USGS
science for a changing world

National Water-Quality Assessment Project

Prepared in cooperation with The Nature Conservancy and Trout Unlimited

Estimating Natural Monthly Streamflows in California and the Likelihood of Anthropogenic Modification

Open-File Report 2016–1189

U.S. Department of the Interior
U.S. Geological Survey

Estimating Natural Monthly Streamflows in California and the Likelihood of Anthropogenic Modification

By Daren M. Carlisle, David M. Wolock, Jeanette K. Howard,
Theodore E. Grantham, Kurt Fesenmyer, Michael Wieczorek

National Water-Quality Assessment Project

Prepared in cooperation with The Nature Conservancy and Trout Unlimited

Open-File Report 2016–1189

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Director

U.S. Geological Survey, Reston, Virginia: 2016

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit http://www.usgs.gov or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit http://store.usgs.gov.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:
Carlisle, D.M., Wolock, D.M., Howard, J.K., Grantham, T.E., Fesenmyer, Kurt, and Wieczorek, Michael, 2016, Estimating natural monthly streamflows in California and the likelihood of anthropogenic modification: U.S. Geological Survey Open-File Report 2016–1189, 27 p., https://doi.org/10.3133/ofr20161189.

ISSN 2331-1258 (online)

Contents

Figures

Tables

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
	Length	
inch (in.)	2.54	centimeter (cm)
foot (ft)	0.3048	meter (m)
	Area	
acre	0.4047	hectare (ha)
	Volume	
acre-foot (acre-ft)	1,233.48	cubic meter (m^3)
	Flow rate	
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)
	Mass	
ton, short (2,000 lb)	0.9072	metric ton (t)
ton, long (2,240 lb)	1.016	metric ton (t)

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

$$°F = (1.8 \times °C) + 32.$$

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:

$$°C = (°F - 32) / 1.8.$$

International System of Units to U.S. customary units

Multiply	By	To obtain
	Length	
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)
kilometer (km)	0.5400	mile, nautical (nmi)
meter (m)	1.094	yard (yd)
centimeter (cm)	0.3937	inch (in.)
millimeter (mm)	0.03937	inch (in.)
	Volume	
cubic centimeter (cm^3)	0.06102	cubic inch (in^3)
	Flow Rate	
meter per second (m/s)	3.281	foot per second (ft/s)
	Area	
square kilometer (km^2)	247.1	acre
square kilometer (km^2)	0.3861	square mile (mi^2)
	Mass	
gram (g)	0.0353	ounce (oz)
	Application rate	
kilograms per square kilometer per year ([kg/km^2]/yr)	0.0089	pounds per acre per year ([lb/acre]/yr)

Datum

Vertical coordinate information is referenced to North American Vertical Datum of 1988 (NAVD 88).

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).

Estimating Natural Monthly Streamflows in California and the Likelihood of Anthropogenic Modification

Daren M. Carlisle,[1] David M. Wolock,[1] Jeanette K. Howard,[2] Theodore E. Grantham,[1] Kurt Fesenmyer,[3] Michael Wieczorek[1]

Abstract

Because natural patterns of streamflow are a fundamental property of the health of streams, there is a critical need to quantify the degree to which human activities have modified natural streamflows. A requirement for assessing streamflow modification in a given stream is a reliable estimate of flows expected in the absence of human influences. Although there are many techniques to predict streamflows in specific river basins, there is a lack of approaches for making predictions of natural conditions across large regions and over many decades. In this study conducted by the U.S. Geological Survey, in cooperation with The Nature Conservancy and Trout Unlimited, the primary objective was to develop empirical models that predict natural (that is, unaffected by land use or water management) monthly streamflows from 1950 to 2012 for all stream segments in California. Models were developed using measured streamflow data from the existing network of streams where daily flow monitoring occurs, but where the drainage basins have minimal human influences. Widely available data on monthly weather conditions and the physical attributes of river basins were used as predictor variables. Performance of regional-scale models was comparable to that of published mechanistic models for specific river basins, indicating the models can be reliably used to estimate natural monthly flows in most California streams. A second objective was to develop a model that predicts the likelihood that streams experience modified hydrology. New models were developed to predict modified streamflows at 558 streamflow monitoring sites in California where human activities affect the hydrology, using basin-scale geospatial indicators of land use and water management. Performance of these models was less reliable than that for the natural-flow models, but results indicate the models could be used to provide a simple screening tool for identifying, across the State of California, which streams may be experiencing anthropogenic flow modification.

Introduction

Natural variability in flow is a fundamental physical property of streams and therefore has major relevance to water quality and the health of riverine ecosystems (Poff and others, 1997). In the absence of human influence, the magnitude and duration of streamflows vary seasonally and annually, which constitutes the natural flow regime. The importance of the natural flow regime to maintaining ecological health in rivers and streams is well documented (Poff and Zimmerman, 2010). Modification of watershed hydrology and streamflows from human activity is pervasive in the United States (Poff and others, 2007; Eng and others, 2013b), and quantitative tools are needed to better understand the natural flow regime and to protect stream health.

Central to understanding the causes of poor stream health is the ability to determine the expected natural (we use the term "natural" to indicate the baseline or background condition unaffected by land use or water management) levels of physical and chemical characteristics of a stream, so that an objective assessment can be made as to which factors have been modified by human activities (Hawkins and others, 2010). For contaminants such as synthetic organic chemicals, natural levels in a stream are zero, so the presence of these chemicals can be unambiguously linked to anthropogenic sources. In contrast, anthropogenic modification of streamflows can be difficult to quantify because the natural background conditions are often highly variable temporally (for example, inter-annual) and spatially (for example, across a region or stream network). As a result, streamflow modification has been characterized in a wide variety of ways (Poff and Zimmerman, 2010), which limits the ability to synthesize and generalize how modified streamflows affect stream health and hinders development of standards aimed at restoration and protection of streams. The ability to estimate natural streamflows in a given region is therefore a critical tool for managers and decision makers, particularly in the face of increased water demand and a changing climate (Sabo and others, 2010).

Estimating flows in unmonitored streams (and by extension, estimating natural flows at monitored sites affected by hydrologic modification) is a major frontier in hydrological

[1]U.S. Geological Survey.

[2]The Nature Conservancy.

[3]Trout Unlimited.

science (Sivapalan, 2003; Sivapalan and others, 2003) and is accomplished with two general approaches: mechanistic and statistical models. Mechanistic models are not considered here, but there is a large amount of literature on published models typically developed for single river basins using process-based understanding. Such models are data intensive and likely are not practical as a predictive tool across large geographic regions. There is much less literature on statistical models (Farmer and Vogel, 2013), which include a wide range of methods reviewed elsewhere (He and others, 2011; Li and Sankarasubramanian, 2012; Shu and Ouarda, 2012; Farmer and Voge,l 2013; Shupe and Potter, 2014), than on mechanistic models.

Another needed management tool is the ability to identify where, across a state or other large geographic area, streamflows are likely to be modified, particularly in areas with sparse streamgaging networks. In most regions, streamflow monitoring is limited to a small subset of the stream network (Poff and others, 2006), largely because of the resources required for gage maintenance. An estimate of the probability of streamflow modification, given readily measured characteristics of a stream basin, would be a useful tool for screening all ungaged stream segments across a region (Eng and others, 2013a). Such a tool would allow decision makers to identify where modified flow, among the many other potential causes, is a likely contributor to poor stream health and where efforts to naturalize streamflows can have the greatest positive ecological outcome.

A study was conducted by the U.S. Geological Survey (USGS) in cooperation with The Nature Conservancy and Trout Unlimited, with the goal of developing statistical models for use in estimating natural streamflow. The purpose of this report is to describe the development of a series of statistical models that (1) predict natural monthly flows each year from 1950 to 2012 for California's streams and (2) predict the likelihood that monthly streamflows are modified by human activity.

Methods

Selection of Spatial Domain

The spatial domain of the study includes aggregated Level 3 Ecoregions (Commission for Environmental Cooperation, 2014) that are present partly or entirely within California. Level 3 Ecoregions represent contiguous geographic areas with similar climate, topography, and natural land cover, which are factors that affect spatial variation in natural streamflow regimes. Prior experience (Carlisle and others, 2010) indicates that statistical models developed at spatial scales for increasingly homogenous environmental settings (for example, similar climate and topography) were less sensitive to broad-scale climatic patterns and more sensitive to catchment-scale physical features, such as soils and geology, than models developed at spatial scales over heterogeneous environmental settings (for example, widely varying climate). In order to achieve balance between an adequate number of reference sites (see section "Identification of Reference Sites") and the environmental homogeneity of a region, Level 3 Ecoregions were aggregated by similar climatic conditions into three large regions (fig. 1): xeric (California Coastal Sage, Chaparral, and Oak Woodlands; Southern Baja California Pine-Oak Mountains; Central California Valley; Mojave Basin and Range; Sonoran Desert; and Central Basin and Range), interior mountains (Sierra Nevada, Eastern Cascades Slopes and Foothills), and north coastal mountains (Klamath Mountains, Coast Range).

The unit of observation for the models in this report is the stream segment and its entire upstream contributing watershed. As defined by the National Hydrography Dataset Version 1.0 (Horizon Systems, 2015), a segment is generally a section of stream bounded by a node (for example, a tributary) on each end. Most segments are less than (<)1 kilometer (km) in total length, and 135,119 segments were identified within the State of California.

General Modeling Approach

Two general principles guided model development. First, we used the reference-condition concept (Bailey and others, 2004), wherein a set of reference sites (that is, least disturbed by human influences) is used to develop models that are subsequently applied to non-reference sites (for example, where hydrologic disturbance is known or suspected) with the goal of predicting expected natural conditions. Second, the approach is based on statistical models of related observed data rather than mechanistic, process-based models (for example, Spruill and others, 2000; Croke and others, 2005). The statistical models contain two general types of predictor variables: (1) static variables that describe watershed features, such as topography, geology, and soils and (2) time-series variables of antecedent precipitation and air temperature. We emphasize that the period of hydrologic and climatic record for this report is 1950–2012. Specifically, the models "learned" the relations among watershed physical features, precipitation, air temperature, and streamflow using observed conditions at reference sites from 1950 to 2012, which has important implications for attempts to use these models in the context of climate variability and change. Finally, monthly mean flows were selected for modeling because they are easily communicated and represent magnitude and timing, which are attributes of the natural flow regime that are relevant to ecosystems and management (Kendy and others, 2012).

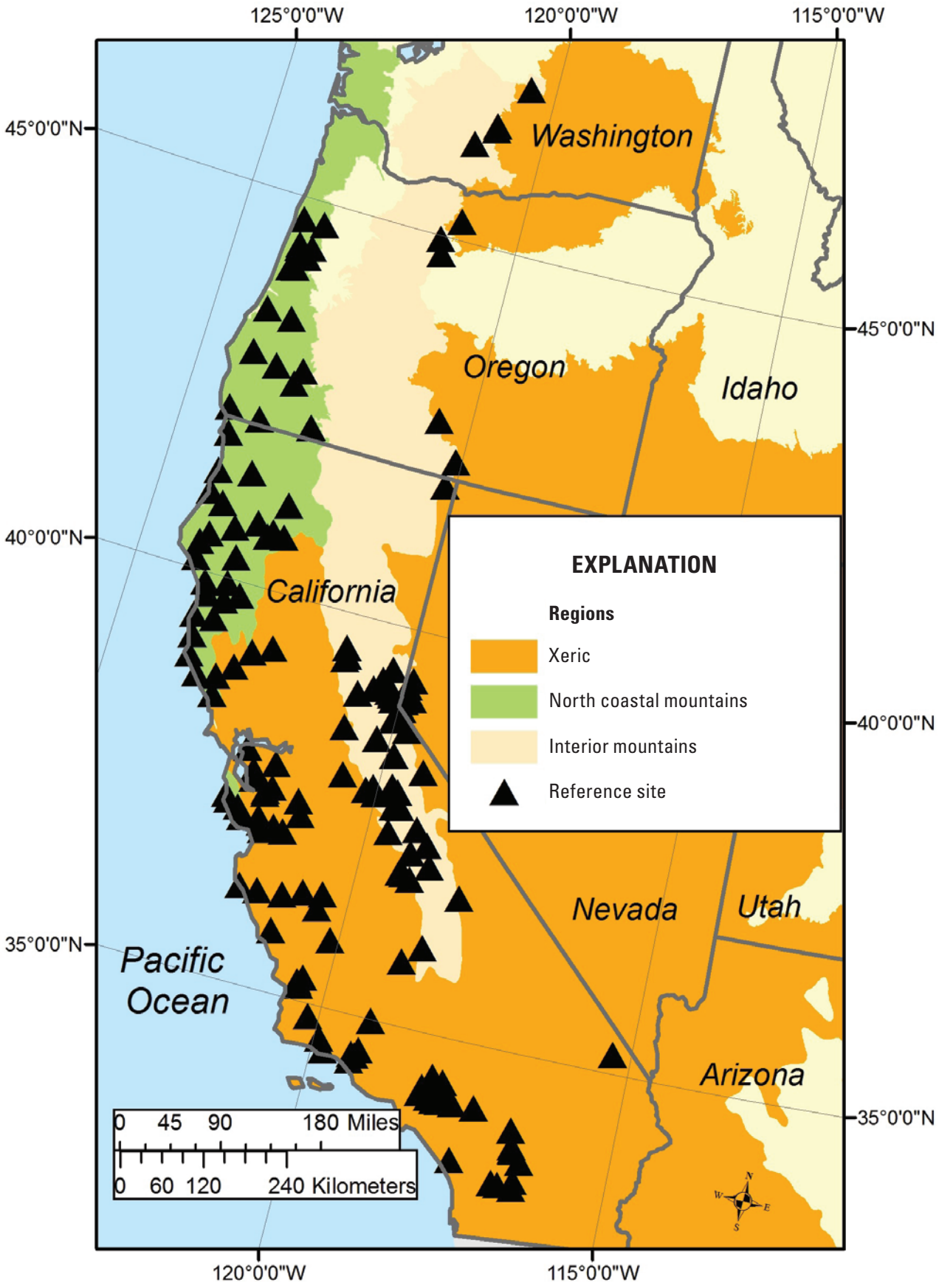

Figure 1. Locations of reference sites in gaged basins and extent of regions used to model streamflows in California.

Identification of Reference Sites

Reference sites were determined to be those river basins that are hydrologically least disturbed (see Stoddard and others, 2006) and where USGS streamgages measured daily streamflow for at least 20 years. Identification of reference sites was accomplished using a three-tiered approach. First, hydrologic disturbance was estimated for each gaged basin using an index that combined several geospatially derived indicators, including total upstream reservoir storage, freshwater withdrawal, pollution discharge, and impervious land cover (Falcone and others, 2010). All gaged basins within the geographic domain of the study were ranked on the value of this index score, and only those within the lower 25th percentile were considered as candidates for reference sites (see Falcone and others, 2010, for details of calculations).

The second tier of reference-site screening was examination of published site-description records. An annual data report is typically produced for each USGS streamgaging station and often contains information about anthropogenic influences on natural streamflow at that site. Notations indicating anthropogenic streamflow modification were considered a reason to classify a site as non-reference.

The third tier of screening was examination of the imagery of each site and its contributing drainage basin. Publicly available satellite imagery and topographic maps were examined for any indication of human activity with the potential to modify streamflows, such as diversions, irrigated agriculture, and wastewater inflows in close proximity to the streamgage. The screening process resulted in 50, 52, and 61 gaged reference basins and 86, 314, and 334 gaged non-reference basins for the north coastal mountains, interior mountains, and xeric regions, respectively. Reference basins had lower levels of water management and land development than disturbed basins but tended to have smaller drainage areas than non-reference basins (table 1).

Representativeness of Reference Sites

Because the streamgaging network in the United States was created by targeting basins where specific water information was needed, there is a legitimate concern as to whether gaged river basins are representative of all river basins within the stream network (Poff and others, 2006). This issue was addressed in two ways. First, three natural basin characteristics known to be important predictors (Carlisle and others, 2010) of flows (basin mean slope, mean precipitation, soil texture) were selected, then the data distributions of these variables for gaged reference sites (that is, those used in natural flow model development) were compared with those of the basins of all stream segments within each region. Second, three basin characteristics indicative of human disturbance (impervious land cover, irrigated agriculture, total reservoir storage) were selected, then the data distributions of these variables for the gaged non-reference sites were compared to those of the basins of all stream segments within each region. Overlap in the distributions of these variables between gaged sites and all stream segments provides a sense of confidence that models developed at gaged sites can reasonably be applied to all river basins in California.

Table 1. Ranges of environmental characteristics, as 1st and 99th percentiles, at reference and non-reference sites in gaged river basins within the north coastal mountains, interior mountains, and xeric regions, California.

[*n*, number; km^2, square kilometer]

Attribute	North coastal mountains region		Interior mountains region		Xeric region	
	Reference site (*n*=50)	Non-reference site (*n*=86)	Reference site (*n*=52)	Non-reference site (*n*=314)	Reference site (*n*=61)	Non-reference site (*n*=334)
Area (km^2)	12–1,962	10–8,382	3–1,758	6–21,145	5–656	7–19,779
Reservoir storage[a]	0–2	0–2,256	0–17	0–1,709	0–2	0–1,663
Impervious[b]	0–1	0–5	0–2	0–4	0–1	0–46
Crop land[c]	0–1	0–7	0–2	0–9	0–3	0–22

[a]Megaliters per square kilometer.

[b]Percent of basin land cover.

[c]Percent of basin land cover consisting of row crops.

Two limitations to the comparisons of basin characteristics were imposed. First, the comparisons were limited to non-gaged basins similar in size to those of gaged basins (table 1). This resulted in the exclusion of many small headwater streams that are present in the stream network but are not represented in the streamgaging network. The second limitation is that the comparisons were qualitative and univariate. Although formal quantitative methods are available for comparing multivarate distributions (for example, Bowman and Somers, 2006), these seemed inappropriate, given that the resulting thousands of statistical tests (for each segment in the stream network) would have limited interpretability.

In all regions, distributions of the six key variables (basin slope, mean precipitation, coarse soils, imperviousness, irrigated agriculture, and reservoir storage) overlapped considerably between reference basins and those of the stream network (appendix 1, figs. 1–1 to 1–3). These findings indicate that, from a univariate perspective, reference basins are largely representative of the natural and human-modified environmental settings of all stream basins in California that are 10–20,000 km^2 in total area.

Statistical Modeling Approach

Because a variety of machine-learning methods (Kuhn, 2008) and linear regression have been used to develop statistical models in hydrology (Farmer and Vogel, 2013), alternative modeling approaches were evaluated to determine which would be most optimal for use in this study. Within each region, predictive models were developed (detailed methods below) using reference sites and six different types of statistical models—five different machine-learning models and multiple linear regression. Detailed descriptions of each machine-learning model are provided in Kuhn and Johnson (2013); brief descriptions are provided here. Random forest (RF), general boosted regression (GBM), and Cubist (CUB) are rule-based methods related to classification and regression trees (Hastie and others, 2001). The major difference among these techniques is in how the tree-based models are constructed. RF and GBM build an ensemble of individual tree-based models that are collectively used to make predictions. In RF, each of these individual models is treated independently and contributes equally to the final predictions of the model. In contrast, GBM builds these individual models in sequence and weights their predictions according to their predictive ability. CUB generates a multiple linear regression equation for each partition of the independent variables identified via simple tree-based methods. Support vector machines are a form of nonlinear regression that are robust to outliers and provide flexible model-evaluation rules (Kuhn and Johnson, 2013). Neural networks are a form of nonlinear regression but with the outcome simulated by a set of unobserved variables that are constructed as linear combinations of observed variables (Kuhn and Johnson, 2013).

Most machine-learning models require user-selected settings of various fitting parameters, so we selected a wide range of possible parameter values (appendix 1, table 1–1) and tuned each model with 10-fold cross validation using the caret library (Kuhn, 2008) in R (R Core Team, 2014). For CUB, support vector machines, and neural network models, independent variables were first centered and rescaled, and highly ($|r| > 0.80$) collinear variables were removed (as recommended and described in Kuhn and Johnson, 2013). The tuned models were then re-applied to the reference sites in each region using leave-one-out cross validation. From the resulting data, model performance was measured with the squared correlation and root mean square error of observed and predicted values. Additional measures of model performance (the mean observed (O)/expected (E), and the standard deviation (sd) of O/E) were also computed.

Across all monthly models and regions, RF and CUB models performed substantively better than all other modeling approaches (fig. 2). Because RF predictions consistently exhibited slightly better precision (that is, lower sd of mean O/E) than CUB, we selected RF to generate predictions of natural flows, after additional refinement as described below.

Tree-based methods, such as RF, are a desirable modeling approach because they are free of assumptions that limit linear methods, and they accommodate complex interactions and non-linear relations among independent and dependent variables. Detailed descriptions of RF are given elsewhere (Cutler and others, 2007). In the interest of parsimony, we evaluated how the performance of RF models varied with increasing numbers of predictor (independent) variables—that is, model complexity. First, a full RF model was developed using all predictor variables (table 1–2). RF evaluates predictor variable importance by randomly permuting each predictor in turn, then measuring loss in model performance (Cutler and others, 2007). The relative loss in model performance is used to rank predictor variable importance; variables that cause the greatest loss in model performance, when randomized, are of highest importance. The top 20 important predictors were selected after running the full model. Then beginning with the highest ranking variable, a new RF model was constructed after successively adding each of the top 20 predictors, in turn. Model performance was examined (see description in section "Predicting Natural Flows: Model Development and Performance") for each of the 20 RF models; one was selected that balanced model performance with the least number of predictor variables, thus providing the most parsimonious model.

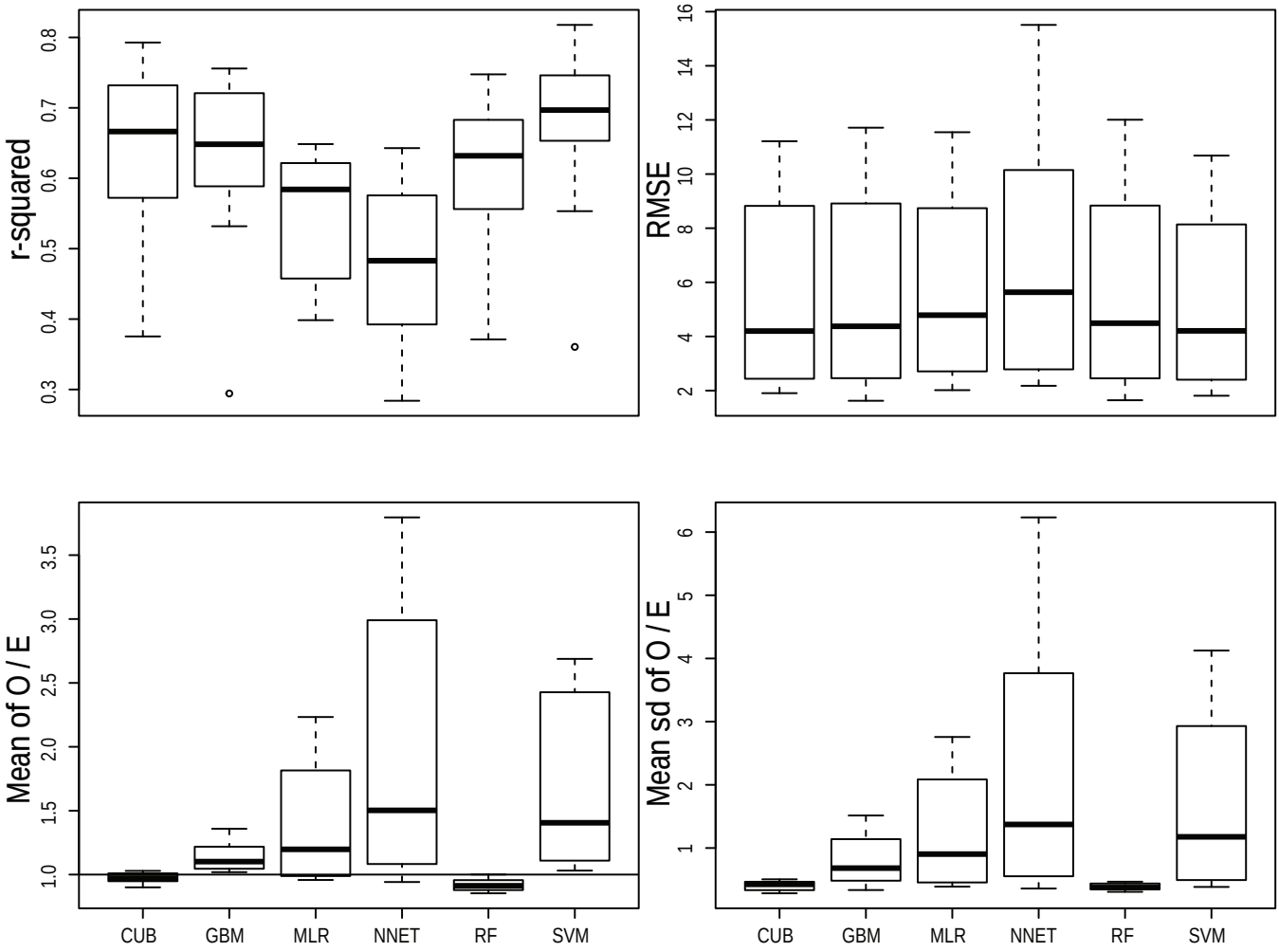

Figure 2. Performance of various machine-learning models for predicting natural monthly streamflows in California's xeric region. Plots for other regions are not shown but exhibited the same patterns of relative performance. Better models have higher r-squared, lower root mean square error (RMSE) and mean standard deviation (sd) of observed/expected (O/E) values, and a mean O/E value near 1. (CUB, cubist; GBM, boosted regression; MLR, multiple linear regression; NNET, neural network; RF, random forest; SVM, support vector machine.) r-squared is the correlation of predicted and observed values of monthly flows. RMSE is the root mean square error. Mean O/E is the average ratio of observed and predicted flows at each reference site. Mean sd of O/E is the mean of the standard deviation of O/E at each reference site.

Predicting Natural Flows: Model Development and Performance

For our first objective, a separate model for each month in each region (36 models) was developed to predict natural monthly flows for any specific year from 1950 to 2012. Measured monthly flow for each year was the dependent variable (U.S. Geological Survey, 2015). The predictor variables included a set of static, physical watershed characteristics and corresponding weather data (table 1–2; Falcone, 2011; Olson and Hawkins, 2014; PRISM Climate Group, 2014). These year-specific weather data included precipitation and air temperature for the month of interest and each of the previous 12 months (Daly and others, 2008). Estimated monthly runoff data from national-scale grids (McCabe and Wolock, 2011) were also used because these estimates indicate the balance between precipitation and evapotranspiration. In summary, the final data matrix for developing models of natural monthly flows included every year for which each reference site had a measured monthly flow value, the set of weather data and modeled runoff associated with each year's measured monthly flow and previous 12 months (39 predictors), and the set of static physical watershed characteristics (113 predictors, Falcone, 2011). The relations between the most influential predictors and the simulated outcome were graphically examined using partial-dependence plots (Cutler and others, 2007). This procedure evaluates how variation in each predictor affects the outcome while holding all other predictors constant (Hastie and others, 2001).

Model performance was evaluated by calculating several statistics (Moriasi and others, 2007) using the observed data and the expected (that is, predicted) monthly data generated by the internal bootstrapping performed by the RF model (Cutler and others, 2007). The squared correlation coefficient (r^2) between observed and predicted monthly flows across all sites was computed. The Nash-Sutcliffe coefficient of model efficiency (NSE) measures the total residual variance (that is, generated from model predictions) relative to the total variance within the data. NSE values near unity indicate that most of the total variance is accounted for by the model, indicating good model performance. Percent bias (PBIAS) estimates the model's tendency to over predict (PBIAS>0) or under predict (PBIAS<0). The root mean square error normalized by the standard deviation of all observations provides a standardized measure of model error. Finally, summary statistics for each site were calculated, including the mean (among years) O/E and the standard deviation (among years) of monthly O/E values.

Predicting the Likelihood of Modified Flows: Model Development and Performance

Objective two was to predict, using geospatial variables, the likelihood of anthropogenic modification of monthly streamflows. Models predicting modified flow were developed with a single dataset of all regions combined because by doing so we maximized the observed variation in affects from human activity, as well as the overall size of the dataset. Initial models for individual regions showed only marginal success in some regions, likely because of small ranges of several geospatial predictor variables. Finally, we had no reason to hypothesize that the relations between human activity factors (for example, freshwater withdrawal) and streamflows would vary by region.

Models described above were applied to all non-reference sites (total n=558) with recent flow records (1990–2010, which generally overlap the time periods of geospatial predictors) to generate a time series of natural monthly flows. Then, O/E was computed and averaged across years to produce a single value for the mean deviation of observed and expected natural flows for each month. Finally, each non-reference site was classified into one of three categories for each month on the basis of that month's mean O/E value: depleted, inflated, or unaltered. Depleted (O/E <0.75) indicates monthly flows that, on average, are reduced relative to natural conditions. Inflated (O/E >1.25) indicates monthly flows that, on average, are augmented relative to natural conditions. Unaltered (all other O/E values) indicates monthly flows that, on average, are similar to natural conditions. Thresholds for defining these categories are arbitrary but based upon a combination of statistical and interpretive reasoning. First, this threshold was within the range of precision (that is, average sd of O/E) of models predicting natural flows. Second, we evaluated model performance at a variety of thresholds and found that ±0.25 O/E units provided the best performance. Finally, a consistently applied threshold defined as a 25-percent reduction/addition of monthly flows is simple to comprehend and communicate.

For each month, two separate RF classification models were developed. One predicted depleted versus non-depleted flows (includes unaltered and inflated flows), and another predicted inflated versus non-inflated flows (includes unaltered and depleted flows). Predictor variables were limited to geospatial indicators of land and water management (table 1–3; Falcone, 2011; USGS, 2008a; U.S. Department of Agriculture, 2012; California Department of Water Resources, 2000; Grantham and others, 2014; USGS, 2008b; USGS 2013). As was done with the models of natural flow, parsimonius models were developed by evaluating model performance at varying levels of model complexity. Model performance was measured using the confusion matrix constructed with observations that were not used in model development (Cutler and others, 2007). The confusion matrix is the summary of the observed versus expected (predicted) classes of each observation used for model validation. Many measures have been proposed to summarize confusion matrices (Kuhn and Johnson, 2013),

each with its own strengths and weaknesses. Given our modeling objective, we saw no reason to favor one type of error over another. Failure to detect anthropogenic modification when it actually exists has negative consequences that may be no worse than the consequences of making false detections. Therefore, the percentage of observations that were correctly classified as altered (sensitivity), the percentage correctly classified as unaltered (specificity), and the kappa statistic as a measure of overall classification performance are reported. Kappa accounts for accuracy that would be generated simply by chance given the frequencies of each class in the data.

Results

Predicting Natural Flows

Model performance was marginally higher in both mountainous regions than in the xeric region and relatively consistent among months (fig. 3). For the xeric region (fig. 3*A*), typically more than 60 percent of the variation in observed flows was explained by the model (r^2, 0.41–0.88; NSE, 0.41–0.87), and bias was no more than 5 percent (PBIAS, -5 to -1). Mean O/E values were typically near unity (mean O/E, 0.90–0.98), and the sd of O/E indicated precision was typically 40 percent (sd O/E, 0.31–0.48). For the north coastal mountains (fig. 3*B*), typically more than 80 percent of the variation in observed flows was explained by the model (r^2, 0.84–0.96; NSE, 0.83–0.96), and bias was less than 5 percent (PBIAS, -3 to 2). Mean O/E values were typically near unity (mean O/E, 0.94–0.98), and the sd of O/E indicated precision was typically 29 percent (sd O/E, 0.24–0.34). For the interior mountains (fig. 3*C*), typically more than 70 percent of variation in observed flows was explained by the model (r^2, 0.79–0.96; NSE, 0.79–0.96) and bias was less than 5 percent (PBIAS, -4 to 4). Mean O/E was typically near unity (0.91–0.97), and sd of O/E indicated precision was typically 32 percent (sd O/E, 0.26–0.41).

The performance of statistical models was comparable to that of a wide range of other mechanistic and statistical approaches for monthly flow prediction. The NSE and PBIAS of the models were within the range of those achieved with statistical transfer methods (Farmer and Vogel, 2013). In addition, the r^2 and NSE of the models for the interior mountains region were comparable or slightly better than those (0.67 and 0.65, respectively) of a published mechanistic model for the Sierra Nevada Mountains (Shupe and Potter, 2014) and models for the Sacramento River (NSE, 0.48–0.82) (Ficklin and others, 2013).

Water balance-based runoff of the current month was an important predictor for all months and in all regions (figs. 1–4 to 1–6). Runoff (wb0-wb6) and precipitation (p0-p6, p2sum-p6sum) in the previous 1–6 months were also among the most important predictors for most models and in all regions. In addition to climatic variables, a variety of other physical attributes were important predictors of monthly flows. In the xeric region (fig. 1–4), basin mean slope (SLOPE), soil texture (NO10AVE), and elevation (ELEVATION) were important predictors, particularly for months when precipitation is typically low or nonexistent. In the north coastal mountains (fig. 1–5), precipitation intensity (RFACT) and overland flow (PERHOR), as well as sedimentary geology (sedimentary), were important predictors, particularly for dry months. Precipitation intensity and geologic properties frequently were important predictors for models in the interior mountains region in most months (fig. 1–6).

As expected, precipitation was the most important predictor of streamflow, but the affects of other watershed attributes is evidence that local physical factors affect the relation between precipitation and streamflow (fig. 4). Predicted flow typically increased monotonically with precipitation intensity (R-Factor), as well as with increasing precipitation (antecedent precipitation) and runoff (estimated runoff) in the target and preceding months. Predicted flow increased with increasing basin slope, which reflects the greater tendency for runoff than for infiltration on steeper slopes. In contrast, predicted flow decreased monotonically with the increasing extent of coarse soils, which indicates that greater infiltration in coarser soils results in lower runoff. Predicted flows tended to increase with increasing compressive strength of basin lithology, which indicates that rocks more resistant to weathering allow limited infiltration of precipitation to groundwater sources.

The models for natural flows lacked predictor variables that are direct measures of groundwater contributions to streamflow, but several surrogate variables frequently were important predictors, indicating that the models managed to capture part of this natural process. Antecedent monthly precipitation (2, 3, and 6 months) was an important variable for most months in all regions and may represent the lag time between precipitation and streamflow as a result of shallow groundwater recharge. Similarly, the average base-flow index (BFI) was an important predictor in the north coastal and interior mountains regions. The BFI was generated by a nationwide interpolation of observed streamflow data and represents a broad indicator of the degree to which groundwater contributes to streamflows (Wolock, 2003).

Models predicting natural streamflows could provide a useful baseline for future studies of how streamflows in California respond to changes in land use, water management, and climate. For example, a recent study (Grantham and others, 2014) used statistical models of natural flows combined with geospatial information about sensitive species to prioritize dams where targeted release strategies are likely to have the greatest ecological benefits. In addition, the ability to generate year-specific predictions of natural monthly streamflows will provide a foundation for examining how human activities influence streamflows and stream health, and how those effects may vary in time. For example, if natural monthly flows back to 1950 were generated for streams with long-term flow monitoring stations, trends in streamflow modification can be associated with trends in land use and water management over the last 60 years.

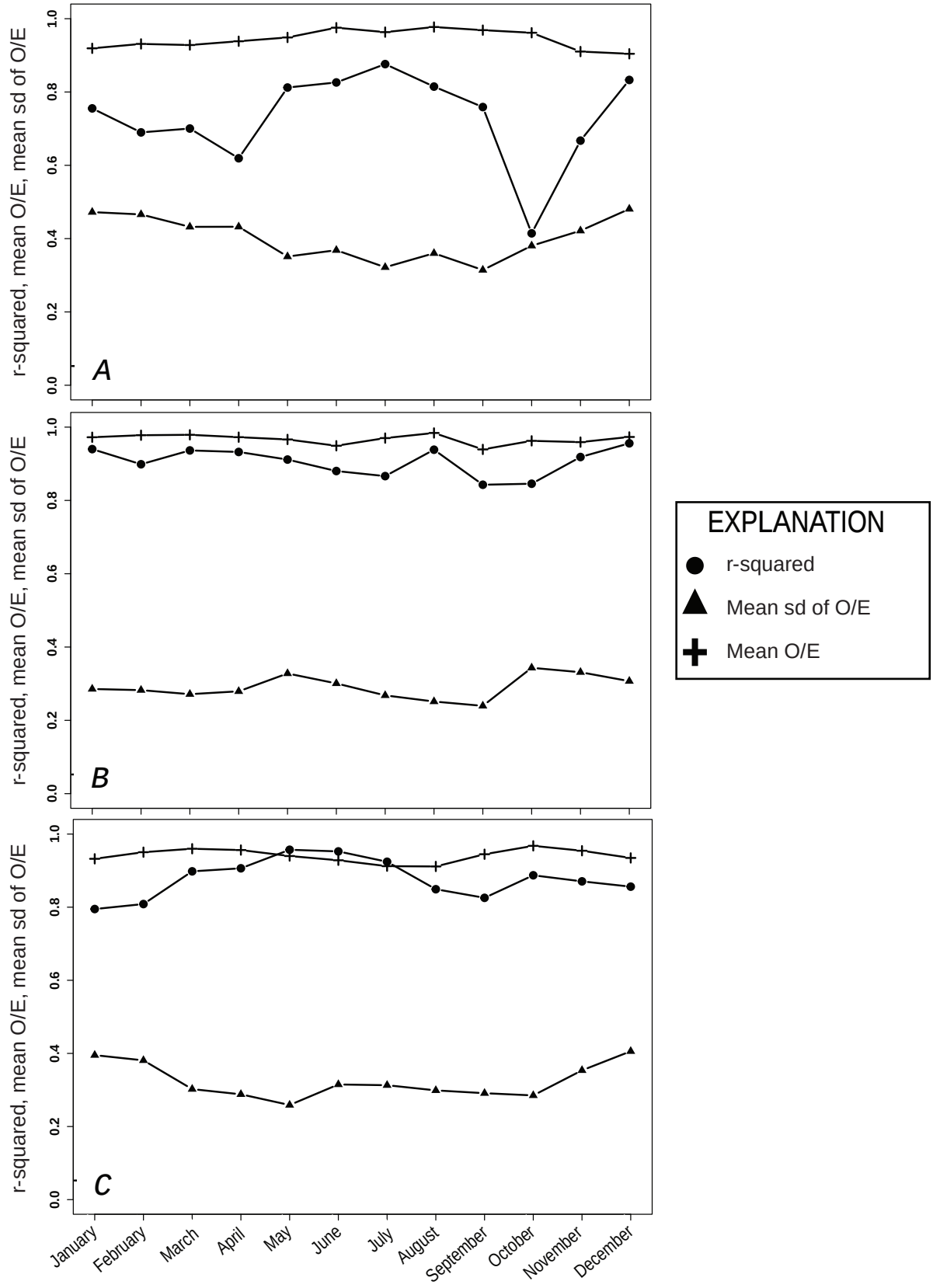

Figure 3. Performance statistics for models predicting natural monthly streamflows in the *A*, xeric; *B*, north coastal mountains; and *C*, interior mountains regions of California (r-squared is the correlation of predicted and observed values of monthly flows; sd, standard deviation; O/E, observed/expected).

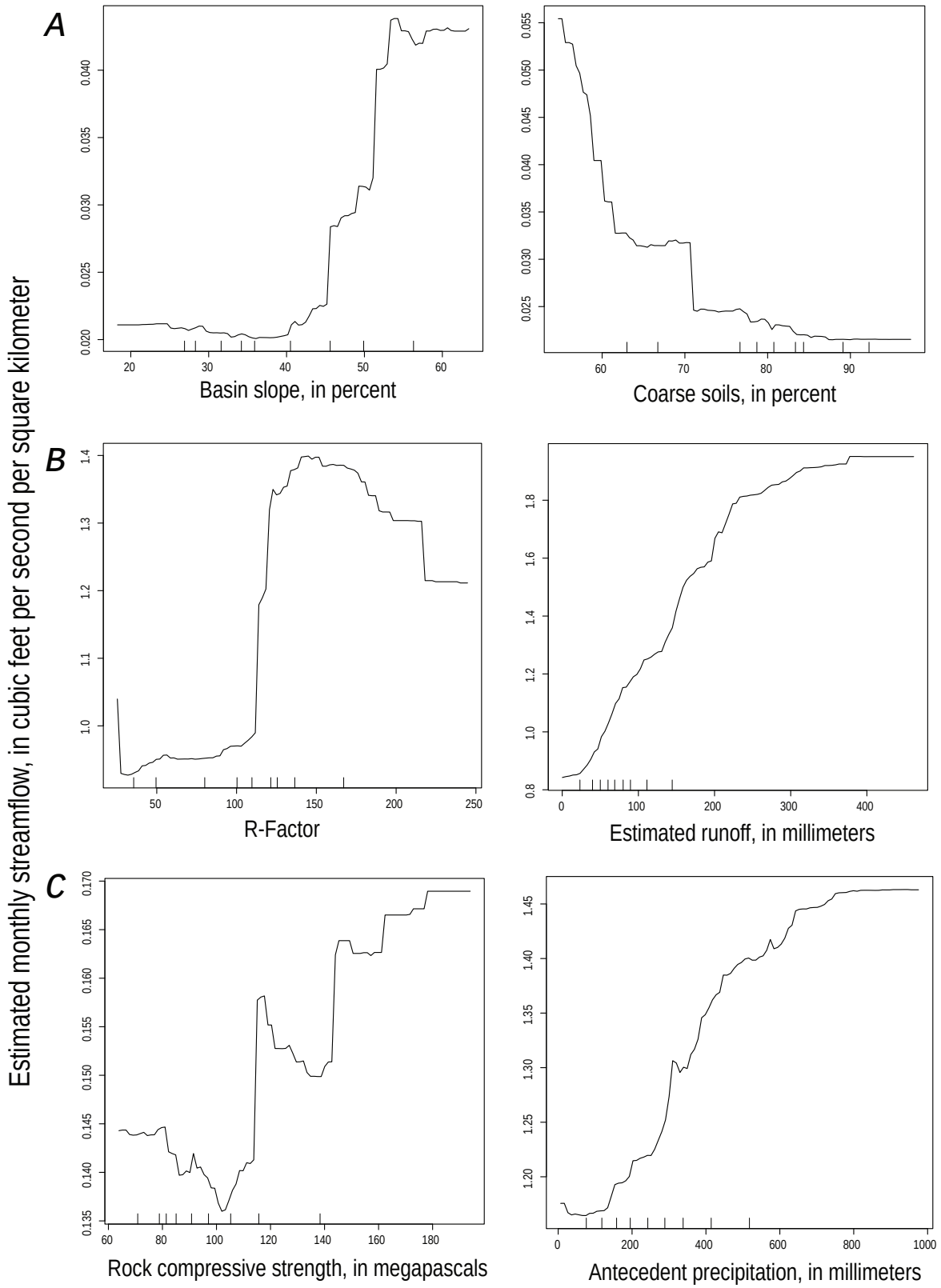

Figure 4. Partial dependence plots showing the relation of predicted natural monthly streamflow to selected predictor variables for the *A*, xeric; *B*, interior mountains; and *C*, north coastal mountains regions, California. Units for R-factor are in hundreds of foot-ton-inches per hour per acre. Estimated runoff is for the target month, and antecedent precipitation is the sum for the previous 6 months.

Predicting Modified Flows

Models predicting modified streamflows had a wide range of performance (fig. 5). Models predicting inflated monthly flows correctly classified, on average, 39 percent of altered sites (that is, sensitivity). The best model was for September (56 percent), and the worst model was for March (13 percent). Ninety percent of unaltered sites were correctly classified (that is, specificity), on average. The average kappa statistic was 0.34 (range 0.15–0.45), and the best models were those for May and June. Various measures of urban development (road stream crossings, impervious area) in the basin or riparian buffer were important predictors of inflated flows in all months (fig. 1–7).

Models predicting depleted monthly flows correctly classified 61 percent of altered sites, on average. The best models were for April (78 percent), and worst were for October (35 percent) (fig. 5). On average, 59 percent of unaltered sites were correctly classified. The average kappa statistic was 0.33 (range 0.25–0.46), and the best models were those for April and May. Various measures of urbanization in the basin were important predictors in all months, but riparian-buffer urban land cover, riparian vegetation height (riparian ht.), fertilizer application (P and N application; phosphate and nitrate, respectively) and freshwater withdrawal (withdrawal) were important predictors of depleted flows for 10 of 12 months (fig. 1–8).

Indicators of urbanization and water use were associated with inflated and depleted streamflows in opposite ways (fig. 6). The probability of inflated monthly flows increased dramatically with increasing impervious cover, which has been abundantly demonstrated in the literature (Paul and Meyer, 2001; Roy and others, 2005; Eng and others, 2013b), but tended to decrease with increasing freshwater withdrawal, which is an indicator of consumptive water use (Maupin and others, 2014). In contrast, the probability of depleted monthly flows increased with increasing freshwater withdrawal, which is also supported by a large body of literature (Jackson and others, 2001), but decreased with increasing urbanization.

The models predicting modified flows using geospatially derived indicators of influences from human activity at the watershed scale have one major limitation. The estimates of water use were based on State records of permitted diversions, which do not reflect the actual quantities of water that are consumptively used (for example, evaporation or export to other river basins) versus quantities returned to the stream or shallow groundwater. As a result, models performed relatively poorly, and typically various surrogates of actual water use (for example, agricultural intensity, impervious land cover) were found to be the best predictors of streamflow modification. Models likely would be improved with future enhancements of geospatially derived indicators of groundwater/surface-water interactions, actual consumptive water use, and return flows. Such data are notoriously difficult to obtain and quantify across wide geographic areas, but pilot programs in arid regions could be used to demonstrate the utility of such data collection efforts.

Although models for some months performed poorly, those for some months performed reasonably well and represent ecologically relevant hydrologic events such as May (spring flows as in Yarnell and others, 2010) and September flows (typically annual low flow). Potentially powerful management tools could be developed by combining predictions of modified streamflows across a large geographic area with other geospatial information, such as water use, sensitive species, or anticipated changes in precipitation owing to climate change.

Published sources of data used in this study and provided in tables 1–2 and 1–3 include the following: Falcone, 2011; Olson and Hawkins, 2014; PRISM Climate Group, 2014. In addition, monthly natural flow data for California stream segments (National Hydrography Dataset, Version 1) generated with models developed in this study are available at Carlisle and others, 2016. In addition, data used to develop models predicting modified flows (objective two) are available at the same source.

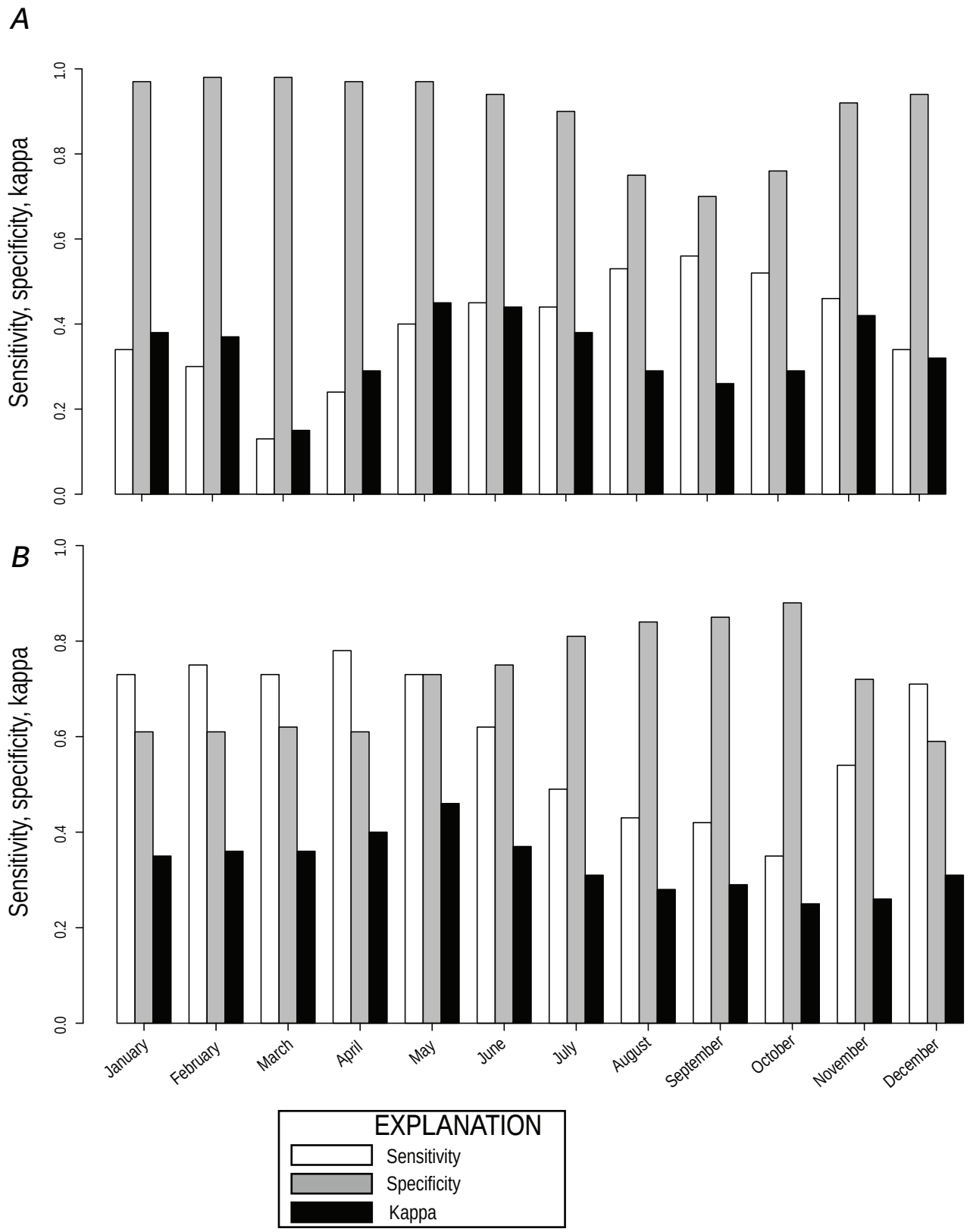

Figure 5. Performance measures for models predicting the probability of *A*, inflated and *B*, depleted flows in California streams.

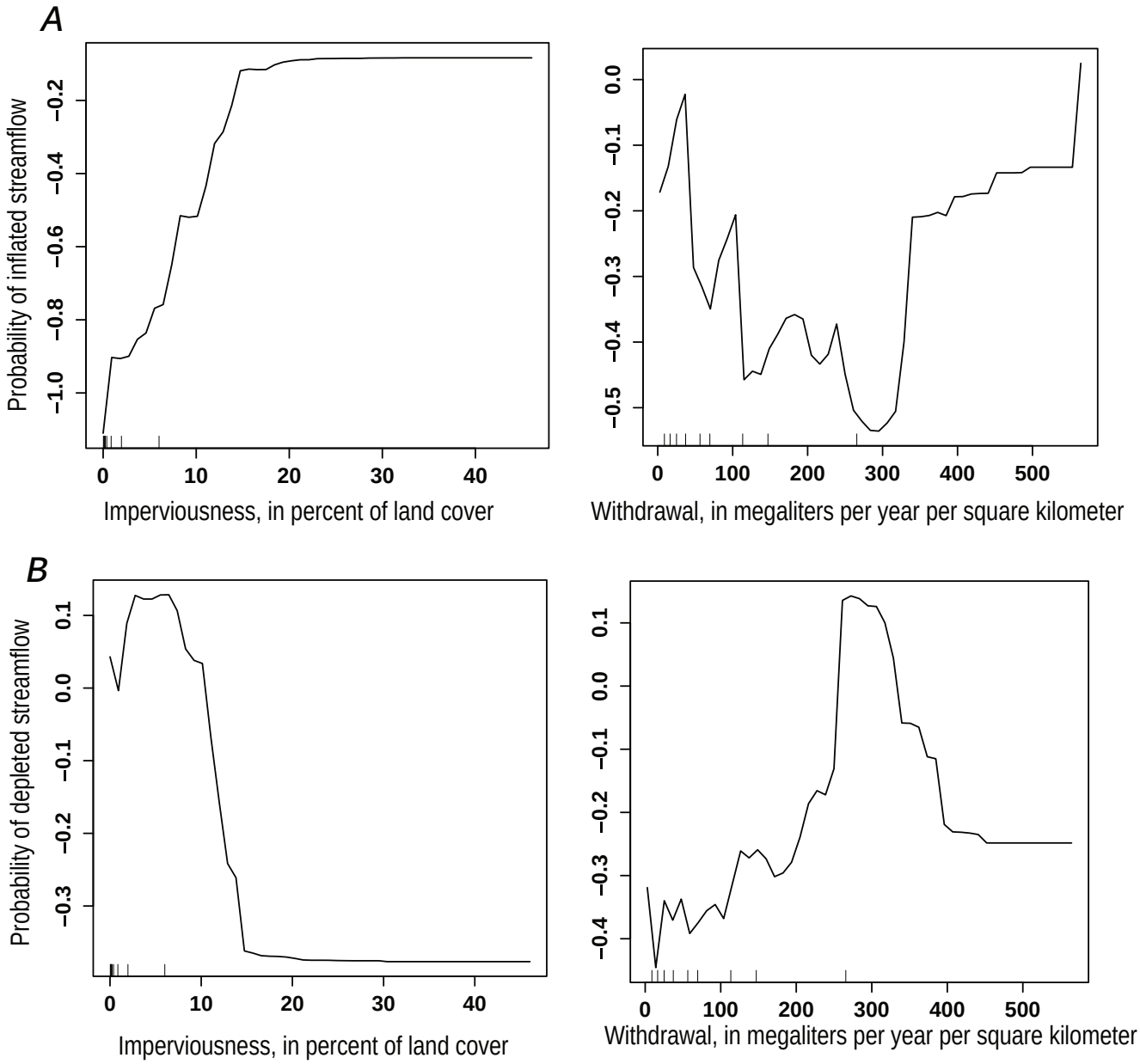

Figure 6. Partial dependence plots showing how the probability (logit of probability / 2) of *A*, inflated and *B*, depleted June streamflows are related to key predictor variables.

Summary

In a study conducted by the U.S. Geological Survey, in cooperation with The Nature Conservancy and Trout Unlimited, models developed to estimate natural monthly flows performed well and should provide a useful baseline for future studies of how streamflows in California respond to changes in land use, water management, and climate. For example, a recent study used statistical models of natural flows combined with geospatial information about sensitive species to prioritize dams where targeted release strategies are likely to have the greatest ecological benefits. In addition, the ability to generate year-specific predictions of natural monthly streamflows will provide a foundation for examining how human activities influence streamflows and stream health, and how those effects may vary in time. For example, if natural monthly flows back to 1950 were generated for streams with long-term flow monitoring stations, trends in streamflow modification can be associated with trends in land use and water management over the last 60 years.

The models that predict the likelihood of modified streamflows performed less reliably than those for natural streamflows but may nevertheless be useful as a general screening tool. Although models for some months performed poorly, those for selected months performed reasonably well and represent ecologically relevant hydrologic events such as May (spring flows) and September flows (typically annual low flow). Potentially powerful management tools could be developed by combining predictions of modified streamflows across a large geographic area with other geospatial information, such as water use, sensitive species, or anticipated changes in precipitation owing to climate change.

Models predicting natural and modified flows likely would be improved with future enhancements of geospatially derived indicators of groundwater/surface-water interactions, actual consumptive water use, and return flows. Such data are notoriously difficult to obtain and quantify across wide geographic areas, but pilot programs in arid regions could be used to demonstrate the utility of such data collection efforts.

Acknowledgments

The authors thank James Falcone of the U.S. Geological Survey (USGS) for providing advice on the use of a geographic information system throughout the project. Jason May and Rodney Knight of the USGS are thanked for providing comments that improved the clarity of the report.

References Cited

Bailey, R.C., Norris, R.H., and Reynoldson, T., 2004, Bioassessment of freshwater ecosystems—Using the Reference Condition Approach: New York, Springer.

Bowman, M.F., and Somers, K.M., 2006, Evaluating a novel Test Site Analysis (TSA) bioassessment approach: Journal of the North American Benthological Society, no. 25, p. 712–727.

California Department of Water Resources, 2000, 2000 Irrigated crop acres and water use, http://www.water.ca.gov/landwateruse/anaglwu.cfm.

Carlisle, D.M., Falcone, M.J., Wolock, D.M., Meador, M.R., and Norris, R.H., 2010, Predicting the natural flow regime—Models for assessing hydrological alteration in streams: River Research and Applications, v. 26, p. 118–136.

Carlisle, D.M., Wolock, D.M., Howard, J.K., Grantham, T.E., Fesenmeyer, K.A., and Wieczorek, M.E., 2016, Empirical Models for Estimating Baseline Streamflows in California and their Likelihood of Anthropogenic Modification: U.S. Geological Survey data release, http://dx.doi.org/10.5066/F7MP51DS.

Commission for Environmental Cooperation, 2014, Ecological regions of North America: Quebec, Canada, Commission for Environmental Cooperation, 71 p.

Croke, B., Andrew, F., Spate, J., and Cuddy, S., 2005, IHACRES User Guide, Technical Report 2005/19: Canberra, Australia, The Australian National University, iCAM, School of Resources, Environment and Society, p. 45.

Cutler, D.R., Edwards, T.C., Jr., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., and Lawler, J.J., 2007, Random forests for classification in ecology: Ecology, v. 88, p. 2783–2792.

Daly, C., Halbleib, M., Smith, J.I., Gibson W.P., and Doggett M.K., 2008, Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States: International Journal of Climatology, v. 28, p. 2031–2064.

Eng, K., Carlisle, D.M., Wolock, D.M., and Falcone, J.A., 2013a, Predicting the likelihood of altered streamflows at ungauged rivers across the conterminous United States: River Research and Applications, v. 29, p. 781–791.

Eng, K., Wolock, D.M., and Carlisle, D.M., 2013b, River flow changes related to land and water management practices across the conterminous United States: Science of Total Environment, v. 463–464, p. 414–422.

Falcone, J.A., Carlisle, D.M., Wolock, D.M., and Meador, M.R., 2010, GAGES—A stream gage database for evaluating natural and altered flow conditions in the conterminous United States: Ecology, v. 91, p. 621.

Falcone, J.A., 2011, GAGES-II—Geospatial Attributes of Gages for Evaluating Streamflow, http://water.usgs.gov/GIS/metadata/usgswrd/XML/gagesII_Sept2011.xml.

Farmer, W.H., and Vogel, R.M., 2013, Performance-weighted methods for estimating monthly streamflow at ungauged sites: Journal of Hydrology, v. 477, p. 240–250.

Ficklin, D.L., Luo, Y., and Zhang, M., 2013, Watershed modelling of hydrology and water quality in the Sacramento River watershed, California: Hydrological Processes, v. 27, p. 236–250.

Grantham, T.E., Viers, J.H., and Moyle, P.B., 2014, Systematic screening of dams for environmental flow sssessment and implementation: BioScience, v. 64, p. 1006–1018.

Hastie, T., Tibshirani, R., and Friedman, J., 2001, The elements of statistical learning: New York, Springer, p. 745.

Hawkins, C.P., Olson, J.R., and Hill, R.A., 2010, The reference condition—Predicting benchmarks for ecological and water-quality assessments: Journal of the North American Benthological Society, v. 29, p. 312–343.

He, Y., Bárdossy, A., and Zehe, E., 2011, A review of regionalisation for continuous streamflow simulation: Hydrology and Earth System Sciences, v. 15, p. 3539–3553.

Horizon Systems, 2015, National Hydrography Dataset Plus: Horizon Systems Corporation, accessed June 1, 2015, at http://www.horizon-systems.com/nhdplus/.

Jackson, R.B., Carpenter, S.R., Dahm, C.N., McKnight, D.M., Naiman, R.M., Postel, S.L., and Running, S.W., 2001, Water in a changing world: Ecological Applications, v. 11, p. 1027–1045.

Kendy, E., Apse, C., and Richardson, A., 2012, A practical guide for environmental flows for policy and planning—With nine case studies in the United States: The Nature Conservency, Charlottesville, Virginia, p. 72.

Kuhn, M., 2008, Building predictive models in R using the caret package: Journal of Statistical Software, v. 28, p. 1–26.

Kuhn, M., and Johnson, K., 2013, Applied Predictive Modeling: New York, Springer, 600 p.

Li, W., and Sankarasubramanian, A., 2012, Reducing hydrologic model uncertainty in monthly streamflow predictions using multimodel combination: Water Resources Research, v. 48, W12516.

Maupin, M.A., Kenny, J.F., Hutson, S.S., Lovelace, J.K., Barber, N.L., and Linsey, K.S., 2014, Estimated use of water in the United States in 2010: U.S. Geological Survey Circular 1405, 56 p.

McCabe, G.J., and Wolock, D.M., 2011, Independent effects of temperature and precipitation on modeled runoff in the conterminous United States: Water Resources Research, v. 47, p. W11522.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., and Veith, T.L., 2007, Model evaluation guidelines for systematic quantification of accuracy in watershed simulations: Transactions of the Asabe, v. 50, p. 885–900.

Olson, J.R., and Hawkins, C.P., 2014, Geochemical and Geophysical Characteristics of the Conterminous United States, http://dx.doi.org/10.1029/2011WR011088.

Paul, M.J., and Meyer, J.L., 2001, Streams in the urban landscape: Annual Review of Ecology and Systematics, v. 32, p. 333–365.

Poff, N.L., Allan, J.D., Bain, M.B., Karr, J.R., Prestegaard, K.L., Richter, B.D., Sparks, R.E., and Stromberg. J.C., 1997, The natural flow regime: BioScience, v. 47, p. 769–784.

Poff, N.L., Bledsoe, B.P., and Cuhaciyan, C.O., 2006, Hydrologic variation with land use across the contiguous United States: Geomorphic and ecological consequences for stream ecosystems: Geomorphology, v. 79, p. 264–285.

Poff, N.L., Olden, J.D., Merritt, D.M., and Pepin, D.M., 2007, Homogenization of regional river dynamics by dams and global biodiversity implications—Proceedings of the National Academy of Sciences of the United States of America, v. 104, p. 5732–5737.

Poff, N. L., and Zimmerman, J.K.H., 2010, Ecological responses to altered flow regimes—A literature review to inform the science and management of environmental flows: Freshwater Biology, v. 55, p. 194–205.

PRISM Climate Group, Oregon State University, accessed January 2014, at http://prism.oregonstate.edu.

R Core Team, 2014, R—A language and environment for statistical computing: Vienna, Austria, R Foundation for Statistical Computing.

Roy, A.H., Freeman, M.C., Freeman, B.J., Wenger, S.J., Ensign, W.E., and Meyer, J.L., 2005, Investigating hydrologic alteration as a mechanism of fish assemblage shifts in urbanizing streams: Journal of the North American Benthological Society, v. 24, p. 656–678.

Sabo, J.L., Sinha, T., Bowling, L.C., Schoups, G.H., Wallender,W.W., Campana,M.E., Cherkauer, K.A., Fuller, P.L., Graf, W.L., Hopmans, J.W., Kominoski, J.S., Taylor, C., Trimble, S.W., Webb, R.H., and Wohl, E.E., 2010, Reclaiming freshwater sustainability in the Cadillac Desert: Procedures of the National Academy of Sciences of the United States, v. 107, p. 21263–21270.

Shu, C., and Ouarda, T.B.M.J., 2012, Improved methods for daily streamflow estimates at ungauged sites: Water Resources Research, v. 48.

Shupe, J., and Potter, C., 2014, Modeling discharge rates using a coupled modeled approach for the Merced River in Yosemite National Park: Journal of the American Water Resources Association, v. 50, p. 153–162.

Sivapalan, M., 2003, Prediction in ungauged basins—A grand challenge for theoretical hydrology: Hydrological Processes, v. 17, p. 3163–3170.

Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., Mendiondo, E.M., O'Connell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D., Uhlenbrook, S., and Zehe, E., 2003, IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012—Shaping an exciting future for the hydrological sciences: Hydrological Sciences Journal, v. 48, p. 857–880.

Spruill, C.A., Workman, S.R., and Taraba, J.L., 2000, Simulation of daily and monthly stream discharge from small watersheds using the SWAT model: Transactions of the ASAE, v. 43, p. 1431–1439.

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

U.S. Department of Agriculture, National Agricultural Statistics Service, 2012, CropScape—Cropland Data Layer, http://nassgeodata.gmu.edu/CropScape/.

U.S. Geological Survey, 2008a, http://sagemap.wr.usgs.gov/HumanFootprint.aspx.

U.S. Geological Survey, 2008b, LANDFIRE Existing Vegetation Height (Refresh—LF 1.1.0), http://www.landfire.gov/viewer/.

U.S. Geological Survey, 2013, Square-mile cells that represent proprietary gas-producing wells from shale intervals in the United States, http://energy.usgs.gov/OilGas/AssessmentsData/NationalOilGasAssessment.aspx.

U.S. Geological Survey, 2014, The National Map, http://nationalmap.gov/.

U.S. Geological Survey, 2015, National Water Information System—Web interface, accessed September 28, 2015, at http://dx.doi.org/10.5066/F7P55KJN.

Wolock, D.M., 2003, Base-flow index grid for the conterminous United States: U.S. Geological Survey Open-File Report 2003–263.

Yarnell, S.M., Viers, J.H., and Mount, J.F., 2010, Ecology and Management of the Spring Snowmelt Recession: BioScience, v. 60, p. 114–127.

Appendix 1. Supplemental Information

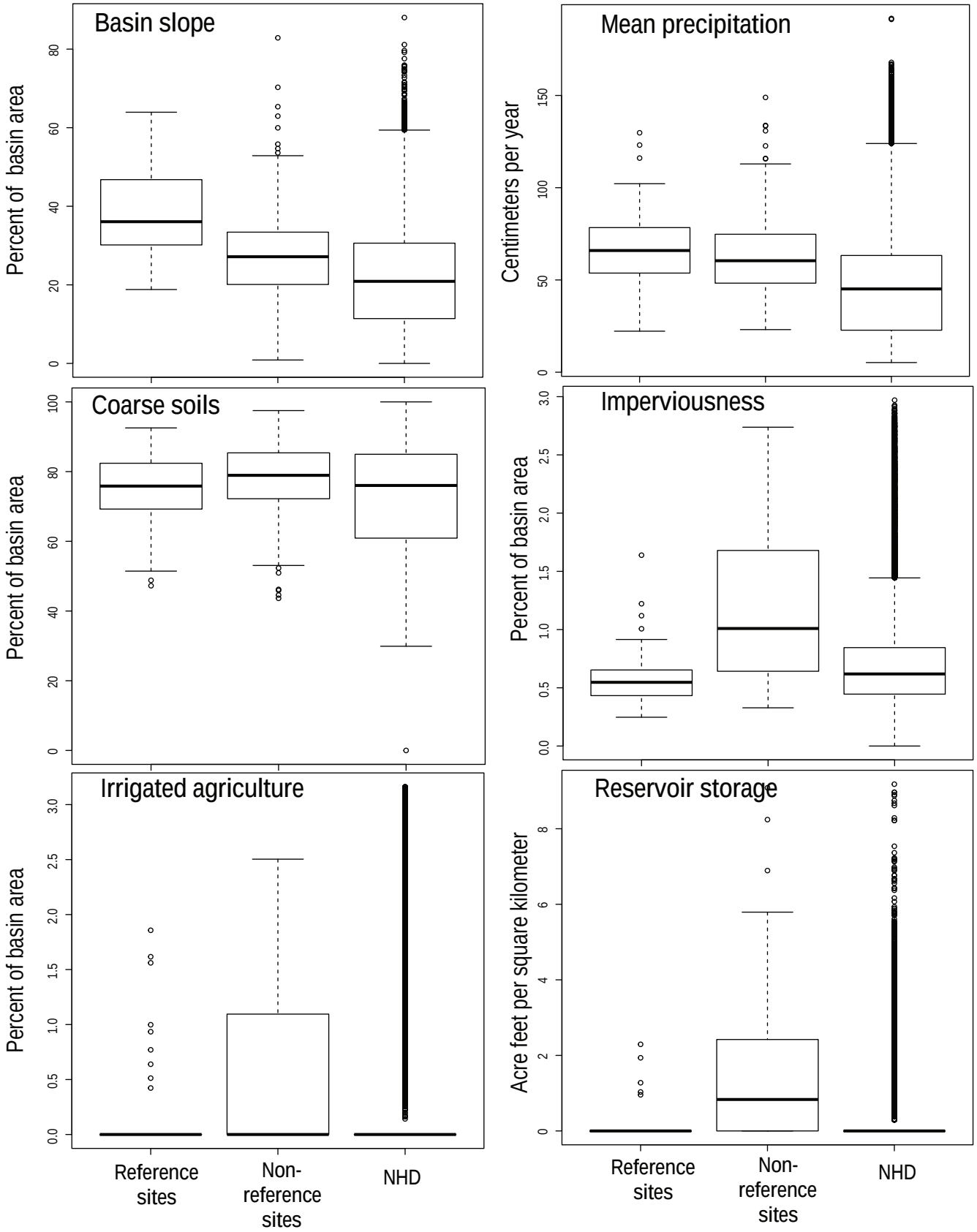

Figure 1–1. Representativeness of gaged basins used in model development relative to all stream segments (NHD= all segments in the National Hydrography Dataset) in the xeric region of California. Descriptions of variables are provided in tables 1–1 and 1–2.

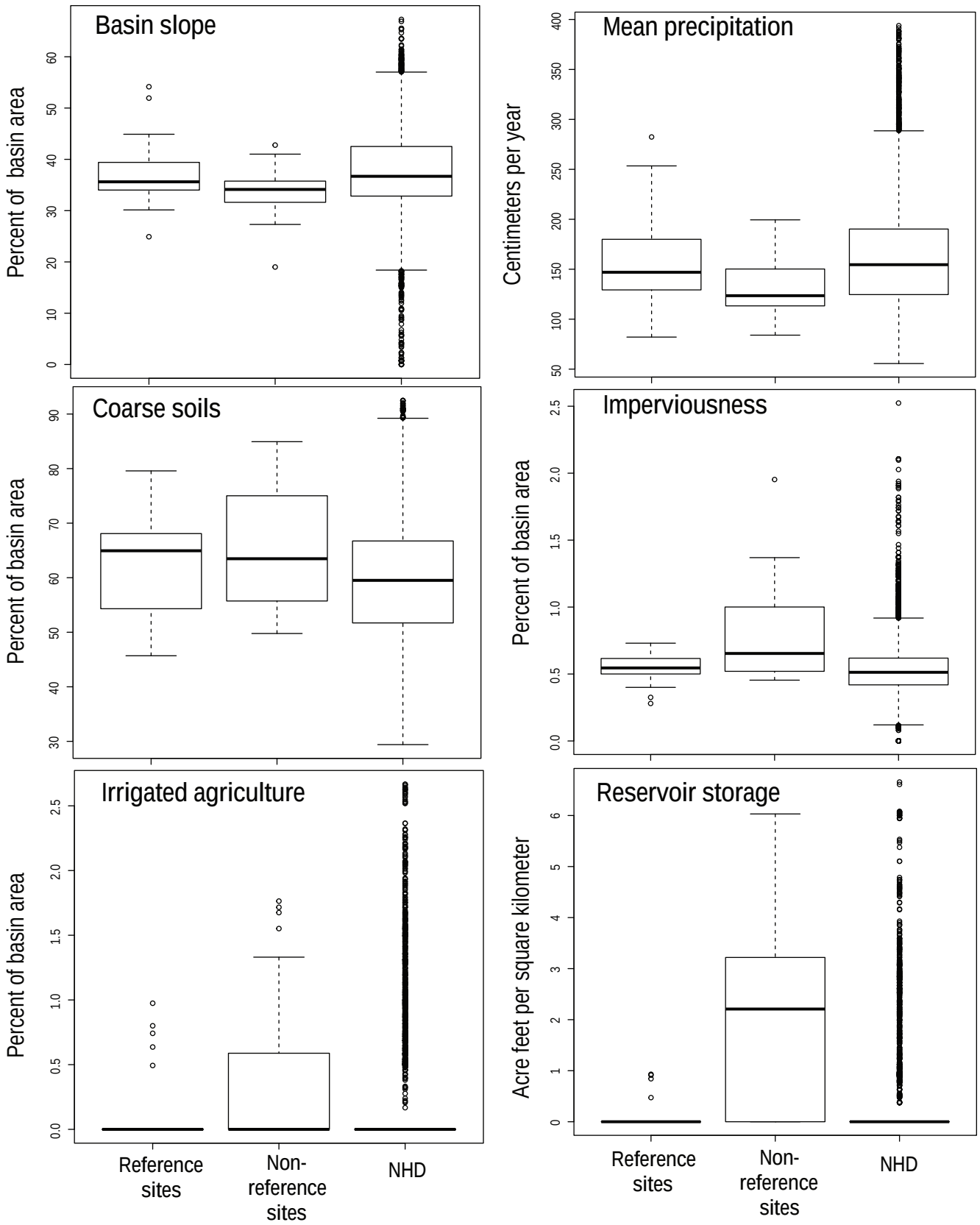

Figure 1–2. Representativeness of gaged basins used in model development relative to all stream segments (NHD) in the north coastal mountains region of California. Descriptions of variables are provided in tables 1–1 and 1–2.

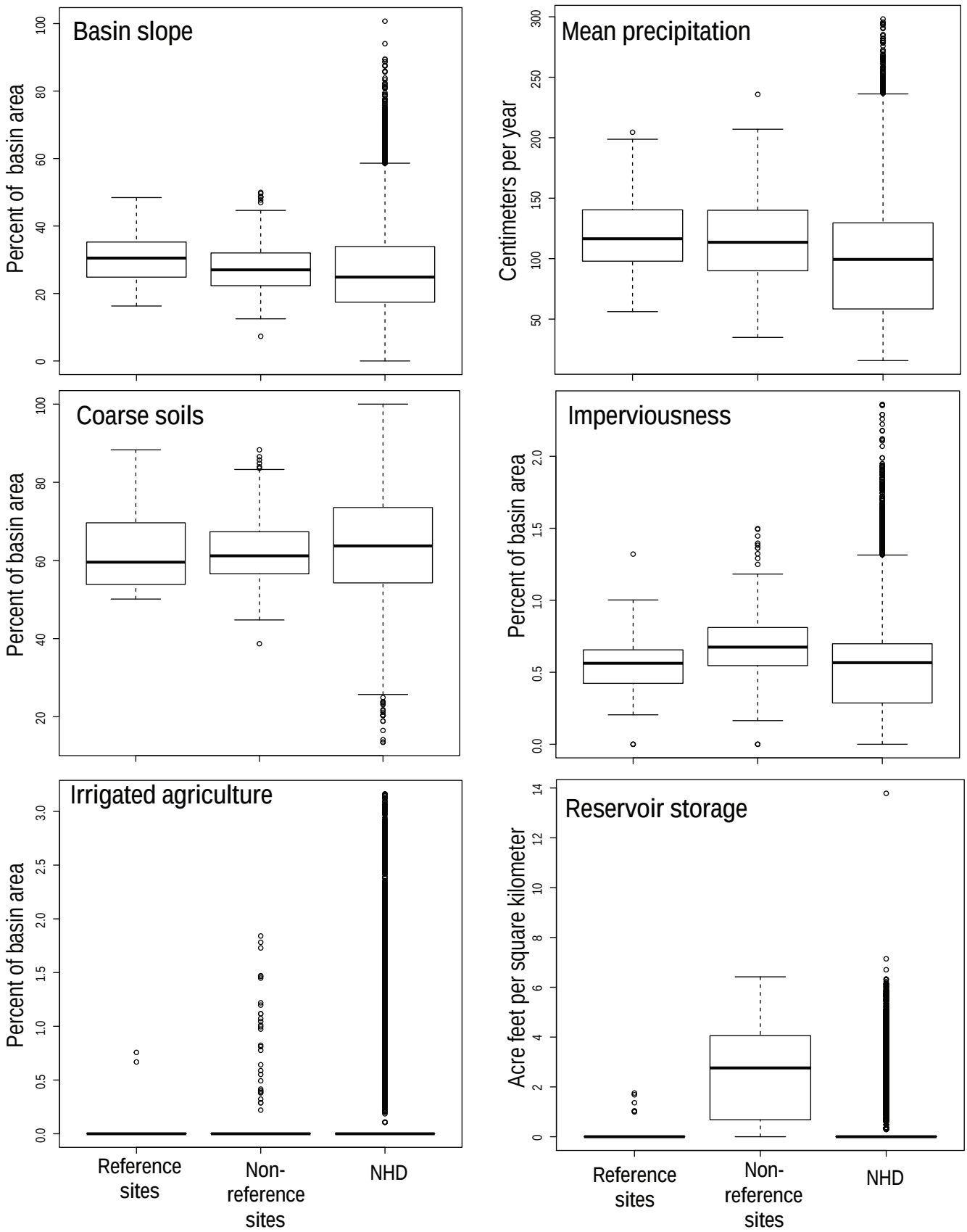

Figure 1–3. Representativeness of gaged basins used in model development relative to all stream segments (NHD) in the interior mountains region of California. Descriptions of variables are provided in tables 1–1 and 1–2.

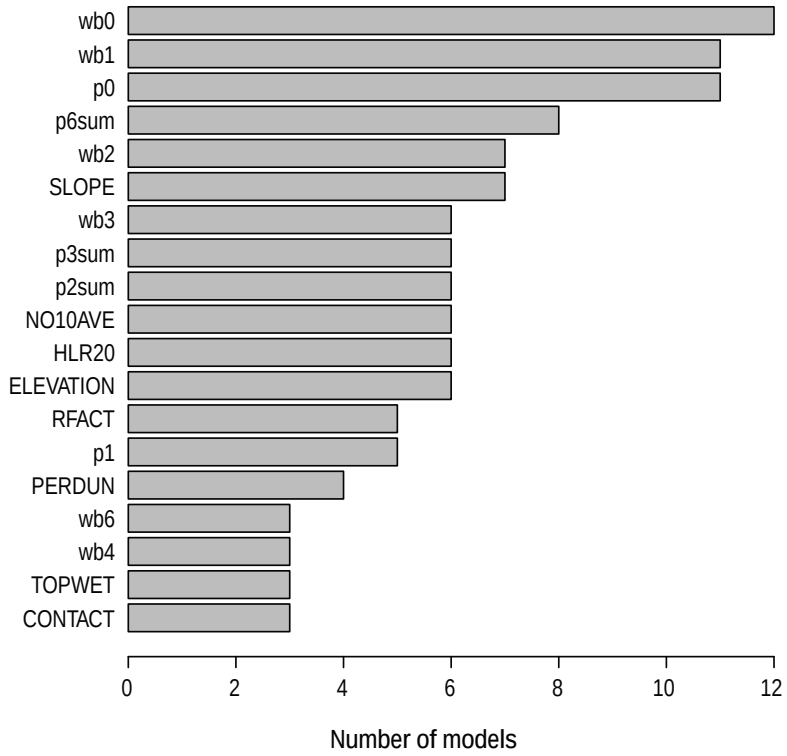

Figure 1–4. Occurrence of variables as important predictors in models of monthly streamflows in the xeric region of California. Descriptions of variables are provided in tables1–1 and 1–2.

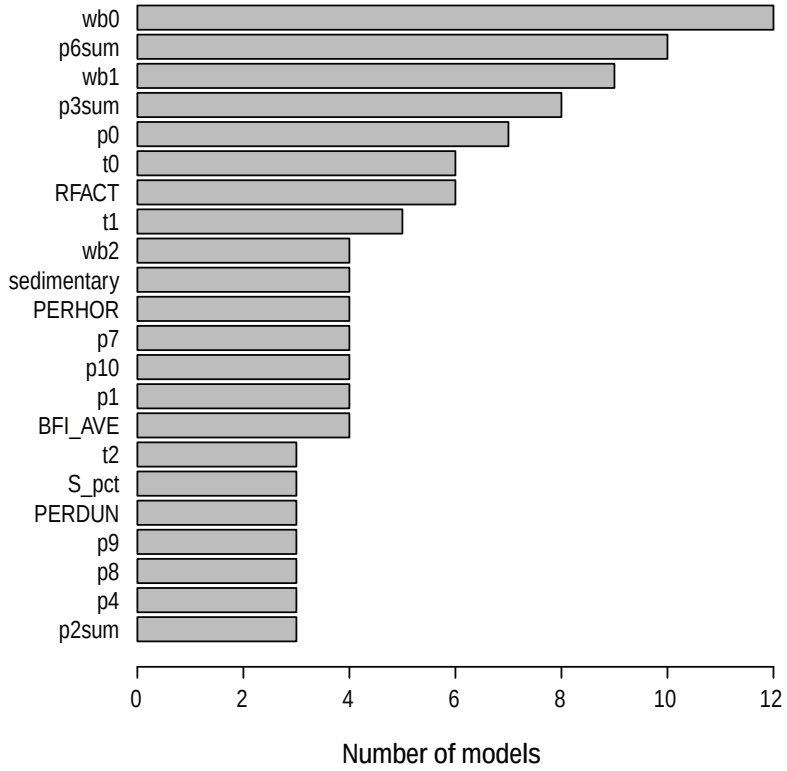

Figure 1–5. Occurrence of variables as important predictors in models of monthly streamflows in the north coastal mountains region of California. Descriptions of variables are provided in tables1–1 and 1–2.

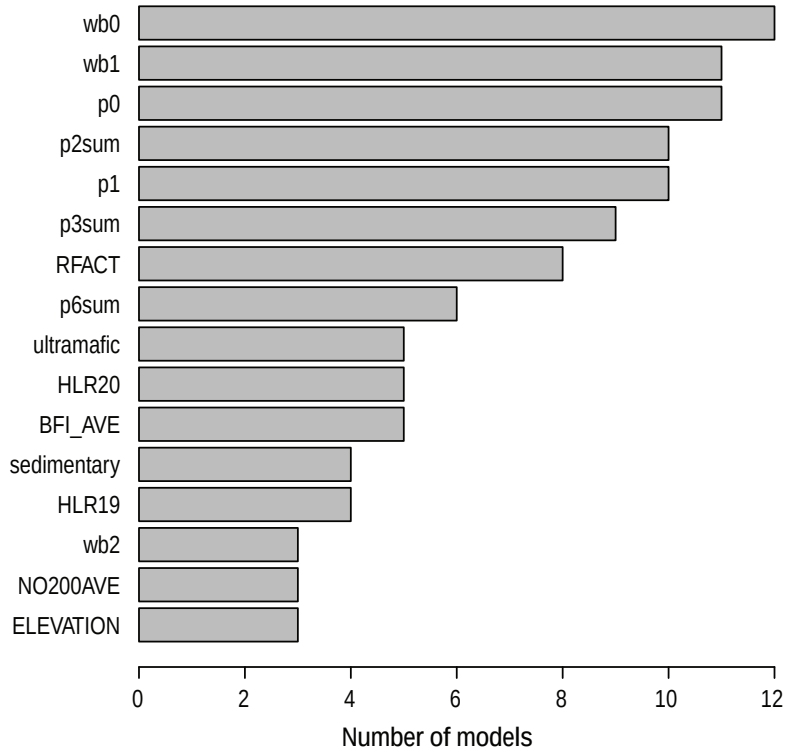

Figure 1–6. Occurrence of variables as important predictors in models of monthly streamflows in the interior mountains region of California. Descriptions of variables are provided in tables 1–1 and 1–2.

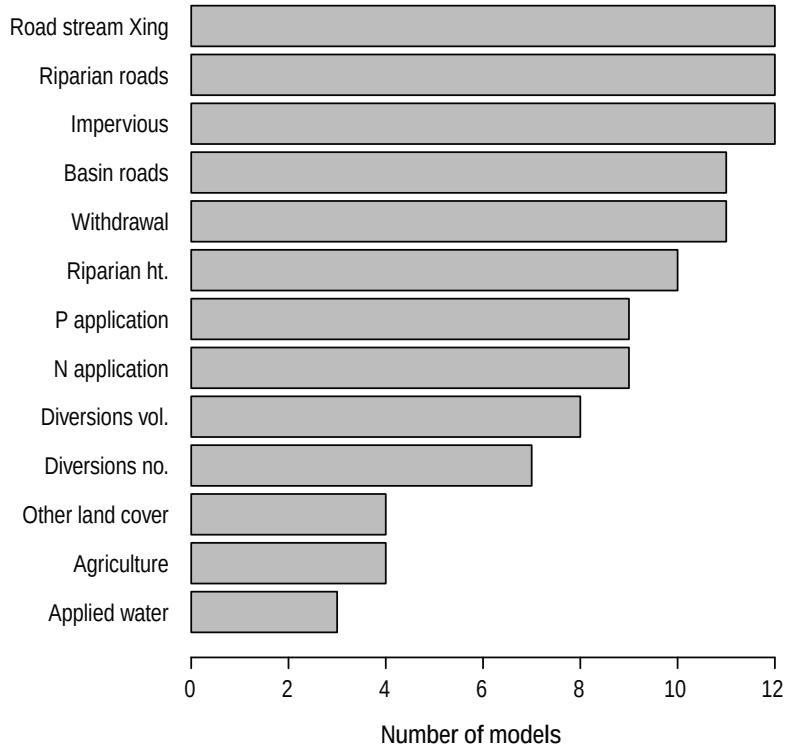

Figure 1–7. Occurrence of variables as important predictors in models predicting the likelihood of inflated monthly streamflows in California. Descriptions of variables are provided in tables 1–1 and 1–2. (ht., height; vol., volume; no., number; P, phosphate; N, nitrate)

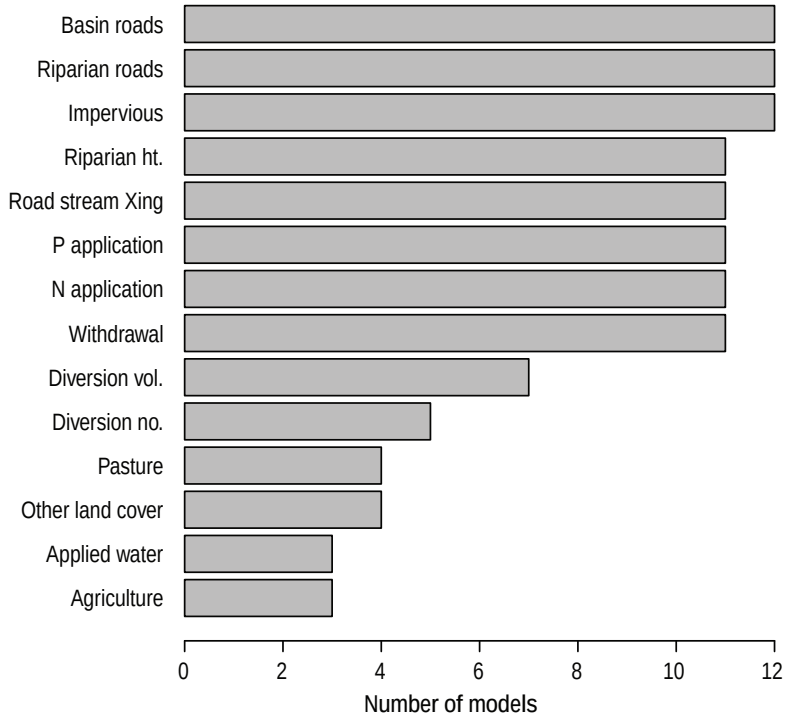

Figure 1–8. Occurrence of variables as important predictors in models predicting the likelihood of depleted monthly streamflows in California. Descriptions of variables are provided in tables 1– 1 and 1–2. (ht., height; vol., volume; no., number; P, phosphate; N, nitrate)

Table 1–1. Machine-learning models and associated tuning parameter settings evaluated for predicting monthly flows in California streams. Settings indicate tuning parameter values that were evaluated. Tuning parameter details provided in Kuhn and Johnson (2013).

Model	Tuning parameter	Settings
Neural network	Decay	0, 0.1, 0.1
	Size	1–9
Support vector machine	Degree	1–3
	Scale	0.01, 0.1, 1
	Cost	0.25, 0.5, 1, 2, 4
Random forest	Number of predictors evaluated at each node	33–57
Boosted regression	Interaction depth	2-12
	Shrinkage	0.01, 0.1
Cubist	Committees	0–100
	Neighbors	0–9

Table 1–2. Watershed physical features considered as potential predictors in statistical models of natural monthly flows in California streams.

[cm, centimeter; hr, hour; m, meter; CaO, calcium oxide; MgO, magnesium oxide; S, sulfur. Data source indicates published source of geospatial data, where 1 = Falcone, 2011; 2 = Olson and Hawkins, 2014; and 3 = PRISM Climate Group, 2014]

Variable name	Description	Units	Data source
DRAIN_SQKM	Drainage area	square kilometers	1
CaO_pct	Rock mean CaO content	percent	2
LPerm	Rock hydraulic conductivity	$x10^6$ meters/second	2
MgO_pct	Rock mean MgO content	percent	2
S_pct	Rock mean S content	percent	2
UCS	Rock compressive strength	megaPascals	2
PERDUN	Dunne overland flow	percent of streamflow	1
PERHOR	Horton overland flow	percent of streamflow	1
CONTACT	Subsurface flow contact time	days	1
TOPWET	Topographic wetness index	log(meters)	1
BFI_AVE	Base flow index	percent of streamflow	1
CLAYAVE	Soil clay content	percent by weight	1
SILTAVE	Soil silt content	percent by weight	1
AWCAVE	Soil water capacity	unitless	1
PERMAVE	Soil permeability	inches/hour	1
BDAVE	Soil bulk density	grams/cubic cm	1
OMAVE	Soil organic matter	percent by weight	1
HGA	Soil hydrologic group A	percent by weight	1
HGB	Soil hydrologic group B	percent by weight	1
HGC	Soil hydrologic group C	percent by weight	1
HGD	Soil hydrologic group D	percent by weight	1
HGAC	Soil hydrologic groups A and C	percent by weight	1
HGAD	Soil hydrologic groups A and D	percent by weight	1
HGBC	Soil hydrologic groups B and C	percent by weight	1
HGBD	Soil hydrologic groups B and D	percent by weight	1
HGCD	Soil hydrologic groups C and D	percent by weight	1
HGVAR	Soil hydrologic group VAR	percent by weight	1
KFACT_UP	Soil erodibility	unitless	1
ROCKDEPAVE	Soil thickness	inches	1
NO4AVE	Soil material <5 millimeters	percent by weight	1
NO10AVE	Soil material <2 millimeters	percent by weight	1
NO200AVE	Soil material <0.1 millimeters	percent by weight	1

Table 1–2. Watershed physical features considered as potential predictors in statistical models of natural monthly flows in California streams.—Continued

[cm, centimeter; hr, hour; m, meter; CaO, calcium oxide; MgO, magnesium oxide; S, sulfur. Data source indicates published source of geospatial data, where 1 = Falcone, 2011; 2 = Olson and Hawkins, 2014; and 3 = PRISM Climate Group, 2014]

Variable name	Description	Units	Data source
WTDEPAVE	Depth to water table	feet	1
RFACT	Rainfall & runoff erosivity	100s foot-ton inches/hr/acre	1
ELEVATION	Mean watershed elevation	m above sea level	1
SLOPE	Mean watershed slope	percent	1
PPTAVG_BASIN	Mean basin precipitation (1971–2000)	centimeters/year	1
Gneiss	Gneiss	percent of basin	1
Granitic	Granitic	percent of basin	1
Ultramafic	Ultramafic	percent of basin	1
Quarternary	Quarternary	percent of basin	1
Sedimentary	Sedimentary	percent of basin	1
Volcanic	Volcanic	percent of basin	1
Anorthositic	Anorthositic	percent of basin	1
Intermediate	Intermediate	percent of basin	1
SGEO1–SGEO45	Surficial geology classes	percent of basin	1
HLR1–HLR 20	Hydrologic landscape regions	percent of basin	1
BEDROCK_PERM	Bedrock permeability	ordinal rank	1
wb 0-12	Monthly runoff estimates from water balance model, for months at time t=0 through t-12	millimeters	3
p 0-12	Monthly precipitation for months at time t=0 through t-12	millimeters	3
t 0-12	Monthly air temperature for months at time t=0 through t-12	degrees Celsius	3
p 2,3,6 sum	Sum of precipitation from previous 2, 3, or 6 months.	millimeters	3

Table 1–3. Geospatial indicators of human activities used as potential predictors in statistical models predicting monthly streamflow modification in California.

[km², square kilometer; NPDES, National Pollution Discharge Elimination System; m, meter. Data source indicates published source of geospatial data, where 1 = Falcone, 2011; 2 = U.S. Geological Survey (USGS), 2008a; 3 = U.S. Department of Agriculture, 2012; 4 = California Department of Water Resources, 2000; 5 = Grantham and others, 2014; 6 = USGS, 2008b; 7 = USGS, 2014; 8 = USGS, 2013]

Variable name	Description	Units	Data source
ARTIFPATH_PCT	Stream length classified as artificial channel	percent of total length	1
ARTIFPATH_MAINSTEM_PCT	Stream length classified as artificial channel	percent of main stem length	1
HIRES_LENTIC_PCT	Lakes, ponds, and reservoirs	percent of basin	1
HIRES_LENTIC_DENS	Lakes, ponds, and reservoirs	number per km²	1
DDENS_2009	Dam density	number per km²	1
MAJ_DDENS_2009	Major dam density	number per km²	1
STOR_NOR_2009	Total reservoir storage	volume per km²	1
pre1990_DDENS	Dam density prior to 1990	number per km²	1
pre1990_STOR	Total reservoir storage prior to 1990	volume per km²	1
CANALS_PCT	Stream length classified as canals	percent of total length	1
CANALS_MAINSTEM_PCT	Stream length classified as canals	percent of main stem length	1
NPDES_MAJ_DENS	NPDES point dischargers	number per km²	1
FRESHW_WITHDRAWAL	Freshwater withdrawal	volume per km²	1
PCT_IRRIG_AG	Irrigated agriculture	percent of basin	1
FRAGUN_BASIN	Fragmentation of undeveloped land	unitless	1
DEVNLCD06	Developed land	percent of basin	1
FORESTNLCD06	Forested land	percent of basin	1
PLANTNLCD06	Crop land	percent of basin	1
WATERNLCD06	Open water	percent of basin	1
NITR_APP_KG_SQKM	Nitrogen application	kilograms per km²	1
PDEN_2000_BLOCK	Population density	persons per km²	1
rd_km_tot	Road density	kilometer per km²	2
rd_km_rip	Road density in riparian corridor	kilometer per km²	2
rd_st_int	Road-stream intersection	number per km²	2
canal_km	Length of canals	kilometer per km²	1
canal_st_i	Canal-stream intersections	number per km²	1
applied_wa	Applied agricultural water	acre feet per year per km²	3, 4
ag_sqkm	Agricultural lands	percent of basin	1
cnt_stor	Storage reservoirs	number per km²	5
cnt_hydro	Hydroelectric reservoirs	number per km²	5
cnt_other	All other reservoirs	number per km²	5

Table 1–3. Geospatial indicators of human activities used as potential predictors in statistical models predicting monthly streamflow modification in California.—Continued

[km^2, square kilometer; NPDES, National Pollution Discharge Elimination System; m, meter. Data source indicates published source of geospatial data, where 1 = Falcone, 2011; 2 = U.S. Geological Survey (USGS), 2008a; 3 = U.S. Department of Agriculture, 2012; 4 = California Department of Water Resources, 2000; 5 = Grantham and others, 2014; 6 = USGS, 2008b; 7 = USGS, 2014; 8 = USGS, 2013]

Variable name	Description	Units	Data source
ht_stor	Mean height of storage reservoir dams	meters	5
ht_hydro	Mean height of hydroelectric reservoirs	meters	5
ht_other	Mean height of all other reservoirs	meters	5
vol_stor	Total volume of storage reservoirs	acre feet per km^2	5
vol_hydro	Total volume of hydroelectric reservoirs	acre feet per km^2	5
vol_other	Total volume of all other reservoirs	acre feet per km^2	5
rip_ht	Riparian vegetation height within 100 m buffer of stream	meters	6
mine_cnt	Active mines	number per km^2	7
og_well	Oil and gas wells	number per km^2	8
divert_cnt	Water diversions	number per km^2	5
diver_fval	Total volume of diversions	acre feet per year per km^2	5
AnnualFACE_VALUE	Total diversions reported value	acre feet per year per km^2	5
JAN_USE–DEC_USE	Monthly water use	acre feet per month per km^2	5

Publishing support provided by:
Rolla Publishing Service Center

For more information concerning this report, contact:
Chief, National Water-Quality Assessment Program
U.S. Geological Survey
413 National Center
12201 Sunrise Valley Drive
Reston, VA 20192
http://water.usgs.gov/nawqa/

ISSN 2331-1258 (online)
https://doi.org/10.3133/ofr20161189