

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

Marcus W. Beck^{1,7}, Raphael D. Mazor^{1,8}, Scott Johnson^{2,9}, Karin Wisenbaker^{2,10}, Joshua Westfall^{3,11}, Peter R. Ode^{4,12}, Ryan Hill^{5,13}, Chad Loflen^{6,14}, Martha Sutula^{1,15}, and Eric D. Stein^{1,16}

¹Southern California Coastal Water Research Project, 3535 Harbor Boulevard, Costa Mesa, California 92626 USA

²Aquatic Bioassay & Consulting Laboratories, Inc. 29 North Olive Street, Ventura, California 93001 USA

³Sanitation Districts of Los Angeles County, 1955 Workman Mill Road, Whittier, California 90601 USA

⁴California Department of Fish and Wildlife, Office of Spill Prevention and Response, 2005 Nimbus Road, Rancho Cordova, California 95670 USA

⁵Department of Forest Engineering, Resources, and Management, Oregon State University c/o Western Ecology Division, 140 Peavy Hall, 3100 Southwest Jefferson Way, Corvallis, Oregon 97333 USA

⁶San Diego Regional Water Quality Control Board, Healthy Waters Branch, Monitoring and Assessment Research Unit, 2375 Northside Drive, San Diego, California 92108 USA

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 model predictions. 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, that compares observed scores with predictions 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 predicted ranges, and were therefore given a low priority for restoration. In contrast, 2 low-scoring sites in the undeveloped upper watershed were prioritized for causal assessment and possible future restoration, whereas 3 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

The widespread use of bioassessment data to assess the ecological condition of aquatic environments is a significant advance over chemical or physical methods of assess-

ment, yet managers and stakeholders require contextual information for synthesizing and interpreting biological information. The reference condition concept that is built into

E-mail addresses: ⁷marcusb@sccwrp.org; ⁸raphaelm@sccwrp.org; ⁹scott@aquaticbioassay.com; ¹⁰karin@aquaticbioassay.com; ¹¹jwestfall@lacsds.org; ¹²peter.ode@wildlife.ca.gov; ¹³hill.ryan@epa.gov; ¹⁴chad.loflen@waterboards.ca.gov; ¹⁵marthas@sccwrp.org; ¹⁶erics@sccwrp.org

DOI: 10.1086/705996. Received 3 March 2019; Accepted 15 July 2019; Published online 11 September 2019.

Freshwater Science. 2019. 38(4):000–000. © 2019 by The Society for Freshwater Science. All rights reserved. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0), which permits non-commercial reuse of the work with attribution. For commercial use, contact journalpermissions@press.uchicago.edu.

000

many biological indices provides a broad context for interpreting observed biological condition relative to that occurring in unaltered habitats for a particular region (Reynoldson et al. 1997, Stoddard et al. 2006). 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) 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, Chessman 2014). 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, Beechie et al. 2007, Paul et al. 2008). In developed landscapes, the majority of stream miles are in poor biotic condition (USGS 1999, Finkenbine et al. 2000, Morgan and Cushman 2005). Restoring streams in urban or agricultural settings can be costly; success is not universally defined, and achieving regionally-defined reference-like conditions may be difficult (Bernhardt et al. 2007, Kenney et al. 2012, Shoredits and Clayton 2013). Conventional approaches to protect and restore biological integrity have commonly focused on both direct improvements at the site level to mitigate instream stressors (Carline and Walsh 2007, Lester and Boulton 2008, Roni and Beechi 2012, Loflen et al. 2016) and implementation of upstream preventive measures. These approaches can lead to improvements in ecological condition (e.g., Bernhardt et al. 2007), but there is no universal remedy for achieving biological integrity in streams. In urban areas, the protective thresholds needed to maintain biological integrity have been debated (Cuffney et al. 2011). For biological integrity, several states have implemented a tiered aquatic life use or alternative use designation as potential approaches to account for shifts in ecosystem baseline conditions caused by channel modification (e.g., FDEP 2011, USEPA 2013, MBI 2016, permitted under section 303(c)(2) of the Clean Water Act, 33 U.S.C. §1251 - 1376). Other approaches may include site-specific criteria or alternative thresholds with specific guidelines for implementation (e.g., SDRWQB 2016).

Herein, we define constrained streams as those where present landscape conditions are likely to limit management options for restoring biological integrity. This definition describes a biological expectation and is distinct from the definition of constrained 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 caused by factors that may be difficult to manage, managers could prioritize sites for protection or restoration where factors influencing condition are more easily managed. For example, a monitoring site with an ob-

served biological index score that is above its predicted range could be assigned a higher management priority (i.e., for protection) relative to a site that is scoring within the expected range given its type and amount of landscape development. A predictive model of bioassessment scores that is based on landscape metrics (e.g., imperviousness) could identify likely 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, Chessman 2014). Similar concepts have been applied in a landscape context to describe variation in biological communities and metrics at different spatial scales (Waite 2013, Waite et al. 2014), although they have not been developed to identify constraints as defined above.

Consistent and empirical relationships between land use and biotic integrity have been identified in many cases (Allan et al. 1997, Wang et al. 1997, Clapcott et al. 2011), and previous modeling efforts have successfully used geospatial data to predict local stream biological condition across either regional or national scales (Vølstad et al. 2004, Carlisle et al. 2009, Brown et al. 2012, Hill et al. 2017). Many of these models are based on our understanding of relationships between stream biota and watershed characteristics (Hynes 1975, Johnson et al. 1997, Richards et al. 1997), 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). However, past efforts have primarily focused on predicting the most likely biological condition occurring at unsampled locations. Alternative modeling approaches, such as quantile-based methods (e.g., Cade and Noon 2003), could be used to predict the range of condition scores likely to occur given a site's environmental setting. Once the responses of macro-invertebrate assemblages to landscape changes that occur across large spatial scales are modeled, predictions can be compared to observed conditions and sites can be prioritized by local managers based on the degree of deviation of observed from the expected conditions.

The goal of this study was to present the development and application of a model to predict the lower and upper bioassessment scores that would be expected at a stream reach given its surrounding land use. Our specific objectives were to 1) develop and validate a model applicable to most sites occurring within a large and environmentally diverse landscape, 2) use the model 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 management actions at the regional scale in conjunction with guidance from a local stakeholder group from a heavily-urbanized 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 km² of land and is extremely diverse in terms of elevation, geology, and climate (Fig. 1A; Ode et al. 2016b). Temperate rainforests occur in the north (North Coast region), deserts and plateaus in the northeast and southeast (Deserts and Modoc Plateau region), and Mediterranean climates in coastal regions (Chaparral and South Coast regions). The Central Valley region is largely agricultural and receives water from multiple rivers that

drain a large mountainous area in the east-central region of the state (Sierra Nevada region). Urban development is concentrated in coastal areas in the central (San Francisco Bay Area, Chaparral region) and southern (Los Angeles, San Diego metropolitan area, South Coast) regions of the state. Developed lands (i.e., low- to high-density urban areas) have increased in California by 38% from 1973 to 2000 (Sleeter et al. 2011). 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).

The predictive model was developed with land-use data, national stream hydrography layers, and biological assessment data. Our predictive model followed the DPSIR framework where the general assumption was human-caused alterations to the landscape influence water quality, which in turn influences biotic integrity (Fig. 2; Smeets and Weterings 1999). The National Hydrography Dataset Plus (NHDPlus; McKay et al. 2012) was used to identify stream segments in California for modeling biological integrity. The NHDPlus is a surface water framework that maps drainage networks

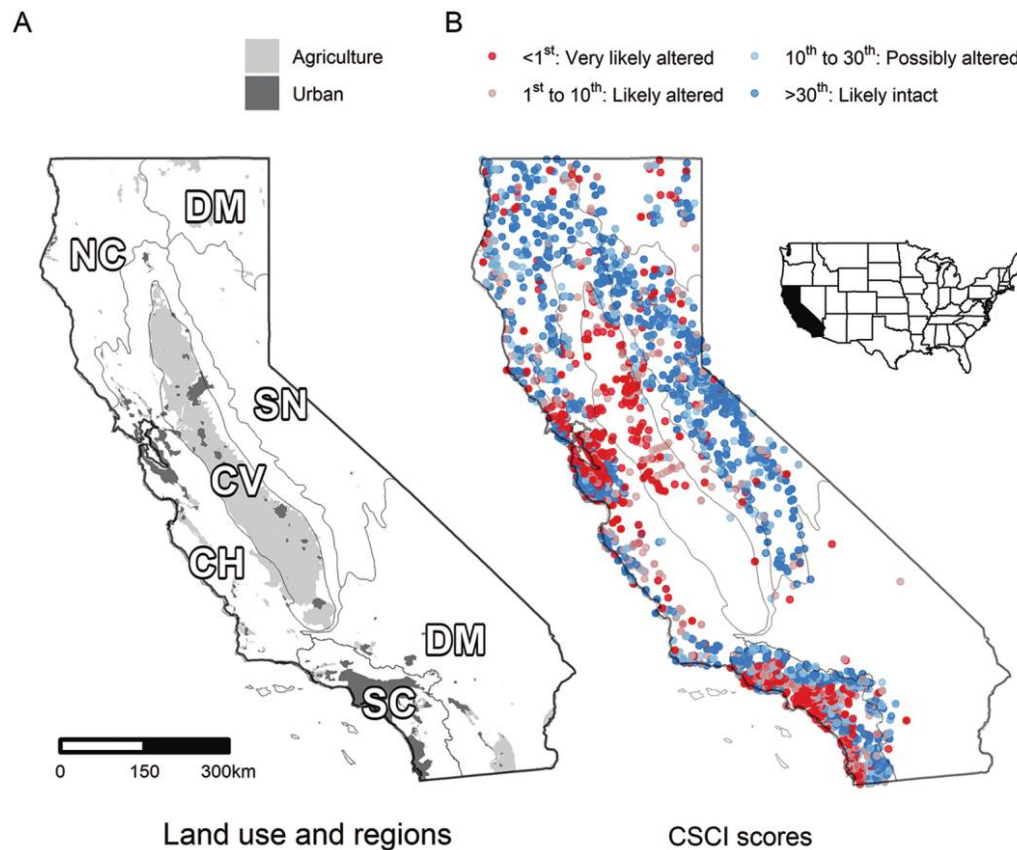


Figure 1. Map of California showing areas with extensive urban and agricultural land use (A) and distribution of observed stream CSCI scores (B). Cover of urban and agricultural land use in stream watersheds was used to develop a model to predict the range of bioassessment scores expected at a stream segment. Breakpoints for defining classes of CSCI scores are the 1st, 10th, and 30th percentile of scores observed at least-disturbed reference sites throughout the state. Altered and intact refer to biological condition (Mazor et al. 2016). Grey lines delineate 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.

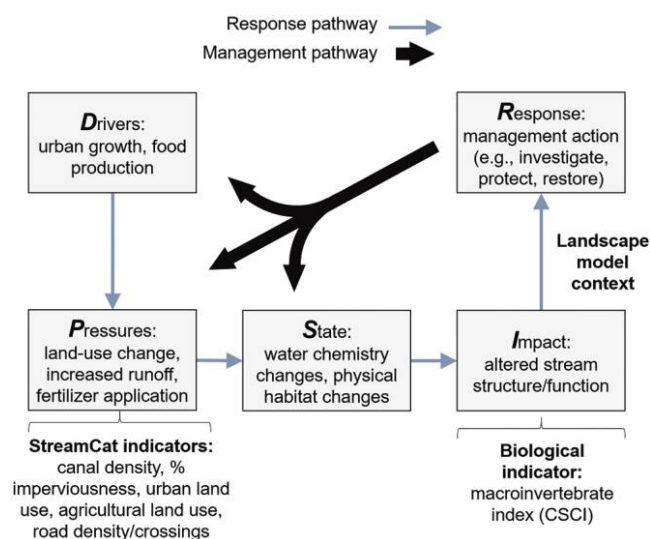


Figure 2. Response and management pathways as conceptualized in the Driver-Pressure-Stress-Impact-Response (DPSIR) framework (Smeets and Weterings 1999). The predictive model was developed to quantify relationships between pressures and impacts and inform potential responses. Landscape predictors provided in StreamCat (Hill et al. 2017) were used to describe pressures from urban and agricultural development that could alter macroinvertebrate assemblages in streams by modifying physical and chemical habitat. Biological impact was measured with the CSCI (Mazor et al. 2016) and then evaluated relative to ranges of CSCI scores that were predicted at each site by the 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.

and associated features (e.g., streams, lakes, canals, etc.) in the United States. Stream segments designated in the NHDPlus were used as discrete spatial units for modeling biological integrity. Here and throughout, segment is defined based on NHDPlus flowlines. Hydrography data were combined with landscape metrics available from the StreamCat dataset (Hill et al. 2016), which provided estimates of land use within the riparian zone (i.e., a 100-m buffer on each side of the stream segment), the local catchment (i.e., nearby landscape flowing directly into the immediate stream segment, excluding upstream segments), and the entire upstream watershed for each NHDPlus segment. Many of the metrics in StreamCat were derived from the 2006 National Land Cover Database (Fry et al. 2011).

The California Stream Condition Index (CSCI; Mazor et al. 2016) was used as a measure of biological condition in California streams. The CSCI is based on a comparison of the taxa and metrics observed at a site to those expected under least disturbed reference conditions (Stoddard et al. 2006). Expected values at a site are derived from models that estimate the likely macroinvertebrate assemblage in

relation to factors that naturally influence biology (e.g., watershed size, elevation, climate, etc.; Moss et al. 1987, Cao et al. 2007). The index score at a site can vary from 0 to ~1.4, with values near 1 indicating less deviation from the reference state. Because the index was developed to minimize the influence of natural gradients on index scores, the scores have consistent meaning across the state (Mazor et al. 2016). A CSCI threshold of 0.79, the 10th percentile of scores from all reference calibration sites for the original index, has been proposed as a threshold below which a site does not meet designated biological uses (SDRWQB 2016). As described below, the expected CSCI scores obtained from the predictive model were compared to this threshold to identify different constraint classes.

Benthic macroinvertebrate data were used to calculate 6270 individual CSCI scores at ~3400 unique sites between 2000 and 2016 (Fig. 1B). We aggregated data collected under >20 federal, state, and regional bioassessment programs (Mazor 2015, Rehn 2015, Ode et al. 2016a). 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), particularly in more arid southern California. Most regions and stream-types where perennial wadeable streams are located were represented in the calibration data set because these programs are so spatially extensive.

We collected field samples during base flow conditions typically between May and July following methods in Ode et al. (2016a). Bioassessment sites were snapped to the closest NHDPlus stream segment in ArcGIS (ESRI ArcGIS v10.5, Redlands, California). In cases where multiple sites were located on the same segment, we selected the most downstream site 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 sampling dates so that 1 CSCI score was matched to a site. This procedure created a final dataset of 2620 unique field observations that we used to calibrate and validate the model.

Building and validating the predicted CSCI score model

We modeled expected CSCI scores based on estimates of canal/ditch density, imperviousness, road density/crossings, and urban and agricultural land use for each stream segment (Table 1, Fig. S1). We used StreamCat as the only source for predictor variables to maintain consistency in methods and linkages to NHDPlus flowlines (Hill et al. 2016). Preliminary analyses indicated that these variables produced a predictive model with comparable performance to a 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. We used landscape condition variables instead of in-stream data because we specifically wanted to

Table 1. Land-use variables used to develop the predictive model of stream bioassessment scores. All variables were obtained from StreamCat (Hill et al. 2016) and applied to stream segments in the National Hydrography Dataset Plus (NHDPlus; McKay et al. 2012). The measurement scales for each variable are at the riparian (100 m buffer), local catchment, 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 local 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. CanalDens = canal density, PctImp = % imperviousness, TotUrb = total urban, TotAg = total agriculture, RdDens = road density, RdCrs = road crossings, 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/km ²
PctImp2006	Cat, Ws, Cat + Rp100, Ws + Rp100	Mean imperviousness of anthropogenic surfaces (Fry et al. 2011)	%
TotUrb2011	Cat, Ws, Cat + Rp100, Ws + Rp100	Total urban land use as sum of developed open, low, medium, and high intensity (Fry et al. 2011)	%
TotAg2011	Cat, Ws, Cat + Rp100, Ws + Rp100	Total agricultural land use as sum of hay and crops (Fry et al. 2011)	%
RdDens	Cat, Ws, Cat + Rp100, Ws + Rp100	Density of roads (2010 Census Tiger Lines)	km/km ²
RdCrs	Cat, Ws	Density of roads-stream intersections (2010 Census Tiger Lines-NHD stream lines)	crossings/km ²

quantify how biological condition was related to these types of landscape alterations, which can be challenging to manage. We also did not use data on presence or absence of in-channel modifications because landscape predictors were more broadly inclusive of the problem (e.g., modified channels are often but not always constrained, and constrained channels are not always modified). Overall, we developed a correlative model by design, which was not intended to identify specific mechanisms of biological alteration. We assumed that the importance of in-stream factors to biological condition scores could be assessed with follow-up analyses as desired or needed.

We used quantile regression forests (QRF) to estimate ranges of likely CSCI scores in different landscapes (Meinshausen 2006, 2017). Random forests modeling is an ensemble learning approach to predictive modeling that aggregates information from a large number of regression trees and have been used extensively in bioassessment applications (Carlisle et al. 2009, Chen et al. 2014, Mazor et al. 2016, Fox et al. 2017). 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, Hastie et al. 2009). Quantile models, such as QRF, evaluate the range of values of the response variable that are expected, in contrast to conventional models that provide only an estimate of the mean response (Cade and Noon 2003). This modeling approach can estimate a lower and an upper limit of likely scores that might be expected at a site given its surrounding land use, which can be therefore used to identify sites where that range includes management targets. We used a statewide, validated QRF to

predict CSCI scores in each stream segment where predictors were available at 5% increments (i.e., 5th, 10th, etc.) from the 5th to 95th percentile of expectations. For example, the 50th percentile prediction was the most likely score for a stream segment given observed values of landscape variables, whereas lower (e.g., 5th percentile) and upper (95th percentile) conditional quantiles identified the 90% (range) of scores likely to occur at a site. We used the *quantregForest* package with default settings to develop the model (Meinshausen 2017).

We stratified sample data to ensure sufficient representation of landscape gradients across major regions in the state (Fig. 1). We used data from 75% of the segments with observed CSCI scores to calibrate the model, and we ensured data were representative of the full range of landscape conditions by randomly drawing segments from 4 quartiles (strata) of the full data defined by increasing watershed imperviousness relative to each region ($n = 1965$ segments). We used the remaining 25% of sites for model validation ($n = 655$). Where replicate samples were available at a single site, we selected 1 sample at random for both calibration and validation purposes. We used out-of-bag estimates for the calibration data to prevent bias and over-fitting. Out-of-bag predictions are based on the subsets of trees that are excluded in random forest models during calibration (Breiman 2001).

We evaluated the model for both predictive ability and bias, both of which may vary depending on land-use setting. We compared differences between observed CSCI scores and median predictions at the same locations (Pearson correlations and root mean square errors [RMSE]) to evaluate performance for both the entire statewide dataset and each

major region. High correlation coefficients and low RMSE values indicate good predictive ability. We assessed potential bias by regressing observed scores on predicted scores and assessing if intercepts and slopes differed from 0 and 1, respectively.

Statewide application of the landscape model

We applied the model to 138,716 stream segments statewide to estimate the extent of streams in 1 of 4 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. We assigned sites to constraint classes by comparing a CSCI threshold representing a management goal with the predicted range or predicted median score at a segment (Fig. 3). We used a CSCI threshold of 0.79 following previous examples (Mazor et al. 2016, SDRWQB 2016) and a lower and upper bound estimated as the 10th and 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 (Fig. 3C). We assigned the remaining sites to possibly unconstrained or possibly constrained classes depending on whether the median expectation was above or below the threshold, respectively (Table 2).

We evaluated the influence of the key decision points on the extent of segment classifications created by the model. Stream segment classifications depend on the percentile range of score expectations (or certainty) from the landscape model (Fig. 3B) and the CSCI threshold for evaluating the overlap extent (Fig. 3C). With respect to the certainty range, these bounds do not describe statistical certainty in the traditional sense (e.g., prediction 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. We evaluated 8 different ranges of values for the score expectations from wide to increasingly narrow at 5% intervals (i.e., 5th–95th, 10th–90th, . . . , 45th–55th). We also evaluated different CSCI

thresholds of 0.63, 0.79, and 0.92, which correspond to the 1st, 10th, and 30th percentile of scores at reference calibration sites used to develop the CSCI (Fig. 1B; Mazor et al. 2016). We estimated the percentage of stream segments in each class statewide and by major regions based on each of the 24 scenarios (prediction interval by threshold combinations). We expected differences among regions based on differences in land use, although some of the results can be assumed (e.g., increasing CSCI thresholds causes more sites to be classified as constrained).

We developed a categorization scheme to assess how observed CSCI scores compared with the range of CSCI scores predicted from the model (Fig. 3D). This post-hoc classification was used 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 model's predictions differently than those that have scores within the prediction intervals. Sites with observed scores above the upper prediction limit (e.g., above the 90th percentile of predicted scores) were considered over-scoring, and sites with scores below the lower limit (e.g., 10th percentile) were considered under-scoring. Otherwise, the site was considered to be in the condition expected given its landscape setting.

RESULTS

Model performance

Model performance varied across regions (Table 3, Fig. S2). For the statewide 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 prediction bias of median scores. 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, Figs S2, S3). Performance for the Chaparral and South Coast regions was similar to that for the statewide dataset for both the calibration ($r = 0.71$,

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 both the cutoff range of likely scores predicted by the model and a chosen threshold for the objective.

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

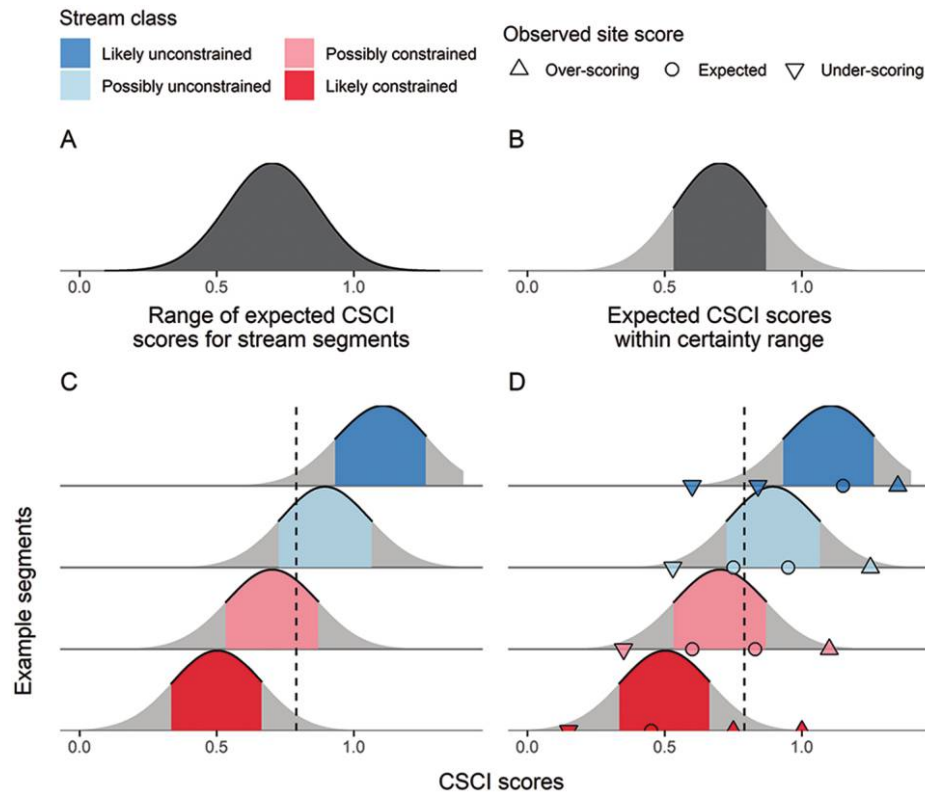


Figure 3. Application of the landscape model to identify site expectations and assess bioassessment performance for 16 example sites (points in D). A range of CSCI scores is predicted from the model (A) and the lower and upper prediction limits are used to define a certainty range of expected CSCI scores (B). Overlap of the certainty range at each segment with a chosen CSCI threshold (C; stream segment class by CSCI threshold) 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 4 stream classes (D; observed CSCI scores by stream class).

0.75, RMSE = 0.19, 0.16, respectively) and validation ($r = 0.74, 0.72, RMSE = 0.19, 0.17$, respectively) datasets. Model predictions for the Central Valley, Desert/Modoc, and North Coast regions had worse performance compared with the statewide data, with correlations of $\sim 0.66, 0.50$, and 0.55 (RMSE = 0.15, 0.20, 0.17) between observed and predicted values for the calibration dataset and $0.49, 0.55$, and 0.55 (RMSE = 0.19, 0.17, 0.16) for the validation dataset. Model performance was weakest for the Sierra Nevada region (calibration $r = 0.45$, RMSE = 0.16, validation $r = 0.21$, RMSE = 0.17), 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 (Fig. S2).

Statewide patterns in stream constraints

Statewide, spatial patterns in the predicted limits of biological integrity were similar to patterns in land use (Fig. 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, both constrained and unconstrained streams were found in every region (Fig. 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). The correspondence between observed scores and predicted ranges from the model further indicates minimal bias in the calibration data. 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 be evidence of a slight bias of over-predicting or lack of model precision in this region. Over-scoring sites were slightly more common in the South

Table 3. Performance of the predictive model as measured with calibration (Cal) and validation (Val) datasets in predicting CSCI scores. The statewide dataset (Fig. 4) and individual regions of California (Fig. 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 square errors (RMSE), intercepts, and slopes are for comparisons of predicted and observed values were used to evaluate model performance. All correlations, intercepts, and slopes were 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

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 (Fig. 5). As expected, narrowing the quantile range (from 5th-95th to 45th-55th) shifted a number of streams from the possible to likely category, and changing the CSCI threshold from relaxed to more conservative (from 0.63 to 0.92) increased the number of streams classified as possibly or likely constrained. 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 based on a high CSCI threshold with the narrowest range of predictions, whereas <1% of segments were in this category based on 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 based on a low CSCI threshold and narrow prediction interval.

DISCUSSION

Managing for biological integrity requires the use of 1) assessment tools that can accurately evaluate condition and 2) tools that can estimate the range of attainable conditions given the landscape settings. We developed our predictive model 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 and inform management actions within a landscape context. The model can inform how landscape conditions can constrain 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 developed predictions of biological condition from geospatial data. For example, Hill et al. (2017) developed a national model to predict stream site condition that correctly classified sites as poor, fair, or good at ~75% of locations, depending on region. For continuous predictions of biointegrity index scores, Carlisle et al. (2009) 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 assignment to condition class at ~85% of sites, which was similar in performance to models that were developed to predict classes. Our model had similar performance to those developed in other studies based on a comparison of the percentage of correctly classified sites that were above the 10th percentile of reference site scores (0.79) and observed site scores. The 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 the most likely bioassessment scores at unsampled locations, but rather to quantify the

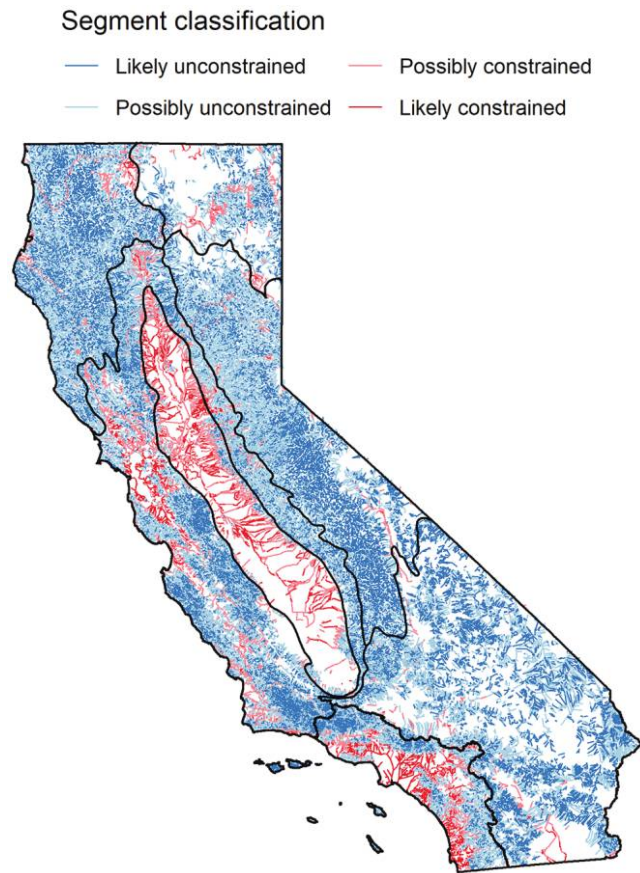


Figure 4. Statewide application of the landscape model showing the stream segment classifications. Major regional boundaries are also shown (see Fig. 1).

range of scores likely to occur given specific land-use settings and, thus, help identify likely constraints on biological condition. Interpretation of differences in the predictive accuracies of different models should consider the goals that informed the development of each model.

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

We applied the statewide model in a regional context through collaboration with a stakeholder group from the San Gabriel River watershed (Los Angeles County, California; Fig. S4). The statewide model provides a range of expected scores for a given stream segment. Comparison of observed index scores derived from biological samples with model predictions can provide a basis for how managers prioritize sites (Fig. S5). For example, managers may prioritize sites for protection if the observed scores are above the modeled predictions and for restoration if the observed scores are below the modeled predictions. Alternatively, a site scoring within the prediction interval in an unconstrained segment could be a higher priority for management actions (e.g., restoration or protection) than a similarly scor-

ing site in a constrained segment. The latter site may require more resources to achieve comparable changes in biotic condition. The lower San Gabriel watershed is a useful case study because it is heavily urbanized with many modified channels (Fig. S4B), and managers require prioritization tools to identify where efforts should be focused among many sites that vary in landscape and land use setting.

The stakeholder group had 3 management priorities for individual sites (Table S1, Fig. S6). First, investigate sites by either continuing to monitor or review supplementary data (e.g., field visits, review aerial imagery). Second, protect sites by recommending additional scrutiny of any proposed development or projects that could affect a site. Third, restore sites by targeting action for either causal assessment or restoration activity at a site.

These priority actions were first identified independently without knowledge of model prediction condition classes. The priority actions were then assigned to each site 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 3 actions to sites in likely unconstrained segments where CSCI scores were either over- or under-scoring or at sites that were possibly unconstrained but the observed CSCI scores were below the biological threshold (Table S1, dotted line in Fig. S6). Constrained sites were given lower priority overall or restoration actions were recommended as a lower priority despite low CSCI scores. Stakeholders also identified continuing current practices (e.g., routine monitoring, neither of the above actions) as a necessary action 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 model is primarily a screening tool to help identify patterns among monitoring sites where more intensive analyses may be appropriate. This application was tested through engagement with our local stakeholder group. Rather than identifying individual sites in need of specific management actions, the group used the 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 by broad categories of protect, restore, and investigate was based on group discussions meant 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, Kroll et al. 2019). However, the recommended actions have relevance only in the context of the interests of the San Gabriel Regional Monitoring Program. Localized applications of the statewide model

Table 4. Summary of stream length for each stream class statewide and within major regions of California (Figs 1, 4). Lengths are in km 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 percentiles of scores predicted by the model. CV = Central Valley, CH = Chaparral, DM = Deserts and Modoc Plateau, NC = North Coast, SN = Sierra Nevada, SC = South Coast.

Region	Constrained		Unconstrained	
	Likely	Possibly	Possibly	Likely
Statewide	8150 (4)	24,735 (11)	101,591 (46)	85,317 (39)
CV	3356 (22)	8010 (52)	3202 (21)	951 (6)
CH	1642 (3)	7840 (13)	30,693 (50)	21,206 (35)
DM	255 (0)	3395 (6)	27,194 (47)	26,479 (46)
NC	108 (0)	1442 (5)	14,152 (49)	13,286 (46)
SN	20 (0)	1067 (3)	18,228 (48)	19,032 (50)
SC	2770 (15)	2981 (16)	8122 (45)	4363 (24)

must engage stakeholders in a similar process to develop recommendations that are specific to regional needs at the watershed scale (Brody 2003, Reed 2008).

Engagement with the stakeholder group was facilitated through creation of an interactive and online application, SCAPE (Fig. S7; Beck 2018b). The SCAPE application can be used to select and visualize management priorities for all monitoring sites in the San Gabriel watershed (Fig. S8) and was also critical for demonstrating how results from the statewide model could be used at a regional scale. This application allowed the stakeholder group 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., Fig. S5A), stakeholders had limited information to prioritize among sites (i.e., no context for scores; Fig. S5B).

The SCAPE application also demonstrated core concepts of the model and allowed stakeholders to explore the key decision points that affected model output. Specifically, drop-down menus and sliders allowed users to change prediction intervals 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. Our results (Fig. 5) also demonstrate the broader implications of how key decision points can affect model results at regional and statewide scales. These results and the functionality provided by SCAPE demonstrate flexibility of the landscape model and the considerations of appropriate decision points that could 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

Table 5. Summary of CSCI scores by relative expectations for each stream class statewide and within each major region of California (Figs 1, 4). Average CSCI scores (standard deviation) and counts (%) 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 upper prediction interval for a segment, expected if within the lower and upper prediction interval, or under-scoring if below the lower prediction interval. CV = Central Valley, CH = Chaparral, DM = Deserts and Modoc Plateau, NC = North Coast, SN = Sierra Nevada, SC = South Coast.

Region	Under-scoring		Expected		Over-scoring	
	CSCI	<i>n</i> (%)	CSCI	<i>n</i> (%)	CSCI	<i>n</i> (%)
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)

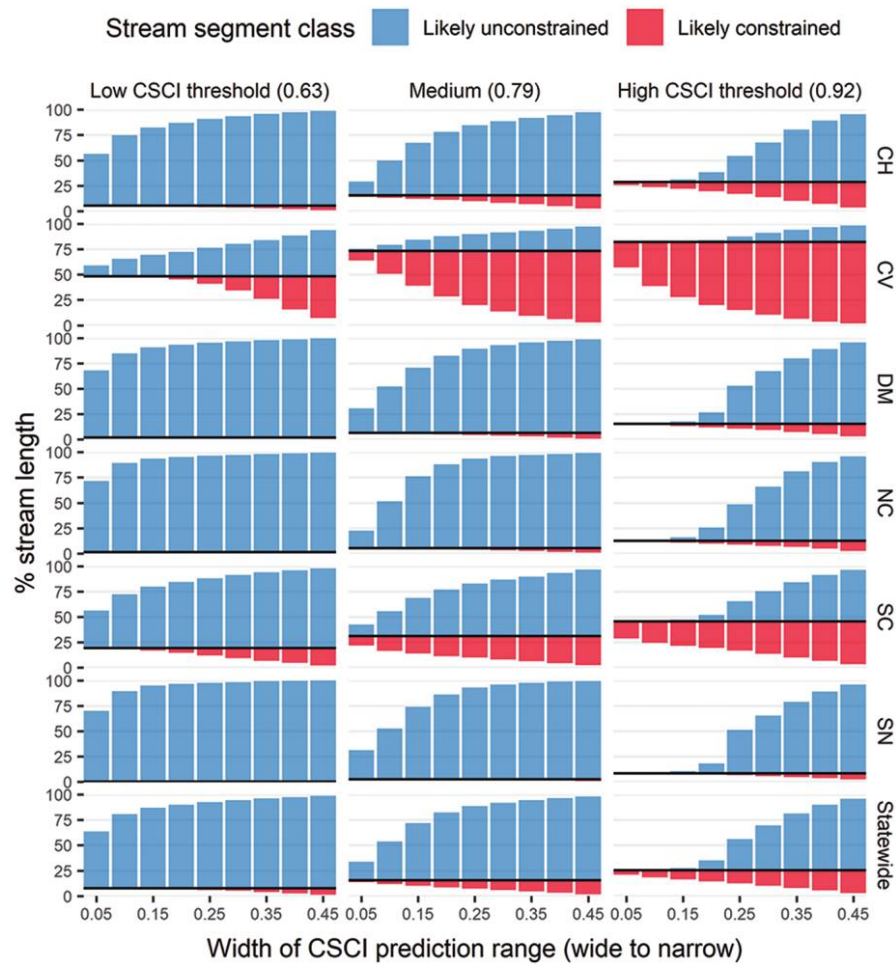


Figure 5. Differences in stream segment class assignments between different scenarios used to define biological constraints by region and statewide. Twenty-seven scenarios were tested that evaluated different combinations of prediction interval in the CSCI predictions (9 scenarios from wide- to narrow-prediction ranges as identified by the tail cutoff for the expected quantiles) and potential CSCI thresholds (3 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 from 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.

chosen prediction interval defining the lower and upper expectation of biological integrity determines how constraint classes are assigned to stream segments in 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 in general.

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 their physical and chemical condition concurrent with biological assessment scores. Our model can place observed biological condition 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 help 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 elements 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 (Davies and Jackson 2006) may be needed. However, the model is not intended, nor is it 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 model are also possible by identifying where additional restoration, monitoring, or protection may have the most benefit. For example, these types of 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). Results from our 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). In many cases, including California, bioassessment programs are sufficiently large in spatial extent that they 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). Our 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 (FDEP 2011, USEPA 2013, MBI 2016). Although our results generally support the link between degraded 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 our model, a constrained channel may or may not be engineered (see supplement for Tecolote Creek example, Fig. 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 water-

shed is largely undeveloped. For example, Stein et al. (2013) observed reference-like bioassessment index scores in armored reaches within national forest lands in southern California. A classification framework for biological constraints based only on channel modification would provide incomplete and potentially misleading information on stream condition and likelihood for needing management. Ideally, context for evaluating biological condition from a 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. Our modeling approach could be applied to other bioassessment methods, such as a multi-metric index (the most common bioassessment approach within the United States; Buss et al. 2014), O/E assessments (Moss et al. 1987), biological condition gradients (Davies and Jackson 2006), or with other biological endpoints (e.g., fish or diatoms). More importantly, our use of national geospatial datasets (i.e., NHDPlus, StreamCat; McKay et al. 2012, Hill et al. 2016) means that these methods could be applied across the United States. National-scale bioassessment indices have been developed, and the modeling approach described here could be applied at a national scale to estimate constraints on biological condition that complement recent work that predicted probable biological conditions with the National Rivers and Streams Assessment (Hill et al. 2017). 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).

Extension of these types of 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) 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 United States. In their paper, Hill et al. (2017) 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 can help establish priorities at the local-scale, and a similar process could be used for applying different 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 other stressor gradients are needed. For example, our results suggest that streams in the North Coast and Sierra Nevada regions are largely unconstrained, but the observed and predicted scores had the lowest correlation coefficients among all regions (calibration $r = 0.45$, validation $r = 0.21$). The dominant stressors likely to affect stream condition in these regions originate from sources that are less common in developed landscapes, such as silviculture, cannabis cultivation, or livestock grazing. Our current model does not adequately capture these impacts. Moreover, poor model predictions are compounded by low sensitivity of the CSCI to some stressor gradients in these regions (Mazor et al. 2016). 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). 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 by applying 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, Chessman 2014, Waite et al. 2014). 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 alter benthic macroinvertebrate assemblages for a decade or more after harvesting activities have stopped (Stone and Wallace 1998, Quinn and Wright-Stow 2008). Channel and riparian modifications through historical splash damming or railroad tie driving can also have effects lasting several decades (Young et al. 1994, Miller 2010, Wohl 2019). 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 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.

Our model identifies potential 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). 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). More comprehensive assessments at individual sites may be needed to diagnose the immediate causes of degraded condition.

Finally, there are a few concerns regarding using a modeling approach for bioassessment based on the NHDPlus flowlines as a base layer. We applied our model to the entire network of the NHDPlus represented in StreamCat, which included a large number of intermittent or ephemeral streams, as well as non-wadeable rivers. The application of model results to these stream-types is open to question and would be valid only to the degree that the CSCI and its response to landscape disturbance describe biological integrity in these locations. 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 the range of biological index scores that could be expected at a stream segment given the amount and type of landscape development surrounding a stream. Although random forest models have been increasingly used in bioassessment applications, our approach is the first to use quantile models to estimate the range of biological condition scores likely to be observed at a site. Additional work could build on this initial approach to apply such models in different locations, apply them to alternative biological response endpoints, or explore different predictors that represent 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 if the relative importance of predictors varies depending on the quantiles being

predicted. 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.

This modeling approach can be used to characterize the extent of biologically constrained channels in developed landscapes, and it provides a tool to determine how managers can best prioritize resources for stream management by identifying what landscape factors might constrain the biological condition of each stream 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 given current land use conditions. Such integration of information 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.

ACKNOWLEDGMENTS

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.

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. We acknowledge feedback received from a Science Advisory Panel that included Clifford Dahm, Chuck Hawkins, Ken Reckhow, Paul Stacey, Jan Stevenson, and Lester Yuan. 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.

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

LITERATURE CITED

- 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. Zenodo. (Available from: <http://doi.org/10.5281/zenodo.1218121>)
- Beck, M. W. 2018b. Constrained streams for biological integrity in California. Knowledge Network for Biocomplexity. (Available from: <https://knb.ecoinformatics.org/view/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. Follstad-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.
- 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., 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.
- FDEP (Florida Department of Environmental Protection). 2011. Development of aquatic life use support attainment thresholds for Florida's stream condition index and lake vegetation index. Technical report DEP-SAS-003/11. Florida Department of Environmental Protection Standards and Assessment Section and Bureau of Assessment and Restoration Support, Tallahassee, Florida. (Available from: <https://floridadep.gov/sites/default/files/attainment-thresholds-sci-and-lvi.pdf>)
- 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.
- 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*. 2nd edition. Springer, New York, 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. Mazon, 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. *Internationale Vereinigung für Theoretische und Angewandte Limnologie: Verhandlungen* 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, Campaign, Illinois*.
- Kennedy, 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.
- Kroll, S. A., R. J. Horwitz, D. H. Keller, B. W. Sweeney, J. K. Jackson, and L. B. Perez. 2019. Large-scale protection and restoration programs aimed at protecting stream ecosystem integrity: The role of science-based goal-setting, monitoring, and data management. *Freshwater Science* 38:23–39.
- 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. Technical Report 844. Southern California Coastal Water Research Project Authority, Costa Mesa, California. (Available from: http://ftp.sccwrp.org/pub/download/DOCUMENTS/TechnicalReports/844_SoCalStrmAssess.pdf)
- 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.
- MBI (Midwest Biodiversity Institute). 2016. Identification of predictive habitat attributes for Minnesota streams to support tiered aquatic life uses. Technical Report. Midwest Biodiversity Institute, Columbus, Ohio. (Available from: <https://www.pca.state.mn.us/sites/default/files/wq-s6-39.pdf>)
- 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*. (Available from: <https://cran.r-project.org/web/packages/quantregForest/quantregForest.pdf>)
- Miller R. R. 2010. Is the past present? Historical splash-dam mapping and stream disturbance detection in the Oregon Coastal Province. MS Thesis. Oregon State University, Corvallis, Oregon.
- 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. (Available from: https://www.nemi.gov/protocols/protocol_summary/655/)
- 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, U.S. Environmental Protection Agency, Sacramento, California. (Available from: https://www.waterboards.ca.gov/water_issues/programs/swamp/docs/reports/psa_smmry_rpt.pdf)
- 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.
- 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. (Available from: https://www.waterboards.ca.gov/water_issues/programs/swamp/bioassessment/docs/psa_memo_121015.pdf)
- 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. 1st edition. John Wiley & Sons, Hoboken, New Jersey.
- SDRWQB (San Diego Regional Water Quality Board). 2016. Clean Water Act sections 305(B) and 303(D). California Environmental Protection Agency, Sacramento, California.
- 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. Souldard, 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. 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.
- USEPA (US Environmental Protection Agency). 2013. Technical support document for EPA's action on the state of Oregon's Revised water quality standards for the west division main canal. USEPA Region 10, Seattle, Washington. (Available from: https://www3.epa.gov/region10/pdf/water/wqs/or/epa_approval_revised_oregon_wqs_08042015.pdf)
- USGS (US Geological Survey). 1999. The quality of our nation's waters: Nutrients and pesticides. Circular 1225. US Geological Survey, 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.
- Wohl, E. 2019. Forgotten legacies: Understanding and mitigating historical human alterations of river corridors. *Water Resources Research* 55:5181–5201.
- Young, M. K., D. Haire, and M. A. Bozek. 1994. The effect and extent of railroad tie drives in streams of southeastern Wyoming. *Western Journal of Applied Forestry* 9:125–130.