Fall 2022

Estimating Natural Background Water Quality in California Rivers

Emma A. Debasitis
California State University, Monterey Bay, edebasitis@csumb.edu

ESTIMATING NATURAL BACKGROUND WATER QUALITY IN CALIFORNIA RIVERS

A THESIS

Presented to the

Faculty of

Applied Environmental Science

California State University, Monterey Bay

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in

Environmental Science

by

Emma A. Debasitis

Fall 2022

CALIFORNIA STATE UNIVERSITY MONTEREY BAY

The Undersigned Faculty Committee

Approves the Thesis of Emma A. Debasitis:

ESTIMATING NATURAL BACKGROUND WATER QUALITY

IN CALIFORNIA RIVERS

John R. Olson
Chair California State University, Monterey Bay

Judith Canner

Judith Canner
California State University, Monterey Bay

Raphael Mazor

Raphael Mazor
Southern California Coastal Water Resource Project

Doug Smith

Doug Smith CSUMB Interim Dean
Graduate Studies & Research

12/6/2022

Approval Date

ABSTRACT

ESTIMATING NATURAL BACKGROUND WATER QUALITY

IN CALIFORNIA RIVERS
by
Emma A. Debasitis
Master of Science in Environmental Science
California State University Monterey Bay, 2022

Water chemistry affects organisms at all levels of the food web in aquatic habitats. Water quality alteration from natural conditions can seriously degrade habitat quality, human health, and the survival rate of native species. Estimating background (i.e., historic or baseline) water quality of impaired streams can be difficult due to the effects of anthropogenic involvement in aquatic systems over decades to centuries. The ability to model natural background water quality levels would aid in overall stream management. The natural background predictions provided by a model would allow for increased understanding of stream condition and would allow us to determine the amount of divergence between current stream conditions and natural background.

To better predict baseline water chemistry levels, we created random forest models for ionic concentrations and integrated water quality measures of ionic balance, including chloride, calcium, magnesium, sodium, sulfate, alkalinity, hardness, total dissolved solids, and specific conductivity. Water quality measurements from minimally disturbed reference sites across the United States were used as response variables for model training. We developed these models using both static (e.g., geology, soils, etc.) and dynamic (i.e., monthly evapotranspiration, precipitation, and temperature) EPA StreamCat and PRISM predictor variables. The models explained 66% to 98% of the variation in samples from California streams and 55% to 85% of the variation across the US. The top predictors across models include yearly temperature averages, yearly precipitation averages, percent lithological sulfur, and base flow index. The baseline water chemistry estimates produced by these models will help California establish site-specific water quality standards and manage habitat in various situations, including urban development projects, habitat restoration, and endangered species monitoring.

Research impact statement:

Natural background water chemistry is needed to understand the underlying processes of aquatic ecosystems. These predictions will aid in helping set water quality thresholds for aquatic species survival.

Key words: California water chemistry, natural background, aquatic habitats

Dedication

I dedicate this thesis to my family who have been my support system throughout many challenges life has thrown my way.

Acknowledgments

I am incredibly grateful for my thesis advisor Dr. John Olson for his invaluable support on this project. I also would like to thank my additional committee members, Dr. Raphael Mazor and Dr. Judith Canner, for their advice and input throughout the life of this project.

This project would not have been accomplished without funding from the Regional Water Quality Board – Santa Ana Region and the Surface Water Ambient Monitoring Program. I would also like to thank Janet Walker for being so generous with her time and advice.

Lastly, I would like to thank my family for supporting me throughout these last two and a half years, providing much needed encouragement throughout this process. I'd also like to thank my dogs. They always needed to go on walks, and they forced me to leave the computer for a while.

TABLE OF CONTENTS

List of Tables

List of Figures

Introduction

Water quality has significant potential to impact aquatic species. Fish, due to their position in the food chain and longer lifespans, are sensitive to changes in water quality (Amizi and Rocher 2016). For example, an Egyptian study found that the occurrence of fish organ lesions and blood parameters were increased in populations present in the El-Rahway river due to increased agricultural, industrial, and domestic waste dumping in the region (Gaber et al. 2013). Invertebrates are also used as water quality indicators through community condition metrics such as the California Stream Condition Index, CSCI, that looks at how invertebrate community composition is affected by stream conditions (Rhen et al 2015).

The Environmental Protection Agency's (EPA) water quality standards are generally determined by human health standards and aquatic species impacts. For example, California's water quality standards include maintaining balanced ecological communities where the communities are diverse, self-sustaining, food for all chain components is present, and the area is not dominated by tolerant species (Santa Ana Water Quality Objectives 1995). In the Santa Ana region of California, there are several protected aquatic species which may be impacted by upstream urbanization and its effect on water quality. The effect of changes in water quality parameters and ionic concentrations on these species has not been well studied. The habitat needs of these native species, such as the Arroyo Chub, Santa Ana Sucker, and the California Newt, must be considered when regional habitat and water resource managers make decisions regarding crucial habitat in and around the Santa Ana Basin.

Changes in physiological stressors may lead to species adaptation, migration, or decline. For example, increased salinity concentration led to a significant decrease in the western pond turtle body mass and increased energy output into osmoregulation (Agha et al. 2019). Similarly, the

Santa Ana Speckled Dace, considered at a significantly increased risk of extinction within the next 50 years, is often located in areas with increased fires, debris flows, urban populations, and invasive species (CA Dept. of Fish and Wildlife March 10, 2022). In many cases, researchers only focused on dissolved oxygen and the physical properties needed to make a habitat suitable for native species. They often neglected the effects other water chemistry components had on survivability and reproduction (Thompson et al. 2009). Thompson et al. (2009) did mention, however, that water chemistry had significant potential to influence the aquatic species.

To gain a better understanding of how the alteration of water chemistry affects aquatic species, natural background levels need to be estimated to determine how far current stream conditions have shifted from natural levels. The variability of water chemistry suggests that watershed topographic, geological, and climatic variables need to be considered when predicting natural background water quality (Olson and Hawkins 2012). Once we can determine the amount of alteration in the current water chemistry, we can quantify the ecological effects caused by the alteration. Human impacts may also lead to ecosystem degradation both locally and downstream due to the connectivity of stream systems. The addition of dams and urban and industrial areas has changed the natural flow regime in downstream channels and has the potential to change water quality levels throughout the stream systems (Poff et al. 1997). Upstream influences such as agriculture, impervious surfaces, and population density influence how chemical processes occur in an aquatic system.

Increases in temperature often lead to evaporation and increased concentrations of solutes in stream systems by altering the amount of freshwater. These altered distributions include changes to the amounts and timing of evaporation, precipitation and water vapor transport (Carpenter et al. 1992). The changes in the global climate are affecting the timing of snowmelt runoff as well as the volume and distribution of snowpacks (Manning et al. 2013). While a precipitation increase generally leads to an overall dilution effect, increasing concentrations may be noted due to unique local factors, such as the lithological makeup of the surrounding area. Areas with large amounts of limestone, for example, may be expected to have greater natural calcium concentrations than an area surrounded by granite. Peters (1984) found that annual precipitation and rock type were the most important factors affecting the yield of ions in the basins studied. Static predictors (e.g., geology and soil composition) and dynamic predictors (e.g., evapotranspiration and precipitation) are important components when trying to analyze the main variables affecting water chemistry. Olson and Cormier (2019) used both static and dynamic predictors in their previous study modeling specific conductivity. Their results have been used by various regional managers to analyze shifts in specific conductivity away from natural levels and they have provided data for these agencies to create restoration plans (J. Olson, personal communication, 21 July 2022).

Estimates of natural background water quality are necessary in order to understand where human alterations of water quality parameters have occurred. Natural background water quality levels are not well known, and the variability of natural levels depends greatly on the specific geologic and climatic components of individual stream systems (Olson and Cormier 2019). For example, a natural background water quality model of specific conductivity was used by Vander Laan et al. (2013) to estimate first the divergence from natural background due to human activity and then estimate the impact of this divergence on biological conditions. These unknowns make it difficult for land managers to understand the impact human sources are having on stream systems, since there isn't a baseline value for comparison.

We modeled natural temporal and spatial variation to predict variation in major ion concentrations and integrated measures for streams throughout California, allowing management stakeholders to better understand California's natural water quality parameters and assist management agencies with their conservation and restoration efforts. We modeled five ionic parameters and four integrated parameters using natural environment factors that influence natural variation in water quality (Table 1). The modeled estimates of these parameters allow us to determine the amount of alteration from natural levels by comparing them to current measured concentrations throughout California. These comparisons and natural estimates can be used to establish water quality thresholds for aquatic life and to restore habitat health. Conservation agencies, such as California Department of Fish and Wildlife, may also use these models to determine habitats suitable for the expansion of endangered species and any potential water quality issues needing to be addressed within the expanded range. We then evaluated the predictive metrics to determine what environmental factors drive natural background levels. This research will improve management practices and aid in targeting areas in the greatest need of restoration efforts, allowing revenue resources, such as the California Department of Fish and Wildlife (CDFW) Service Based Budget (CDFW SSB 2021), to be used in the most cost-effective ways, saving stakeholders and the taxpayers' financial resources.

Methods

Data Acquisition

We modeled how California's natural water quality varies among individual stream segments over time, using water quality data from the Contiguous United States obtained from bioassessment databases (Figures 1 & 2, Table 2). We used data from the contiguous United States to aid in the development of our water quality models for a better representation of the range of environments across spatially heterogeneous California. These databases included the California Environmental Data Exchange Network (CEDEN, ceden.waterboards.ca.gov/), the National Rivers and Streams Assessment (NRSA, https://www.epa.gov/national-aquatic-resource-surveys/nrsa), the National Water-Quality Assessment Project (NAQWA, https://www.usgs.gov/mission-areas/water-resources/science/national-water-quality-assessment-nawqa), and the Storm Monitoring Coalition (SMC, smc.sccwrp.org, Olson & Cormier 2019). From these five sources we extracted measurements of nine water quality parameters (Table 1). This provided data from 8598 potentially reference stream segments, 11% of which were in California (Table 2). We withheld 20% of sites at both the national and California scale for model validation following Olson & Cormier (2019).

Table 1: Ionic Strength and Integrated measure parameter. All parameters were measured in mg/L, except Specific Conductivity which was measured in µS/cm.

Ionic concentration	Integrated measures
Calcium	Alkalinity
Chloride	Hardness as CaO3
Magnesium	Total Dissolved Solids
Sulfate	Specific Conductivity
Sodium	

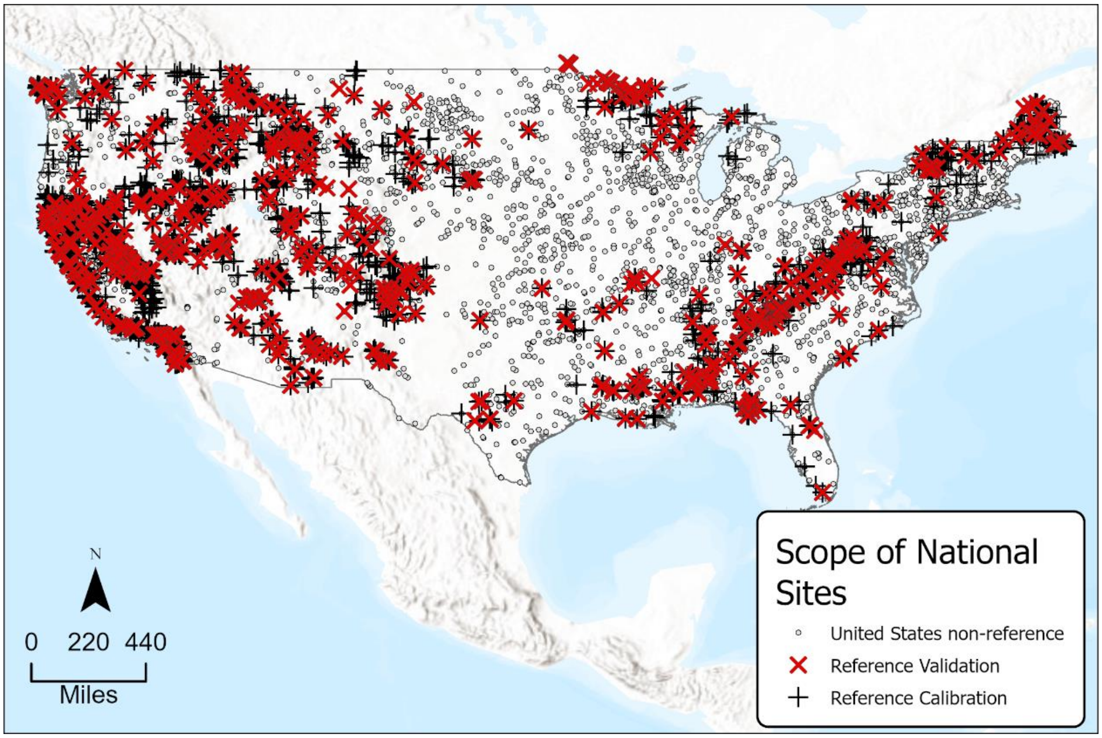


Figure 1: Non-reference and reference calibration and validation site data at the national scale.

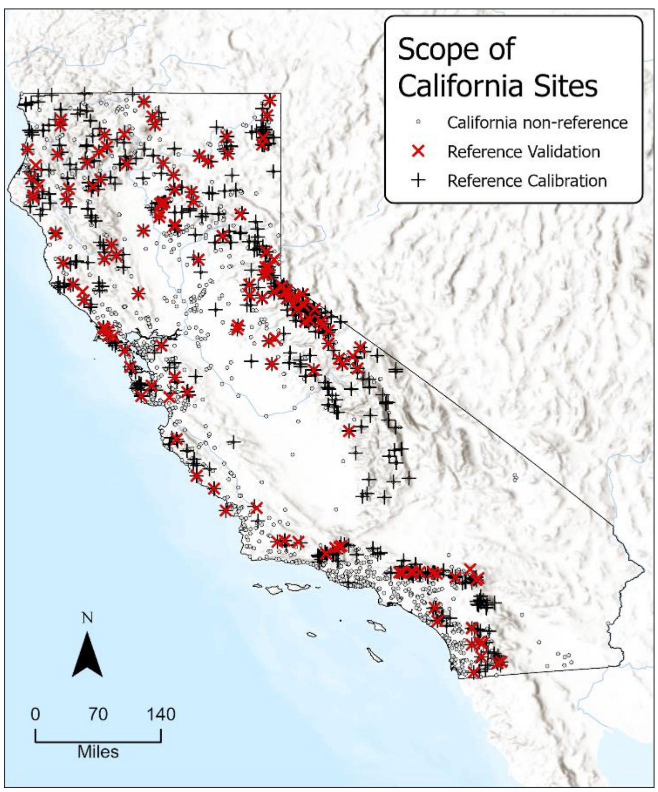


Figure 2: Non-reference and reference calibration and validation site data at the California scale.

*Table 2: Data sources and number of sites available for model development. *NWIS and Olson are 0 because they were duplicates of CEDEN datapoints*

Program	Extent	# of Unique Stream Segments	# Reference sites in California	# Reference sites Nationwide
NAQWA/NWIS	Nationwide	112	0	1378
NRSA	Nationwide	4166	23	457
SWAMP/CEDEN	California	2433	976	1517
Olson & Cormier (2019)	Nationwide	1877	0	6864

Response and Predictor Data

In order to associate sampling locations with geospatial data used as predictors in our water quality models, we matched each site to a catchment in the National Hydrography Dataset Plus (NHD+; (McKay et al. 2012). Sites were matched by overlaying our sample locations with NHD+ catchments to determine the catchment's unique identifier (FEATUREID). Sites could then be matched with the corresponding unique identifier in the StreamCat dataset (COMID; (Hill et al. 2016)) to obtain additional geospatial data. Sites that were unable to be matched to a catchment were excluded from further analysis.

Once matched to a catchment, we extracted geospatial data from the StreamCat dataset (Hill et al. 2016). We acquired two kinds of geospatial data from StreamCat: 1) metrics that characterize natural gradients that could influence ionic concentrations (predictors) and 2) metrics that characterize human disturbance (human activities). Natural gradient metrics were used as predictors in our models, and human disturbance metrics were used to identify minimally disturbed reference sites. All StreamCat variables evaluated in this study are presented in Appendix E. Additionally, we determined the Omernik Level 3 ecoregion (Omernik and Griffith 2014) of each site based on its location (Appendix D). StreamCat defines a catchment as an area that drains directly to an NHD stream segment and defines a watershed as a set of connected catchments that flow to a focal point (Hill et al 2016).

Table 3:Selected StreamCat variables grouped by category - Abbreviations are listing in parentheses. For metric abbreviations, Ws indicates watershed-scale calculations, and Cat indicates catchment-scale calculations. A complete list is provided in Appendix E.

Category	**Selected metrics**
Predictors	
Lithology	Aluminum oxide content (Al203Ws), Sulfur content (SWs), Lithological ferric oxide (Fe2O3Ws), Silicic residual (PctSilicicWs; PctSilicicCat), Sodium oxide (Na2OWs), lithological hydraulic conductivity (HydrlCondWs)
Groundwater	Groundwater discharge into streams in catchment ratio (BFICat)
Soil	Soil erodibility, soil permeability
Precipitation	3-month, 6-month, and 1 year mean precipitation (PrecipWs; PrecipCat)
Temperature	3-month, 6-month, and 1 year maximum temperature (TmaxWs; TmaxCat); 3 month, 6-month, and 1 year mean temperature (TmeanWs; TmeanCat)
Landcover *not an important metric in any models	Evergreen forest landcover (PctConif2016Cat; PctConif2016Ws), Mixed deciduous/evergreen forest cover (PctMxFst2016Ws)
Topography	Composite topographic index (WetIndexWs), Catchment elevation (ElevCat)
Human activity	
Land use	Crop land use, Hay land use, Developed high, mid, and low intensity land use, Developed open-space land use, Density of road-stream intersections, Density of roads
Reservoirs	Natural and possible volume
Dams	Dam density
Canals	Canal density
Mining	Mine density

Water quality varies over time depending on factors such as temperature, precipitation and evapotranspiration. In order to incorporate this variability in our models, we created a dynamic component, following Olson & Cormier (2019), in addition to the spatial analysis. The dynamic component involved matching the sample dates to the corresponding temporal values regarding: mean and maximum temperature, mean and maximum evapotranspiration, and minimum, mean, and maximum precipitation. Values for these dynamic components were calculated for the month the sample was collected, one month prior, the average over the 3 months prior, the average over the 6 months prior, and finally the average over the 12 months prior. This added dynamic component allowed for the analysis of temporal variables over time and spatial scales (Table 4).

We took steps to eliminate sample bias and duplication. We removed duplicate site observations sampled during the same month and in the same stream catchment. The retained sample was selected at random to prevent a bias caused by over-representing sites that were repeatedly sampled. Sites with values below the method detection limit were replaced with a value of zero since minimum recording limits were not provided by all databases.

Our final dataset contained static predictors, dynamic predictors, and human influence factors. Static predictors included factors such as geology, soils, and landcover, while the dynamic predictors included factors such as precipitation, evapotranspiration, and temperature at each segment over time. The human impact predictors were dam and canal presence, reservoir presence, land use, and mining.

Table 4: PRISM and StreamCat variables used to create the spatial dynamic model.

Dynamic parameter	**Months from sample date**
Evapotranspiration	0,1,2,3,6, and 12
Temperature	0,1,2,3,6, and 12
Precipitation	0,1,2,3,6, and 12

Screening reference sites

We used StreamCat data to identify reference sites minimally affected by human activity, following the procedure in Ode et al. (2016). Any site that failed one or more of the thresholds in

Table 5 was not considered reference. Once screened, sites with exceptionally high ionic parameter values were further evaluated in Google Earth. This screen often provided evidence of human impact or natural, but unusual, sources of ions (e.g., hot springs or evaporite deposits) not detected in the initial reference screening. Disturbances included tidal influence, cattle grazing, industrial complexes, and excessive erosion in an area. If evidence of a disturbance was found, the site was no longer considered reference and was removed.

Table 5:Criteria used to identify reference sites based on Ode et al. (2016)

StreamCat Predictor	Threshold	Unit
Agricultural Use in Catchment	3	%
Agricultural Use in Watershed	3	%
Dam Density in Watershed	2	#/km²
Dam Density in Watershed	2	#/km²
Mine Density in Watershed	0.1	#/km²
Canal Density in Watershed	10	#/km²
NLCD CODE 21	5	%
Low Urban Presence in Catchment	3	%
Med. Urban Presence in Catchment	3	%
High Urban Presence in Catchment	3	%
Low Urban Presence in Watershed	3	%
Med. Urban Presence in Watershed	3	%
High Urban Presence in Watershed	3	%
Road Crossings per Catchment	5	#/km²
Road Crossings per Watershed	10	#/km²
Road Density per Catchment	2	#/km²
Road Density per Watershed	2	#/km²

Reference sites for the United States and California were identified for all parameters (Table 6). Large representative datasets were assembled for all analytes from multiple data sources (Table 2). We used nationwide data to ensure we were encompassing the broad diversity of California's geology and topography. Given our large datasets, encompassing nationwide data for model training, we are more likely to represent a large proportion of the environmental characteristics found in streams in California.

Model training

Once we established a reference dataset, we developed random forest models to predict ionic concentrations of reference sites based on natural landscape-level factors. Models were built using R software (R-Core-Team 2020) and the 'randomForest' (Liaw and Wiener 2002) and 'caret' (Kuhn 2008) packages.

To train our models, we used a subset of the reference sites to identify the most important predictors. We then validated the model using the withheld sites. To create these reference datasets, we subset the reference dataset into calibration (80%) and validation (20%) datasets (Figures 1, 2, & Appendix C). The subsets were stratified by the eighty-five Level 3 Ecoregions, ensuring that major ecoregions were equally represented in the calibration and validation data sets. Next, we built random forest models for each parameter using the calibration data to train the model. Each model was initially run with all predictors present (Supplement 1). Multiple model selection techniques were attempted, including the Variable Selection Using Random Forest (VSURF) package, an RFE (recursive feature elimination) approach, and variable removal using variable importance. Due to the extensive processing power demands associated with the VSURF and RFE approaches, we performed a preliminary variable selection using variable removal by variable importance: The least-important variables were removed from the models, the new models were calibrated, determining if the overall variance explained improved. These models were created by first removing all but the top fifty predictor variables and then removing the five least important variables in each subsequent model until we obtained the greatest percent variable explained. An evaluation of the stability of variable selection using this method was completed by Fox et al. (2017) determining that the "out-of-bag" (OOB) performance remained steady until minimum variables remained. A sudden increase in variable importance explained occurred when only the most important variables remained, then significantly decreased as too many variables were removed.

Model performance

We assessed model performance by looking at the accuracy and precision of our modeled results. To determine the accuracy and precision values, we used pseudo-R^2, root mean squared error (RMSE), and out-of-bag predictions for each random-forest model. Predicted values from the random forest models were compared to observed values in the calibration and validation reference data sets. Out-of-bag predictions, generated when random forest created a subset of trees withholding the sites in question, were used to assess calibration performance. This validation approach allowed us to use the OOB predictions to measure model performance independent of the sample data used to train the model. Linear regressions comparing observed to expected values were calculated. Model precision was estimated with the regression's R^2 value; larger values indicated better precision. Model accuracy was assessed using the slope and intercept of the regression; intercepts close to zero indicated higher accuracy, and slopes close to 1 indicated that model performance is consistently accurate across a range of conditions. These model performance parameters were summarized for the entire contiguous United States. Model performance was also assessed for California sites alone following this procedure.

Model performance was also evaluated for spatial bias by plotting model residuals on a map, depicting any geographic patterns in the model errors. The residuals were calculated by subtracting the predicted values from the sites' observed values. If the areas with insufficient predictive power were all located in similar geographic regions, that indicates potential geographic bias in our model. If the sites with insufficient predictive power were equally spread throughout geographic locations, our model had less likelihood of geographical bias.

Partial Dependence Plot Analysis

In order to understand why environmental characteristics of a stream's surroundings cause different analytic concentrations in varying stream systems, we used partial dependence plots (PDP). These plots interpret correlation relationships between our response variables and predictor variables using the PDP package in R (Greenwell BM 2017). The graphical results from a PDP analysis show how the response variable responds to variation in a single predictor variable while holding all others constant. We used a PDP analysis to determine if the presence of the individual analytes were positively or negatively correlated with the individual predictor variables.

Mapping Water Quality Natural Background and Alteration

In order to obtain natural background estimates for California streams, we applied our models to all stream segments in California. The model application, and map generation of expected natural levels, were possible because StreamCat data is available for almost every NHD+ segment in California. The symbology was then adjusted to show graduated colors from low to high concentrations of each parameter's expected value. In order to understand the divergence of current stream conditions from natural background estimates, we mapped estimated water quality alteration from natural conditions as the difference between observations at non-references sites and model-estimated natural background. These predictions are available in a GitHub repository maintained by the Southern California Coastal Water Research Project (https://github.com/SCCWRP/RB8_biointegrity_impacts).

Inter- and Intra-Annual Variability in California Predictions

We mapped both the intra- and inter-annual variation of each analyte in California to examine how the spatial patterns in natural background levels vary temporally. We calculated averages for each analyte for each season in wet, dry, and normal years. We then compared the average values by subtracting dry years from wet years (i.e., inter-annual variation) and dry seasons from wet seasons (i.e., intra-annual variation).

Results

Reference Sites

Our reference screening provided us enough sites to create the models at a nationwide scale and to validate at the California scale. Chloride, sulfate, and most of the integrated measures had >400 reference sample points whereas sodium, calcium, magnesium, and TDS all had <300 sites within California (Table 6).

Table 6: Summary of sample data assembled for modeling. N: number of sample segments

	California		Nationwide		
Parameter	*N total*	*N reference*	*N total*	*N reference*	*Unique Segments*
Ions					
Cl	9643	611	15426	2238	6140
SO4	4478	502	6834	1704	1952
Na	8244	232	12786	1496	4675
Ca	8562	258	13506	1558	4918
Mg	8558	258	13504	1558	4839
Integrated measures					
TDS	8873	247	9069	429	645
Hardness	5271	415	5375	493	1745
Alkalinity	2053	546	3337	1106	1891
Sp.Cond	10880	944	28470	9575	8508

Model Performance

Over half of our models explained more than 80% of the variation in water chemistry (Figures 3 and 4). In addition, the low RMSE values suggested that there was little error in our models. The nationwide sodium model had an r^2 value below our 70% threshold. The model does, however, explain over half of the variation and is still likely a useful model. The models performed very well when validated at the California scale, with validation r^2 values above 70% for all analytes (Figures 3, 4, & Table 7). The slope values close to 1 showed that the model predictions are precise. Since our slopes are close to one, our intercepts are close to zero, and our standard errors are low, there is a minimal chance of the models under or overpredicting natural background values.

Partial Dependence Plot Analysis

The partial dependence plot (PDP) analysis of each model showed that the % Sulfur in the watershed, and both the 12-month average max and mean temperatures were associated with an increase in analyte concentrations across all models (Appendix B). This PDP analysis also indicated that the 3-, 6-, and 12-month precipitation averages, base flow index, and the % Sodium Oxide in the watershed were associated with a decrease in analyte concentrations across all models (Appendix B).

Mapping Modeled Predictions

As seen in Figure 5, alteration was most severe in the Central Valley and Southern California. Alkalinity was an exception with a large amount of alteration occurring in the Eastern Sierras. Some areas with a large percent change had little shifts in the absolute change, showing that the relationship of observed over expected reacts differently than the absolute change of observed minus expected, underscoring the value of analyzing changes in water chemistry both ways in order to grasp larger patterns in alteration. Overall, the integrated parameters had greater shifts in both the percent change and absolute change than the individual ionic parameters. Integrated parameters have multiple water chemistry and environmental variables associated with them that would likely lead to greater increases from natural background values.

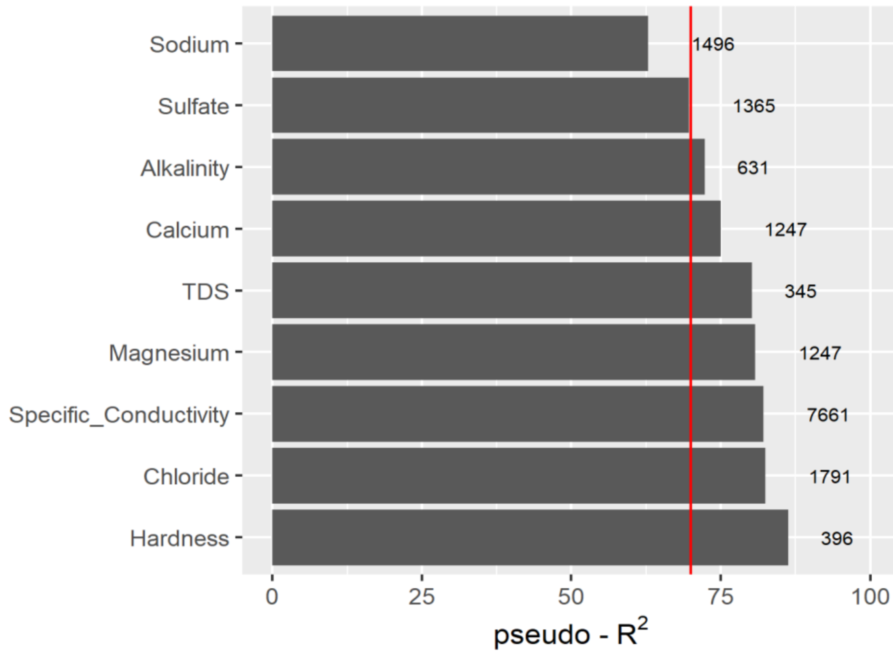


Figure 3: Random forest r2 values for all models, the numbers at the end of the bars are the amount of reference sites for each model at the national scale.

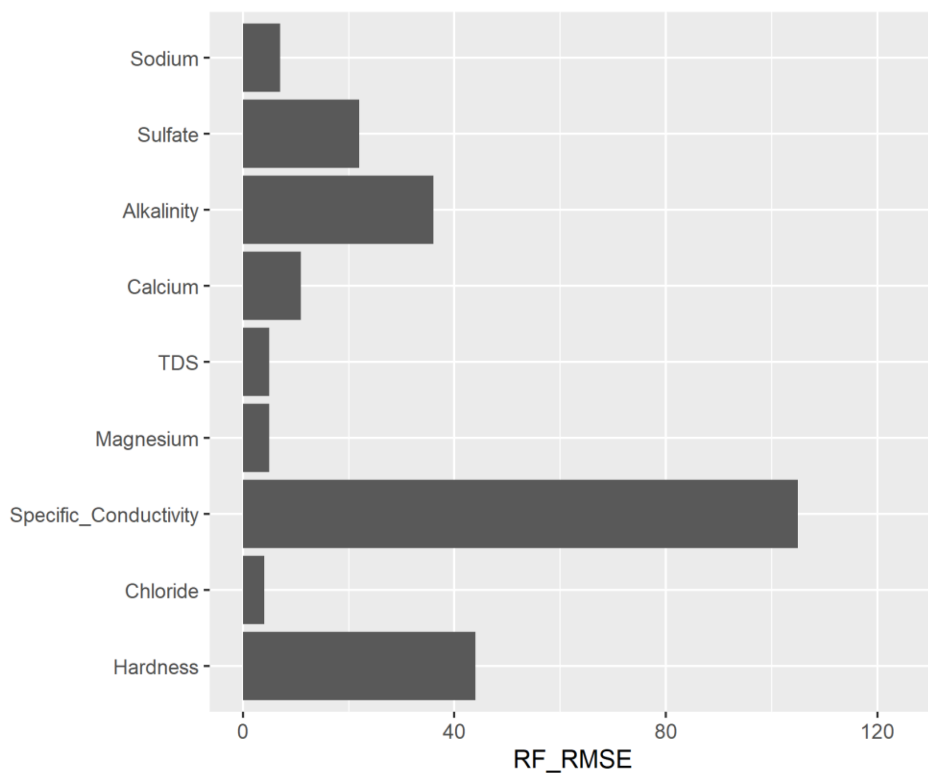


Figure 4: Random forest root mean squared error for all models. These values show the degree of error in our models. Specific Conductivity has greater variation in observed concentrations due to the units of measure being µS/cm, leading to a larger RMSE.

Table 7:Model performance across all analytes at the Nationwide and California scale. OE r2 values were calculated by the regression line created by plotting the data. These values are unique to the scale being analyzed (CA or Nationwide).

Parameter	Set	Scale	N ref	OE r2	OE Slope	OE Slope SE	OE Intercept	OE Intercept SE
Ions								
Mg	Cal	CA	209	0.98	1.06	0.01	-0.9	0.25
	Val	CA	49	**0.95**	1.01	0.03	-0.6	0.91
	Cal	Nationwide	1247	0.81	1.55	0.01	-0.45	0.17
	Val	Nationwide	311	0.88	0.98	0.02	-0.08	0.25
Ca	Cal	CA	211	0.97	1.06	0.01	-1.85	0.57
	Val	CA	48	**0.94**	1.02	0.04	-1.63	2
	Cal	Nationwide	1247	0.75	1.03	0.02	-0.75	0.45
	Val	Nationwide	311	0.88	0.99	0.02	-0.61	0.6
Na	Cal	CA	188	0.97	1.07	0.01	-1.38	0.31
	Val	CA	44	**0.92**	1.09	0.05	-1.09	1.26
	Cal	Nationwide	1198	0.63	1.05	0.02	-0.45	0.25
	Val	Nationwide	298	0.66	1.05	0.04	-0.32	0.51
Cl	Cal	CA	497	0.98	1.11	0.01	-0.93	0.12
	Val	CA	114	**0.8**	1.14	0.05	-1.11	0.7
	Cal	Nationwide	1791	0.83	1.09	0.01	-0.48	0.1
	Val	Nationwide	447	0.76	1.05	0.03	-0.4	0.2
SO4	Cal	CA	403	0.95	1.14	0.01	-4.39	0.88
	Val	CA	99	**0.73**	1.48	0.09	-10.75	5.14
	Cal	Nationwide	1365	0.7	1.07	0.02	-0.74	0.66
	Val	Nationwide	339	0.75	1.41	0.05	-3.81	1.43
Integrated measures								
TDS	Cal	CA	201	0.95	1.08	0.02	-23.22	5.76
	Val	CA	46	**0.78**	1.08	0.09	-2.85	26.37
	Cal	Nationwide	345	0.8	1.01	0.03	-0.9	6.66
	Val	Nationwide	84	0.85	1.1	0.05	-9	11.6
Sp.Cond	Cal	CA	745	0.97	1.08	0.01	-21.41	2.66
	Val	CA	119	**0.78**	1.01	0.04	3.43	0.04
	Cal	Nationwide	7661	0.82	1.03	0.01	-6.22	1.54
	Val	Nationwide	1914	0.81	1.03	0.01	-4.35	3.33
Hardness	Cal	CA	334	0.98	1.04	0.01	-6.36	1.61
	Val	CA	81	**0.73**	1.09	0.07	0.12	14.02
	Cal	Nationwide	396	0.86	1	0.02	-1.7	3.28
	Val	Nationwide	97	0.76	1.09	0.06	-1.65	11.01
Alkalinity	Cal	CA	438	0.96	1.13	0.01	-14.38	1.37
	Val	CA	108	**0.73**	1.21	0.07	-18.32	8.84
	Cal	Nationwide	631	0.73	1.08	0.03	-7.16	2.59
	Val	Nationwide	221	0.65	1.13	0.06	-14.12	5.49

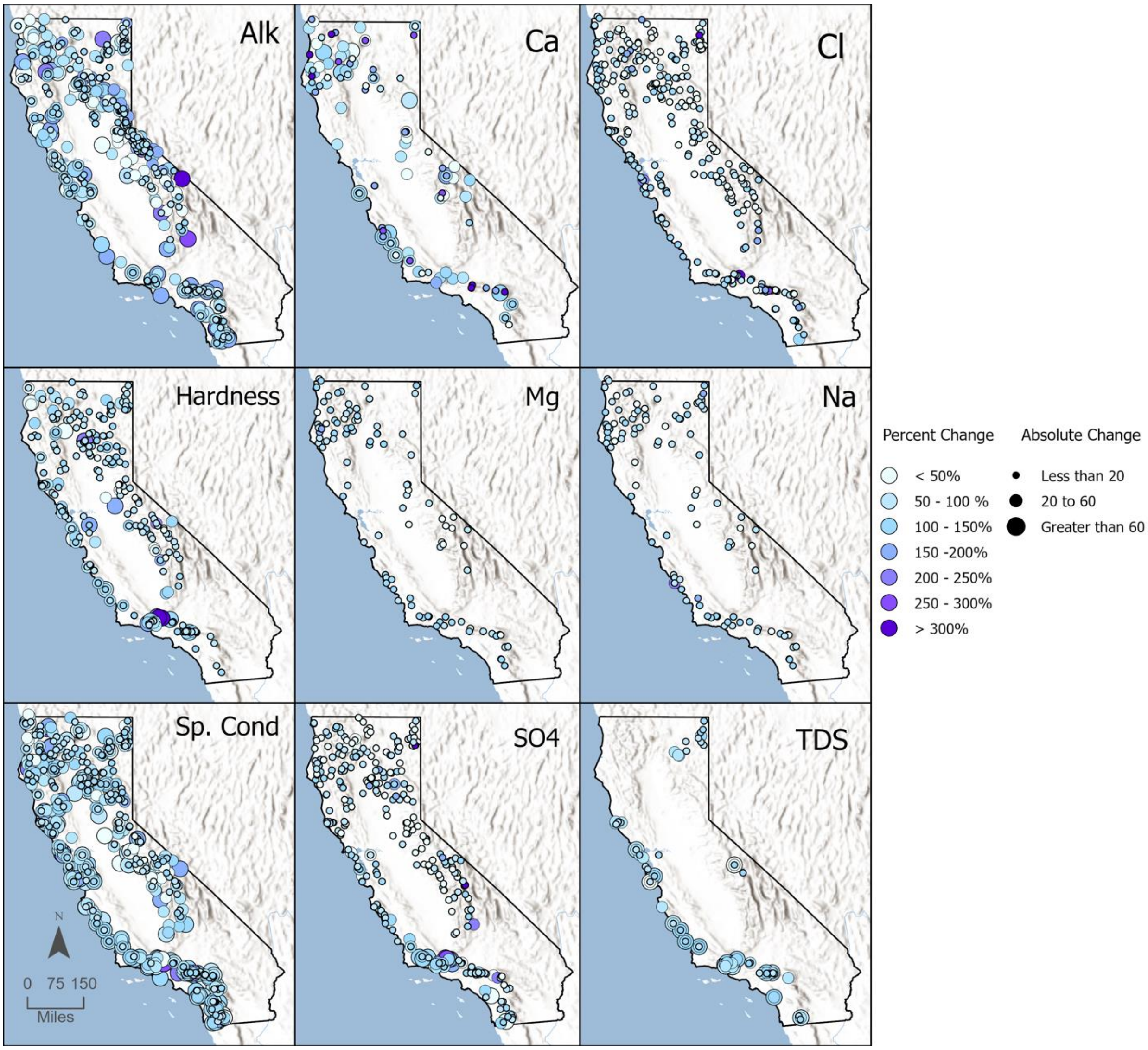


Figure 5: Alteration of water chemistry estimated as the difference in parameter values between model predictions (expected) and observations at non-reference sites. Percent change is the ratio of "Observed/Expected". Absolute change is "|Observed - Expected|".

Intra- and Inter-annual Analysis

This study's results show a difference in natural background water quality depending on seasonality and precipitation for most of the analytes modeled, shown by the intra- and inter-annual maps (Figures 6 & 7). The intra-annual results depict that ionic and integrated concentrations will be greater in aquatic systems during the fall and winter months. This pattern is likely due to the dilutionary effect of increased water flow in the spring and summer months due to snow melt in higher elevations. The inter-annual results show that dry years will have higher concentrations of ions throughout the year compared to wet years. These results follow the general pattern of ionic concentration and dilution with an increase or decrease in precipitation. In years with greater rainfall, there is more water flowing through the aquatic systems creating a dilutionary effect. While years with less rainfall have a greater evaporative effect and this leads to an increase in aquatic ionic concentrations.

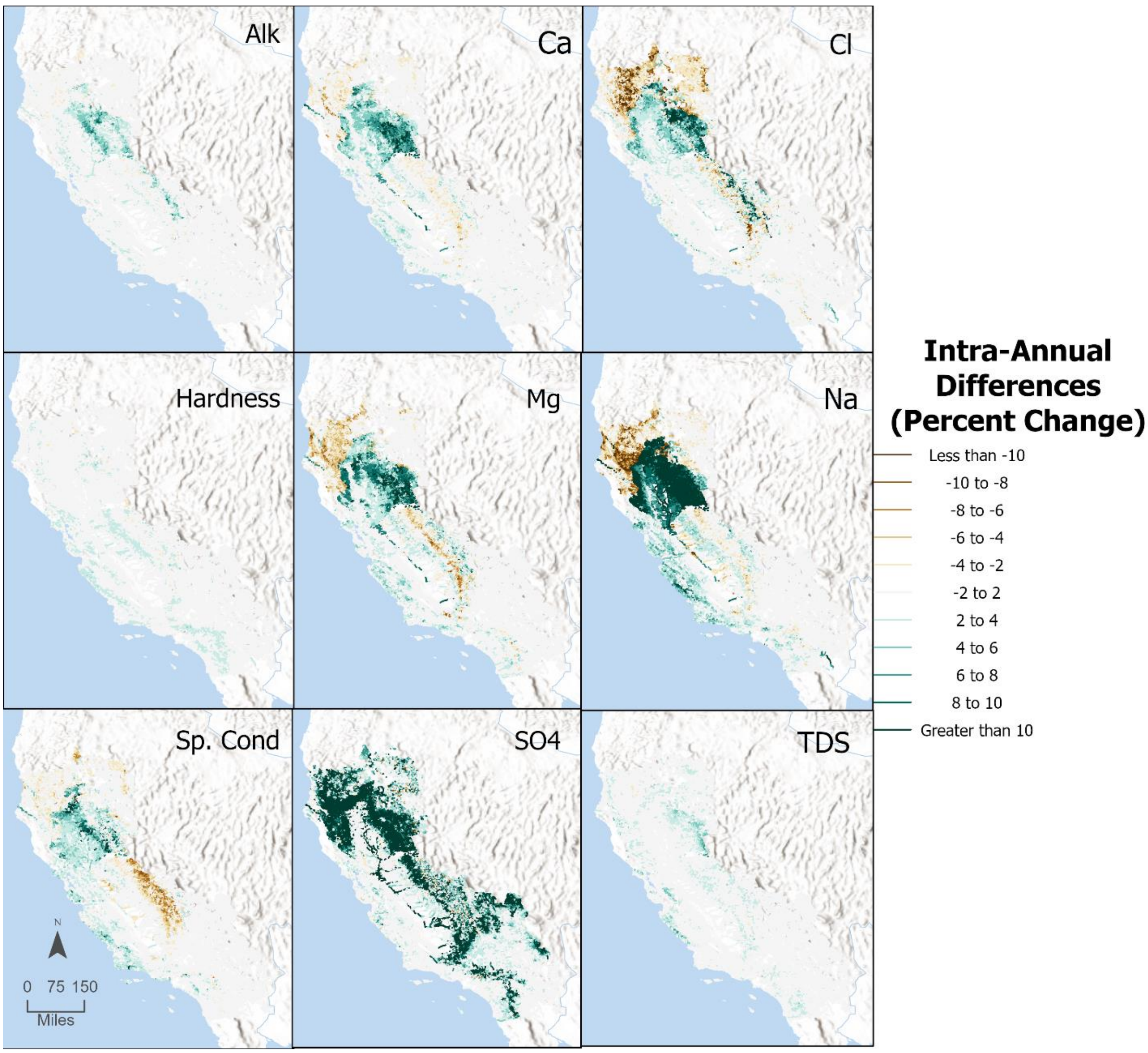


Figure 6: Intra-annual differences across all analytes, displayed by percent change.

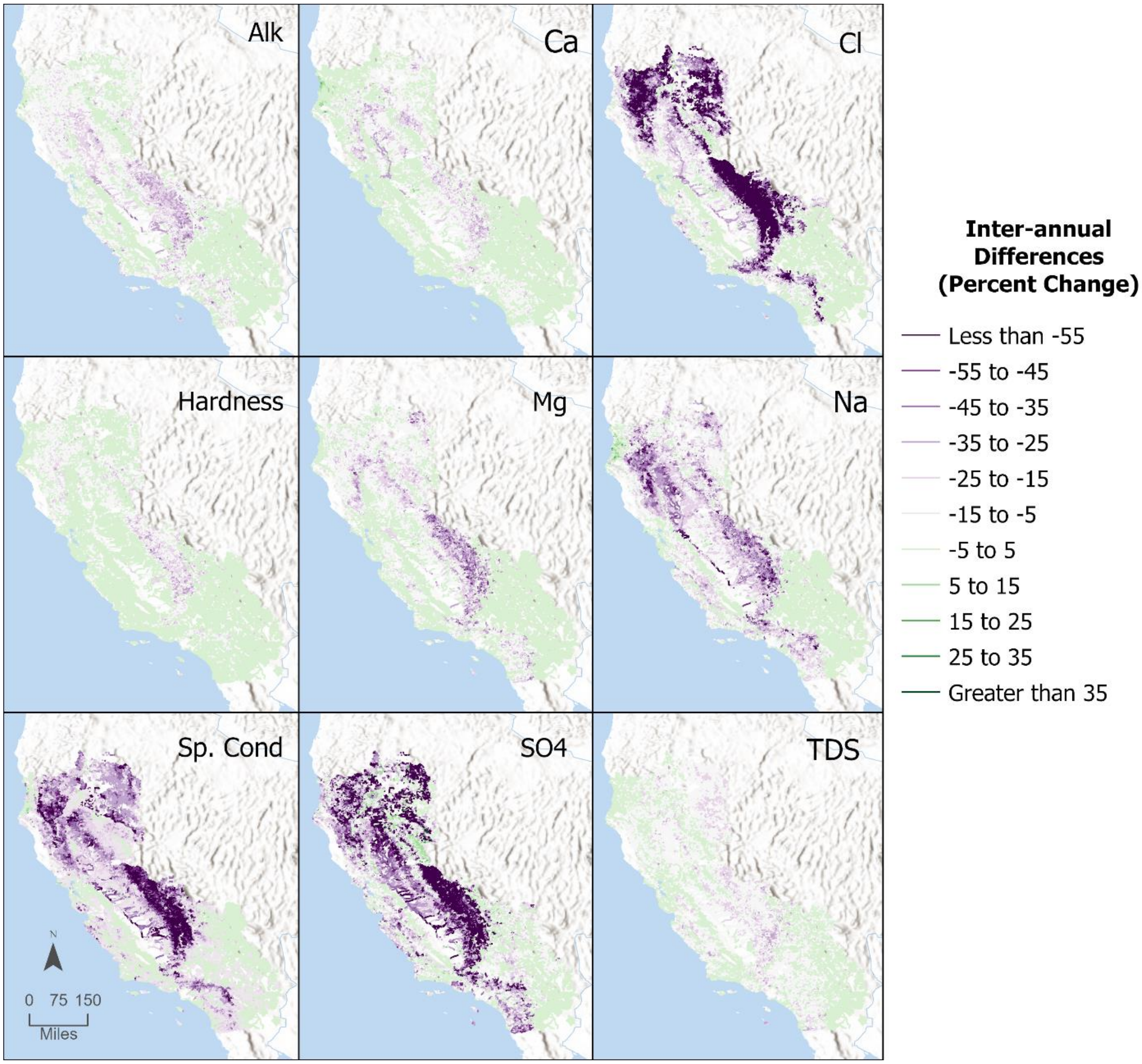


Figure 7: Inter-annual differences across all analytes, displayed by percent change.

Discussion

Our models performed well, explaining more than 70 percent of the variation in each water quality parameter across California. Since we created multiple models, including almost all major ions and integrated measures, we could look at water quality in greater detail than previous studies. The Olson and Cormier (2019) paper provided a national view of how specific conductivity reacts with differing temporal and spatial scales, but its training dataset had fewer

California sites than ours did, and its performance within California was not assessed. The use of StreamCat allowed us to create predictions for each stream segment in California with associated StreamCat data. This provided us with a more detailed view of how analyte concentrations interact and vary over space and time. This study demonstrates that national data sets can be used to develop models with high levels of performance suitable for state or regional applications.

Our models' spatial and temporal sensitivity will aid in adjusting aquatic life and water quality thresholds for current climate patterns and will allow future water quality responses to climatic shifts to be estimated. Our model predictions have the potential to help us understand the effects of future climates on California streams. For example, these models can be used to estimate the future shifts in natural background water quality in California streams due to climate change, similar to work done by Olson and Cormier (2019). In addition to estimating future climatic shifts, our model predictions may help researchers study the effects of aquatic stress, metabolic rates, and oxygen demand by providing natural background levels to calculate the amount of divergence and its impact on aquatic species.

The seasonal prediction capabilities of our models provide land managers and researchers with a way of identifying streams suffering from freshwater salinization syndrome (FSS). Kaushal et al. (2018) state that freshwater salinization syndrome has many far-reaching effects regarding ground and surface water, the quality of drinking water, and ecosystem functions. Given the seasonal predictions our models can make, we suggest water quality monitoring should occur multiple times a year and consider the sampling year's precipitation levels. These patterns will help researchers and managers target areas that need the greatest conservation and restoration efforts to decrease the occurrence of freshwater salinization syndrome and the metabolic stress on native aquatic organisms.

Our model predictions have the potential to be a valuable tool for ensuring that water quality targets are both protective and achievable. To illustrate how natural background levels can inform the development of water quality objectives, we plotted current objective values and our predictions (Fig. 8). The plot suggests that the current Basin Plan may be under protecting many streams, such as the Santa Ana River Reach 3, while over protecting others similar to the San Jacinto Reach 6 (for SO_4 and TDS, SWRCB Resolution No. R8-2004-0001). The predictions created by our models encompass all seasons and a wide range of precipitation levels, providing an in-depth analysis of what natural background analytic levels should be in California streams that managers can use to set thresholds appropriate for each stream and season.

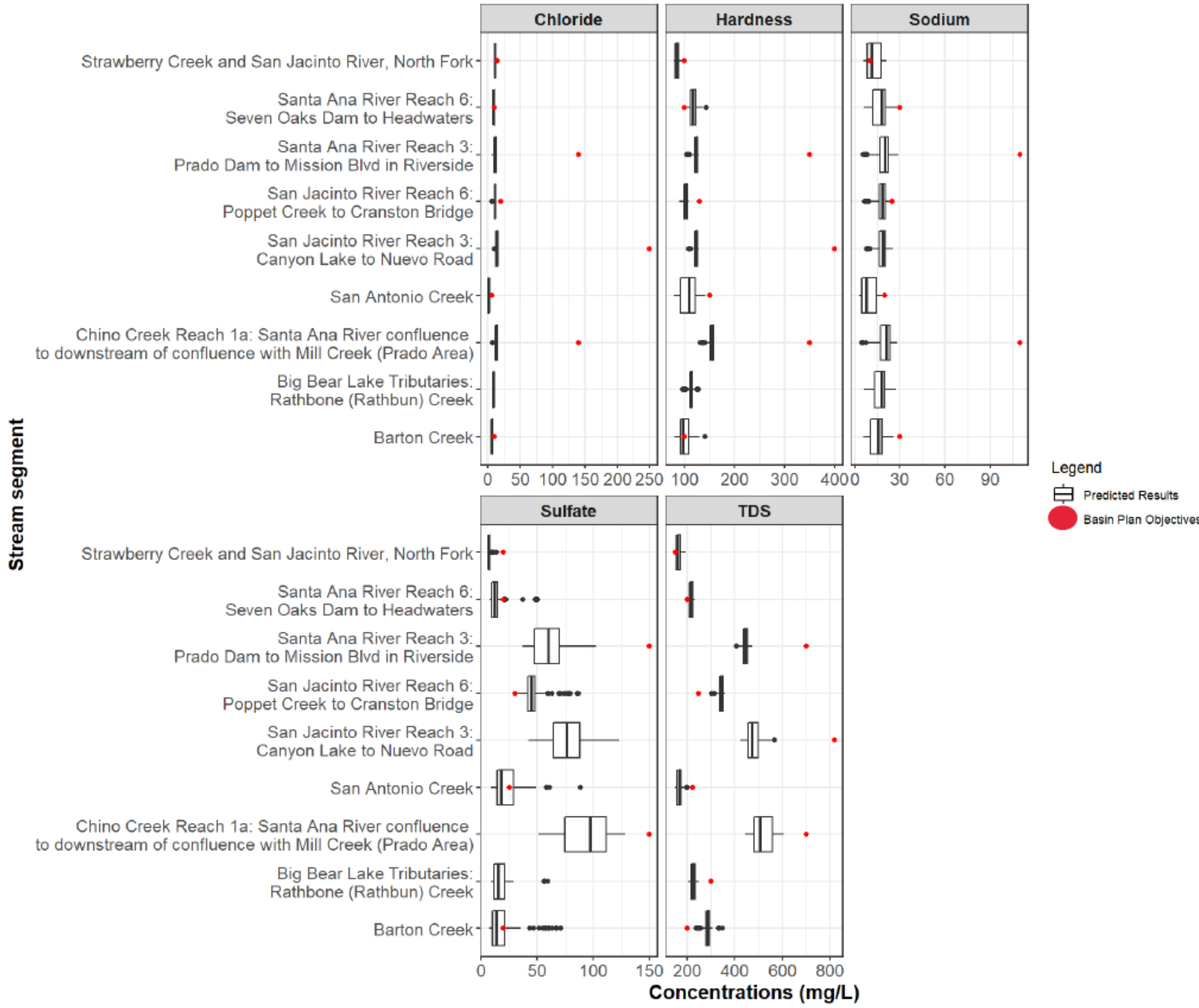


Figure 8:Comparison of the Santa Ana Basin Plan objective values to our model estimations at 9 streams in the Santa Ana area.

Natural background water quality predictions can provide much needed insight regarding alteration from natural levels. Alteration can be calculated by comparing observations to natural background predictions. These alteration calculations can then be modeled as response to human activities, allowing the relative impact of different activities to be assessed. The use of models allows for a broader understanding of potential drivers and the predictions also allows for individual stream comparisons.

REFERENCES

Agha, Mickey, Yuzo R Yanagitsuru, Nann A Fangue, A Justin Nowakowski, Laura V Kojima, Joseph J Cech, Melissa K Riley, Janna Freeman, Dennis E Cocherell, and Brian D Todd. “Physiological Consequences of Rising Water Salinity for a Declining Freshwater Turtle.” *Conservation Physiology* 7, no. 1 (August 21, 2019): coz054. https://doi.org/10.1093/conphys/coz054.

Akcil A, Koldas S. Acid Mine Drainage (AMD): causes, treatment and case studies. Journal of Cleaner Production. 2006;14(12):1139–1145. (Improving Environmental, Economic and Ethical Performance in the Mining Industry. Part 2. Life cycle and process analysis and technical issues). doi:10.1016/j.jclepro.2004.09.006

Azimi S, Rocher V. Influence of the water quality improvement on fish population in the Seine River (Paris, France) over the 1990-2013 period. The Science of the Total Environment. 2016;542(Pt A):955–964. doi:10.1016/j.scitotenv.2015.10.094

Basin Plan | Santa Ana Regional Water Quality Control Board. [accessed 2022 Nov 27]. https://www.waterboards.ca.gov/santaana/water_issues/programs/basin_plan/

“California Environmental Data Exchange Network (CEDEN) | San Francisco Estuary Institute.” Accessed December 9, 2022. https://www.sfei.org/projects/california-environmental-data-exchange-network-ceden.

Carpenter, Stephen R., Stuart G. Fisher, Nancy B. Grimm, and James F. Kitchell. “Global Change and Freshwater Ecosystems.” *Annual Review of Ecology and Systematics* 23, no. 1 (1992): 119–39. https://doi.org/10.1146/annurev.es.23.110192.001003.

Daniel WM, Infante DM, Hughes RM, Tsang Y-P, Esselman PC, Wieferich D, Herreman K, Cooper AR, Wang L, Taylor WW. Characterizing coal and mineral mines as a regional source of stress to stream fish assemblages. Ecological Indicators. 2015;50:50–61. doi:10.1016/j.ecolind.2014.10.018

Fox et al., “Assessing the Accuracy and Stability of Variable Selection Methods for Random Forest Modeling in Ecology.”

Gaber HS, El-Kasheif MA, Ibrahim SA, Authman MMN. Effect of water pollution in El-Rahawy drainage canal on hematology and organs of freshwater fish Clarias gariepinus. World Applied Sciences Journal. 2013;21(3):329–341.

Greenwell BM (2017). “pdp: An R Package for Constructing Partial Dependence Plots.” *The R Journal*, **9**(1), 421–436. https://journal.r-project.org/archive/2017/RJ-2017-016/index.html.

Hawkins CP, Olson JR, Hill RA. The reference condition: predicting benchmarks for ecological and water-quality assessments. Journal of the North American Benthological Society. 2010;29(1):312–343. doi:10.1899/09-092.1

Hill RA, Weber MH, Leibowitz SG, Olsen AR, Thornbrugh DJ. The Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the Conterminous United States. JAWRA Journal of the American Water Resources Association. 2016;52(1):120–128. doi:10.1111/1752-1688.12372

Ingram J. Impacts of environmental and hydrologic factors on urban stream water quality. Electronic Theses and Dissertations. 2020 Jan 1. https://ir.library.louisville.edu/etd/3563. doi:10.18297/etd/3563

Kaushal SS, Likens GE, Pace ML, Utz RM, Haq S, Gorman J, Grese M. Freshwater salinization syndrome on a continental scale. Proceedings of the National Academy of Sciences. 2018 [accessed 2022 Nov 27];115(4). https://pnas.org/doi/full/10.1073/pnas.1711234115. doi:10.1073/pnas.1711234115

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, *28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Liaw A, Wiener M (2002). "Classification and Regression by randomForest." *R News*, **2**(3), 18-22. https://CRAN.R-project.org/doc/Rnews/.

Links between climate change, water-table depth, and water chemistry in a mineralized mountain watershed - ScienceDirect. [accessed 2022 Nov 5]. https://www.sciencedirect.com/science/article/pii/S0883292713001741

Lohani MB, Singh A, Rupainwar DC, Dhar DN. Seasonal variations of heavy metal contamination in river Gomti of Lucknow city region. Environmental Monitoring and Assessment. 2008;147(1):253–263. doi:10.1007/s10661-007-0117-1

Ma L, Yates SR, Ashworth D. Parent and conjugated estrogens and progestagens in surface water of the Santa Ana River: Determination, occurrence, and risk assessment. Environmental Toxicology and Chemistry. 2016;35(11):2657–2664. doi:10.1002/etc.3447

Maps throughout this thesis were created using ArcGIS® software by Esri. ArcGIS® and ArcMap™ are the intellectual property of Esri and are used herein under license. Copyright © Esri. All rights reserved. For more information about Esri® software, please visit www.esri.com.

Mazor RD. Bioassessment of Perennial Streams in Southern California: A Report on the First Five Years of the Stormwater Monitoring Coalition's Regional Stream Survey. :2.

McGrane SJ. Impacts of urbanisation on hydrological and water quality dynamics, and urban water management: a review. Hydrological Sciences Journal. 2016;61(13):2295–2311. doi:10.1080/02626667.2015.1128084

McKay L, et. al. NHDPlus Version 2: User Guide. 2012. 2015. [accessed 2021 Nov 12];(3). https://wildlife.ca.gov/Conservation/SSC/Fishes

Ode PR, Rehn AC, Mazor RD, Schiff KC, Stein ED, May JT, Brown LR, Herbst DB, Gillett D, Lunde K, et al. Evaluating the adequacy of a reference-site pool for ecological assessments in environmentally complex regions. Freshwater Science. 2016;35(1):237–248. doi:10.1086/684003

Olson JR, Cormier SM. Modeling Spatial and Temporal Variation in Natural Background Specific Conductivity. Environmental Science & Technology. 2019;53(8):4316–4325. doi:10.1021/acs.est.8b06777

Omernik JM, Griffith GE. Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework. Environmental Management. 2014;54(6):1249–1266. doi:10.1007/s00267-014-0364-1

Paul MJ, Meyer JL. Streams in the Urban Landscape. Annual Review of Ecology and Systematics. 2001;32(1):333–365. doi:10.1146/annurev.ecolsys.32.081501.114040

Peters, Norman E. "Evaluation of Environmental Factors Affecting Yields of Major Dissolved Ions of Streams in the United States." USGS Numbered Series. *Evaluation of Environmental Factors Affecting Yields of Major Dissolved Ions of Streams in the United States*. Vol. 2228. Water Supply Paper. U.S. G.P.O., 1984. https://doi.org/10.3133/wsp2228.

Physiological consequences of rising water salinity for a declining freshwater turtle | Conservation Physiology | Oxford Academic. [accessed 2022 Nov 5]. https://academic.oup.com/conphys/article/7/1/coz054/5552357

Poff, N. L., Allan, J. D., Bain, M. B., Karr, J. R., Prestegaard, K. L., Richter, B. D., Sparks, R. E., & Stromberg, J. (1997). The natural flow regime: A paradigm for river conservation and restoration. *BioScience*, *47*(11), 769-784. https://doi.org/10.2307/1313099

Pond GJ, Passmore ME, Borsuk FA, Reynolds L, Rose CJ. Downstream effects of mountaintop coal mining: comparing biological conditions using family- and genus-level macroinvertebrate bioassessment tools. Journal of the North American Benthological Society. 2008;27(3):717–737. doi:10.1899/08-015.1

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Robin Genuer, Jean-Michel Poggi, Christine Tuleau-Malot. VSURF: An R Package for Variable Selection Using Random Forests. The R Journal, 2015, 7 (2), pp.19-33. ffhal-01251924f

"Service Based Budgeting." https://wildlife.ca.gov/Budget/Service-Based-Budgeting

Southern California Stormwater Monitoring Coalition." Accessed December 9, 2022. https://smc.sccwrp.org/.

Tang, "The California Stream Condition Index (CSCI): A New Statewide Biological Scoring Tool for Assessing the Health of Freshwater Streams."

Thompson AR, Baskin JN, Swift CC, Haglund TR, Nagel RJ. Influence of habitat dynamics on the distribution and abundance of the federally threatened Santa Ana Sucker, Catostomus santaanae, in the Santa Ana River. Environmental Biology of Fishes. 2010;87(4):321–332. doi:10.1007/s10641-010-9604-2

U.S. Department of the Interior, U.S. Geological Survey. (1999). National Water-Quality Assessment (NAWQA) Program. [Washington, D.C.] :USGS NAWQA, 1999.

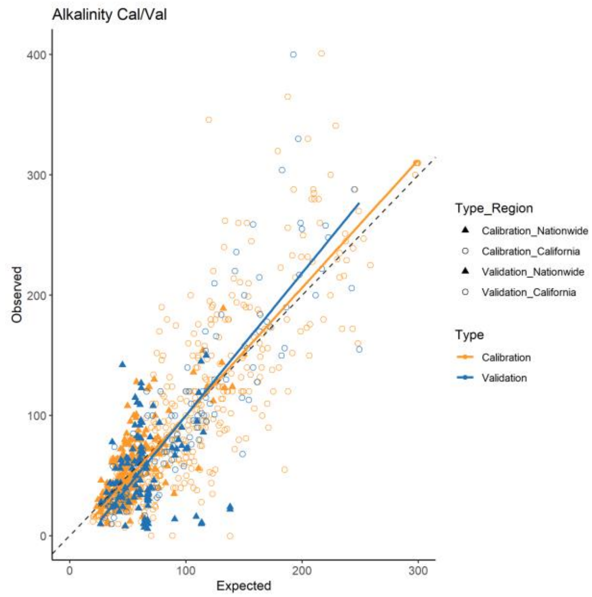
US EPA, OW. "National Aquatic Resource Surveys." Collections and Lists, August 29, 2013. United States. https://www.epa.gov/national-aquatic-resource-surveys.

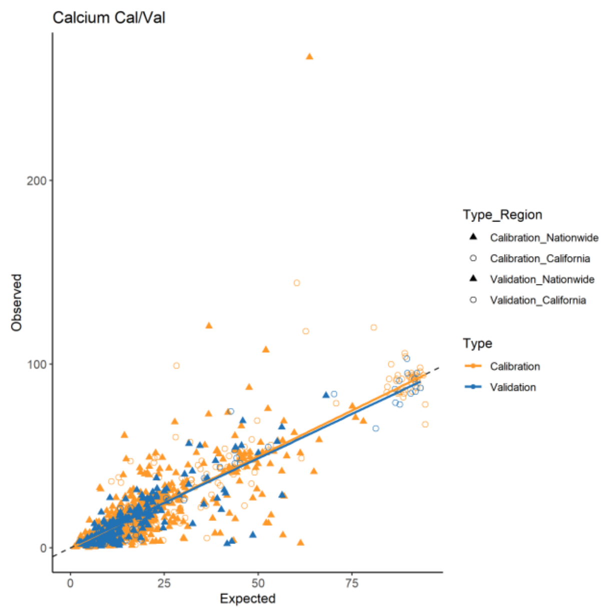
USGS Open-File Report 2009-1097: Effects of Wastewater Discharges on Endocrine and Reproductive Function of Western Mosquitofish (Gambusia spp.) and Implications for the Threatened Santa Ana Sucker (Catostomus santaanae). [accessed 2022 Nov 5]. https://pubs.usgs.gov/of/2009/1097/

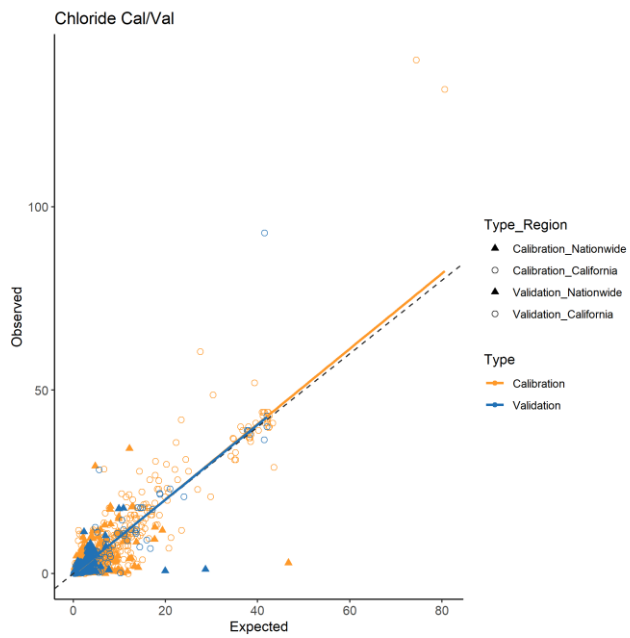
Vander Laan, Jacob J., Charles P. Hawkins, John R. Olson, and Ryan A. Hill. "Linking Land Use, in-Stream Stressors, and Biological Condition to Infer Causes of Regional Ecological Impairment in Streams." *Freshwater Science* 32, no. 3 (2013): 801–20. https://doi.org/10.1899/12-186.1.

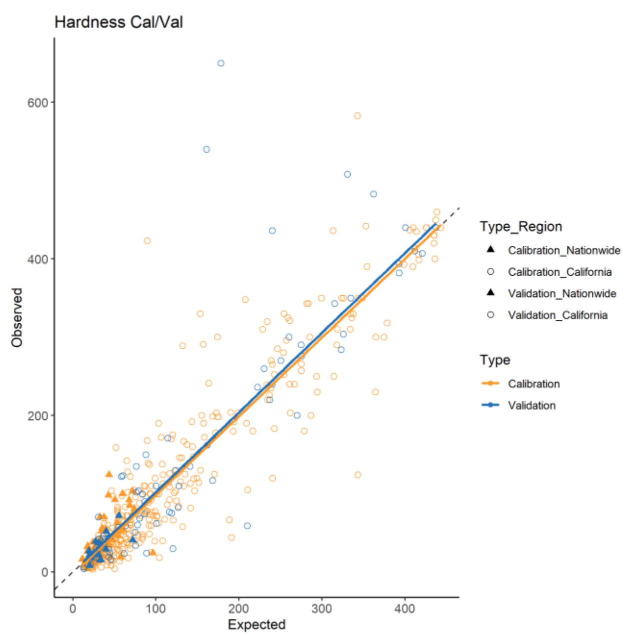
Walsh CJ, Roy AH, Feminella JW, Cottingham PD, Groffman PM, Morgan RP. The urban stream syndrome: current knowledge and the search for a cure. Journal of the North American Benthological Society. 2005;24(3):706–723. doi:10.1899/04-028.1

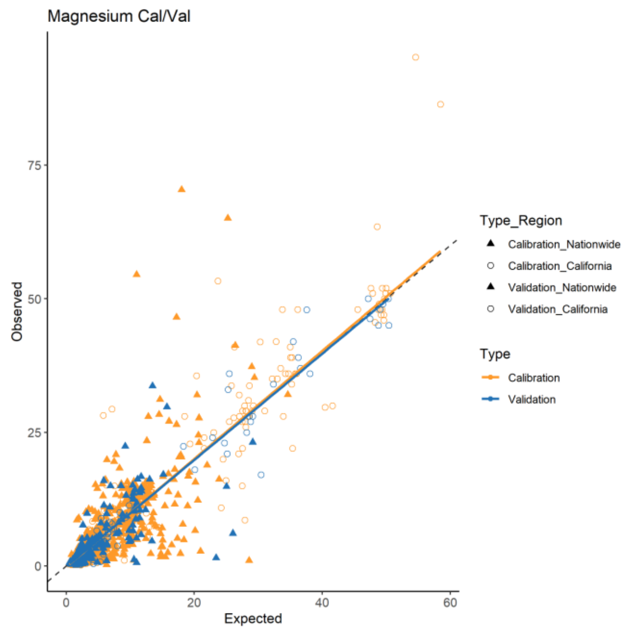
APPENDIX A
MODEL PERFORMANCE PLOTS

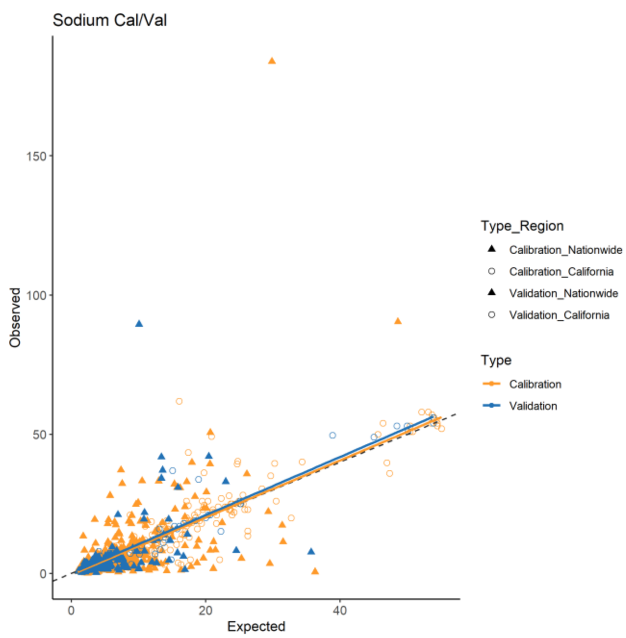


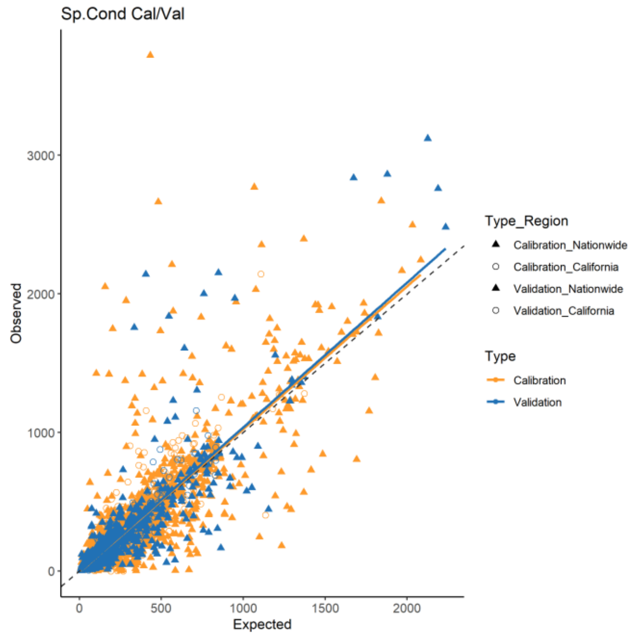


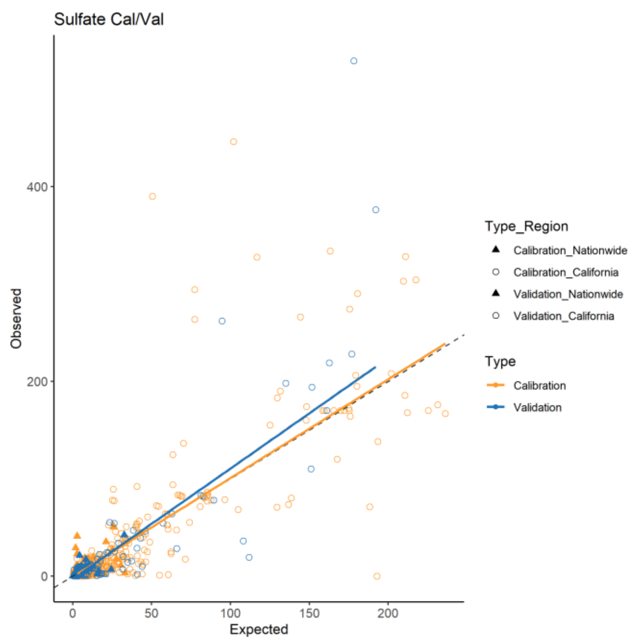

Chloride Cal/Val
Observed
Expected
0
50
100
0
20
40
60
80
Type_Region
Calibration_Nationwide
Calibration_California
Validation_Nationwide
Validation_California
Type
Calibration
Validation

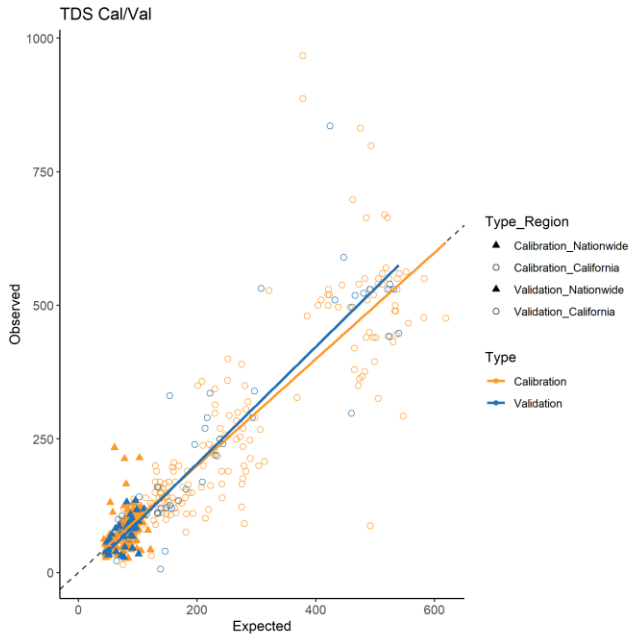

Hardness Cal/Val
Observed
Expected
0
200
400
600
0
100
200
300
400
Type_Region
Calibration_Nationwide
Calibration_California
Validation_Nationwide
Validation_California
Type
Calibration
Validation


Magnesium Cal/Val
Observed
Expected
0
25
50
75
0
20
40
60
Type_Region
Calibration_Nationwide
Calibration_California
Validation_Nationwide
Validation_California
Type
Calibration
Validation


Sodium Cal/Val
Observed
Expected
0
50
100
150
0
20
40
Type_Region
Calibration_Nationwide
Calibration_California
Validation_Nationwide
Validation_California
Type
Calibration
Validation


Sp.Cond Cal/Val
Observed
Expected
0
500
1000
1500
2000
3000
Type_Region
Calibration_Nationwide
Calibration_California
Validation_Nationwide
Validation_California
Type
Calibration
Validation


Sulfate Cal/Val
Observed
Expected
0
50
100
150
200
400
Type_Region
Calibration_Nationwide
Calibration_California
Validation_Nationwide
Validation_California
Type
Calibration
Validation



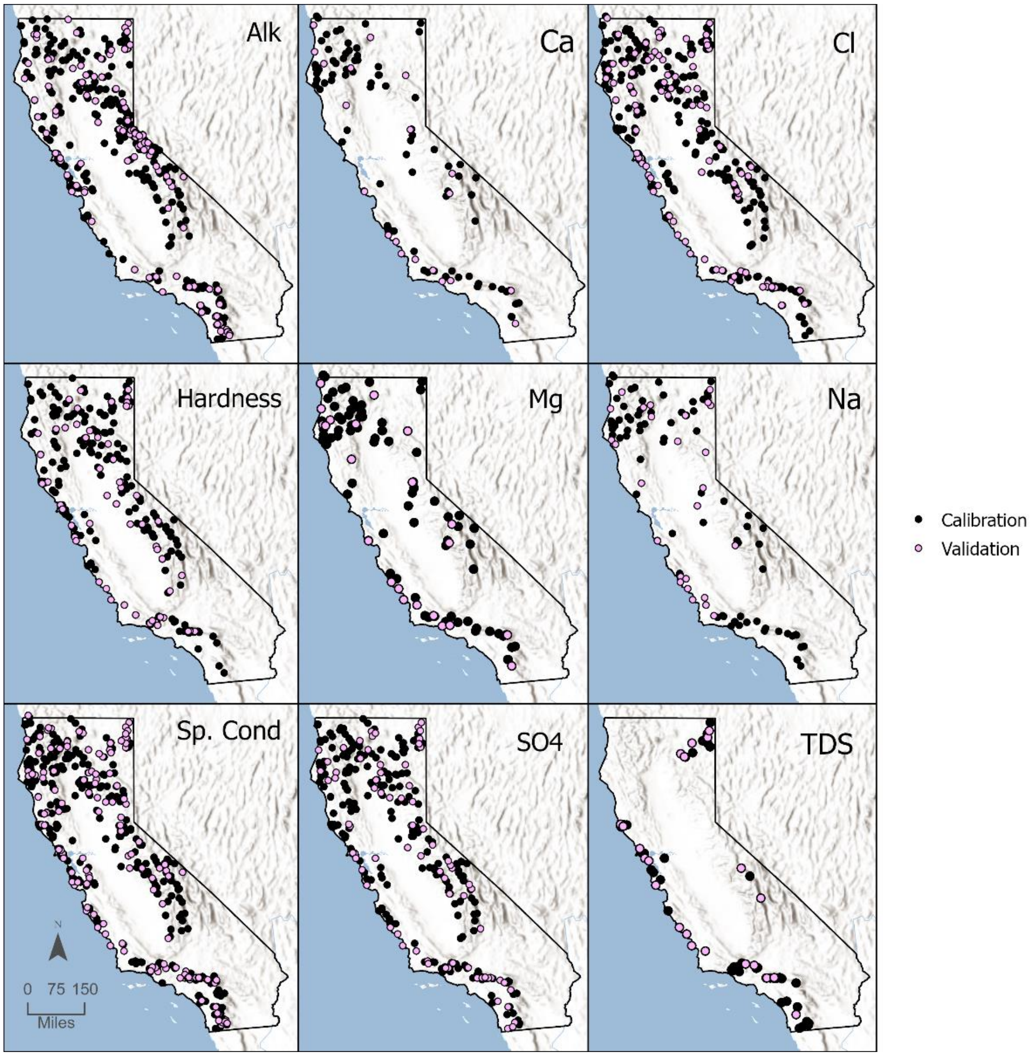
APPENDIX B

PARTIAL DEPENDENCE PLOTS ANALYSIS

		Calcium model	Chloride model	Magnesium model	Sodium model	Sulfate model	Hardness model	TDS model	Sp. Conductivity model	Alkalinity model
Geology	% Aluminum Oxide in Watershed	NA	- / +	-	NA	-	-	NA	- / +	- / +
	% Aluminum Oxide in Catchment	NA	NA	NA	+	NA	NA	NA	NA	NA
	% Calcium Oxide in Catchment	NA	NA	NA	- / +	NA	NA	NA	NA	NA
	% Calcium Oxide in Watershed	+	NA	+	NA	NA	+	NA	+	NA
	% Colluvial Sediment in Watershed	NA	NA	NA	NA	NA	NA	NA	NA	NA
	% Iron Oxide in Catchment	NA	NA	NA	+	NA	NA	NA	NA	NA
	% Iron Oxide in Watershed	- / +	NA	NA	NA	-	-	-	NA	- / +
	% Magnesium Oxide in Catchment	NA	NA	NA	NA	NA	NA	NA	NA	NA
	% Magnesium Oxide in Watershed	- / +	+	+	NA	NA	NA	NA	- / +	NA
	% Nitrogen in Catchment	+	NA	+	NA	NA	NA	NA	NA	NA
	% Nitrogen in Watershed	NA	NA	NA	NA	-	NA	+	- / +	NA
	% Phosphorus Pentoxide in Catchment	NA	NA	NA	NA	NA	NA	NA	NA	NA
	% Phosphorus Pentoxide in Watershed	NA	NA	NA	NA	- / +	NA	NA	NA	NA
	% Potassium Oxide in Watershed	- / +	NA	-	NA	NA	-	NA	- / +	- / +
	% Silicic Residual Matereial in Watershed	-	NA	NA	NA	NA	NA	-	NA	NA
	% Silicon Dioxide in Watershed	-	NA	-	NA	NA	NA	NA	-	NA
	% Sodium Oxide in Catchment	- / +	- / +	NA	- / +	NA	- / +	NA	NA	NA
	% Sodium Oxide in Watershed	-	- / +	-	-	-	-	-	- / +	- / +
	% Sulfur in Catchment	NA	+	+	+	+	- / +	NA	+	+
	% Sulfur in Watershed	+	+	+	+	+	+	+	+	+
Soils	% Sand in Catchment	NA	NA	NA	NA	NA	NA	NA	NA	NA
	% Sand in Watershed	NA	- / +	NA	NA	NA	NA	NA	- / +	- / +
	% Clay in Catchment	NA		NA	NA	NA	NA	NA	NA	NA
	% Clay in Watershed	NA	+	NA	NA	NA	+	+	NA	+
	Soil Erodability in Watershed (K Factor)	NA	- / +	NA	NA	NA	+	NA	NA	NA
	Soil Erodability in Catchment (K Factor)	NA	- / +	NA	NA	NA	NA	NA	NA	NA
Climatic	12 month Maximum Temperature Average	NA	+	+	+	+	+	+	+	+
	12 Month Mean Temperature Average	+	+	+	+	+	+	NA	+	+
	Tmean0	NA	NA	NA	NA	- / +	NA	NA	NA	NA
	12 month Monthly Evapotranspiration Avg	NA	NA	NA	NA	NA	NA	NA	- / +	- / +
	12 month Precipitation Average	-	-	NA	-	-	-	- / +	-	-
	3 month Precipitaion Average	-	- / +	-	-	- / +	-	-	-	-
	6 month Precipitation Average	-	-	-	-	- / +	NA	-	-	-
Topographic	Base Flow Index in Catchment	- / +	- / +	- / +	- / +	- / +	- / +	NA	-	- / +
	Base Flow index in Watershed	- / +	-	- / +	- / +	- / +	- / +	-	-	-
	Compressive Strength of Catchment	NA	NA	NA	NA	NA	NA	+	NA	NA
	Compressive Strength of Watershed	NA	NA	NA	NA	NA		NA	- / +	NA
	Elevation in Catchment	NA	-	NA	NA	NA	+	+ / -	-	+ / -
	Elevation in Watershed	+ / -	NA	-	NA	- / +	+ / -	+ / -	NA	+ / -
	Hydrologic Conductivity in Catchment	NA	NA	NA	+	NA	NA	NA	NA	NA
	Hydrologic Conductivity in Watershed	NA	-	NA	NA	NA	NA	NA	NA	NA
	Wetness Index in Catchment	NA	NA	NA	NA	NA	-	NA	NA	- / +
	Wetness Index in Watershed	- / +	NA	- / +	NA	-	-	NA	NA	- / +

APPENDIX C

CALIBRATION AND VALIDATION ANALYTE MAPS AT THE CALIFORNIA SCALE



APPENDIX D

NUMBER OF REFERENCE SITES IN EACH ECOREGION BY ANALYTE

Ecoregion	Type	Alkalinity	Calcium	Chloride	Hardness	Magnesiur	Sodium	Sp.Cond	Sulfate	TDS
1	Calibration	38	90	152	39	89	94	586	126	47
4	Calibration	29	30	80	22	30	28	78	66	NA
5	Calibration	198	230	311	103	231	187	287	340	65
6	Calibration	83	79	148	108	78	69	181	89	84
7	Calibration	2	1	1	2	1	1	4	1	NA
8	Calibration	84	16	69	44	16	20	130	73	34
9	Calibration	34	3	30	30	3	11	81	29	7
13	Calibration	116	73	124	18	74	63	452	114	94
14	Calibration	1	NA	NA	NA	NA	NA	4	1	NA
78	Calibration	47	25	46	27	25	20	74	44	NA
80	Calibration	12	38	38	2	38	37	130	10	10
85	Calibration	29	NA	5	1	NA	1	31	13	4
2	Calibration	NA	2	2	NA	2	2	2	1	NA
10	Calibration	NA	8	9	NA	8	9	7	NA	NA
11	Calibration	NA	8	8	NA	8	7	22	NA	NA
12	Calibration	NA	1	1	NA	1	2	1	NA	NA
15	Calibration	NA	22	25	NA	22	22	76	10	NA
16	Calibration	NA	10	12	NA	10	13	90	NA	NA
17	Calibration	NA	52	53	NA	52	50	273	20	NA
18	Calibration	NA	9	7	NA	9	8	15	NA	NA
19	Calibration	NA	9	12	NA	9	14	87	NA	NA
20	Calibration	NA	3	3	NA	3	2	61	NA	NA
21	Calibration	NA	29	33	NA	29	31	484	1	NA
22	Calibration	NA	5	5	NA	5	3	44	NA	NA
23	Calibration	NA	8	7	NA	8	7	235	NA	NA
25	Calibration	NA	5	4	NA	5	6	8	NA	NA
26	Calibration	NA	5	6	NA	5	6	233	NA	NA
30	Calibration	NA	5	5	NA	5	3	13	NA	NA
35	Calibration	NA	1	2	NA	1	2	76	NA	NA
38	Calibration	NA	1	NA	NA	1	2	7	NA	NA
39	Calibration	NA	1	1	NA	1	NA	15	NA	NA
40	Calibration	NA	1	1	NA	1	1	1	NA	NA
41	Calibration	NA	3	3	NA	3	2	32	NA	NA
42	Calibration	NA	1	1	NA	1	1	28	NA	NA
43	Calibration	NA	12	11	NA	12	11	58	NA	NA
44	Calibration	NA	50	75	NA	50	50	105	78	NA
45	Calibration	NA	6	7	NA	6	3	388	5	NA
49	Calibration	NA	3	3	NA	3	4	59	NA	NA
50	Calibration	NA	153	168	NA	153	151	282	143	NA
58	Calibration	NA	202	207	NA	202	211	142	178	NA
60	Calibration	NA	3	NA	NA	3	3	1	NA	NA
62	Calibration	NA	5	51	NA	5	6	46	NA	NA
65	Calibration	NA	8	7	NA	8	7	273	NA	NA
66	Calibration	NA	5	5	NA	5	5	1390	NA	NA
67	Calibration	NA	3	29	NA	3	4	191	22	NA
68	Calibration	NA	3	2	NA	3	1	78	NA	NA
69	Calibration	NA	1	1	NA	1	1	131	NA	NA
70	Calibration	NA	1	2	NA	1	NA	4	NA	NA
75	Calibration	NA	1	1	NA	1	1	226	NA	NA
77	Calibration	NA	5	5	NA	5	4	17	NA	NA

79	Calibration	NA	1	1	NA	1	1	43	NA	NA
81	Calibration	NA	2	2	NA	2	2	32	1	NA
82	Calibration	NA	8	9	NA	8	8	84	NA	NA
83	Calibration	NA	1	1	NA	1	1	1	NA	NA
28	Calibration	NA	NA	NA	NA	NA	NA	16	NA	NA
29	Calibration	NA	NA	NA	NA	NA	NA	4	NA	NA
31	Calibration	NA	NA	NA	NA	NA	NA	1	NA	NA
37	Calibration	NA	NA	NA	NA	NA	NA	2	NA	NA
46	Calibration	NA	NA	NA	NA	NA	NA	4	NA	NA
51	Calibration	NA	NA	NA	NA	NA	NA	59	NA	NA
52	Calibration	NA	NA	NA	NA	NA	NA	1	NA	NA
55	Calibration	NA	NA	NA	NA	NA	NA	5	NA	NA
63	Calibration	NA	NA	NA	NA	NA	NA	100	NA	NA
71	Calibration	NA	NA	NA	NA	NA	NA	22	NA	NA
72	Calibration	NA	NA	NA	NA	NA	NA	1	NA	NA
73	Calibration	NA	NA	NA	NA	NA	NA	24	NA	NA
74	Calibration	NA	NA	NA	NA	NA	NA	12	NA	NA
76	Calibration	NA	NA	NA	NA	NA	NA	2	NA	NA
84	Calibration	NA	NA	NA	NA	NA	NA	9	NA	NA
1	Validation	21	26	31	9	27	22	130	26	7
4	Validation	16	8	15	7	8	4	26	11	NA
5	Validation	45	54	90	24	53	51	67	95	20
6	Validation	18	21	32	27	22	23	53	21	21
8	Validation	27	5	16	11	5	NA	36	18	5
9	Validation	9	NA	9	6	NA	3	21	7	4
10	Validation	1	1	NA	NA	1	NA	4	NA	NA
13	Validation	26	16	36	4	15	18	108	32	24
15	Validation	1	5	2	NA	5	5	17	4	NA
17	Validation	6	12	11	NA	12	14	63	7	NA
44	Validation	7	13	22	NA	13	13	27	16	NA
45	Validation	1	1	NA	NA	1	3	84	1	NA
50	Validation	16	34	46	NA	34	36	71	33	NA
67	Validation	5	2	4	NA	2	1	47	6	NA
78	Validation	7	2	16	7	2	4	18	7	NA
80	Validation	5	8	11	2	8	9	35	5	2
85	Validation	10	1	2	NA	1	NA	10	4	1
11	Validation	NA	1	1	NA	1	2	7	NA	NA
12	Validation	NA	1	1	NA	1	NA	1	NA	NA
16	Validation	NA	6	4	NA	6	3	27	NA	NA
18	Validation	NA	1	3	NA	1	2	3	NA	NA
19	Validation	NA	7	4	NA	7	2	20	NA	NA
21	Validation	NA	12	8	NA	12	10	137	NA	NA
25	Validation	NA	1	2	NA	1	NA	3	NA	NA
26	Validation	NA	1	NA	NA	1	NA	48	NA	NA
30	Validation	NA	1	1	NA	1	3	7	NA	NA
35	Validation	NA	1	NA	NA	1	NA	19	NA	NA
38	Validation	NA	1	2	NA	1	NA	1	NA	NA
43	Validation	NA	3	4	NA	3	4	11	NA	NA
49	Validation	NA	1	1	NA	1	NA	21	NA	NA
58	Validation	NA	58	54	NA	58	49	30	44	NA

62	Validation	NA	2	8	NA	2	1	14	NA	NA
66	Validation	NA	1	1	NA	1	1	352	NA	NA
70	Validation	NA	1	NA	NA	1	2	3	NA	NA
82	Validation	NA	3	2	NA	3	3	16	NA	NA
7	Validation	NA	NA	2	NA	NA	NA	NA	1	NA
23	Validation	NA	NA	1	NA	NA	1	57	NA	NA
60	Validation	NA	NA	3	NA	NA	NA	2	NA	NA
65	Validation	NA	NA	1	NA	NA	1	66	NA	NA
68	Validation	NA	NA	1	NA	NA	2	18	NA	NA
20	Validation	NA	NA	NA	NA	NA	1	18	NA	NA
22	Validation	NA	NA	NA	NA	NA	2	9	NA	NA
39	Validation	NA	NA	NA	NA	NA	1	6	NA	NA
41	Validation	NA	NA	NA	NA	NA	1	12	NA	NA
77	Validation	NA	NA	NA	NA	NA	1	1	NA	NA
14	Validation	NA	NA	NA	NA	NA	NA	3	NA	NA
28	Validation	NA	NA	NA	NA	NA	NA	12	NA	NA
31	Validation	NA	NA	NA	NA	NA	NA	1	NA	NA
37	Validation	NA	NA	NA	NA	NA	NA	1	NA	NA
42	Validation	NA	NA	NA	NA	NA	NA	7	NA	NA
46	Validation	NA	NA	NA	NA	NA	NA	3	NA	NA
51	Validation	NA	NA	NA	NA	NA	NA	12	NA	NA
52	Validation	NA	NA	NA	NA	NA	NA	1	NA	NA
55	Validation	NA	NA	NA	NA	NA	NA	2	NA	NA
63	Validation	NA	NA	NA	NA	NA	NA	16	NA	NA
69	Validation	NA	NA	NA	NA	NA	NA	31	NA	NA
71	Validation	NA	NA	NA	NA	NA	NA	5	NA	NA
73	Validation	NA	NA	NA	NA	NA	NA	8	NA	NA
74	Validation	NA	NA	NA	NA	NA	NA	5	NA	NA
75	Validation	NA	NA	NA	NA	NA	NA	63	NA	NA
76	Validation	NA	NA	NA	NA	NA	NA	3	NA	NA
79	Validation	NA	NA	NA	NA	NA	NA	9	NA	NA
81	Validation	NA	NA	NA	NA	NA	NA	4	NA	NA
84	Validation	NA	NA	NA	NA	NA	NA	2	NA	NA
2	Validation	NA	NA	NA	NA	NA	NA	NA	1	NA

APPENDIX E

STREAMCAT AND PRISM VARIABLES EVALUATED IN MODEL CREATION AND ANALYSIS

Description	Variable name	Unit
Mean soil erodibility of soils within catchment on agricultural land	AgKffactCat	Unitless

Mean soil erodibility of soils within watershed on agricultural land	AgKffactWs	Unitless
Mean % aluminum oxide within catchment	Al2O3Cat	%
Mean % aluminum oxide within watershed	Al2O3Ws	%
Ground water discharge into streams in catchment ratio	BFICat	%
Ground water discharge into streams in watershed ratio	BFIWs	%
Density of canals, ditches, or pipelines within catchment	CanalDensCat	km/km2
Density of canals, ditches, or pipelines within watershed	CanalDensWs	km/km2
Mean percent Calcium Oxide within catchment	CaOCat	%
Mean percent Calcium Oxide within watershed	CaOWs	%
Mean percent clay content within catchment	ClayCat	%
Mean percent clay content within watershed	ClayWs	%
Mean lithological uniaxial compressive strength within catchment	CompStrgthCat	MPa
Mean lithological uniaxial compressive strength within watershed	CompStrgthWs	Mpa
Density of georeferenced dams within catchment	DamDensCat	#/km2
Density of georeferenced dams within watershed	DamDensWs	#/km2
Total possible volume of all reservoirs in catchment	DamNIDStorCat	m3/km2
Total possible volume of all reservoirs in watershed	DamNIDStorWs	m3/km2
Normal volume of all reservoirs in catchment	DamNrmStorCat	m3/km2
Normal volume of all reservoirs in watershed	DamNrmStorWs	m3/km2
Mean catchment elevation	ElevCat	m
Mean watershed elevation	ElevWs	m
Mean lithological ferri oxide in catchment	Fe2O3Cat	%
Mean lithological ferric oxide in watershed	Fe2O3Ws	%
Mean lithological hydraulic conductivity in catchment	HydrlCondCat	um/s
Mean lithological hydraulic conductivity in watershed	HydrlCondWs	um/s
Mean lithological potassium oxide in catchment	K2OCat	%
Mean lithological potassium oxide in watershed	K2OWs	%
Mean soil erodibility within catchment	KffactCat	Unitless
Mean soil erodibility within watershed	KffactWs	Unitless
Mean magnesium oxide in catchment	MgOCat	%
Mean magnesium oxide in watershed	MgOWs	%
Density of mines in catchment	MineDensCat	#/km2
Density of mines in watershed	MineDensWs	#/km2
Mean sodium oxide in catchment	Na2OCat	%

Mean sodium oxide in watershed	Na2OWs	%
Mean nitrogen in catchment	NCat	%
Mean nitrogen in watershed	NWs	%
Mean organic matter in catchment	OmCat	%/weight
Mean organic matter in watershed	OmWs	%/weight
Mean phosphorous oxide in catchment	P2O5Cat	%
Mean phosphorous oxide in watershed	P2O5Ws	%
Alkaline intrusive volcanic rock in catchment	PctAlkIntruVolCat	%
Alkaline intrusive volcanic rock in watershed	PctAlkIntruVolWs	%
Alluvium and fine textured coastal sediment in catchment	PctAlluvCoastCat	%
Alluvium and fine textured coastal sediment in watershed	PctAlluvCoastWs	%
Barren land cover in catchment	PctBl2016Cat	%
Barren land cover in watershed	PctBl2016Ws	%
Carbonate residual material in catchment	PctCarbResidCat	%
Carbonate residual material in watershed	PctCarbResidWs	%
Coarse coastal zone sediment in catchment	PctCoastCrsCat	%
Coarse coastal zone sediment in watershed	PctCoastCrsWs	%
Colluvial sediment in catchment	PctColluvSedCat	%
Colluvial sediment in watershed	PctColluvSedWs	%
Evergreen forest landcover in catchment	PctConif2016Cat	%
Evergreen forest landcover in watershed	PctConif2016Ws	%
Crop land use in catchment	PctCrop2016Cat	%
Crop land use in watershed	PctCrop2016Ws	%
Deciduous forest land cover in catchment	PctDecid2016Cat	%
Deciduous forest land cover in watershed	PctDecid2016Ws	%
Coarse eolian sediment in catchment	PctEolCrsCat	%
Coarse eolian sediment in watershed	PctEolCrsWs	%
Fine eolian sediment in catchment	PctEolFineCat	%
Fine eolian sediment in waterhsed	PctEolFineWs	%
extrusive volcanic rock in catchment	PctExtruVolCat	%
extrusive volcanic rock in watershed	PctExtruVolWs	%
Coarse textured glacial outwash and glacial lake sediment in catchment	PctGlacLakeCrsCat	%
Coarse textured glacial outwash and glacial lake sediment in watershed	PctGlacLakeCrsWs	%
Fine textured glacial lake sediment in catchment	PctGlacLakeFineCat	%
Fine textured glacial lake sediment in watershed	PctGlacLakeFineWs	%
Glacial till, clayey in catchment	PctGlacTilClayCat	%
Glacial till, clayey in watershed	PctGlacTilClayWs	%
Coarse textured glacial till in catchment	PctGlacTilCrsCat	%
Coarse textured glacial till in watershed	PctGlacTilCrsWs	%
Loamy glacial till in catchment	PctGlacTilLoamCat	%
Loamy glacial till in watershed	PctGlacTilLoamWs	%

Grassland/herbaceous landcover in catchment	PctGrs2016Cat	%
Grassland/herbaceous landcover in watershed	PctGrs2016Ws	%
Hay land use in catchment	PctHay2016Cat	%
Hay land use in watershed	PctHay2016Ws	%
Herbaceous wetland cover in catchment	PctHbWet2016Cat	%
Herbaceous wetland cover in watershed	PctHbWet2016Ws	%
Peat and much hydric soils in catchment	PctHydricCat	%
Peat and much hydric soils in watershed	PctHydricWs	%
Ice/snow land cover in catchment	PctIce2016Cat	%
Ice/snow land cover in watershed	PctIce2016Ws	%
Mean impervious surfaces in catchment	PctImp2011Cat	%
Mean impervious surfaces in watershed	PctImp2011Ws	%
Mixed deciduous/evergreen forest cover in catchment	PctMxFst2016Cat	%
Mixed deciduous/evergreen forest cover in watershed	PctMxFst2016Ws	%
non-carbonate residual material in catchment	PctNonCarbResidCat	%
non-carbonate residual material in watershed	PctNonCarbResidWs	%
Open water land over in catchment	PctOw2016Cat	%
Open water land over in watershed	PctOw2016Ws	%
Saline like sediment in catchment	PctSalLakeCat	%
Saline like sediment in watershed	PctSalLakeWs	%
Shrub/scrub land cover in catchment	PctShrb2016Cat	%
Shrub/scrub land cover in watershed	PctShrb2016Ws	%
Silicic residual material in catchment	PctSilicicCat	%
Silicic residual material in watershed	PctSilicicWs	%
Developed, high-intensity land use in catchment	PctUrbHi2016Cat	%
Developed, high-intensity land use in watershed	PctUrbHi2016Ws	%
Developed, low-intensity land use in catchment	PctUrbLo2016Cat	%
Developed, low-intensity land use in watershed	PctUrbLo2016Ws	%
Developed, medium-intensity land use in catchment	PctUrbMd2016Cat	%
Developed, medium-intensity land use in watershed	PctUrbMd2016Ws	%
Developed, open space land use in catchment	PctUrbOp2016Cat	%
Developed, open space land use in watershed	PctUrbOp2016Ws	%
Catchment area that is water	PctWaterCat	%
Watershed area that is water	PctWaterWs	%
Woody wetland land cover in catchment	PctWdWet2016Cat	%
Woody wetland land cover in watershed	PctWdWet2016Ws	%
Mean permeability of soils in catchment	PermCat	cm/hr
Mean permeability of soils in watershed	PermWs	cm/hr
30 year normal mean precipitation in catchment	PrecipCat	mm
30 year normal mean precipitation in watershed	PrecipWs	mm

Mean depth to bedrock in catchment	RckDepCat	cm
Mean depth to bedrock in watershed	RckDepWs	cm
Density of road-stream intersections multiplied by the NHD+ slope in catchment	RdCrsSlpWtdCat	(Crossings*slope/km2)
Density of road-stream intersections multiplied by NHD+ slope in watershed	RdCrsSlpWtdWs	(Crossings*slope/km2)
Density of road-stream intersections in watershed	RdCrsWs	crossings/km2
Density of roads in catchment	RdDensCat	km/km2
Density of roads in watershed	RdDensWs	km/km2
Sand content in catchment	SandCat	%
Sand content in watershed	SandWs	%
Sulfur content in catchment	SCat	%
Silicon dioxide content in catchment	SiO2Cat	%
Silicon dioxide content in watershed	SiO2Ws	%
Sulfur content in watershed	SWs	%
Evapotranspiration from PRISM	1 and 2 month values, and 3-,6-, and 12- month averages	˚C
Preciptiation from PRISM	1 and 2 month values, and 3-,6-, and 12- month averages	˚C
Mean Temperature from PRISM	1 and 2 month values, and 3-,6-, and 12- month averages	˚C
Maximum Temperature from PRISM	1 and 2 month values, and 3-,6-, and 12- month averages	˚C
Mean composite topographic index in catchment	WetIndexCat	Unitless
Mean composite topographic index in watershed	WetIndexWs	Unitless
Mean seasonal water table depth in catchment	WtDepCat	cm
Mean seasonal water table depth in watershed	WtDepWs	cm

All variable descriptions and units were obtained from EPA StreamCat website (https://www.epa.gov/national-aquatic-resource-surveys/streamcat-metrics-and-definitions) and Oregon State University's PRISM database (https://www.prism.oregonstate.edu/)