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 R. Ode ([peter.ode@wildlife.ca.gov](mailto:peter.ode@wildlife.ca.gov)), Ryan Hill ([hill.ryan@epa.gov](mailto:hill.ryan@epa.gov)), Chad Loflen ([Chad.Loflen@waterboards.ca.gov](mailto:Chad.Loflen@waterboards.ca.gov)), Martha Sutula ([marthas@sccwrp.org](mailto:marthas@sccwrp.org)), Eric D. Stein ([erics@sccwrp.org](mailto:erics@sccwrp.org))

# Unclassifiable reaches

Some stream reaches were not classified following application of the landscape model to the statewide hydrography dataset. Unclassifiable reaches occurred if StreamCat data were unavailable to predict bioassessment expectations using the landscape model or if a reach was excluded from the NHD-plus dataset (typically, small headwater streams). The former was more common, particularly in developed areas where canals and ditches were sometimes excluded 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. As described below, approximately 15% of the reaches in the San Gabriel River watershed were unclassifiable and a method for assigning a classification to these reaches was desired by the stakeholder group.

An approach for assigning biological expectations to unclassified reaches was developed for “typically” urban or agricultural reaches that was based on the range of expectations for reaches with similar land use. This analysis was conducted statewide and stratified by major regions to account for statewide variation in land use. The approximate range of CSCI scores in unclassifiable reaches were defined for three different gruops: reaches dominated by either 1) urban, 2) agricultural, or 3) open (i.e., lack of urban or agricultural land use). The three groups were identified using kmeans clustering of percentage land use estimates that were available across reaches (MacQueen [1967](#ref-MacQueen67)). This created groups of reaches with similar land use types, where membership of a reach within a particular group was based on the minimum difference in land use estimates for a reach from the group average for each land use type (within-group centroid). The two groups that were dominated by agricultural or urban land use were identified based on the largest centroid average of the clusters for each land use type. The third “open” group that was defined by a lack of urban and agricultural land use was identified by the minimum sum of the centroid values for the two land use types. The expected range of CSCI scores for the three groups were based on averages from the landscape model for reaches with available predictions.

Ranges of expected CSCI scores for typical reaches in urban, agricultural, and open (neither urban, nor agriculture) are shown in table 1. 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 open, 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) is much smaller 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 open. The range of scores in urban and agricultural streams were similar in the Central Valley where agriculture is the dominant land use.

Table 1 Ranges of expected CSCI scores for sites that are typically urban, agricultural, or open (neither urban nor agricultural) 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 and Modoc Plateau, NC: North Coast, SN: Sierra Nevada, SC: South Coast.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Region | Land use | High certainty (10th - 90th) | Moderate (25th - 75th) | Low certainty (40th - 60th) |
| Statewide | Urban | 0.34 - 0.83 | 0.45 - 0.72 | 0.53 - 0.64 |
|  | Ag | 0.38 - 0.93 | 0.47 - 0.77 | 0.54 - 0.66 |
|  | Open | 0.80 - 1.15 | 0.91 - 1.08 | 0.97 - 1.03 |
| DM | Urban | 0.53 - 1.07 | 0.68 - 0.98 | 0.78 - 0.89 |
|  | Ag | 0.39 - 0.96 | 0.48 - 0.78 | 0.56 - 0.67 |
|  | Open | 0.79 - 1.15 | 0.90 - 1.08 | 0.96 - 1.03 |
| SN | Urban | 0.51 - 1.07 | 0.65 - 0.97 | 0.76 - 0.88 |
|  | Ag | 0.41 - 1.03 | 0.53 - 0.87 | 0.62 - 0.75 |
|  | Open | 0.80 - 1.16 | 0.92 - 1.09 | 0.98 - 1.04 |
| NC | Urban | 0.72 - 1.17 | 0.87 - 1.10 | 0.94 - 1.04 |
|  | Ag | 0.41 - 1.04 | 0.51 - 0.86 | 0.60 - 0.72 |
|  | Open | 0.82 - 1.14 | 0.92 - 1.07 | 0.97 - 1.03 |
| CH | Urban | 0.32 - 0.80 | 0.42 - 0.69 | 0.50 - 0.60 |
|  | Ag | 0.40 - 0.98 | 0.51 - 0.84 | 0.60 - 0.72 |
|  | Open | 0.80 - 1.15 | 0.91 - 1.08 | 0.97 - 1.03 |
| CV | Urban | 0.39 - 0.90 | 0.51 - 0.79 | 0.60 - 0.71 |
|  | Ag | 0.36 - 0.89 | 0.45 - 0.73 | 0.52 - 0.63 |
|  | Open | 0.67 - 1.11 | 0.80 - 1.02 | 0.87 - 0.96 |
| SC | Urban | 0.30 - 0.76 | 0.40 - 0.66 | 0.48 - 0.57 |
|  | Ag | 0.41 - 1.01 | 0.53 - 0.90 | 0.63 - 0.78 |
|  | Open | 0.83 - 1.15 | 0.93 - 1.08 | 0.98 - 1.04 |

# Supplemental Figures and Tables



Figure 1 Screenshots from the Stream Classification and Priority Explorer (SCAPE) tool used by the stakeholder group to interact with and use results from the landscape model. The application allowed users to visualize results of reach classifications, relative site scores for the CSCI based on the expectation, and recommend management actions for each reach type. The app is accessible at <http://shiny.sccwrp.org/scape/> (Beck [2018](#ref-Beck18c)).

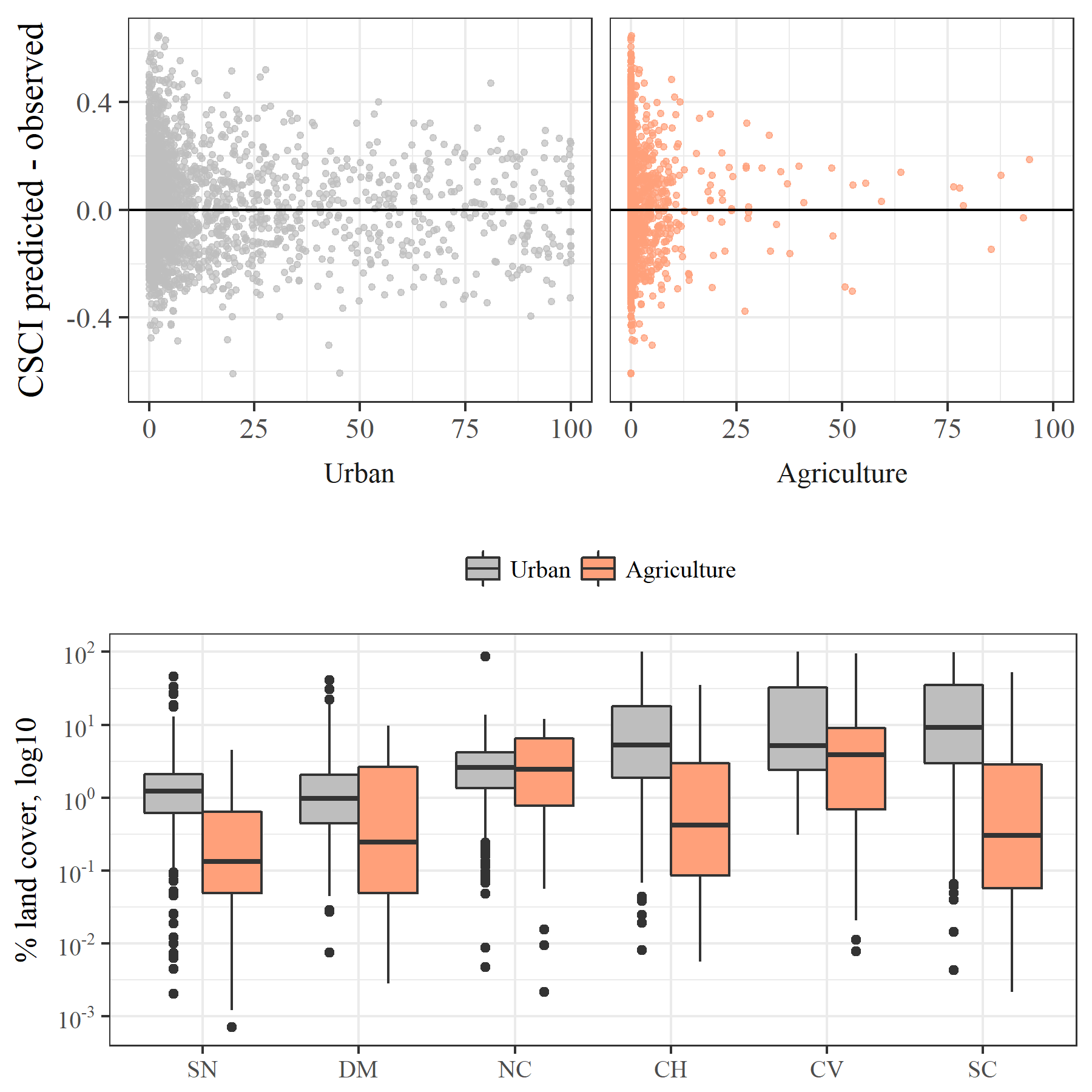


Figure 2 Model performance in relation to land cover and land cover by major regions in California. Model residuals (CSCI predicted - observed) were smaller in regions with more urban or agricultural land use (e.g., SC, CV) and larger in regions with less anthropogenic land use (e.g., SN, DM). CV: Central Valley, CH: Chaparral, DM: Deserts and Modoc Plateau, NC: North Coast, SN: Sierra Nevada, SC: South Coast.

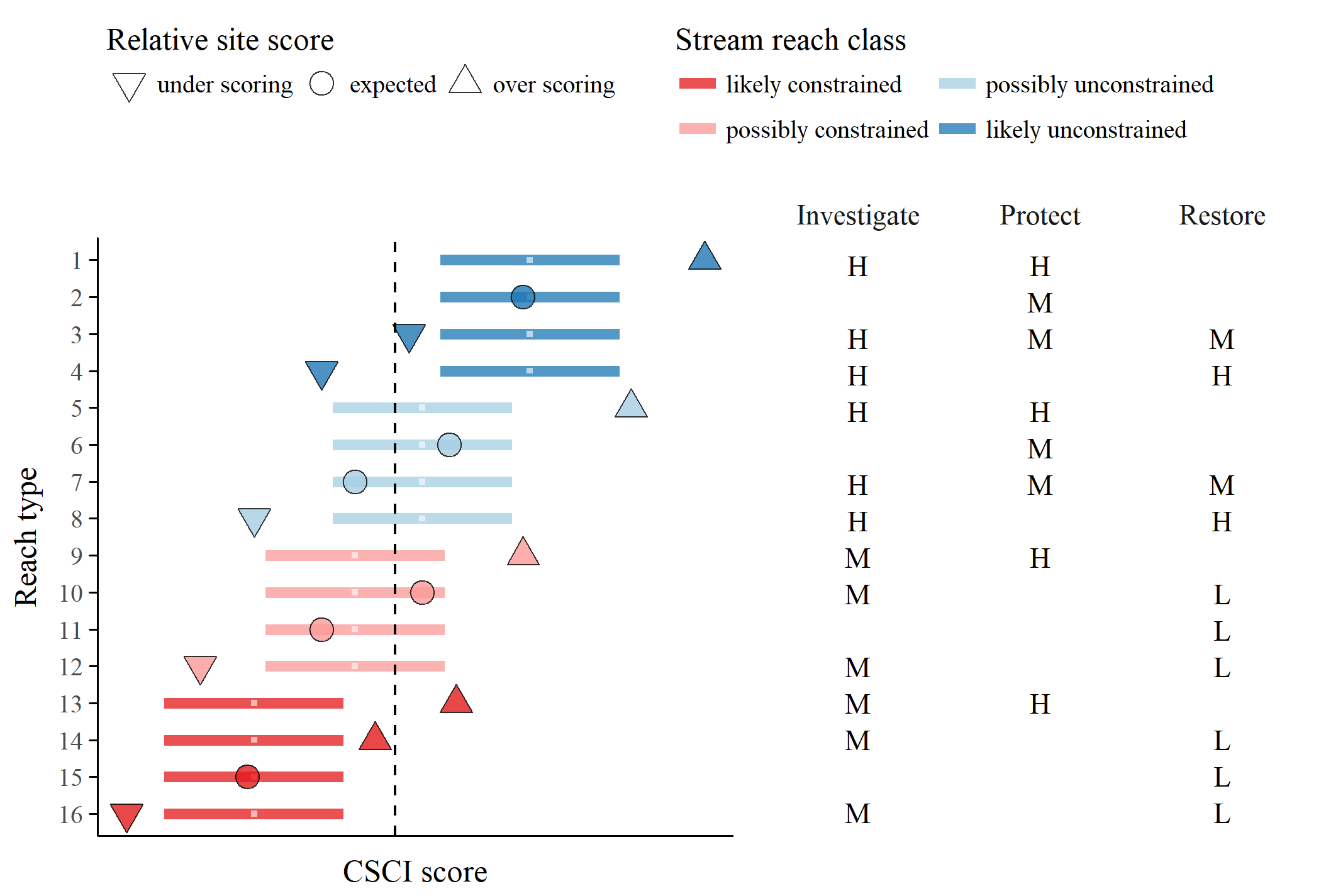


Figure 3 Template provided to stakeholders for prioritization of recommended actions for each stream type. The reach types (Table 2) relate to the stream class for the biological expectation (likely unconstrained, possibly unconstrained, possibly constrained, likely constrained), relative site score 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.

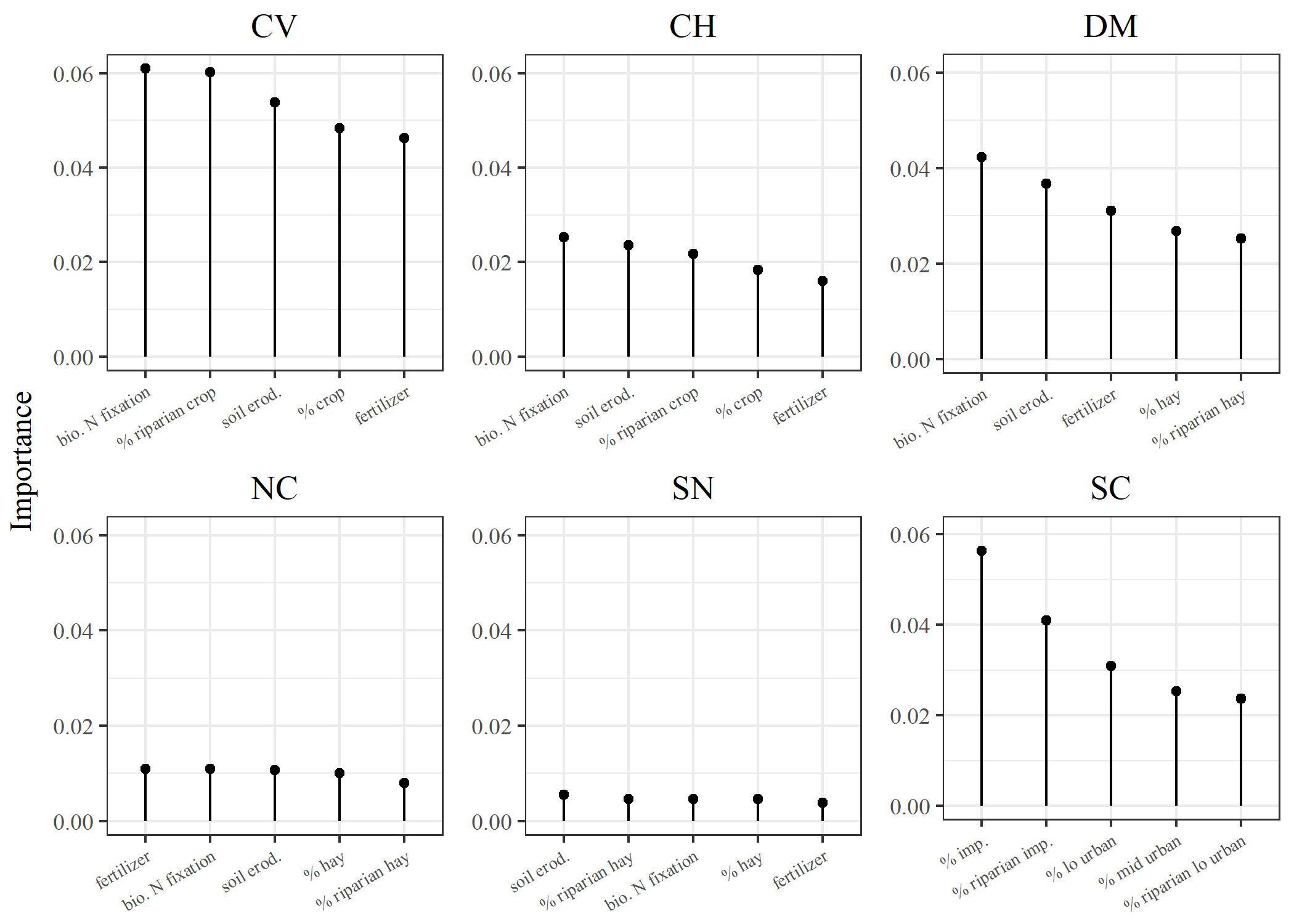


Figure 4 Factors associated with constrained and unconstrained stream reaches by major regions in California. Importance measures were obtained from random forest models of 130 watershed and riparian measures of landscape and geological characteristics from the StreamCat dataset (Hill et al. [2016](#ref-Hill16)). 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 and Modoc Plateau, NC: North Coast, SN: Sierra Nevada, SC: South Coast.



Figure 5 Importance measures for landscape variables used to develop the landscape model of expected stream bioassessment scores in California. Values were obtained from quantile regression models of twenty landscape measures shown in Table 2 obtained from the StreamCat dataset (Hill et al. [2016](#ref-Hill16)). The importance measures describe the percent increase in mean square error and the increase in node impurity with exclusion of a variable across all random trees for each model (Meinshausen [2017](#ref-Meinshausen17)).

Table 2 Land use variables used to develop the landscape model of stream bioassessment scores. All variables were obtained from StreamCat (Hill et al. [2016](#ref-Hill16)). The measurement scale for each variable is at the catchment, watershed, and/or riparian scale (100 m buffer) relative to a stream reach. Total urban and agriculture land use variables were based on sums of indvidual variables in StreamCat as noted in the desciption.

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Scale | Description | Unit |
| CanalDensCat | catchment | Density of NHDPlus line features classified as canal, ditch, or pipeline | km/sq km |
| CanalDensWs | watershed | Density of NHDPlus line features classified as canal, ditch, or pipeline | km/sq km |
| PctImp2006Cat | catchment | Mean imperviousness of anthropogenic surfaces (NLCD 2006) | % |
| PctImp2006Ws | watershed | Mean imperviousness of anthropogenic surfaces (NLCD 2006) | % |
| PctImp2006CatRp100 | catchment, riparian | Mean imperviousness of anthropogenic surfaces (NLCD 2006) | % |
| PctImp2006WsRp100 | watershed, riparian | Mean imperviousness of anthropogenic surfaces (NLCD 2006) | % |
| TotUrb2011Ws | watershed | Total urban land use as sum of developed open, low, medium, and high intensity (NLCD 2011) | % |
| TotUrb2011Cat | catchment | Total urban land use as sum of developed open, low, medium, and high intensity (NLCD 2011) | % |
| TotUrb2011WsRp100 | watershed, riparian | Total urban land use as sum of developed open, low, medium, and high intensity (NLCD 2011) | % |
| TotUrb2011CatRp100 | catchment, riparian | Total urban land use as sum of developed open, low, medium, and high intensity (NLCD 2011) | % |
| TotAg2011Ws | watershed | Total argricultural land use as sum of hay and crops (NLCD 2011) | % |
| TotAg2011Cat | catchment | Total argricultural land use as sum of hay and crops (NLCD 2011) | % |
| TotAg2011WsRp100 | watershed, riparian | Total argricultural land use as sum of hay and crops (NLCD 2011) | % |
| TotAg2011CatRp100 | catchment, riparian | Total argricultural land use as sum of hay and crops (NLCD 2011) | % |
| RdDensCat | catchment | Density of roads (2010 Census Tiger Lines) | km/sq km |
| RdDensWs | watershed | Density of roads (2010 Census Tiger Lines) | km/sq km |
| RdDensCatRp100 | catchment, riparian | Density of roads (2010 Census Tiger Lines) | km/sq km |
| RdDensWsRp100 | watershed, riparian | Density of roads (2010 Census Tiger Lines) | km/sq km |
| RdCrsCat | catchment | Density of roads-stream intersections (2010 Census Tiger Lines-NHD stream lines) | crossings/sq km |
| RdCrsWs | watershed | Density of roads-stream intersections (2010 Census Tiger Lines-NHD stream lines) | crosssings/sq km |

Table 3 Performance of the landscape model by calibration and validation datasets in predicting CSCI scores. The statewide dataset (Figure 4) and individual regions of California (Figure 1) are evaluated. Averages and standard deviations (in parentheses) for observed and predicted CSCI values of each dataset are shown. Pearson correlations (r), root mean squared errors (RMSE), intercept, and slopes are for comparisons of predicted and observed values to evaluate model performance. All correlations, intercepts, and slopes are significant at alpha = 0.05.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset | Location | n | Observed | Predicted | r | RMSE | Intercept | Slope |
| Cal | Statewide | 1965 | 0.82 (0.26) | 0.83 (0.20) | 0.75 | 0.17 | 0.34 | 0.60 |
|  | CH | 512 | 0.76 (0.27) | 0.79 (0.21) | 0.71 | 0.19 | 0.38 | 0.54 |
|  | CV | 116 | 0.51 (0.18) | 0.57 (0.15) | 0.66 | 0.15 | 0.29 | 0.54 |
|  | DM | 86 | 0.87 (0.22) | 0.91 (0.14) | 0.50 | 0.20 | 0.63 | 0.31 |
|  | NC | 208 | 0.92 (0.20) | 0.94 (0.13) | 0.55 | 0.17 | 0.61 | 0.36 |
|  | SC | 631 | 0.79 (0.24) | 0.78 (0.21) | 0.75 | 0.16 | 0.27 | 0.65 |
|  | SN | 412 | 0.98 (0.18) | 0.98 (0.09) | 0.45 | 0.16 | 0.75 | 0.23 |
| Val | Statewide | 655 | 0.82 (0.25) | 0.84 (0.20) | 0.72 | 0.18 | 0.36 | 0.58 |
|  | CH | 172 | 0.76 (0.27) | 0.81 (0.21) | 0.74 | 0.19 | 0.39 | 0.56 |
|  | CV | 40 | 0.52 (0.19) | 0.59 (0.16) | 0.49 | 0.19 | 0.38 | 0.40 |
|  | DM | 28 | 0.84 (0.17) | 0.93 (0.11) | 0.55 | 0.17 | 0.63 | 0.36 |
|  | NC | 71 | 0.94 (0.19) | 0.96 (0.11) | 0.55 | 0.16 | 0.67 | 0.31 |
|  | SC | 208 | 0.80 (0.24) | 0.78 (0.21) | 0.72 | 0.17 | 0.27 | 0.63 |
|  | SN | 136 | 0.97 (0.17) | 0.98 (0.09) | 0.21 | 0.17 | 0.88 | 0.11 |

# References

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

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>.

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.

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