodo.1218121>.

Beck, M. W. 2018b. Constrained streams for biological integrity in California. Knowledge Network for Biocomplexity. urn:uuid:75411f50-32ed-42a5-bbfd-26833c7a441f.

Beechie, T., G. Pess, P. Roni, and G. Giannico. 2007. Setting river restoration priorities: A review of approaches and general protocol for identifying and prioritizing actions. North American Journal of Fisheries Management 28:891–905.

Bernhardt, E. S., E. B. Sudduth, M. A. Palmer, J. D. Allan, J. L. Meyer, G. Alexander, J. Follastad-Shah, B. Hassett, R. Jenkinson, R. Lave, J. Rumps, and L. Pagano. 2007. Restoring rivers one reach at a time: Results from a survey of U.S. river restoration practitioners. Restoration Ecology 15:482–493.

Board, S. D. R. W. Q. C. 2016. Clean Water Act Sections 305(B) and 303(D) Integrated Report for the San Diego Region. California Environmental Protection Agency, Sacramento, California.

Breiman, L. 2001. Random forests. Machine Learning 45:5–32.

Brody, S. D. 2003. Measuring the effects of stakeholder participation on the quality of local plans based on the principles of collaborative ecosystem management. Journal of Planning Education and Research 22:407–419.

Brown, L. R., T. F. Cuffney, J. F. Coles, F. Fitzpatrick, G. McMahon, J. Steuer, A. H. Bell, and J. T. May. 2009. Urban streams across the USA: Lessons learned from studies in nine metropolitan areas. Journal of the North American Benthological Society 28:1051–1069.

Brown, L. R., J. T. May, A. C. Rehn, P. R. Ode, I. R. Waite, and J. G. Kennen. 2012. Predicting biological condition in southern California streams. Landscape and Urban Planning 108:17–27.

Buss, D. F., D. M. Carlisle, T. -S. Chon, J. Culp, J. s. Harding, H. E. Keizer-Vlek, W. A. Robinson, S. Strachan, C. Thirion, and R. M. Hughes. 2014. Stream biomonitoring using macroinvertebrates around the globe: A comparison of large-scale programs. Environmental Monitoring and Assessment 187:4132.

Cade, B. S., and B. R. Noon. 2003. A gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environment 1:412–420.

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.

Carline, R. F., and M. C. Walsh. 2007. Responses to riparian restoration in the Spring Creek watershed, central Pennsylvania. Restoration Ecology 15:731–742.

Carlisle, D. M., J. Falcone, and M. R. Meador. 2009. Predicting the biological condition of streams: Use of geospatial indicators of natural and anthropogenic characteristics of watersheds. Environmental Monitoring and Assessment 151:143–160.

Chen, K., R. M. Hughes, S. Xu, J. Zhang, D. Cai, and B. Wang. 2014. Evaluating performance of macroinvertebrate-based adjusted and unadjusted multi-metric indices (MMI) using multi-season and multi-year samples. Ecological Indicators 36:142–151.

Chessman, B. C. 2014. Predicting reference assemblages for freshwater bioassessment with limiting environmental difference analysis. Freshwater Science 33:1261–1271.

Chessman, B. C., M. Muschal, and M. J. Royal. 2008. Comparing apples with apples: Use of limiting environmental differences to match reference and stressor-exposure sites for bioassessment of streams. River Research and Applications 24:103–117.

Chessman, B. C., and M. J. Royal. 2004. Bioassessment without reference sites: Use of environmental filters to predict natural assemblages of river macroinvertebrates. Journal of the North American Benthological Society 23:599–615.

Clapcott, J. E., K. J. Collier, R. G. Death, E. O. Goodwin, J. S. Harding, D. Kelly, J. R. Leathwick, and R. G. Young. 2011. Quantifying relationships between land-use gradients and structural and functional indicators of stream ecological integrity. Freshwater Biology 57:74–90.

Cuffney, T. F., S. S. Qian, R. A. Brightbill, J. T. May, and I. R. Waite. 2011. Response to King and Baker: Limitations on threshold detection and characterization of community thresholds. Ecological Applications 21:2840–2845.

Davies, S. P., and S. K. Jackson. 2006. The biological condition gradient: A descriptive model for interpreting change in aquatic ecosystems. Ecological Applications 16:1251–1266.

Domisch, S., G. Amatulli, and W. Jetz. 2015. Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution. Scientific Data 2:150073.

ESRI (Environmental Systems Research Institute). 2016. ArcGIS v10.5.

Finkenbine, J. K., J. W. Atwater, and D. S. Mavinic. 2000. Stream health after urbanization. Journal of the American Water Resources Association 36:1149–1160.

Florida Department of Environmental Protection. 2011. Development of aquatic life use support attainment thresholds for Florida’s Stream Condition Index and Lake Vegetation Index. Page 39. FDEP Standards; Assessment Section, Bureau of Assessment; Restoration Support, Tallahassee, Florida.

Fox, E. W., R. A. Hill, S. G. Leibowitz, A. R. Olsen, D. J. Thornbrugh, and M. H. Weber. 2017. Assessing the accuracy and stability of variable selection methods for random forest modeling in ecology. Environmental Monitoring and Assessment 189:316.

Fry, J., G. Xian, S. Jin, J. Dewitz, C. Homer, L. Yang, C. Barnes, N. Herold, and J. Wickham. 2011. Completion of the 2006 National Land Cover Database for the conterminous United States. Photogrammetric Engineering and Remote Sensing 77:858–864.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. The elements of statistical learning: Data mining, inference, and prediction. Page 745. 2nd editions. Springer, New York.

Hill, R. A., E. W. Fox, S. G. Leibowitz, A. R. Olsen, D. J. Thornbrugh, and M. H. Weber. 2017. Predictive mapping of the biotic condition of conterminous U.S. rivers and streams. Ecological Applications 27:2397–2415.

Hill, R. A., M. H. Weber, S. G. Leibowitz, A. R. Olsen, and D. J. Thornbrugh. 2016. The Stream-Catchment (StreamCat) dataset: A database of watershed metrics for the conterminous United States. Journal of the American Water Resources Association 52:120–128.

Howard, J. K., K. A. Fesenmyer, T. E. Grantham, J. H. Viers, P. R. Ode, P. B. Moyle, S. J. Kupferburg, J. L. Furnish, A. Rehn, J. Slusark, R. D. Mazor, N. R. Santos, R. A. Peek, and A. N. Wright. 2018. A freshwater conservation blueprint for California: Prioritizing watersheds for freshwater biodiversity. Freshwater Science 37:417–431.

Hynes, H. B. N. 1975. The stream and its valley. SIL Proceedings, 1922-2010 19:1–15.

Johnson, L., C. Richards, G. Host, and J. Arthur. 1997. Landscape influences on water chemistry in Midwestern stream ecosystems. Freshwater Biology 37:193–208.

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, Champaign, Illinois.

Kenney, M. A., P. R. Wilcock, B. F. Hobbs, N. E. Flores, and D. C. Martínez. 2012. Is urban stream restoration worth it? Journal of the American Water Resources Association 48:603–615.

Koenker, R., and J. A. F. Machado. 1999. Goodness of fit and related inference processes for quantile regression. Journal of the American Statistical Association 94:1296–1310.

Lester, R. E., and A. J. Boulton. 2008. Rehabilitating agricultural streams in Australia with wood: A review. Environmental Management 42:310–326.

Loflen, C., H. Hettesheimer, L. B. Busse, K. Watanabe, R. M. Gersberg, and V. Lüderitz. 2016. Inadequate monitoring and inappropriate project goals: A case study on the determination of success for the Forester Creek improvement project. Ecological Restoration 34:124–134.

Mazor, R. D. 2015. Bioassessment of perennial streams in Southern California: A report on the first five years of the Stormwater Monitoring Coalition’s regional stream survey. Southern California Coastal Water Research Project, Costa Mesa, California.

Mazor, R. D., M. W. Beck, and J. Brown. 2018. 2017 report on the SMC regional stream survey. Southern California Coastal Water Research Project, Costa Mesa, California.

Mazor, R. D., A. C. Rehn, P. R. Ode, M. Engeln, K. C. Schiff, E. D. Stein, D. J. 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.

Mazor, R. D., E. D. Stein, P. R. Ode, and K. Schiff. 2014. Integrating intermittent streams into watershed assessments: Applicability of an index of biotic integrity. Freshwater Science 35:459–474.

McKay, L., T. Bondelid, T. Dewald, J. Johnston, R. Moore, and A. Reah. 2012. NHDPlus Version 2: User Guide.

Meinshausen, N. 2006. Quantile regression forests. Journal of Machine Learning Research 7:983–999.

Meinshausen, N. 2017. QuantregForest: Quantile regression forests.

Midwest Biodiversity Institute. 2016. Identification of predictive habitat attributes for Minnesota streams to support tiered aquatic life uses. Midwest Biodiversity Institute, prepared on behalf of the Minnesota Pollution Control Agency, Columbus, Ohio.

Morgan, R. P., and S. E. Cushman. 2005. Urbanization effects on stream fish assemblages in Maryland, USA. Journal of the North American Benthological Society 24:643–655.

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

Nel, J. L., D. J. Roux, R. Abell, P. J. Ashton, R. M. Cowling, J. V. Higgins, M. Thieme, and J. H. Viers. 2009. Progress and challenges in freshwater conversation planning. Aquatic Conservation: Marine and Freshwater Ecosystems 19:474–485.

Ode, P. R., A. E. Fetscher, and L. B. Busse. 2016a. Standard operating procedures (SOP) for the collection of field data for bioassessments of California wadeable streams: Benthic macroinvertebrates, algae, and physical habitat. California State Water Resources Control Board Surface Water Ambient Monitoring Program, Sacramento, California.

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). 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; Game Aquatic Bioassessment Laboratory,; the U.S. Environmental Protection Agency, Sacramento, California.

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. 2016b. Evaluating the adequacy of a reference-site pool for ecological assessments in environmentally complex regions. Freshwater Science 35:237–248.

Paul, M. J., D. W. Bressler, A. H. Purcell, M. T. Barbour, E. T. Rankin, and V. H. Resh. 2008. Assessment tools for urban catchments: Defining observable biological potential. Journal of the American Water Resources Association 45:320–330.

Quinn, J. M., and A. E. Wright-Stow. 2008. Stream size influences stream temperature impacts and recovery rates after clearfell logging. Forest Ecology and Management 256:2101–2109.

RDCT (R Development Core Team). 2018. R: A language and environment for statistical computing, v3.5.1. R Foundation for Statistical Computing, Vienna, Austria.

Reed, M. S. 2008. Stakeholder participation for environmental management: A literature review. Biological Conservation 141:2417–2431.

Rehn, A. C. 2015. The perennial streams assessment (PSA): An assessment of biological condition using the new California Stream Condition Index (CSCI). California State Water Resources Control Board Surface Water Ambient Monitoring Program, Sacramento, California.

Reynoldson, T. B., R. H. Norris, V. H. Resh, K. E. Day, and D. M. Rosenberg. 1997. The reference condition: A comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. Journal of the North American Benthological Society 16:833–852.

Richards, C., R. Haro, L. Johnson, and G. Host. 1997. Catchment and reach-scale properties as indicators of macroinvertebrate species traits. Freshwater Biology 37:219–230.

Roni, P., and T. Beechi. 2012. Stream and watershed restoration guide: A guide to restoring riverine processes and habitats. Page 316. first. John Wiley & Sons, Hoboken, New Jersey.

Shoredits, A. S., and J. A. Clayton. 2013. Assessing the practice and challenges of stream restoration in urbanized environments of the USA. Geography Compass 7:358–372.

Sleeter, B. M., T. S. Wilson, C. E. Soulard, and J. Liu. 2011. Estimation of the late twentieth century land-cover change in California. Environmental Monitoring and Assessment 173:251–266.

Smeets, E., and R. Weterings. 1999. Environmental indicators: Typology and overview. Page 19. European Environmental Agency, Copenhagen, Denmark.

Stein, E. D., M. R. Cover, A. E. Fetscher, C. O’Reilly, R. Guardado, and C. W. Solek. 2013. Reach-scale geomorphic and biological effects of localized streambank armoring. Journal of the American Water Resources Association 49:780–792.

Stein, E. D., A. Sengupta, R. D. Mazor, K. McCune, B. P. Bledsoe, and S. Adams. 2017. Application of regional flow-ecology relationships to inform watershed management decisions: Applications of the ELOHA framework in the San Diego River watershed, California, USA. Ecohydrology 10:e1869.

Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson, and R. H. Norris. 2006. Setting expectations for the ecological condition of streams: The concept of reference condition. Ecological Applications 16:1267–1276.

Stone, M. K., and J. B. Wallace. 1998. Long-term recovery of a mountain stream from clear-cut logging: The effects of forest succession on benthic invertebrate community structure. Freshwater Biology 39:151–169.

US Environmental Protection Agency, Region 10. 2013. Technical Support Document for EPA’s Action on the State of Oregon’s Revised Water Quality Standards for the West Division Main Canal.

USGS (US Geological Survey). 1999. The quality of our nation’s waters: nutrients and pesticides. Reston, Virginia.

Vølstad, J. H., N. E. Roth, G. Mercurio, M. T. Southerland, and D. E. Strebel. 2004. Using environmental stressor information to predict the ecological status of Maryland non-tidal streams as measured by biological indicators. Environmental Monitoring and Assessment 84:219–242.

Waite, I. R. 2013. Development and application of an agricultural intensity index to invertebrate and algal metrics from streams at two scales. Journal of the American Water Resources Association 49:431–448.

Waite, I. R., J. G. Kennen, J. T. May, L. R. Brown, T. F. Cuffney, K. A. Jones, and J. L. Orlando. 2014. Stream macroinvertebrate response models for bioassessment metrics: Addressing the issue of spatial scale. PLOS ONE 9:e90944.

Walsh, C. J., A. H. Roy, J. w. Feminella, P. D. Cottingham, P. M. Groffman, and R. P. Morgan. 2005. The urban stream syndrome: Current knowledge and the search for a cure. Journal of the North American Benthological Society 24:706–723.

Wang, L., J. Lyons, P. Kanehl, and R. Gatti. 1997. Influences of watershed land use on habitat quality and biotic integrity in Wisconsin streams. Fisheries 22:6–12.