Prioritizing management goals for stream biological integrity within the context of landscape constraints

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 D. 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))

# Abstract

Many streams are failing to achieve desired biological condition and require management decisions to restore designated uses. Some management goals may be impractical with limited resources, particularly in streams where large-scale changes on the landscape (e.g., urbanization) impose constraints on the upper limit of biological integrity. A statewide landscape model was developed that sets reasonable expectations for observed conditions within landscape constraints to prioritize management actions. The model provides a context for what is likely to be achieved at a given site independent of an actual bioassessment score. With this approach, sites can be ranked as over- or under-scoring relative to an expectation that is typical for the observed level of landscape alteration. We developed a visualization tool to compare observed bioassessment scores with modelled expectations to rapidly identify reaches that were scoring better or worse than expected. Using this tool, a group of regulators, dischargers, stormwater agencies, and environmental advocates from the San Gabriel River watershed (Los Angeles County, California) identified regions in the watershed with consistent patterns in bioassessment scores relative to expectations. Based on these patterns, they prioritized different management actions for each region. Sites in both developed and undeveloped areas that scored below expectations were prioritized for restoration; in contrast, restoration was not a priority at developed sites where scores were low but within expected ranges. Sites scoring better than expected were prioritized for enhanced protection, as well as additional monitoring. Interactive tools that connect landscape models with observed data can help set management goals appropriate for stakeholder needs and likely constraints on biological integrity. These tools can easily be applied to other locations where biological data are used to assess environmental condition.

# Introduction

* Degraded biological condition in streams can occur from individual or multiple stressors acting at different scales (Novotny et al. [2005](#ref-Novotny05); Townsend, Uhlmann, and Matthaei [2008](#ref-Townsend08); Leps et al. [2015](#ref-Leps15)).
* In many urban and agricultural areas the majority of stream miles are not healthy and in need of some level of management (cite SWAMP, SMC, NRSA)
* Unfortunately, there are not sufficient resources to restore all streams to reference conditions, nor is it practical (e.g., varying costs and challenges of urban stream restoration (Kenney et al. [2012](#ref-Kenney12); Shoredits and Clayton [2013](#ref-Shoredits13)))
* Need a way to comprehensively evaluate streams across large spatial scales for “management potential”. Biological filters act at different scales (Poff [1997](#ref-Poff97)) and we can use this information to describe an expectation for prioritization that is scale-specific. This allows establishment of reasonable expectations and prioritization of limited resources most effectively. Landscape-level constraints are particularly relevant for macroinvertebrate communities in streams (Sponseller, Benfield, and Valett [2001](#ref-Sponseller01))
* Once these large spatial scales are understood, sites can be prioritized by local managers to ensure resources are wisely allocated.
* Goal: demonstrate application of a landscape model to classify and prioritize stream monitoring sites using estimated constraints on biological integrity.
  + Build on knowledge and relationships developed through existing monitoring programs and apply that in a predictive manner across entire landscapes to inform decisions
  + Statewide application of the model - The model provides an estimate of context for biological condition that provides an expectation of what is likely to be achieved at a given site relative to large-scale drivers of stream health. The model was developed and applied to all stream reaches in California.
  + A case study is used to demonstrate how the model can be used to classify and prioritize by watershed using guidance from a regional stakeholder group. Specific questions that were addressed through the case study. Active stakeholder involvement was critical in applying the landscape models to define a framework for decision-making because priorities varied with management objectives.

# Methods

## Study area and data sources

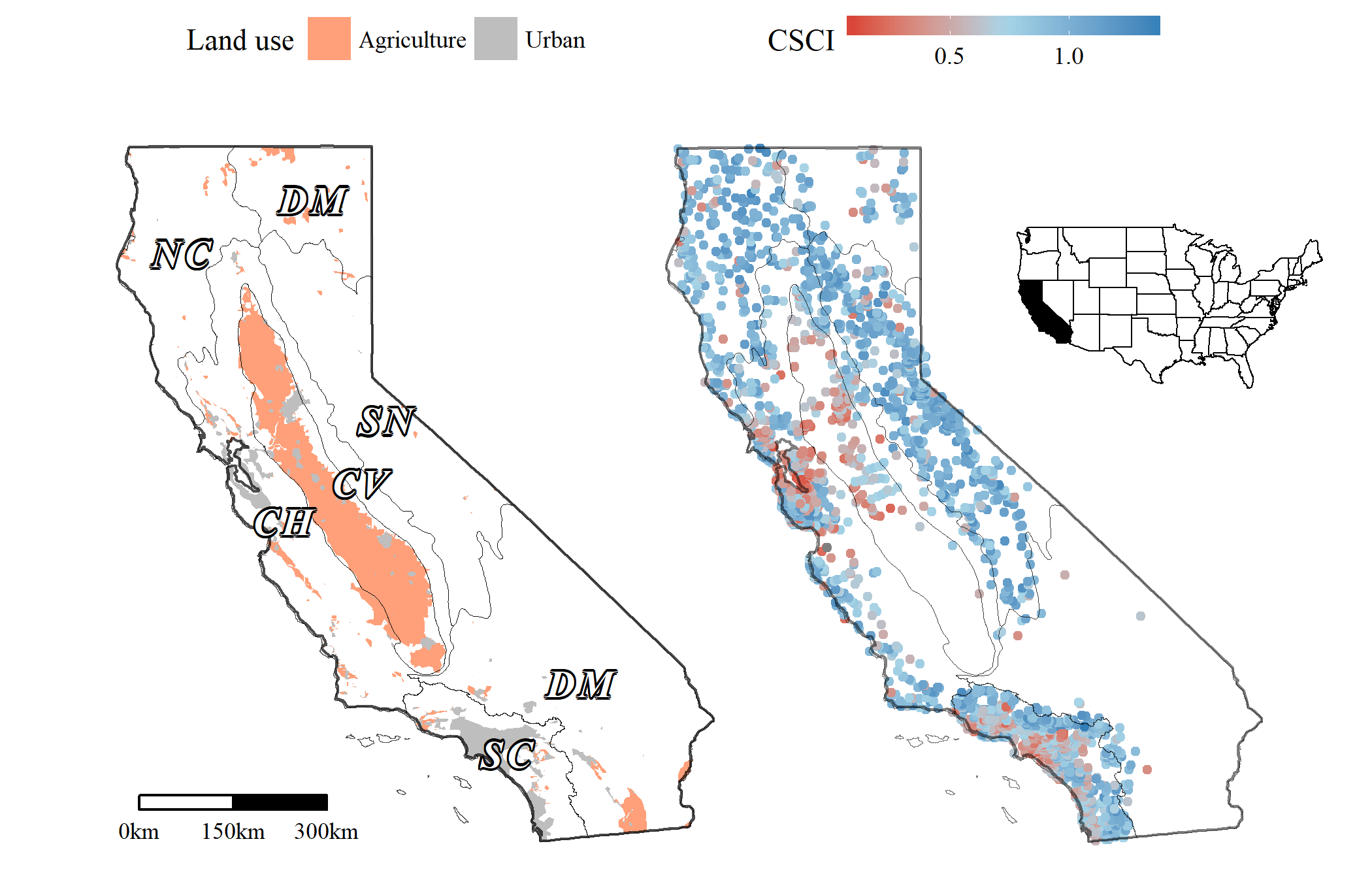


Figure 1 Urban and agricultural land use (left) and distribution of observed stream CSCI scores (right) in California. Cover of urban and agricultural land use in stream catchments was used to develop landscape models for stream reach expectations of bioassessment scores. Grey lines are ecoregions in California, CV: Central Valley, CH: Chaparral, DM: Deserts Modoc, NC: North Coast, SN: Sierra Nevada, SC: South Coast.

Landscape models were developed for California using land use data, stream hydrography, and biological assessments. California covers 424,000 km of land from latitudes 33 to 42N that includes extreme variation in altitude and climate (Figure 1). Temperate rainforests occur in the north, deserts in the northeast and southeast, and Mediterranean climates in coastal regions. California’s stream network is approximately 280,000 km in length and covers all of the major climate zones in the state. A high degree of endemism and biodiversity occurs in these streams including nearly 4000 species of vascular plants, macroinvertebrates, and vertebrates that depend on fresh water during their life history (Howard and Revenga [2000](#ref-Howard09); Howard et al. [2015](#ref-Howard15)). Approximately 30% of streams in California are perennial with the remaining as intermittent or ephemeral for portions of the year. Much of California is publicly owned and is used heavily for recreation. A large portion of the central region of the state is agricultural (i.e., Central Valley), whereas dense areas of urban development are in the southwest (i.e., Los Angeles and San Diego) and central (San Francisco Bay area) coast areas. Developed lands increased in California by 38% from 1973 to 2000 (Sleeter et al. [2011](#ref-Sleeter11)).

Stream data from the National Hydrography Dataset (NHD) (USGS (US Geological Survey) [2014](#ref-USGS14)) were used to identify reaches in California for modelling biological integrity. The NHD is a surface water framework that maps drainage networks and associated features (e.g., streams, lakes, canals, etc.) in the United States. Stream flow lines in the NHD are developed from flow accumulation models that estimate location of a stream given slope and elevation changes from existing elevation datasets. As such, flow lines in California represent both perennial, intermittent, and ephemeral streams that have wide variation in observed flow throughout the year. Stream reaches designated in the NHD were used as the discrete spatial unit for modelling biological integrity. Hydrography data were combined with landscape metrics available from the StreamCat Dataset (Hill et al. [2016](#ref-Hill16)) to estimate land use at the catchment (nearby landscape flowing directly into a stream) and the entire upstream watershed for each reach. The StreamCat Dataset was developed specifically for the NHD to leverage the topology of stream connections to estimate cumulative landscape metrics of all reaches.

The California Stream Condition Index (CSCI) (Ode et al. [2016](#ref-Ode16); Mazor et al. [2016](#ref-Mazor16)) was used as a measure of biological condition in California streams. Benthic macroinvertebrate data used to calculate CSCI scores were collected at nearly 3400 sites (6270 with repeat visits) between 2000 and 2016. Field data were collected during baseflow conditions typically between May and July following methods in Ode ([2007](#ref-Ode07)). The CSCI is a predictive index of stream health that compares the observed taxa and metrics at a site to those expected under reference conditions. Expected conditions 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. The CSCI score at a site is based on an observed-to-expected ratio of taxa and a predictive multimetric index composed of six metrics that describe the structure and function of the macroinvertebrate community. The index score at a site can vary from 0 to 1.4, with higher values indicating an observed community with less deviation from reference conditions. Because the index was developed to minimize the influence of natural gradients, the index scores have consistent meaning across the state (Reynoldson et al. [1997](#ref-Reynoldson97)). A threshold score based on a selected lower percentile of scores (e.g., 10%) at all reference sites is used to define nominally low and high scoring sites.

## Building and validating landscape models

A prediction model of the CSCI was developed to estimate likely ranges of scores associated with land use gradients. Land use as urban and agricultural was quantified for the catchment of each stream reach in California using the StreamCat database (Hill et al. [2016](#ref-Hill16)). CSCI scores were modelled using only the estimates of urban and agricultural land use as the developed portion of the landscape within each stream reach. The model was incomplete by design to describe scores only in relation to large-scale constraints on biological condition that are not easily controlled by management actions or where costs to mitigate are likely to be excessive. The remainder of the variation in scores not related to landscape constraints could be attributed to additional, unmeasured environmental variables that influence stream biointegrity. Deviation of observed scores from the model predictions were considered diagnostic of variation not related to landscape effects.

Models were developed using quantile regression forests to estimate ranges of likely CSCI scores in different landscapes (Meinshausen [2006](#ref-Meinshausen06), [2017](#ref-Meinshausen17)). Quantile models evaluate the conditional response across the range of values that are expected, such as the lower and upper percentiles of the distribution, as compared to only the mean response with conventional models (Cade and Noon [2003](#ref-Cade03)). This allows use of model predictions to describe where bioassessment targets are unlikely to be met or where streams are unlikely to be impacted by placing bounds on the range of expectations relative to landscape constraints. Random forest models also provide robust predictions by evaluating different subsets of observations from random splits of the predictor variables. The final predictions are the averaged response across several models. These models have been used extensively in bioassessment applications (Carlisle, Falcone, and Meador [2009](#ref-Carlisle09); Chen et al. [2014](#ref-Chen14); Mazor et al. [2016](#ref-Mazor16)) and can produce unbiased estimates that are relatively invariant to noisy relationships or non-normal distributions (Breiman [2001](#ref-Breiman01); Hastie, Tibshirani, and Friedman [2009](#ref-Hastie09)). Quantile regression forests were used to predict CSCI scores in each stream reach from the 5th to the 95th percentile of expectations at five percent intervals (i.e., 5th, 10th, etc.).

Calibration data for the landscape models were based on a random selection of 75% of monitoring stations with observed CSCI scores. The random selection was stratified by ecoregion (Figure 1) and relative amounts of impervious surfaces in each catchment based on percentile distributions. The stratification method was chosen to ensure sufficient representation of landscape gradients in each ecoregion. The remaining sites were used for model validation. Model performance was assessed for the statewide dataset and within each major region. Differences between observed CSCI scores and median predictions were evaluated using correlation analysis and root mean squared errors (RMSE). Regression analysis between predicted and observed scores was used to assess potential bias based on intercept and slope values differing from 0 and 1, respectively.

## San Gabriel River watershed case study

Stream reach and bioassessment data from the San Gabriel River (SGR) watershed in southern California were used to develop reach classifications, site performance categories, and management priorities from the landscape models. A strong land use gradient occurs in the SGR watershed (Figure 2). Headwaters begin in the San Gabriel mountains where the land is primarily undeveloped or protected for reacreational use, whereas the lower watershed is in a heavily urbanized region of Los Angeles County. The San Gabriel river is dammed at four locations for flood control in the upper watershed and is hydrologically connected to the Los Angeles river to the west through the Whittier Reservoir in the lower watershed. Spreading grounds are present in the middle of the watershed for groundwater recharge during high flow. Nearly all of the stream reaches in the lower half of the watershed are channelized with concrete or other reinforcements.

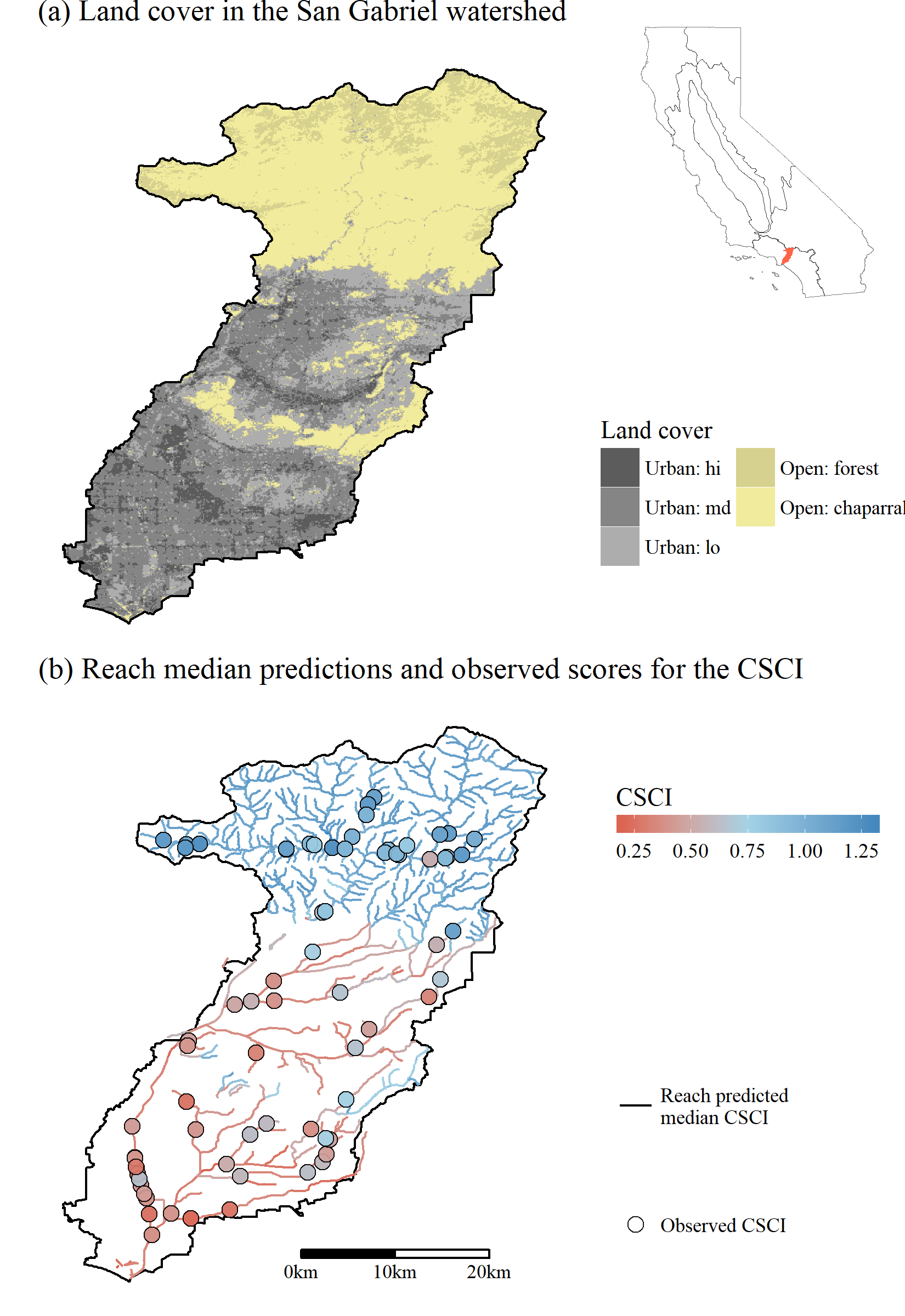


Figure 2 San Gabriel River watershed in southern California. Land cover is shown in plot (a) and the predicted median CSCI scores at each stream reach and observed CSCI scores are shown in (b). The watershed is undeveloped in the north and heavily urbanized in the south.

The SGR watershed contains a diverse group of stakeholders from local municipalities, water districts, water quality regulatory agencies, consulting groups, and non-government organizations. Collectively, the San Gabriel River Regional Monitoring Program (SGRRMP) includes stakeholders from these groups that cooperatively work to increase awareness of issues in the SGR watershed and work to improve coordination of compliance and ambient monitoring efforts. The stakeholder workgroup included individuals from the SGRRMP with interests in water supply, improvements to water quality, habitat protection or creation, and storm water permitting. Individuals were selected for participation to include diverse management interests and based on willingness to adopt tools developed from the landscape models. The stakeholder workgroup met monthly over a six-month period to discuss model applications and to refine the interpretation of results. Stakeholder involement was critical for developing an assessment framework that met the needs of all engaged parties and ensured that final products were more likely to be incorporated into formal decision-making processes.

## Reach classification, site performance, and prioritization

A framework for identifying site priorities for management actions was developed using a three-step process. First, estimates of the range of expected CSCI scores at each stream reach in relation to land use were used to define reach classifications. Second, the relationship between observed CSCI scores and the reach classifications were then used to assign a relative performance value for each monitoring site. Third, site performance categories in relation to reach classification and bioassessment targets were used to define management priorities. This framework was developed through close interaction with the regional stakeholder group to demonstrate how the landscape model can be used as a management tool given that priorities will vary by interests and location. As such, the results are provided as a guide to facilitate decision-making rather than a prescription of targeted actions to manage stream health. The entire process is shown in Figure 3 and Table 1 and is described in detail below.

Identifying site priorities began with defining a classification framework for stream reaches to identify the possible or likely extent of biological constraints. Classifications were developed using the range of CSCI expectations at a reach (Figure 3a,b) relative to a chosen threshold for the CSCI to define nominally low or high scores (Figure 3c). The reach classification was based solely on the intersection of CSCI expectations at a reach with the chosen CSCI threshold, where expectations could be below, above, or overlapping the threshold. Stream reaches with a range of CSCI score expectations entirely below the threshold were considered likely constrained, whereas those with expectations entirely above were considered likely unconstrained. Reaches with score expectations that overlapped the CSCI threshold were considered possibly constrained or possibly unconstrained, where distinction between the two was based on location of the median expectation of a reach relative to the threshold.

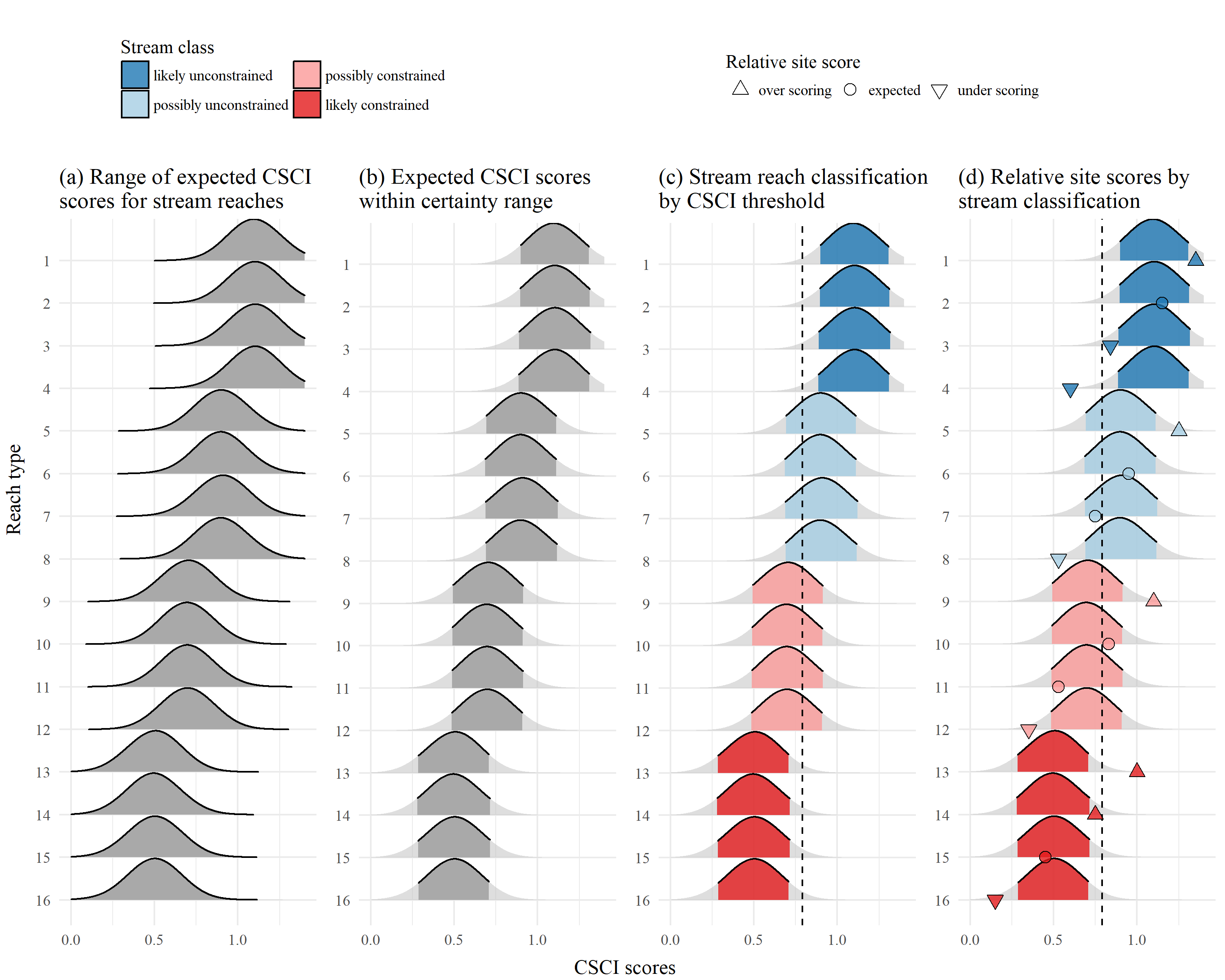


Figure 3 Application of landscape models to identify site expectations and bioassessment performance for sixteen example stream reaches. A range of CSCI scores is predicted from the model (a) and the lower and upper limits of the expectations are cut to define a certainty range for the predictions (b). Overlap of the certainty range at each reach with a chosen CSCI threshold (c) defines the stream reach classification as likely unconstrained (lu), possibly unconstrained (pu), possibly constrained (pc), and likely constrained (lc). 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).

CSCI scores from biomonitoring data were used to define relative sites scores at a sample site given the stream reach classification (Figure 3d). For each of the four reach classifications (likely constrained, possibly constrained, possibly unconstrained, and likely unconstrained), relative site scores were defined based on location to the range of expected CSCI scores. This provided a definition that can be used to understand the observed score relative to the biological context of a reach. Sites with observed scores above the upper limit of the reach expectation (e.g., above the 95th percentile of expected scores) were considered over scoring and sites below the lower limit were under scoring. Sites with CSCI scores within the range of expectations were as expected.

Categories for relative sites scores were further split if they were above or below the selected CSCI threshold. This final split was created with the intent that description of site scores relative to a defined threshold (e.g., impairment threshold or restoration target) should also be considered. Specifically, a fourth category for sites within each reach classification was added to define a site as above or below the threshold. For a likely unconstrained reach, under scoring sites below the minimum expected score were additionally defined as being above or below the CSCI threshold. Similarly, over scoring sites above the maximum expected score in a likely constrained reach were additionally defined as being below or above the CSCI threshold. For possibly constrained and possibly unconstrained reaches, sites that were as expected were additionally defined as being below or above the CSCI threshold. In total, sixteen site types were defined for the reach classifications (Table 1).

Table 1 Possible site types based on stream reach classification, relative site score, and observed CSCI score. The observed score column describes where a CSCI score is observed relative to the lower and upper percentiles (e.g., 5th and 95th) of expected scores for a reach and the chosen CSCI threshold (e.g., 10th percentile of scores at reference sites or 0.79) for nominally low or high values.

|  |  |  |  |
| --- | --- | --- | --- |
| Reach expectation | Relative site score | Observed score | Type |
| **likely unconstrained** | over scoring | 95th | 1 |
|  | expected | 5th to 95th | 2 |
|  | under scoring | 0.79 to 5th | 3 |
|  | under scoring | < 0.79 | 4 |
| **possibly unconstrained** | over scoring | 95th | 5 |
|  | expected | 0.79 to 95th | 6 |
|  | expected | 5th to 0.79 | 7 |
|  | under scoring | < 5th | 8 |
| **possibly constrained** | over scoring | 95th | 9 |
|  | expected | 0.79 to 95th | 10 |
|  | expected | 5th to 0.79 | 11 |
|  | under scoring | < 5th | 12 |
| **likely constrained** | over scoring | 0.79 | 13 |
|  | over scoring | 95th to 0.79 | 14 |
|  | expected | 5th to 95th | 15 |
|  | under scoring | < 5th | 16 |

Each site type was used to define a priority as a demonstration of how results from the landscape model can inform different stream management objectives. This final process relied exclusively on feedback from the stakeholder group. An interactive online application was created for the stakeholder group to facilitate the recommendation of management actions for each site type (Figure 4). Priorities for each site type were defined accordingly with the expectation that site types will have different meanings for prioritization given the interest (e.g., monitoring, regulation, etc.). Stakeholders were tasked with identifying their relevant priorities by ranking each site type from high to low priority using a graphical template of Table 1 for reference. A brief description of the rationale for a site priority was also requested with the feedback. The final priorities were generalized into three categories to recommend actions in addition to baseline monitoring and maintenance. The final priorities also assumed that existing information available for each site was “true” following established practices to account for uncertainty or variation between assessments. A consensus was reached for the following definitions of each action:

* Investigate: Additional monitoring or review of supplementary data (e.g., field visits, review aerial imagery);
* Protect: Additional scrutiny of proposed development and/or projects;
* Restore: Targeted action for causal assessment and/or restoration funds.

Each site type was ranked as high, medium, or low priority for each action. No priority assigned to an action for a stream type was indication that baseline monitoring and maintenance was sufficient for a site type.

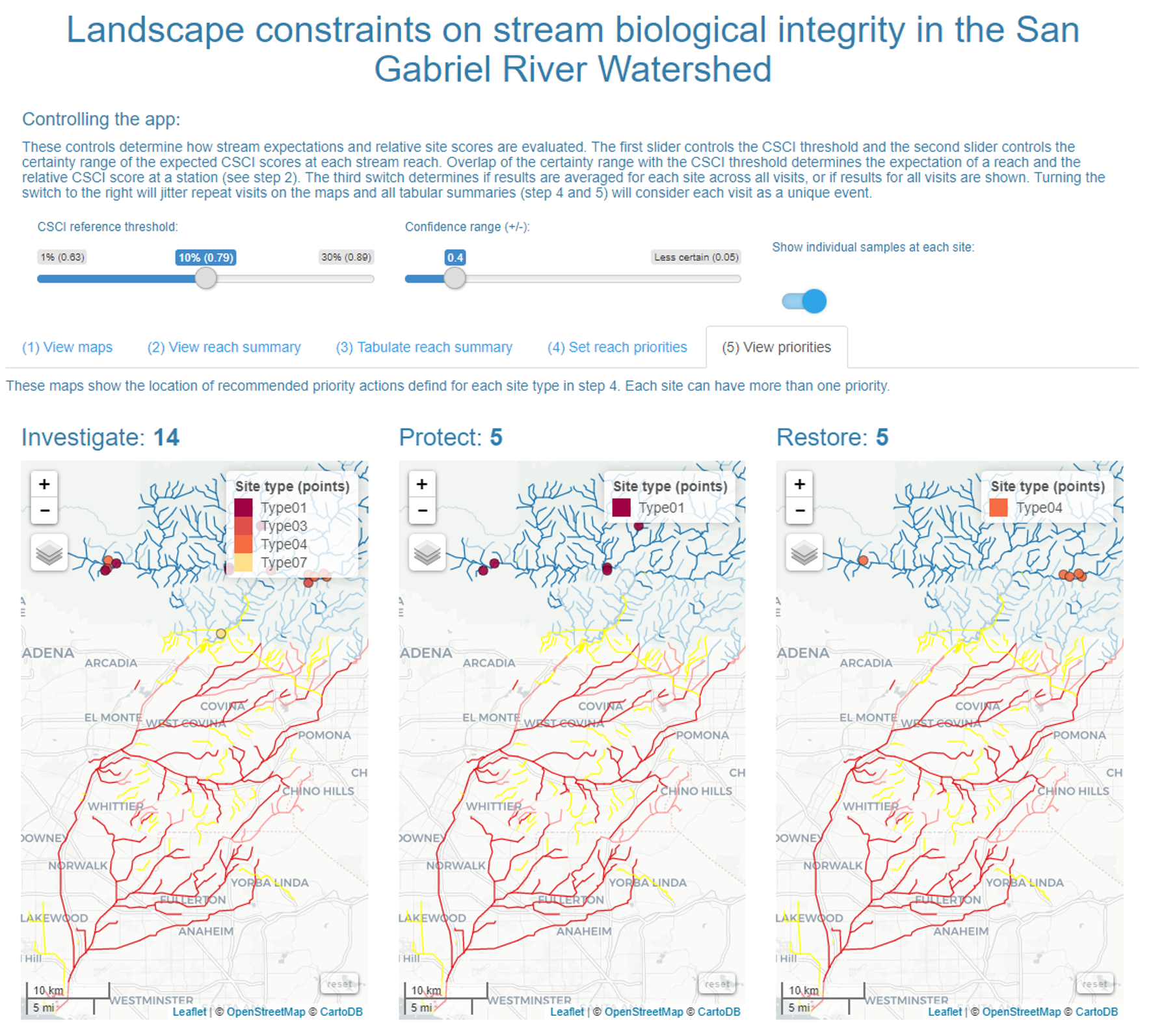


Figure 4 Screenshot of the online application used by the stakeholder group to interact with and use results from the landscape models. The application allowed users to visualize results of reach classications, relative site scores for the CSCI based on the expectation, and recommend management actions for each reach type. The app can be viewed in the supplementary material.

## Factors explaining constraints and sensitivity analysis

Factors explaining variation between constrained and unconstrained stream reaches were evaluated for the major regions in California (Figure 1). Landscape and geological data in StreamCat at the riparian and watershed scale were used to model variation between reach classes using random forest models (Breiman [2001](#ref-Breiman01)). For each region, 1000 regression trees were created and the mean reduction in accuracy was estimated for the exclusion of each variable across all models. This created an estimate of importance of each variable for describing differences between constrained and unconstrained stream classes. Mean reduction in accuracy was estimated for all variables in each model to identify the top five important variables in each region. Reach classes as possibly or likely constrained (or unconstrained) were combined to evaluate the complete dataset. Although the landscape models were developed from land use data derived from StreamCat, relatively coarse measures of land use were used (i.e., combined estimates of low to high intensity urban or agricultural land use). The StreamCat dataset includes additional information at the riparian to watershed scale that could provide greater insight into specific factors within each region that potentially constrain stream integrity.

Stream reach classifications and site performance categories depend on the range of score expectations (or certainty) from the landscape model (Figure 3b) and the CSCI threshold for defining nominally low or high scores (Figure 3c). This framework for identifying priorities was developed to allow flexibility in how the model could be applied. The 5th and 95th percentile of expected scores at a reach are used as a default range in which a high degree of certainty in the model output is assumed. The ability to reduce this range (e.g., 25th to 75th percentile) to assume less certainty in the model is provided. The CSCI threshold can also be changed to assess effects of relaxing or increasing flexibility in a potential definition of a regulatory standard. A threshold of 0.79 is used by default as a measure of the 10th percentile of scores at all reference (non-impacted) sites that were used to calibrate the CSCI index. This value can be increased to examine effects of a more conservative threshold or decreased for a more relaxed threshold. The combined effects of changing both the certainty in the model and the CSCI threshold were evaluated to estimate the changes in stream miles in each classification.

## Unclassified reaches

Finally, some stream reaches were not classifed following application of the landscape model to the statewide hydrography dataset. Unclassified reaches occurred if StreamCat data were unavailable to estimate CSCI predictions or if a stream catchment basin could not be defined for a particular reach. The latter was more common, particularly in developed areas where engineered channels or agricultural ditches were hydrologically removed from the natural stream network. Overall, unclassified reaches were not common in the statewide dataset but they may have regional importance depending on needs of local management groups. An approach for assigning biological expectations to unclassified reaches is demonstrated for “typically” urban and agricultural reaches that relies on the range of expectations for reaches with similar land use by region. The approximate range of scores for reaches dominated by either urban, agricultural, and neither of the two (other) was identified by using kmeans clustering of percentage land use for each reach (MacQueen [1967](#ref-MacQueen67)). Typically urban and agricultural reaches were identified based on the largest centroid average of the clusters for each land use type, whereas the other category was based on the minimum sum of the centroid values for urban and agricultural land use. The expected ranges for each land use type in each region were based on averages of ranges across all reaches in the identified clusters.

# Results

## State-wide patterns

The landscape model was used to predict an expected range of CSCI scores for 138716 stream reaches in California. The bioassessment dataset used to develop the model included 2620 unique field observations assigned to stream reaches in the NHDPlus dataset. By region, the most bioassessment samples were observed in the South Coast (n = 839), followed by the the Chapparal (n = 684) and Sierra Nevada regions (n = 548). Model performance statewide and by region indicated generally good agreement between observed CSCI scores and the median prediction for the associated stream reach (Table 2). Agreement between observed and predicted values for the entire calibration dataset was 0.84 and RMSE = 0.14. The intercept and slope for a regression between observed and predicted values were 0.24 and 0.72, suggesting a slight negative bias of predictions at lower scores and slight positive bias at higher scores. The statewide calibration data showed similar results, with slightly smaller correlation ( 0.72) and larger RMSE (0.18) estimates.

Table 2 Performance of landscape models by calibration and validation datasets in predicting CSCI scores. The statewide dataset (Figure 6) and individual regions of California (Figure 1 are evaluated. Averages and standard deviations (in parentheses) for observed and predicted values of each dataset are shown. Correlations, root mean squared errors (RMSE), intercept, and slopes are for comparisons of predicted and observed values to evaluate model performance.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset | Location | n | Observed | Predicted | Correlation | RMSE | Intercept | Slope |
| Cal | Statewide | 1965 | 0.82 (0.26) | 0.83 (0.22) | 0.84 | 0.14 | 0.24 | 0.72 |
|  | CH | 512 | 0.76 (0.27) | 0.77 (0.23) | 0.85 | 0.14 | 0.22 | 0.73 |
|  | CV | 116 | 0.51 (0.18) | 0.54 (0.15) | 0.76 | 0.12 | 0.21 | 0.64 |
|  | DM | 86 | 0.87 (0.22) | 0.88 (0.17) | 0.77 | 0.14 | 0.38 | 0.58 |
|  | NC | 208 | 0.92 (0.2) | 0.93 (0.14) | 0.74 | 0.13 | 0.45 | 0.52 |
|  | SC | 631 | 0.79 (0.24) | 0.79 (0.21) | 0.83 | 0.14 | 0.23 | 0.71 |
|  | SN | 412 | 0.98 (0.18) | 0.99 (0.1) | 0.59 | 0.14 | 0.66 | 0.34 |
| Val | Statewide | 655 | 0.82 (0.25) | 0.84 (0.21) | 0.72 | 0.18 | 0.36 | 0.59 |
|  | CH | 172 | 0.76 (0.27) | 0.8 (0.21) | 0.73 | 0.19 | 0.38 | 0.57 |
|  | CV | 40 | 0.52 (0.19) | 0.57 (0.17) | 0.51 | 0.19 | 0.34 | 0.45 |
|  | DM | 28 | 0.84 (0.17) | 0.94 (0.12) | 0.57 | 0.17 | 0.61 | 0.39 |
|  | NC | 71 | 0.94 (0.19) | 0.96 (0.11) | 0.55 | 0.16 | 0.67 | 0.31 |
|  | SC | 208 | 0.8 (0.24) | 0.78 (0.21) | 0.71 | 0.17 | 0.28 | 0.62 |
|  | SN | 136 | 0.97 (0.17) | 0.98 (0.09) | 0.18 | 0.18 | 0.90 | 0.09 |

Model performance differed by region. Performance for the Chapparal and South Coast regions were comparable or slightly improved compared to the statewide dataset for both the calibration and validation datasets. Model predictions for the Central Valley, Desert Modoc, and North Coast regions had slightly lower performance compared to the statewide results, with correlations of approximately 0.75 with observed values in the calibration dataset and 0.55 in the validation dataset. Model performance was poor for the Sierra Nevada region where the lowest correlation between predicted and observed values was observed. Regression estimates for the validation data in this region showed an intercept and slope of approximately 1 and 0. Overall, model performance was strongly associated with land use gradients in each region (Figure 5). The landscape model peformed well in regions with a mix of urban, agricultural, and open land, such as the South Coast, where strong gradients occur in many watersheds. Conversely, the model did not perform well in regions where developed landscapes were less common, such as the Sierra Nevada region.

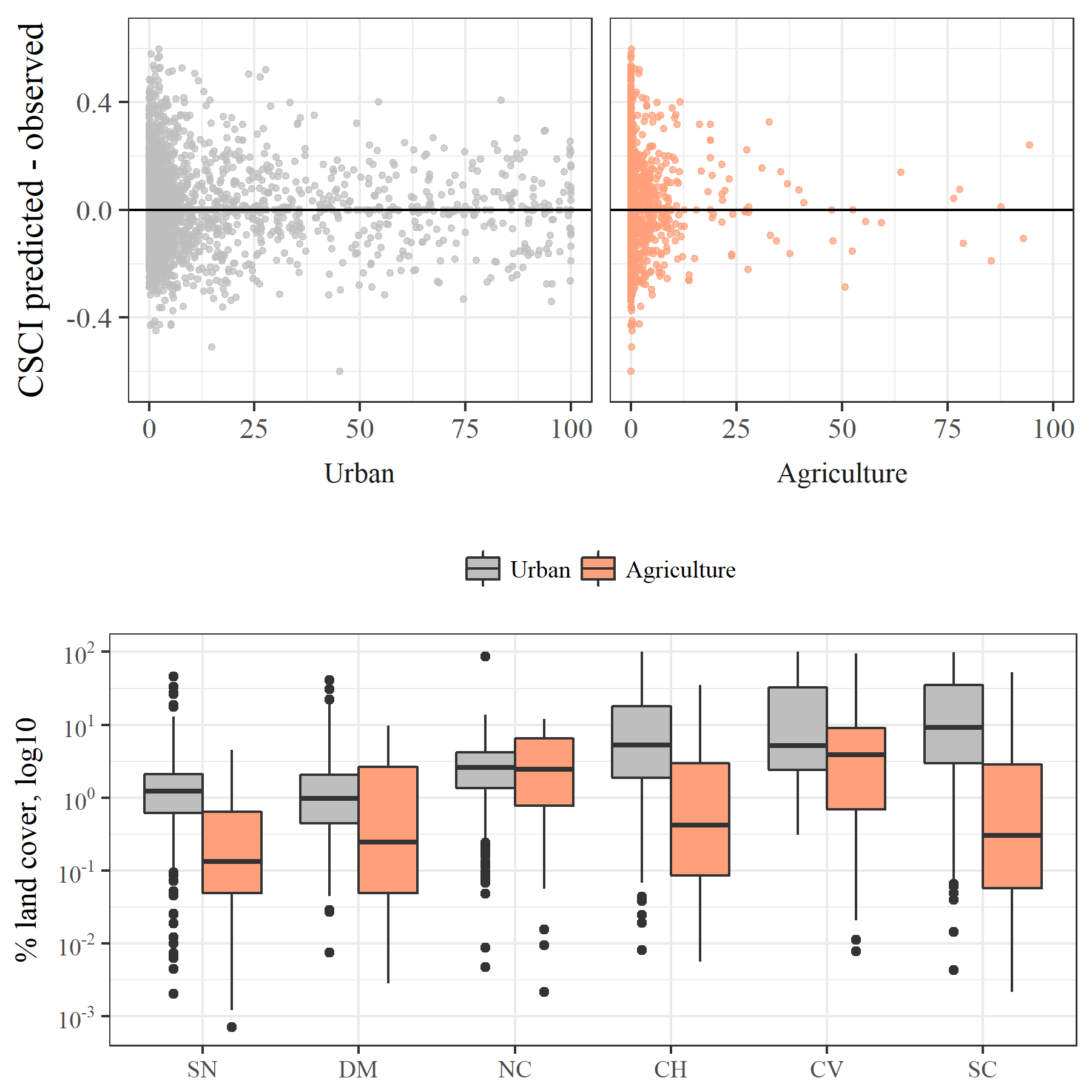


Figure 5 Model performance in relation to land cover and land cover by major regions in California. Model residuals (CSCI predicted - observed) were smaller in regions with more urban or agricultural land use (e.g., SC, CV) and larger in regions with less anthropogenic land use (e.g., SN, DM). CV: Central Valley, CH: Chaparral, DM: Deserts Modoc, NC: North Coast, SN: Sierra Nevada, SC: South Coast.

Statewide patterns in stream constraints were apparent from the results of the landscape models consistent with land use (Figure 6). Stream reaches were more often constrained for biotic integrity in regions with more watershed development, either as urban or agricultural land. For example, likely constrained reaches were apparent from the statewide map in coastal reaches of the South Coast where heavy urbanization occurs and in the Central Valley where agriculure is the dominant land use. Stream reaches were more likely to be unconstrained in regions with less development, with areas in the North Coast and the Sierra Nevada region visible on the map (right, Figure 6). A majority of reaches statewide were classified as possibly constrained (23% of all stream length) or possibly unconstrained (67%), whereas a minority were likely constrained (3%) or likely unconstrained (7%) (Table 4). By region, the most reaches classified as likely unconstrained reaches were in the North Coast (20%) and Sierra Nevada regions (19%) and the most reaches as likely constrained were observed in the Central Valley (25%) and South Coast (9%) regions. Relative CSCI scores compared to reach expectations were as expected for 92% of the sampled locations statewide, whereas a much smaller percentage of sites were equally under or over scoring (Table 4). 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.

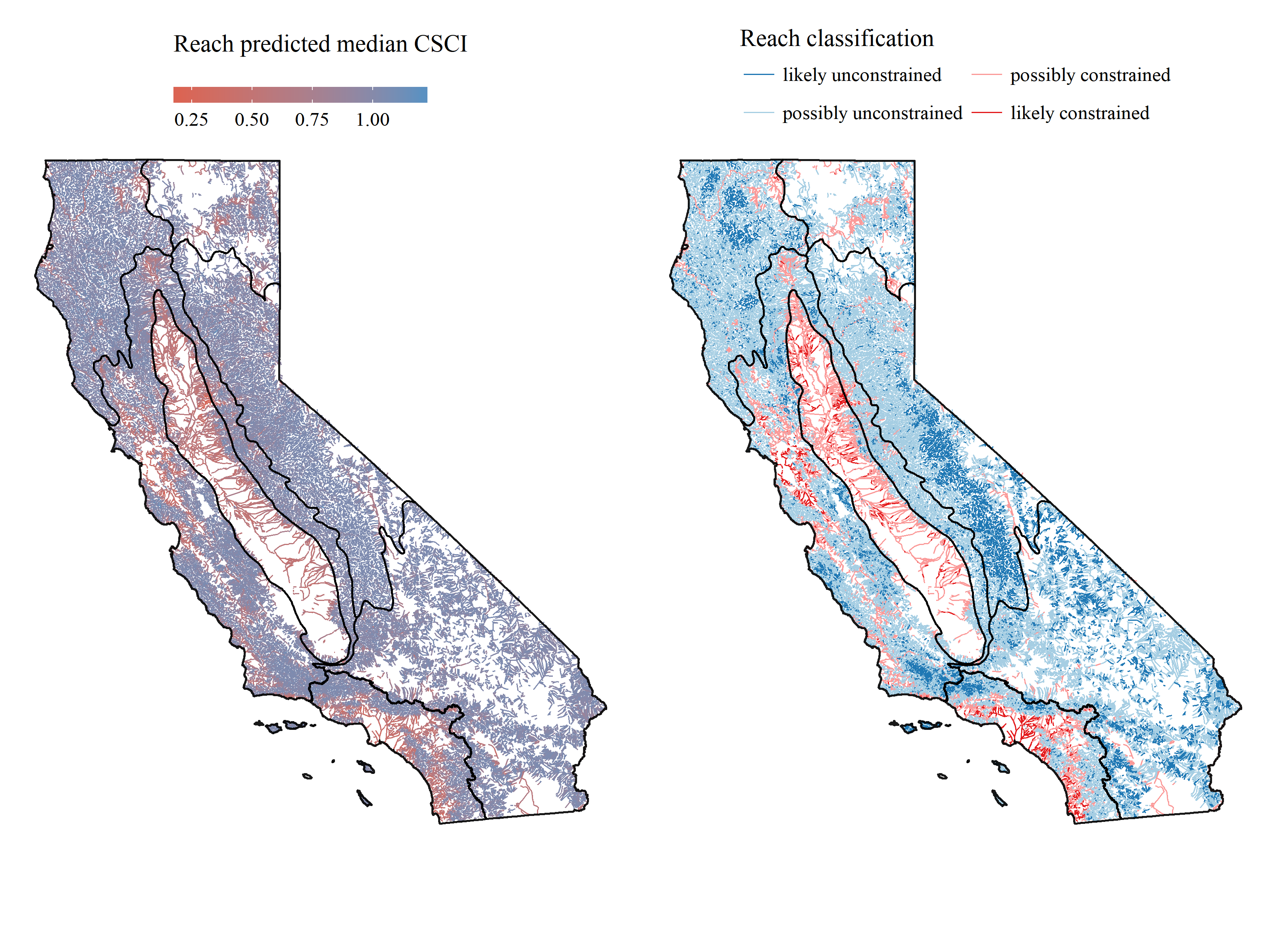


Figure 6 Statewide application of landscape models showing the median predicted CSCI scores for each stream reach (left) and corresponding stream reach classifications (right). Major regional boundaries are also shown (see Figure 1).

*Table 3: (#tab:clstot) Summary of stream length for each stream class statewide and in major regions of California (Figures 1, 6). Lengths are in kilometers with the percentage of the total length in a region in parentheses. CV: Central Valley, CH: Chaparral, DM: Deserts Modoc, NC: North Coast, SN: Sierra Nevada, SC: South Coast.*

|  | constrained | | unconstrained | |
| --- | --- | --- | --- | --- |
| Region | likely | possibly | possibly | likely |
| Statewide | 4225 (2) | 28545 (13) | 144870 (66) | 40216 (18) |
| CV | 1444 (9) | 9884 (64) | 3809 (25) | 331 (2) |
| CH | 957 (2) | 8475 (14) | 42535 (71) | 8083 (13) |
| DM | 23 (0) | 3614 (6) | 39683 (70) | 13719 (24) |
| NC | 25 (0) | 1553 (5) | 22274 (77) | 4949 (17) |
| SN | 1 (0) | 1083 (3) | 26127 (68) | 11128 (29) |
| SC | 1775 (10) | 3936 (22) | 10443 (58) | 2006 (11) |

*Table 4: (#tab:reltot) Summary of CSCI scores by relative expectations for each stream class statewide and in each major region of California (Figures 1, 6). Average (standard deviation) scores and counts (percents) of the number of monitoring stations in each relative expectation and region are shown. Sites are over scoring if the observed scores are above the range of expectations at a reach, expected if within the range, or under scoring if below the range. CV: Central Valley, CH: Chaparral, DM: Deserts Modoc, NC: North Coast, SN: Sierra Nevada, SC: South Coast.*

|  | under scoring | | expected | | over scoring | |
| --- | --- | --- | --- | --- | --- | --- |
| Region | CSCI | n (%) | CSCI | n (%) | CSCI | n (%) |
| Statewide | 0.47 (0.2) | 105 (4) | 0.83 (0.24) | 2333 (92) | 1.14 (0.15) | 99 (4) |
| CH | 0.39 (0.17) | 31 (5) | 0.77 (0.25) | 613 (93) | 1.15 (0.15) | 16 (2) |
| CV | 0.29 (0.06) | 8 (6) | 0.52 (0.17) | 136 (94) | 0.81 (NA) | 1 (1) |
| DM | 0.59 (0.11) | 4 (4) | 0.87 (0.19) | 104 (94) | 1.21 (0.06) | 3 (3) |
| NC | 0.54 (0.17) | 11 (4) | 0.93 (0.16) | 254 (93) | 1.2 (0.04) | 9 (3) |
| SC | 0.44 (0.2) | 24 (3) | 0.79 (0.22) | 744 (92) | 1.09 (0.18) | 41 (5) |
| SN | 0.62 (0.17) | 27 (5) | 0.99 (0.14) | 482 (90) | 1.2 (0.05) | 29 (5) |

## Case study

Application of the landscape model to the San Gabriel River watershed and engagement with a stakeholder group demonstrated how the results can be used to locally prioritize actions for different stream reaches. About 750 reaches in the SGR were identified and classified from NHDPlus, of which 10% were visited for bioassessment sampling. CSCI scores ranged from 0.2 to 1.23 consistent with heavy urban development in the lower watershed and open land use at higher elevation in the upper watershed (Figure 7). Application of the landscape model results to the CSCI scores provided a context of expectations consistent with the strong land use gradient in the watershed. Stream reaches in the upper watershed were a mix of likely and possibly unconstrained, whereas almost all stream reaches in the lower watershed were classified as likely constrained. Several reaches in the lower watershed had ranges that were left-skewed toward very low CSCI scores consistent with extreme landscape pressures.

Most of the sampling stations in the SGR watershed were within the expected ranges of CSCI scores for the defined stream classes (top, Figure 8). Using a hypothetical CSCI threshold of 0.79 and a relatively certain range of expected scores from the 5th to the 95th percentile of the model predictions, only three sites were under scoring (two likely unconstrained and one likely constrained) and two sites were over scoring (both likely constrained). One of the under scoring sites was below the hypothetical CSCI threshold. None of the remaining sites in both possibly and likely classes were above or below the chosen CSCI threshold for the given class.

The three priority management actions identified by the stakeholder group (investigate, protect, restore) were ranked from low, medium, to high for each potential reach type (Figure 9). High priority recommendations were generally given to over and under scoring sites in likely unconstrained reaches or those below the hypothetical threshold with possibly unconstrained classification. Low priority actions were most often recommended for possibly and likely constrained sites, or no action was recommended where it was assumed baseline monitoring and maintenance was sufficient. Recommended actions to investigate were more common for both over scoring and under scoring sites, protect was more common at over scoring sites, and restore was more common at under scoring sites. A clear distinction between low and high priority actions was observed on the watershed map (bottom, Figure 8). Sites in the lower watershed were low priority if an action was recommended, whereas the four high priority sites were in the upper watershed. Several sites that were scoring as expected for likely and possibly unconstrained reaches in the upper watershed were recommended as medium priority for protection.

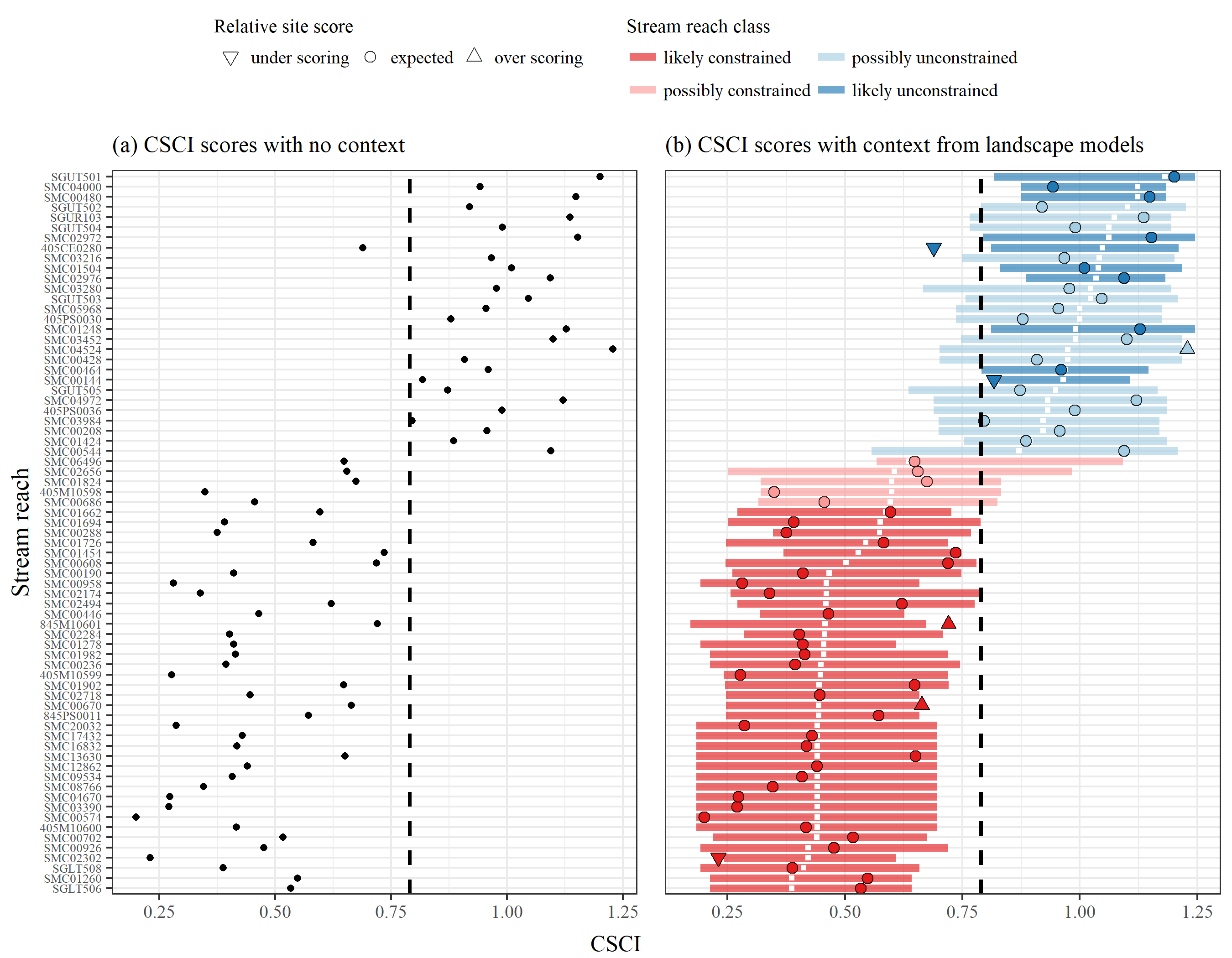


Figure 7 Application of the landscape model to stream reaches in the San Gabriel River watershed, Los Angeles, California. CSCI scores with no context from the model are on the left (a) and scores with context from the model are on the right (b). Relative site scores as under scoring, expected, or over scoring are based on observed scores given the reach class as likely constrained, possibly constrained, possibly unconstrained, and likely unconstrained. Reach classes are based on overlap of the expectations with a hypothetical biological threshold for the CSCI (dashed lined) and location of the median expectation (white ticks).

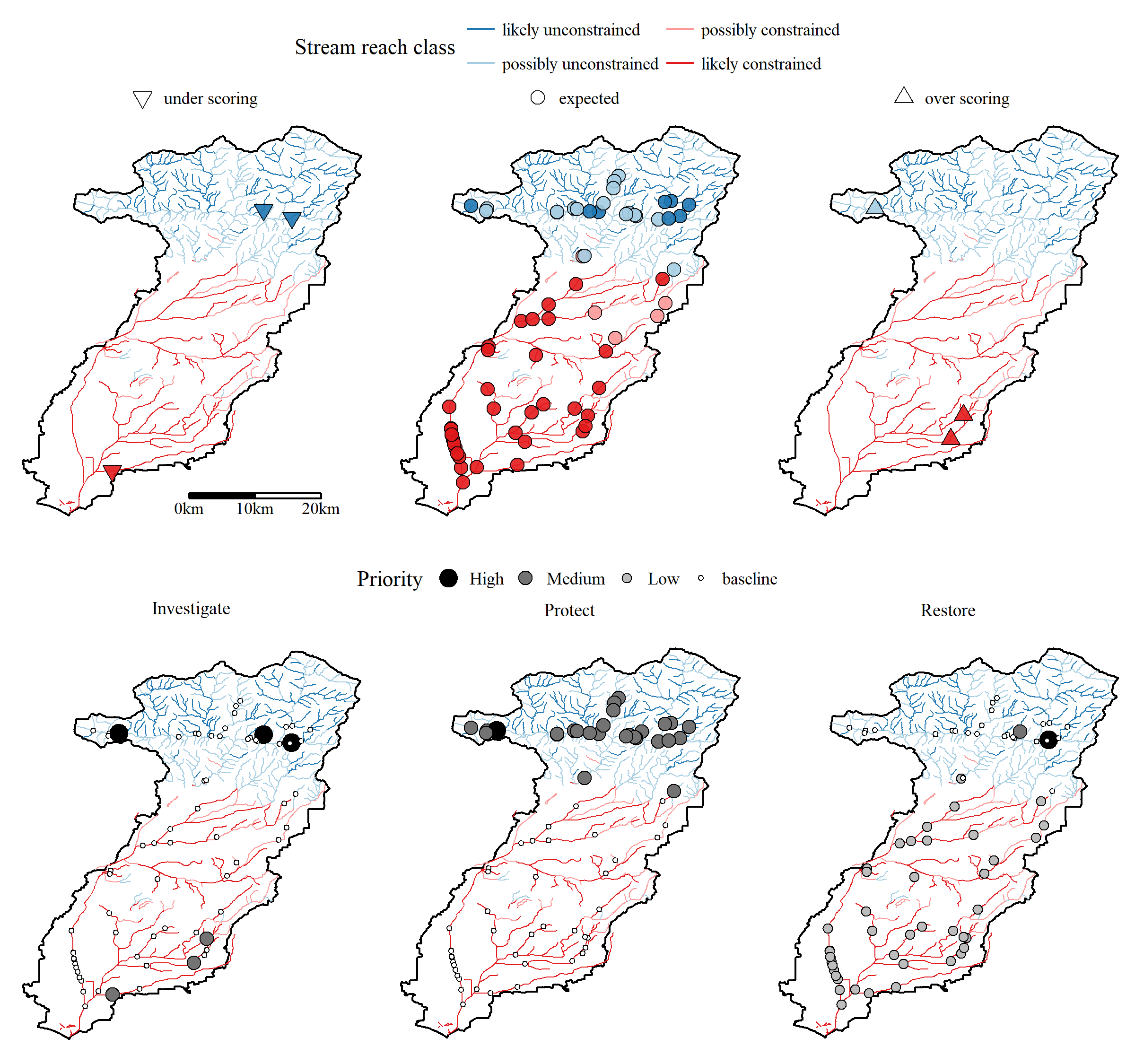


Figure 8 Relative site scores and recommended management actions for locations with CSCI scores in the San Gabriel River watershed. Relative site scores as under scoring, expected, or over scoring are based on observed scores given the reach class as likely constrained, possibly constrained, possibly unconstrained, and likely unconstrained. Recommended management actions are ranked by priority for actions to investigate, protect, and restore a site. No recommended actions assumes baseline maintainence and monitoring is sufficient for a site. Recommended actions were defined by a local stakeholder group (see Figure 9).

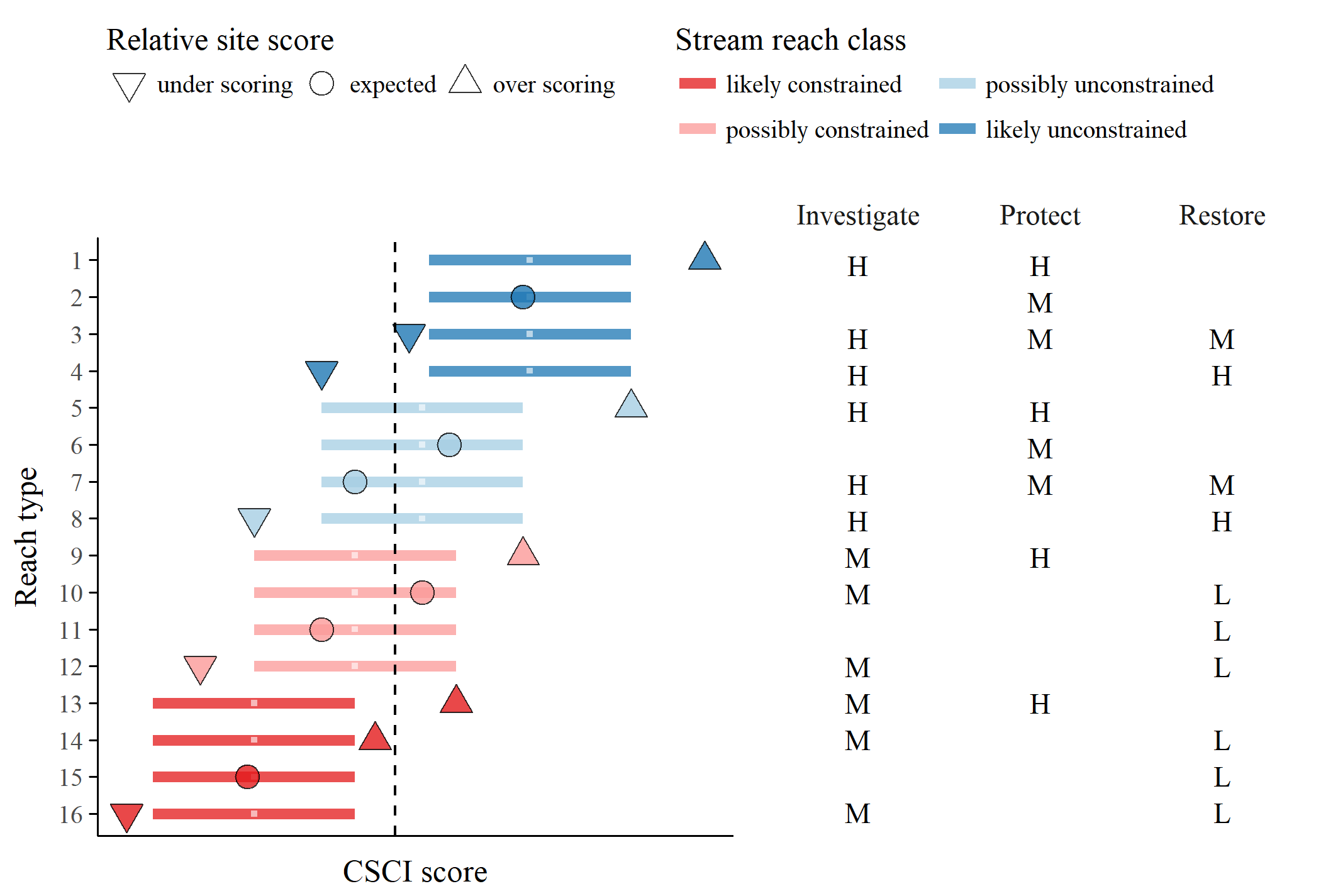


Figure 9 Template provided to stakeholders for priorization of recommended actions for each stream type. The reach types (Table 1) relate to the stream class for the biological expectation (likely unconstrained, possibly unconstrained, possibly constrained, likely constrained), relative site score for the observed CSCI (over scoring, expected, under scoring), and location of the score relative to a hypothetical biological threshold (dashed line, above or below). Horizontal lines are the range of expected CSCI score for a site with tick marks for the median. Priority actions defined by stakeholders are shown on the right for each stream type. Actions are generalized as investigate, protect, or monitor as high (H), medium (M), or low (L) priority. Blank cells indicate that no additional measures are recommended beyond the baseline monitoring and maintenance practiced at all sites.

## Drivers of biological constraints and sensitivity analysis

Importance measures from random forest models identified key variables that explained the differences between constrained and unconstrained reaches between each region (Figure 10). Relative magnitudes of the importance measures between regions confirmed the estimates of model performance, such that regions where the model performed well (e.g., South Coast, Central Valley) had higher importance measures than those where the model did not perform well (e.g., North Coast, Sierra Nevada). The top five most important variables were similar between regions although some specific differences were observed. The amount of biological nitrogen fixation in watershed soils was ranked the most important variable for the Central Valley, Desert Modoc, and North Coast region and second most important for the Chapparal and Sierra Nevada regions. This variable was not in the top five for the South Coast region, which was exclusively described by imperviousness and urbanization. Soil erodibility was the most important variable in the Chaparral and Sierra Nevada regions. Other important variables that were shared between regions (excluding the South Coast) were fertilizer applications and the amount of crops and hay at the riparian and watershed scale.

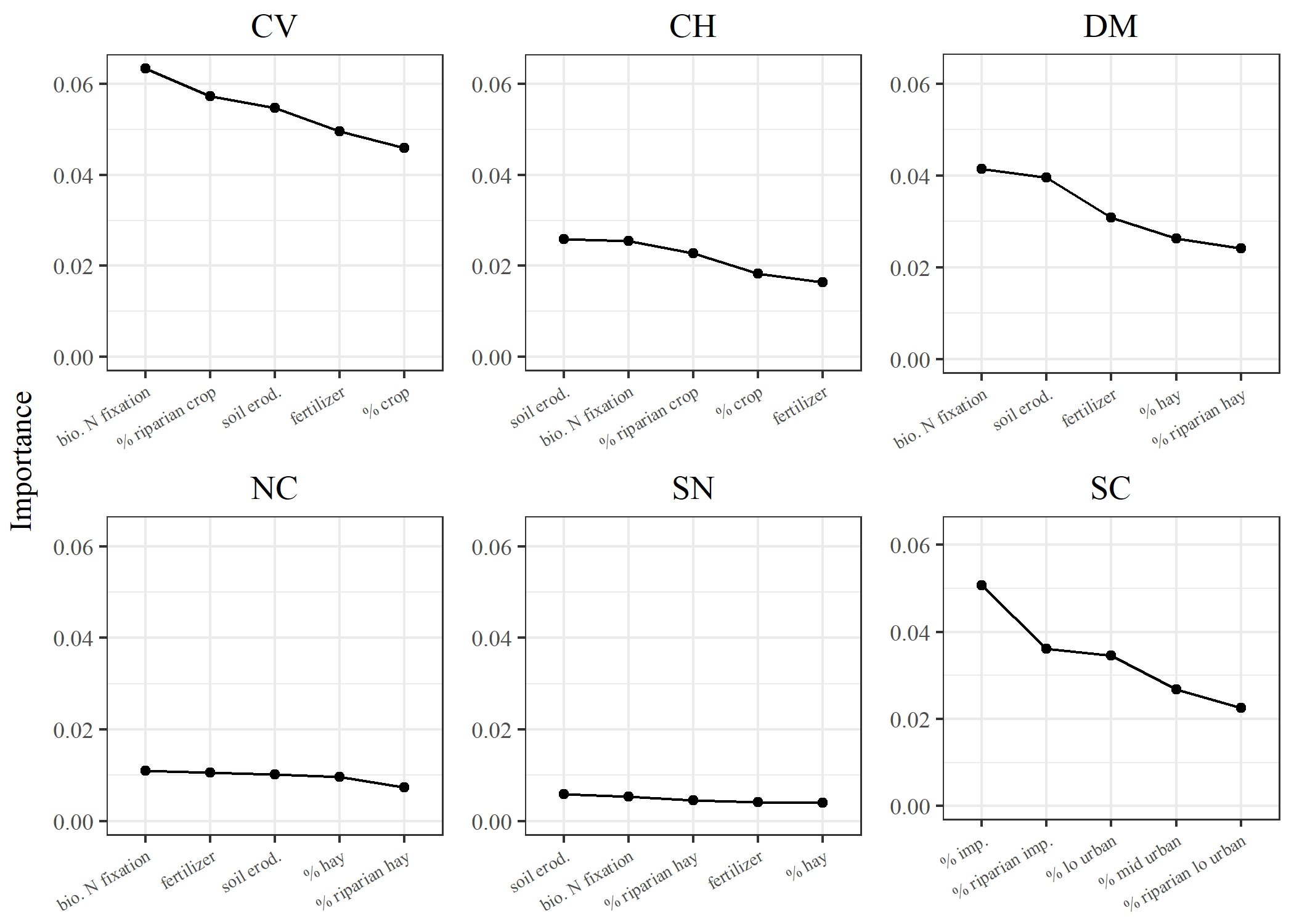


Figure 10 Factors explaining variation between constrained and unconstrained stream reaches by major regions in California. Importance measures were obtained from random forest models of 130 watershed and riparian measures of landscape and geological characteristics from the StreamCat dataset (Hill et al. [2016](#ref-Hill16)). The top five variables for each region are shown. The importance measures describe the mean decrease in prediction accuracy with exclusion of a variable across 1000 random trees for each model. Stream reach classes as possibly or likely were combined for constrained and unconstrained to evaluate the complete dataset. CV: Central Valley, CH: Chaparral, DM: Deserts Modoc, NC: North Coast, SN: Sierra Nevada, SC: South Coast.

The effects of changing certainty in the CSCI predictions and a hypothetical CSCI threshold were evaluated for nine scenarios to identify the change in stream reach classifications (Figure 11). Changes were evaluated as the percentage of stream length statewide and within major regions for each reach classification. Decreasing the certainty of predictions from the landscape model by choosing a narrower range of scores (5th to 95th, 25th to 75th, 45th to 55th) increased the number of streams from the possible to likely category in both constrained and unconstrained reaches. Similarly, changing the CSCI threshold from relaxed to more conservative (0.63, 0.79, 0.92) increased the number of streams classified as possibly or likely constrained and decreased the number of streams as possibly or likely unconstrained. Changes by region with the different scenarios were also observed. For example, over 80% of reaches in the Central Valley were classified as likely constrained using a conservative CSCI threshold with low certainty of predictions, whereas less than 1% of reaches were in this category using a relaxed CSCI threshold with high certainty. Opposite trends were observed in regions with reduced land use pressures. For example, almost all stream reaches in the North Coast and Sierra Nevada regions were classified as likely unconstrained using a relaxed CSCI threshold and low certainty of predictions.

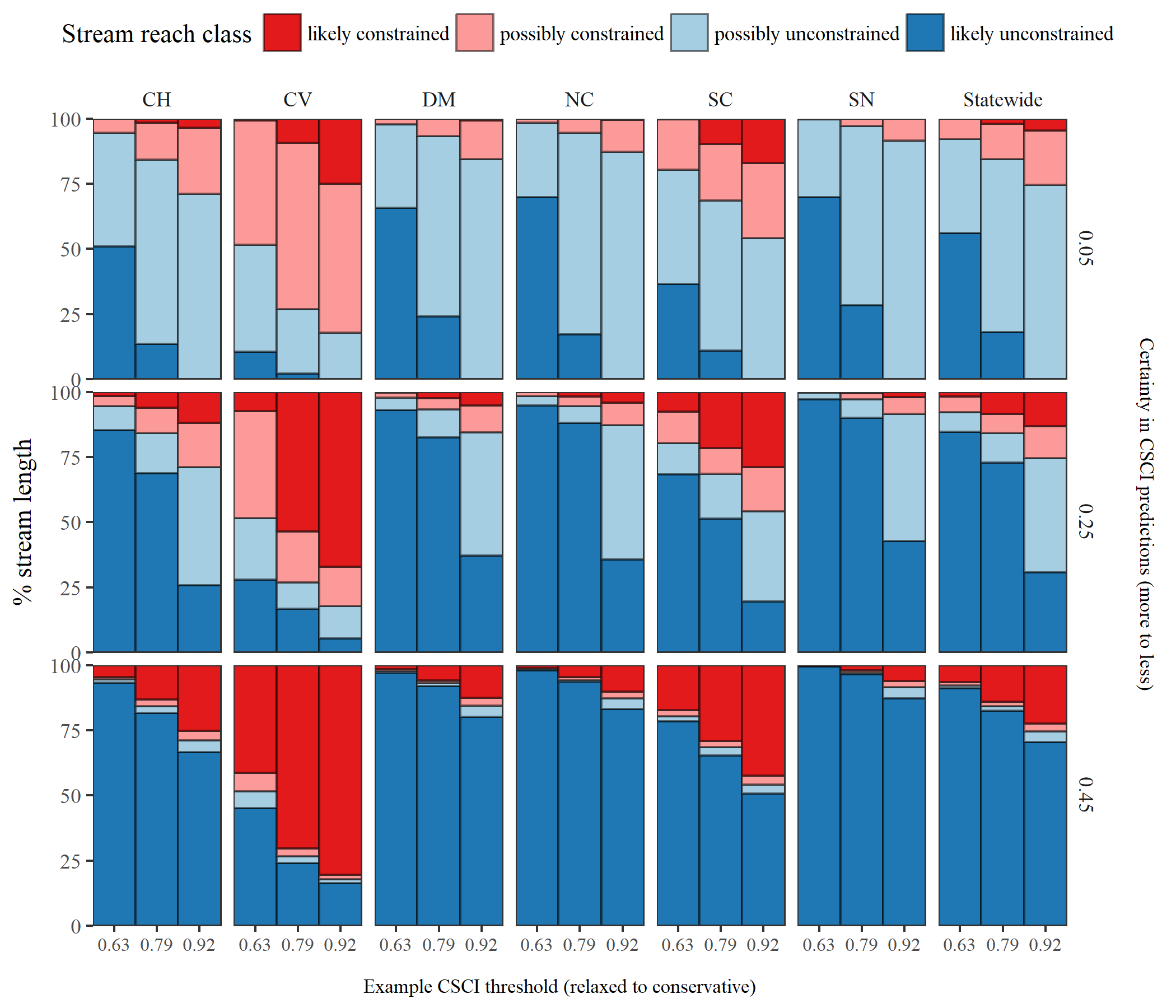


Figure 11 Changes in stream reach classes by region and statweide for different scenarios used to define biological constraints. Nine scenarios were tested that evaluated different combinations of certainty in the CSCI predictions (more certain to less certain as identified by the tail cutoff for the expected range) and potential CSCI threshold (relaxed to conservative). The percentage of total stream length for each classification is shown for each scenario. CV: Central Valley, CH: Chaparral, DM: Deserts Modoc, NC: North Coast, SN: Sierra Nevada, SC: South Coast.

## Unclassified reaches

Ranges of expected CSCI scores for typical reaches in urban, agricultural, and neither of the two (other) are shown in table 5. These typical values are shown for more to less certainty in the range of predictions. Unclassified reaches can be defined by the dominant watershed land use as urban, agricultural, or other, and then matched to the appropriate values in the table. Between regions, the variation in expected scores also provides context for landscape pressures that differ by location. For example, the expected range of scores in regions with heavy urban development (e.g., South Coast) are much lower than streams that are neither urban nor agricultural. The North Coast region in contrast has an expected range of scores in urban streams that is similar to streams that are neither urban nor agricultural. Interestingly, the range of scores in urban and agricultural streams were similar in the Central Valley where agriculture is the dominant land use.

Table 5 Ranges of expected CSCI scores for sites that are typically urban, agricultural, or other land uses by major regions in California and statewide. Ranges can be used to identify approximate expectations for stream reaches with insufficient data for application of the landscape model. CV: Central Valley, CH: Chaparral, DM: Deserts Modoc, NC: North Coast, SN: Sierra Nevada, SC: South Coast.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Region | grps | 5th - 95th (More certain) | 25th - 75th | 45th - 55th (Less certain) |
| Statewide | Urban | 0.3 - 0.89 | 0.45 - 0.72 | 0.56 - 0.61 |
|  | Ag | 0.33 - 1 | 0.47 - 0.77 | 0.57 - 0.63 |
|  | Other | 0.73 - 1.19 | 0.91 - 1.08 | 0.99 - 1.02 |
| DM | Urban | 0.43 - 1.11 | 0.65 - 0.96 | 0.79 - 0.84 |
|  | Ag | 0.34 - 1.03 | 0.48 - 0.79 | 0.59 - 0.65 |
|  | Other | 0.71 - 1.19 | 0.9 - 1.08 | 0.98 - 1.02 |
| SN | Urban | 0.43 - 1.12 | 0.63 - 0.96 | 0.78 - 0.84 |
|  | Ag | 0.35 - 1.08 | 0.53 - 0.87 | 0.65 - 0.72 |
|  | Other | 0.73 - 1.2 | 0.92 - 1.09 | 0.99 - 1.02 |
| NC | Urban | 0.62 - 1.19 | 0.86 - 1.09 | 0.95 - 1 |
|  | Ag | 0.35 - 1.08 | 0.51 - 0.86 | 0.63 - 0.69 |
|  | Other | 0.76 - 1.18 | 0.92 - 1.07 | 0.98 - 1.01 |
| CH | Urban | 0.3 - 0.92 | 0.47 - 0.74 | 0.58 - 0.63 |
|  | Ag | 0.37 - 1.07 | 0.53 - 0.86 | 0.64 - 0.7 |
|  | Other | 0.73 - 1.18 | 0.91 - 1.08 | 0.98 - 1.01 |
| CV | Urban | 0.33 - 0.96 | 0.5 - 0.78 | 0.62 - 0.68 |
|  | Ag | 0.32 - 0.97 | 0.45 - 0.73 | 0.54 - 0.6 |
|  | Other | 0.6 - 1.16 | 0.8 - 1.02 | 0.89 - 0.93 |
| SC | Urban | 0.27 - 0.82 | 0.41 - 0.66 | 0.51 - 0.56 |
|  | Ag | 0.35 - 1.06 | 0.54 - 0.9 | 0.67 - 0.74 |
|  | Other | 0.77 - 1.19 | 0.93 - 1.08 | 0.99 - 1.02 |

# Discussion

* What is the value of identifying constrained channels?
  + Identification of constrained channels allows us to determine how best to spend limited resources and to focus on reaches where we have a decent chance of improving the biological condition.
  + Use of more data to develop context of assessment
  + Targeted management for desired outcomes
  + Informing decisions about future monitoring (i.e., prioritize future monitoring locations)
  + Listing implications, perhaps with a subheader, “Implications for determining biological impairments”. We should enlist Chad to help with this—after presenting him with LACSD’s proposal.
* What is useful about our approach compared to alternatives?
  + Field-based methods to identify constrained channels vs. landscape modelling
  + Related directly to biological condition and regulatory standards
  + Results are widely corraborated by other landscape studies - land use is big determinant of macroinvert assemblage
* What contributed to our success in defining priorities?
  + Stakeholder involvement guided process, contributed to achieving goals
  + An interactive/iterative approach was used - we provided tools to facilitate (web apps) and we did not assume priorities
* Caveats of our aproach
  + This analysis doesn’t truly tell us if a site can be fixed or if the conditions are truly constrained (key message, need to mention in intro)
  + What do priorities really mean? Depends on your interests, needs, values, etc.
  + Constrained may not always mean constrained - CSCI vs other biological indicators
  + Site-specific approaches are warranted in certain cases
  + Changing certainty or CSCI treshold - mechanistic effects and implications. Don’t cook the books.
* Future work
  + Ability to link with other assessment tools besides CSCI
  + Link with engineered channels study
  + Priorities statewide
  + Application to larger regions possible (national-scale), or how it can be applied in other areas

# Supplement

Online application.

# References

Breiman, L. 2001. “Random Forests.” *Machine Learning* 45:5–32.

Cade, B. S., and B. R. Noon. 2003. “A Gentle Introduction to Quantile Regression for Ecologists.” *Frontiers in Ecology and the Environment* 1 (8):412–20.

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 (1-4):143–60. <https://doi.org/10.1007/s10661-008-0256-z>.

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* 36:142–51. <https://doi.org/10.1016/j.ecolind.2013.07.006>.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer.

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 Assocation* 52:120–28. <https://doi.org/10.1111/1752-1688.12372>.

Howard, J. K., K. R. Klausmeyer, K. A. Fesenmyer, J. Furnish, T. Gardali, T. Grantham, J. V. E. Katz, et al. 2015. “Patterns of Freshwater Species Richness, Endemism, and Vulnerability in California.” *PLOS ONE* 10 (7):e0130710. <https://doi.org/10.1371/journal.pone.0130710>.

Howard, J., and C. Revenga. 2000. “California’s Freshwater Biodiveristy in a Continental Context. Science for Conservation Technical Brief Series.” San Francisco, CA: The Nature Conservancy of California.

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 (3):603–15. <https://doi.org/10.1111/j.1752-1688.2011.00635.x>.

Leps, M., J. D. Tonkin, V. Dahm, P. Haase, and A. Sundermann. 2015. “Disentangling Environmental Drivers of Benthic Invertebrate Assemblages: The Role of Spatial Scale and Riverscape Heterogeneity in a Multiple Stressor Environment.” *Science of the Total Environment* 536:546–56. <https://doi.org/10.1016/j.scitotenv.2015.07.083>.

MacQueen, J. 1967. “Some Methods for Classification and Analysis of Multivariate Observations.” In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, edited by L. M. Le Cam and J. Neyman, 281–97. Berkeley, California, USA: University of California Press.

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 (1):249–71.

Meinshausen, N. 2006. “Quantile Regression Forests.” *Journal of Machine Learning Research* 7:983–99.

Meinshausen, Nicolai. 2017. *QuantregForest: Quantile Regression Forests*. <https://CRAN.R-project.org/package=quantregForest>.

Novotny, V., A. Bartosová, N. O’Reilly, and T. Ehlinger. 2005. “Unlocking the Relationship of Biotic Integrity of Impaired Waters to Anthropogenic Stresses.” *Water Research* 39 (1):184–98. <https://doi.org/10.1016/j.watres.2004.09.002>.

Ode, P. R. 2007. “Standard Operating Procedures for Collecting Benthic Macroinvertebrate Samples and Associated Physical and Chemical Data for Ambient Bioassessment in California.” Surface Water Ambient Monitoring Program. Sacramento, CA.

Ode, P. R., A. C. Rehn, R. D. Mazor, K. C. Schiff, E. D. Stein, J. T. May, L. R. Brown, et al. 2016. “Evaluating the Adequacy of a Reference-Site Pool for Ecological Assessments in Environmentally Complex Regions.” *Freshwater Science* 35 (1):237–48.

Poff, N. L. 1997. “Landscape Filters and Species Traits: Towards Mechanistic Understanding and Prediction in Stream Ecology.” *Journal of the North American Benthological Society* 16 (2):391–409.

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 (4):833–52.

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 (5):358–72. <https://doi.org/10.1111/gec3.12039>.

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 (1-4):251–66. <https://doi.org/10.1007/s10661-010-1385-8>.

Sponseller, R. A., E. F. Benfield, and H. M. Valett. 2001. “Relationships Between Land Use, Spatial Scale and Stream Macroinvertebrate Communities.” *Freshwater Biology* 46 (10):1409–24. <https://doi.org/10.1046/j.1365-2427.2001.00758.x>.

Townsend, C. R., S. S. Uhlmann, and C. D. Matthaei. 2008. “Individual and Combined Responses of Stream Ecosystems to Multiple Stressors.” *Journal of Applied Ecology* 45 (6):1810–9. <https://doi.org/10.1111/j.1365-2664.2008.01548.x>.

USGS (US Geological Survey). 2014. “National Hydrography Dataset available on the World Wide Web.”