

California State University, Monterey Bay  
**Digital Commons @ CSUMB**

---

School of Natural Sciences Faculty Publications  
and Presentations

School of Natural Sciences

---

2-4-2012

## Predicting Natural Base-Flow Stream Water Chemistry in the Western United States

John R. Olson

*California State University, Monterey Bay*, [joolson@csumb.edu](mailto:joolson@csumb.edu)

Charles P. Hawkins

Follow this and additional works at: [https://digitalcommons.csumb.edu/sns\\_fac](https://digitalcommons.csumb.edu/sns_fac)

---

### Recommended Citation

Olson, J. R., and C. P. Hawkins (2012), Predicting natural base-flow stream water chemistry in the western United States, *Water Resour. Res.*, 48, W02504, doi:10.1029/2011WR011088.

This Article is brought to you for free and open access by the School of Natural Sciences at Digital Commons @ CSUMB. It has been accepted for inclusion in School of Natural Sciences Faculty Publications and Presentations by an authorized administrator of Digital Commons @ CSUMB. For more information, please contact [digitalcommons@csumb.edu](mailto:digitalcommons@csumb.edu).

# Predicting natural base-flow stream water chemistry in the western United States

John R. Olson<sup>1</sup> and Charles P. Hawkins<sup>1</sup>

Received 27 June 2011; revised 5 December 2011; accepted 22 December 2011; published 4 February 2012.

[1] Robust predictions of stream solute concentrations expected under natural (reference) conditions would help establish more realistic water quality standards and improve stream ecological assessments. Models predicting solute concentrations from environmental factors would also help identify the relative importance of different factors that influence water chemistry. Although data are available describing the major factors controlling water chemistry (i.e., geology, climate, atmospheric deposition, soils, vegetation, topography), geologic maps do not adequately convey how rocks vary in their chemical and physical properties. We addressed this issue by associating rock chemical and physical properties with geological map units to produce continuous maps of percentages of CaO, MgO, S, uniaxial compressive strength, and hydraulic conductivity for western United States lithologies. We used catchment summaries of these geologic properties and other environmental factors to develop multiple linear regression (LR) and random forest (RF) models to predict base flow electrical conductivity (EC), acid neutralization capacity (ANC), Ca, Mg, and SO<sub>4</sub>. Models were derived from observations at 1414 reference-quality streams. RF models were superior to LR models, explaining 71% of the variance in EC, 61% in ANC, 92% in Ca, 58% in Mg, and 74% in SO<sub>4</sub> when assessed with independent observations. The root-mean-square error for predictions on validation sites were all <11% of the range of observed values. The relative importance of different environmental factors in predicting stream chemistry varied among models, but on average rock chemistry > temperature > precipitation > soil = atmospheric deposition > vegetation > amount of rock/water contact > topography.

**Citation:** Olson, J. R., and C. P. Hawkins (2012), Predicting natural base-flow stream water chemistry in the western United States, *Water Resour. Res.*, 48, W02504, doi:10.1029/2011WR011088.

## 1. Introduction

### 1.1. Statement of Problem

[2] Predictive models are needed that account for the natural spatial variation in ecologically important water chemistry constituents [Billett and Cresser, 1992]. Such models could greatly enhance the accuracy and precision of both chemical and biological water quality assessments [Hawkins *et al.*, 2010]. To assess if stream water quality or aquatic biota are supporting designated uses, regulators must be able to compare existing chemical and biological conditions with an appropriate reference condition, i.e., a benchmark representing either a desired or near-natural state. Existing stream conditions can be determined by sampling a stream, but determining the chemical or biological reference condition is a challenge even in catchments with minor human modifications. Because the chemical reference condition is generally unknown, current biological assessments ignore naturally occurring variations in water

chemistry [Hawkins *et al.*, 2010], even though it is known to influence the abundances and distributions of stream biota [Minshall and Minshall, 1978; Townsend *et al.*, 1983]. Predictive water chemistry models are therefore needed to help establish appropriate reference conditions among thousands of individual sites that water quality managers are required to assess. However, most existing water chemistry models require extensive, site-specific parameterization that greatly constrains their use at multiple streams. Furthermore, few models exist for the biologically important water chemistry constituents such as total dissolved solids (TDS) and electrical conductivity (EC). Empirical models based on known drivers of water chemistry could provide predictions of water chemistry constituents needed for chemical and biological assessments across regions. Quantifying relationships between natural base flow water chemistry and potential environmental drivers could also help resolve questions regarding the relative importance of these drivers in controlling natural spatial variations in stream water chemistry [Drever, 1997 p. 283].

### 1.2. Background

[3] Many mass balance and process-based models that predict water chemistry were developed in the 1980s to assess the effects of acid rain on freshwater systems

<sup>1</sup>Western Center for Monitoring and Assessment of Freshwater Ecosystems, Department of Watershed Sciences, and the Ecology Center, Utah State University, Logan, Utah, USA.

(e.g., MAGIC [Cosby *et al.*, 1985] and ILWAS [Goldstein *et al.*, 1984; Gherini *et al.*, 1985]). These models primarily predict temporal dynamics in water chemistry in individual streams, including responses to changes in chemical fluxes associated with some forms of human activity (e.g., atmospheric deposition in MAGIC). Although some process-based models can predict naturally occurring concentrations and fluxes of different chemical constituents, these predictions rely on measured water chemistry for calibration and accurate estimates of human-caused inputs to streams. When water quality assessments are required for thousands of streams, the costs of obtaining calibration data greatly limits the routine use of process-based models. Also, although the fluxes of some types of chemical constituents affected by human activity can be estimated with reasonable accuracy (e.g., atmospheric deposition or water treatment outflows), the fluxes associated with many types of watershed alteration are more difficult to estimate (e.g., nonpoint sources associated with dispersed land use such as livestock grazing or novel sources such as mountain top removal mining). Moreover, few process-based models incorporate the effects of lithology on water chemistry, an important driver of natural spatial variation in water chemistry. To overcome the inherent limitations of process-based approaches in predicting spatial variation in water chemistry, Cresser *et al.* [2000] and Smart *et al.* [2001] developed the empirical G-BASH model to predict water chemistry attributes for the River Dee in Scotland from rock geochemistry. They subsequently underscored the need to also account for variation in climate and atmospheric deposition when applying their model to other catchments [Cresser *et al.*, 2006]. Other empirical models have been developed to predict spatial variation in water chemistry across regions from land use data, but these models primarily predict water chemistry variation associated with differences in land use, not variation in natural background conditions.

[4] The development of models capable of predicting variation in natural water chemistry has been restricted because environmental attributes such as climate and geology that likely influence water chemistry have not been quantified at regional scales. Climate, topography, and vegetation data are now readily available for the entire United States; however, obtaining useful data on geology, perhaps the principal driver of natural variation in water chemistry, presents special challenges. Geologic maps primarily depict geologic spatial variation by classifying the landscape into map units based on similarities in rock age, structure, and formative processes [U.S. Geological Survey (USGS), 2006]. This categorization hinders the use of geologic maps in predicting stream chemistry in three ways. First, map units defined by their similarity in age or formative process may have very different chemical and physical properties (e.g., co-occurring limestone and sandstone). Second, and in contrast, map units differing in their formative process may have similar geochemical effects on streams (e.g., small dissolved loads in streams originating in gneiss or granite). Finally, classifying map units by age or formative process does not inherently provide information on general chemical and physical differences among classes.

[5] Many approaches have been developed to predict stream ecosystem properties from geologic information despite the limitations of current geologic classifications.

Geology is most often associated with either chemical or biological attributes of streams by classifying geology into coarse rock types and then determining which classes are dominant [e.g., Bricker and Rice, 1989; Davy-Bowker *et al.*, 2006]. However, such classification obscures continuous variability among rocks, and applying these geologic groupings to catchments that span multiple rock types can be problematic. Increasing the number of categories and mapping geologic classes at higher-spatial and taxonomic resolutions can improve associations; but the use of many categories of data in predictive models would result in more complicated models with reduced degrees of freedom. To overcome the limitations associated with using geologic classes in predicting stream properties, two approaches have been proposed that extract more useful information from geologic maps. McCartan *et al.* [1998] reclassified geologic map units into lithochemical classes based on the presence of water-reactive rocks. Streams that differed in their solute concentrations were then associated with these new classes. The G-BASH model [Smart *et al.*, 1998; Cresser *et al.*, 2000] relies on maps of rock chemical content (CaO, MgO, K<sub>2</sub>O, and Na<sub>2</sub>O) to predict water chemistry. The maps were created by applying the average whole rock chemistry based on rock samples collected from individual geologic formations to an entire map unit, effectively converting discrete classes of rock types into a series of maps depicting geochemistry as continuous variables. Although these approaches can potentially be used to incorporate geologic information more directly into water chemistry models, they have only seen limited application. Because lithochemical maps still rely on a classification scheme, they may not adequately describe the chemical variation among classes that results from variable amounts of different rock types within a class. Characterizations of geologic formations used by the G-BASH model [i.e., Smart *et al.*, 2001] are data-intensive and may therefore be labor- and cost-prohibitive for regional applications. Also, neither of these approaches addresses other rock characteristics that can affect water chemistry such as physical weathering rate (i.e., rock strength) and the amount of rock/water contact (i.e., rock hydraulic conductivity).

[6] Early water chemistry models predominantly focused on predicting concentrations of major cations and acid neutralization capacity (ANC) because the original impetus for these models was to understand and predict the effects of acid deposition. Although certain taxa are sensitive to some specific ions (e.g., the association of mollusks with Ca), stream biota can also be sensitive to changes in TDS because the amount of TDS determines the osmotic regulatory challenge biota face. Differences in TDS, as measured by EC, have been shown to affect both periphyton [Leland and Porter, 2000] and macroinvertebrates [Minshall and Minshall, 1978]. Because of these effects on biota, TDS/EC is becoming an increasingly important water quality parameter in many areas faced with salinization threats associated with agriculture [Williams, 1987], mountain top mining [Pond *et al.*, 2008], oil and gas extraction processes including hydraulic fracturing [Renner, 2009], and coal bed methane production [U.S. Environmental Protection Agency (USEPA), 2004]. In spite of its importance, few models have been developed to predict either natural

background TDS/EC or changes in TDS/EC associated with land use changes (although see *Hendershot et al.* [1992] and *Ballester et al.* [2003]). An accurate estimate of a stream's naturally occurring water chemistry, including TDS/EC, is a prerequisite for effectively assessing water quality and establishing attainable goals for restoration.

### 1.3. Objectives

[7] Our general objective was to model natural base flow water chemistry in western U.S. streams from catchment geology and other environmental factors. We focused on developing models for Ca, Mg, SO<sub>4</sub>, ANC, and EC because they are known to be associated with the distribution of stream macroinvertebrates [*Leland and Fend*, 1998; *Minshall and Minshall*, 1978], the taxonomic group most often used in biological assessments. We also limited this study to base flow conditions because data on stormflow events and our understanding of the effects of stormflow chemistry on biota are both very limited. Pursuing this objective required that we complete three tasks. We first needed to create maps based on the chemical and physical properties of rocks that can influence stream water chemistry. We then needed to create empirical models to predict natural base flow stream chemistry from these chemical and physical rock properties along with other factors known to influence water chemistry, such as climate and soils. To be useful for water quality and ecological assessments, water chemistry predictions should be at least accurate enough to distinguish sites with high concentrations from low, which we assessed as having a normalized root-mean-square error

(nRMSE) <25%. We defined nRMSE as RMSE expressed as a percentage of the range of observed values [*Wu et al.*, 2011]. Finally, we needed to evaluate the relative strength and direction of effects associated with each predictor variable to both assess the conceptual validity of our models [sensu *Rykiel*, 1996] and determine which factors most strongly influence water chemistry at this scale. There is generally broad agreement about what factors control water chemistry, but little understanding about the relative importance of these factors across regions [*Drever*, 1997]. Our work should therefore add to our understanding of the relative importance of different environmental factors on water chemistry.

## 2. Methods

### 2.1. Geology Characterization

[8] We adapted the approach of *Smart et al.* [2001] to translate standard geologic maps into maps depicting chemical and physical rock properties relevant to water chemistry. To do so we assigned an estimate of each map unit's chemical or physical properties to every occurrence of that map unit in the original geologic map. This estimate was calculated as the average of literature values of the respective property for each lithology contained within the map unit, weighted by the prevalence of each lithology within the map unit (step 1 of Figure 1). The source geologic maps we used were the Preliminary Integrated Geologic Map Databases for the United States (*S. Ludington et al.* (2007), Preliminary integrated geologic map databases for the United States

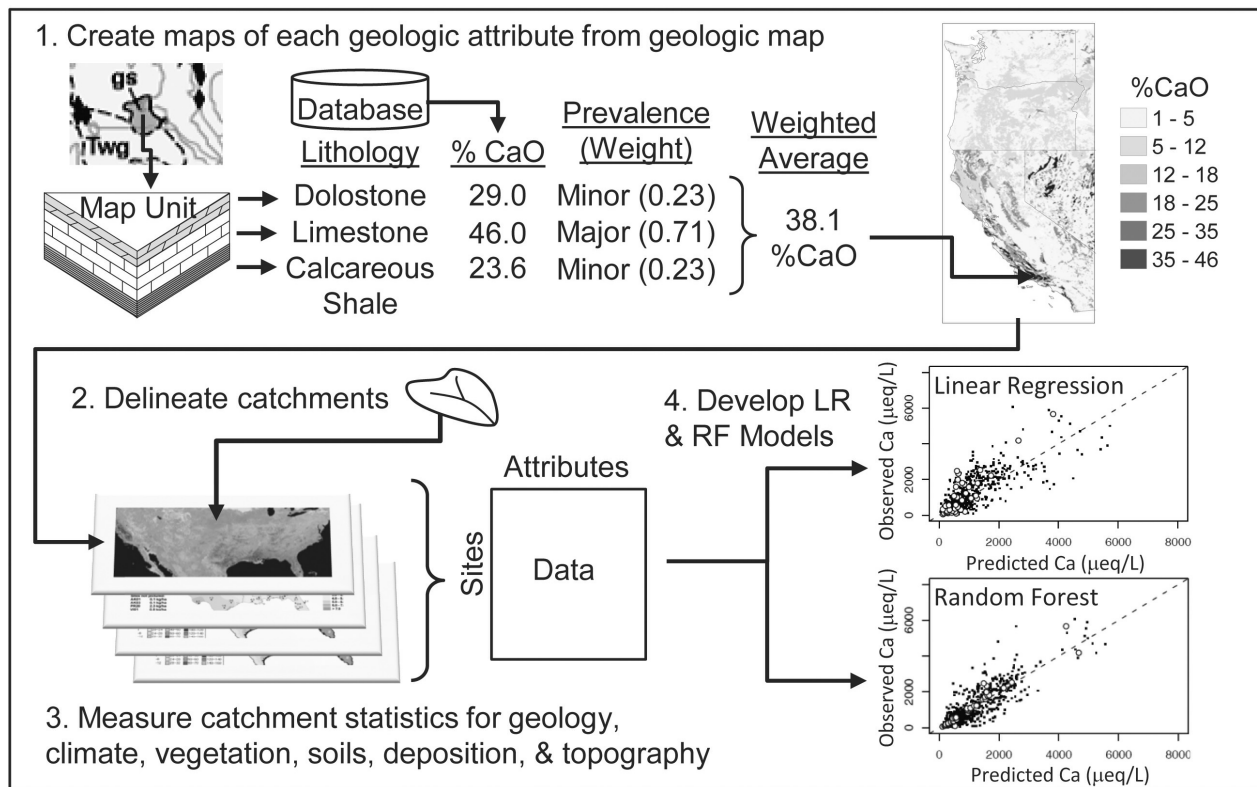


Figure 1. Diagram of work flow.

Western States: California, Nevada, Arizona, Washington, Oregon, Idaho, and Utah, *Open-File Rep. 2005-1305*, and D. B. Stoeser, G. N. Green, L. C. Morath, W. D. Heran, A. B. Wilson, D. W. Moore, and B. S. Van Gosen (2007), Preliminary integrated geologic map databases for the United States: Central States: Montana, Wyoming, Colorado, New Mexico, North Dakota, South Dakota, Nebraska, Kansas, Oklahoma, Texas, Iowa, Missouri, Arkansas, and Louisiana, *Open-File Rep. 2005-1351*, both published by U.S. Geological Survey, Reston, Va., and available only online at <http://pubs.usgs.gov/of/2005/1305/> and <http://pubs.usgs.gov/of/2005/1351/>, respectively), a database of standardized and updated state geologic maps produced by the U.S. Geological Survey (USGS). This database includes information on each geologic map unit's component lithologies, the lithologies' relative volumetric importance within the map unit, and a description of the map unit's associated geologic formations. Although state geologic maps are of relatively coarse resolution (1:500,000 to 1:750,000), preliminary analysis showed that models were not improved when based on data from 1:100,000 scale maps.

[9] We characterized five attributes of each lithology based on the amount of influence we expected these attributes to have on water chemistry and how readily available data were for these attributes across a wide variety of rock types. We characterized chemical attributes in terms of whole rock percentages of CaO, MgO, and S, because these constituents form the principal solutes derived from rock in most stream systems. We also characterized two physical attributes: rock strength, measured as uniaxial compressive strength (UCS), and rock hydraulic conductivity. We used UCS as a measure of rock strength and susceptibility to physical weathering instead of a more direct measure such as tensile strength because of the greater availability of UCS data and its generally high correlation with tensile strength [Hobbs, 1964]. We included rock hydraulic conductivity because of its influence on the amount of rock/water interaction occurring within a catchment, with more permeable rocks having more contact over shorter time frames [Drever, 1997].

[10] We characterized geology based on the 158 different lithologies that the Geologic Map Database lists as occurring in the western U.S. Because some of these lithologies are known to vary widely in their chemical or physical attributes, we created an additional 56 lithologic classes based on common modifiers used in geologic unit descrip-

tions to better parse physical or chemical variability within lithologies (see Table 1). For example, calcareous and non-calcareous sandstones greatly differ in their effect on water chemistry [Hem, 1985; McCartan et al., 1998]. In these situations, we searched the descriptions of both geologic map units and named formations within map units for modifiers listed in Table 1 to assess if the lithology within a particular geologic map unit should be assigned to a separate lithologic class. Descriptions of geologic formations were obtained through either the Lexicon of Geologic Names of the United States (available at <http://ngmdb.usgs.gov/>) or literature searches.

[11] We derived values for each of the five rock attributes for each of the 214 lithologic classes and subclasses from data obtained from the OZCHEM National Whole Rock Geochemistry Database (available at <http://www.ga.gov.au/meta/ANZCW0703011055.html>), Earthchem Geochemical Database (available at <http://www.earthchem.org/>), National Geochemical Database (available at <http://tin.er.usgs.gov/ngdb/rock/>), and literature searches. The information in these data sources ranged from a single sample for rare lithologies to over 20,000 samples for more common rock types. Because only a small proportion of the chemical data described sedimentary rock samples as calcareous or noncalcareous, we used the rocks percentage of CaO to partition samples into three groups representing noncalcareous, partially calcareous, and calcareous sedimentary rocks. The three subsets of calcareous rock content were created by applying a *k*-means clustering algorithm (Euclidian distance and 20 iterations) to the Ca content of each lithology. The group of samples with the lowest Ca content was considered to contain noncalcareous rocks. Our preliminary analysis showed that the partially calcareous and calcareous groups had similar effects on water chemistry, so these two groups were then lumped into a single category describing calcareous rocks. A two-cluster algorithm was also tried, but failed to partition calcareous and noncalcareous rocks as effectively as the three-cluster analysis. We then calculated a measure of central tendency for each attribute for each lithologic class. Mean values were used unless the data were highly skewed, in which case we used the median value. We assessed data as highly skewed if the skew was greater than  $\pm 2$  times the standard error of skew [Cramer and Howitt, 2004]. For generalized rock classes, such as "metamorphic" or "granitic," we used the hierarchical nature of the Geologic Map Databases to identify all subordinate lithologies (e.g., gneiss, schist, slate, etc., for metamorphic rocks) and then calculated their mean. For chemical attributes we weighted the means for each lithology by the number of samples of each subordinate lithology that occurred within the combined database, and used the number of samples as an estimate of the prevalence of any given subordinate rock type within the general rock class. Because the physical characterizations generally had a much lower sample size (often just means reported in the literature), simple averages were used to characterize general rock categories. We could not characterize some lithologic classes because either they were extremely rare and literature values of their properties were unavailable ( $n = 6$ ), or the lithologic class was not actually a specific rock type (e.g., mélange, water, landslides) and could not be characterized ( $n = 62$ ). These

**Table 1.** Modifiers Assigned to Lithology by Chemical or Physical Type and Effect<sup>a</sup>

Chemical	Physical
Alluvial (any coarse or fine detrital)	Alluvial (any coarse or fine detrital)
Lacustrine (sand, silt, or clay)	Lacustrine (sand, silt, or clay)
Landslide (any coarse or fine detrital)	Landslide (any coarse or fine detrital)
Eolian (sand or silt)	Eolian (sand or silt)
Noncalcareous (any clastic sedimentary)	Till (any unsorted glacial deposit)
Calcareous (any clastic sedimentary)	Tuff (any volcanic)
Carbonaceous (any coarse or fine detrital)	

<sup>a</sup>Only applicable lithologies are listed.

**Table 2.** Weights Used to Quantify the Prevalence of Rock Types Within Geologic Map Units

Prevalence	Description	Weight
Major	30%–100% of unit	0.7119
Minor	10%–30% of unit	0.2311
Incidental	<10% of unit	0.0570
Indeterminate	0%–100% of unit	0.5000

classes were coded as “no data” so they would have no influence on the characterization of geologic map units.

[12] Because geologic map units were often mixtures of lithologies, the attribute values we derived for each lithology had to be combined to describe the combined effects of the different lithologies within each geologic map unit. We therefore calculated the rock attribute weighted averages from each component lithology within a map unit. We chose the weights based on the prevalence of each lithology within a map unit. Weights (see Table 2) were derived by rescaling the midpoint of each prevalence category so that all of the weights (except indeterminate) summed to 1. This weighted average characterization was then assigned to every occurrence of the geologic map unit in question in a GIS, producing a continuous raster for that geologic property. We then repeated this process for the other geologic attributes, producing separate rasters of rock percents of CaO, MgO, S, UCS, and hydraulic conductivity.

## 2.2. Other Environmental Predictors of Water Chemistry

[13] *Drever* [1997] outlined five major environmental drivers of natural water chemistry: rock type, climate, relief, vegetation, and amount of rock/water contact. We therefore added characterizations of climate, relief, vegetation, and amount of rock/water contact to our characterization of rock type for all locations within our study area (Table 3). We characterized climate in terms of the long-term temperature and precipitation averages produced by the parameter-elevation regression on independent slopes model (PRISM [*Daly et al.*, 1994]). PRISM data are produced by combining interpolations of point-measured meteorological values from multiple agencies with a digital elevation model (DEM) and other spatial data sets to account for coastal and topographic effects on climate. Although contemporaneous climate and water chemistry measurements are available, our models based on time-specific climate measurements did not perform better than models based on long-term averages. Because we were mainly interested in understanding spatial differences in base flow water chemistry and the importance of environmental factors relative to one another at regional scales, for simplicity we used long-term climate averages as predictors in our models. We also characterized possible spatial interactions between geology and climate by dividing the derived grids of rock chemical properties (section 2.1) by the amount of precipitation within each grid cell. Atmospheric deposition can also be an important driver of stream chemistry, especially near coasts [*Cresser et al.*, 2006] and urban areas [*Chae et al.*, 2004]. We therefore calculated long-term average atmospheric wet deposition from data obtained from the National Atmospheric Deposition

Program National Trends Network. Although the use of soils data has been problematic in predicting water chemistry [*Billett and Cresser*, 1996; *Stutter et al.*, 2004], we wanted to independently assess the effectiveness of soils data in predicting regional variation in water chemistry. We used the State Soil Geographic Database (STATSGO) to characterize soil attributes (other than chemical characteristics, which are incomplete for our study area). We characterized vegetation cover by calculating long-term average MODIS satellite enhanced vegetation index (EVI) values [*Huete et al.*, 2002] from 2000–2009. Although EVI does not capture differences in vegetation composition or structure, it is a good proxy of biomass and so might therefore be associated with differences in water chemistry related to varying amounts of vegetation. To characterize relief and the amount of rock/water contact, we calculated each catchment’s elevation, relief, area, and shape from a DEM. To assess the amount of rock/water contact, we also estimated groundwater velocities with the MRI-Darcy model [*Baker et al.*, 2003], which applies Darcy’s equation within a geographic information system (GIS) environment. The Darcy equation calculates potential groundwater movement from hydraulic conductivity and water table elevation head. The MRI-Darcy model applies the Darcy equation to each grid cell to estimate potential groundwater flux from hydraulic conductivity (derived from our geologic maps as described in section 2.1) and surface slope (derived from DEMs). Potential groundwater flux was estimated at 100 m intervals over 6 km (based on observed groundwater flows in the western U.S.) in 12 directions to determine both discharge and recharge velocities.

## 2.3. Water Chemistry Data and Catchment Assessments

[14] We used base flow water chemistry data collected at 1487 locations across the western U.S. (Figure 2) by multiple agencies (Table 4) to build empirical predictive models. The 13 western states ( $\sim 3.45 \times 10^6$  km<sup>2</sup>) from which we compiled data represent a wide diversity of climatic and geologic environments, ranging from boreal to subtropic biomes and wet to arid climates. These states also represent much (94%) of the lithologic diversity of the continental U.S. Because we wanted to model natural background chemical conditions, we used data only from sites judged by the source agency to have minimal human impacts within their catchments. All data were converted to consistent units (Table 5) and sample concentrations reported as below detection limits were set to half of the reported detection limit. Some agencies measured ANC in the field, whereas others measured it in the laboratory. *Bales et al.* [2002] compared the results obtained from 3–5 water chemistry test kits of the same three varieties used in the field by these agencies against known standards and found that these fixed end-point field titrations were positively biased by 200–500  $\mu\text{eq L}^{-1}$  due to size of the titrant drop and inaccurate titrant concentrations. To assess whether the field and laboratory methods might show bias relative to each other, we compared laboratory and field ANC estimates by regressing each against laboratory-measured Ca concentrations. The intercept for field-measured ANCs was 230  $\mu\text{eq L}^{-1}$  greater than laboratory-measured ANCs ( $p < 0.00001$ , on 342 field and 454 laboratory measurements of

**Table 3.** Predictor Variables Used

Type	Variable	Units	Short Name
Geology <sup>a</sup>	Catchment mean whole rock CaO	(%)	Percentage CaO
	Catchment mean whole rock MgO	(%)	Percentage MgO
	Catchment mean whole rock S	(%)	Percentage S
	Catchment mean unconfined compressive strength	(MPa)	Compressive strength
Climate <sup>b</sup>	Catchment mean log geometric mean hydraulic conductivity	$\times 10^{-6} \text{ m s}^{-1}$	Log hydraulic cond
	Catchment mean of mean 1971–2000 annual precipitation	(mm yr <sup>-1</sup> )	Mean precipitation
	Catchment mean of mean 1971–2000 annual min monthly precipitation	(mm m <sup>-1</sup> )	Minimum precipitation
	Catchment mean of mean 1971–2000 annual max monthly precipitation	(mm m <sup>-1</sup> )	Maximum precipitation
	Catchment mean of mean June–September 1971–2000 monthly precipitation	(mm m <sup>-1</sup> )	Mean summer precipitation
	Catchment mean of mean 1971–2000 annual temperature	(°C)	Mean temperature
	Catchment mean of mean 1971–2000 annual minimum monthly temperature	(°C)	Minimum temperature
	Catchment mean of mean 1971–2000 annual maximum monthly temperature	(°C)	Maximum temperature
	Catchment mean of mean 1961–1990 first and last day of freeze	day of yr	Day last freeze
	Catchment mean of mean 1961–1990 annual number of wet days	(d yr <sup>-1</sup> )	Mean wet days
Atmospheric deposition <sup>c</sup>	Catchment mean of mean 1961–1990 annual relative humidity	(%)	Relative humidity
	Catchment mean of mean 1994–2006 annual precipitation-weighted mean Ca concentration	(mg L <sup>-1</sup> )	Atmospheric Ca
	Catchment mean of mean 1994–2006 annual precipitation-weighted mean Mg concentration	(mg L <sup>-1</sup> )	Atmospheric Mg
	Catchment mean of mean 1994–2006 annual precipitation-weighted mean Na concentration	(mg L <sup>-1</sup> )	Atmospheric Na
	Catchment mean of mean 1994–2006 annual precipitation-weighted mean Cl concentration	(mg L <sup>-1</sup> )	Atmospheric Cl
	Catchment mean of mean 1994–2006 annual precipitation-weighted mean SO <sub>4</sub> concentration	(mg L <sup>-1</sup> )	Atmospheric SO <sub>4</sub>
	Catchment mean of mean 1994–2006 annual precipitation-weighted mean NO <sub>3</sub> concentration	(mg L <sup>-1</sup> )	Atmospheric NO <sub>3</sub>
	Catchment mean of mean 1994–2006 annual total inorganic nitrogen (TN) wet deposition	(kg ha <sup>-1</sup> )	Atmospheric TN
Soil <sup>d</sup>	Catchment mean available water capacity	Fraction	Soil water capacity
	Catchment mean bulk density	(g cm <sup>-3</sup> )	Soil bulk density
	Catchment mean soil erodibility (K factor)	Dimensionless	Soil erodibility
	Catchment mean organic matter content	(% weight)	Soil organic content
	Catchment mean soil permeability	(inches h <sup>-1</sup> )	Soil permeability
	Catchment mean soil depth	(m)	Soil depth
Topography <sup>e</sup>	Catchment mean water table depth	(m)	Water table depth
	Catchment elevation mean, minimum, maximum, and standard deviation	(m)	MCE, MinCE, MaxCE, SDCE
	Catchment elevation relief ratio	Dimensionless	Elevation relief ratio
	Catchment shape ratio (catchment area: length)	Dimensionless	Catchment shape
Vegetation <sup>f</sup>	Catchment area	(km <sup>2</sup> )	Catchment area
	Catchment mean of mean 2000–2009 annual enhanced vegetation index	Dimensionless	Mean EVI
	Catchment maximum of mean 2000–2009 annual enhanced vegetation index	Dimensionless	Max mean EVI
Groundwater <sup>g</sup>	Catchment mean of mean 2000–2009 annual max enhanced vegetation index	Dimensionless	Mean max EVI
	Catchment mean delivery velocity	(m d <sup>-1</sup> )	Mean delivery
	Catchment mean recharge velocity	(m d <sup>-1</sup> )	Mean recharge
	Catchment mean total flux	(m d <sup>-1</sup> )	Mean total flux
Rock/water Interactions <sup>h</sup>	Catchment mean base-flow index	dimensionless	Base-flow index
	Catchment mean percent CaO/mean precipitation	Dimensionless	Percent CaO/precipitation
	Catchment mean percent MgO/mean precipitation	Dimensionless	Percent MgO/precipitation
	Catchment mean percent S/mean precipitation	Dimensionless	Percent S/precipitation

<sup>a</sup>Derived using method described in section 2.1 at a grid resolution of 90 × 90 m.

<sup>b</sup>PRISM climate data [Daly *et al.*, 1994], 2 × 2 km resolution grids were used for the 1961–1990 data, and 800 × 800 m resolution grids were used for the 1971–2000 data.

<sup>c</sup>National Atmospheric Deposition Program National Trends Network (NADP/NTN) 2.5 × 2.5 km resolution grids (obtained from the NADP website available at <http://nadp.sws.uiuc.edu/ntn/>).

<sup>d</sup>Natural Resource Conservation Service State Soil Geographic Database (NRCS STATSGO) 500 × 500 m resolution grids (obtained from the NRCS website available at <http://soils.usda.gov/survey/geography/statsgo/>).

<sup>e</sup>Calculated from National Elevation Database DEMs at 30 × 30 m resolution (obtained from the USGS website available at <http://ned.usgs.gov/>).

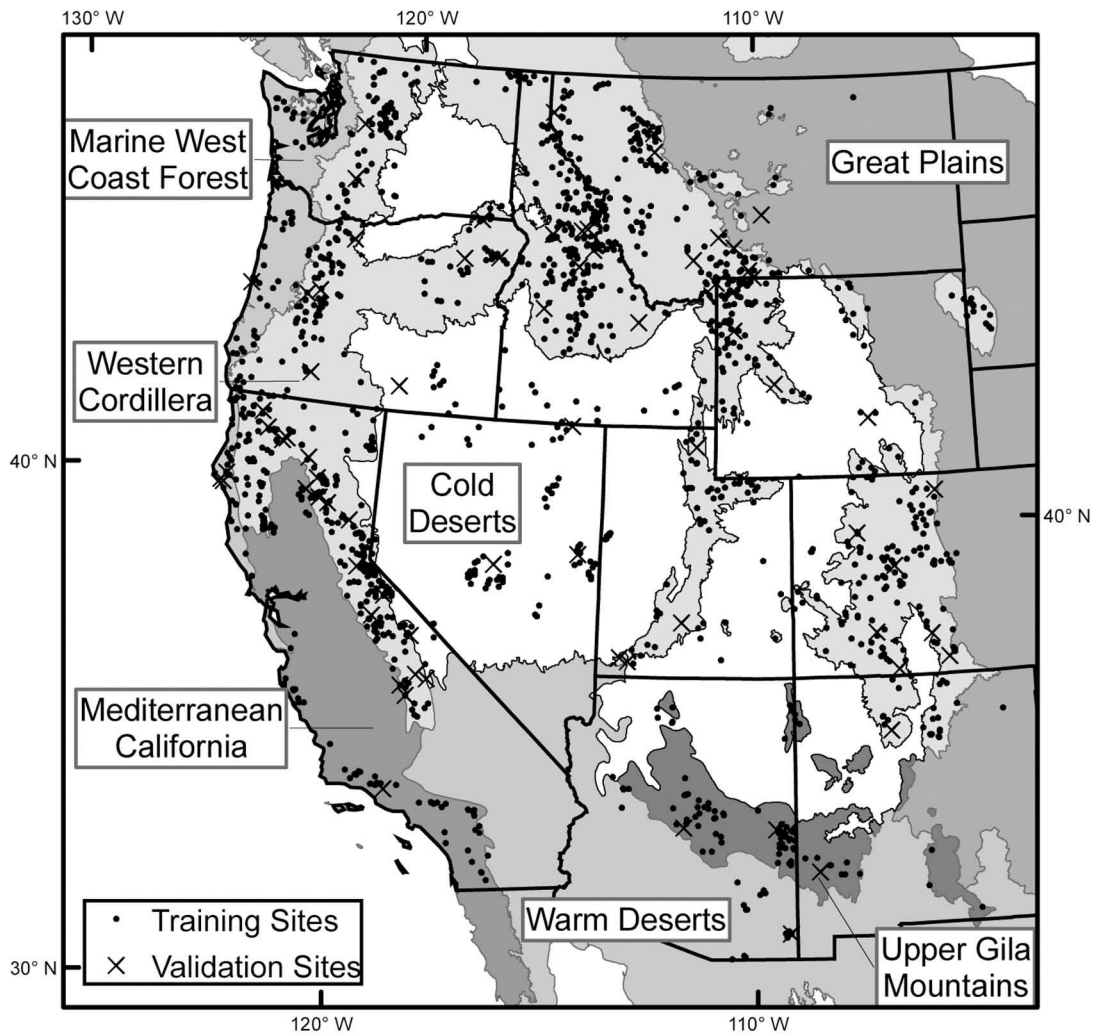
<sup>f</sup>MODIS satellite MOD13A1.V4 data collected every 16 d at 500 × 500 m resolution from 2000–2009 [Huete *et al.*, 2002]. These data are distributed by the Land Processes Distributed Active Archive Center (LP DAAC), located at USGS Earth Resources Observation and Science Center (available at <http://lpdaac.usgs.gov>).

<sup>g</sup>Velocity derived from MRI-Darcy model [Baker *et al.*, 2003], at a 90 × 90 m resolution. Base-flow index values derived from interpolation of the ratio of annual maximum flow to minimum flow for all USGS gage data in the region.

<sup>h</sup>Derived by dividing each rock chemistry grid by the mean precipitation grid to account for spatial interactions.

ANC). Slopes of the two regressions were similar (1.48 for field data and 1.41 for laboratory) but statistically different ( $p < 0.00001$ ). Because the slopes were so similar (<5% different), we corrected field measured ANC values based only on the difference in the intercept.

[15] We used the multi-watershed delineation tool [Chinnayakanahalli, 2006] to delineate catchment boundaries for each water chemistry site from the DEMs (step 2, Figure 1). Catchment averages for all predictive variables were then calculated (step 3, Figure 1). We also calculated



**Figure 2.** Map of 1414 training and 73 validation sites by ecoregion and state.

the coefficient of variation (CV) of each geologic variable as a measure of geologic heterogeneity within catchments.

[16] After delineating and calculating summary statistics for each watershed, we screened out sites with human impacts or replicate samples. To ensure that sites selected by different agencies were all relatively free of human impacts, we inspected any site that had either high values for

conductivity ( $>1000 \mu\text{S cm}^{-1}$ ),  $\text{Cl}^-$  ( $>250 \mu\text{eq L}^{-1}$ ),  $\text{SO}_4^{2-}$  ( $>250 \mu\text{eq L}^{-1}$ ), total phosphate ( $>90 \mu\text{g L}^{-1}$ ), total inorganic nitrogen (TN) ( $>300 \mu\text{g L}^{-1}$ ), or whose catchments contained  $>5\%$  agricultural or urban land use (assessed with the 2001 National Land Cover Data set). These inspection criteria were based on both earlier reference site selection criteria used in the western U.S. [Herlihy *et al.*, 2008;

**Table 4.** Sources of Water Chemistry Data

Data Source	Sites	Years Collected	Location/Contact <sup>a</sup>
Arizona Department of Environmental Quality	46	1992–2008	Patrice Spindler
California Department of Fish and Game	50	2003–2008	Andrew Rehn
Colorado Dept of Public Health and Environment	76	1992–2007	Chris Theel
Sierra Nevada Aquatic Research Laboratory	30	1999–2002	Dave Herbst
USEPA Environmental Monitoring and Assessment Program	339	2000–2004	Available at <a href="http://www.epa.gov/emap2/">http://www.epa.gov/emap2/</a>
USGS National Water-Quality Assessment Program	60	1965–2008	Available at <a href="http://water.usgs.gov/nawqa/">http://water.usgs.gov/nawqa/</a>
New Mexico Environment Department	26	1999–2007	Shann Stringer
Oregon Department of Environmental Quality	71	1992–2002	Shannon Hubler
US Forest Service PACFISH/INFISH Biological Opinion	224	2001–2009	Forestry Sciences Laboratory, Logan UT
Utah State University	401	1998–2003	John Olson
US Forest Service Region 5	148	2000–2001	Joseph Furnish
USGS National Water Information System	16	1973–1995	Available at <a href="http://waterdata.usgs.gov/nwis">http://waterdata.usgs.gov/nwis</a>

<sup>a</sup>People listed are affiliated with organizations listed under Data Source.



**Table 5.** Summary of Water Chemistry Training Data

Constituent <sup>a</sup>	Units	Minimum	Mean	Maximum	<i>n</i> <sup>b</sup>	Transform <sup>c</sup>
EC	( $\mu\text{S cm}^{-1}$ )	7	133	1171	1391	0.20
ANC	( $\mu\text{eq L}^{-1}$ )	-110	1271	7280	1324	0.14
Ca	( $\mu\text{eq L}^{-1}$ )	27	998	7194	796	0.25
Mg	( $\mu\text{eq L}^{-1}$ )	9	509	7108	755	0.16
SO <sub>4</sub>	( $\mu\text{eq L}^{-1}$ )	2	302	9279	450	0.51

<sup>a</sup>EC, electrical conductivity; ANC, acid neutralization capacity.

<sup>b</sup>Number of sites used for model development after removal of outliers and sites with high influence.

<sup>c</sup>Exponent used for power transformations applied to data prior to linear regression (LR) modeling only.

*Herlihy and Sifneos*, 2008] and personal experience. This inspection included examining both aerial photographs (using Google Earth) and maps (USGS 1:24,000 topographic maps) for any evidence of human impacts beyond atmospheric deposition (ranches, mines, agriculture, clearcuts, etc.). We removed sites from the data set that showed probable anthropogenic influence on water chemistry. For those sites that were sampled on multiple dates, we selected a single sampling date at random from those dates with the most complete data (i.e., contained estimates for the most constituents). To minimize spatial replication and autocorrelation within our data set, we considered samples to be from a single site if their catchments overlapped by >90% and were within 1 km of one another.

#### 2.4. Modeling

[17] We split the data into training and validation data sets prior to modeling. Validation sites were chosen by first stratifying all data by level II ecoregion [*Commission for Environmental Cooperation (CEC)*, 2006] and then randomly selecting 5% of the sites within each ecoregion that had observations for each constituent.

[18] Prior to modeling, we inspected Cleveland plots of EC and ANC for extreme values [*Zuur et al.*, 2009] and examined sites with these values for potential human influences as described above. If the extreme values could not be attributed to human influences and there were no indications that the value was due to human error (i.e., the measurement was consistent with other water chemistry values or other measurements from similar sites), then the value was retained.

[19] We used both multiple linear regressions (LR) and random forest (RF) regression [*Breiman*, 2001] to develop predictive models (step 4, Figure 1). We used both methods because we wanted to compare the performance of these two modeling approaches. RF is a nonparametric modeling approach and has been widely applied to a variety of classification and regression problems in genetics, biomedical applications, ecology, and financial forecasting, and often provides better predictions than other methods [*Cutler et al.*, 2007; *Siroky*, 2009]. RF is based on the concept of classification and regression trees (CART [*Breiman et al.*, 1984]) where data are recursively partitioned on one of the predictor variables, such that each partition results in greater homogeneity of the response variable values in the resulting subgroups relative to the unpartitioned data. RF extends CART by creating an ensemble of trees from bootstrapped samples of the data and randomly selected sets of

predictor variables. Predictions are then made by averaging results across the entire ensemble. Model fit is assessed by measuring prediction error of samples not included during the tree creation, i.e., “out of bag” samples [for more details, see *Cutler et al.*, 2007; *Siroky*, 2009]. We developed RF models to take advantage of their abilities to automatically account for nonlinear relationships and interactions among predictors. We also developed LR models because, although often not as robust as nonparametric methods such as RF, they can be easily used to make continuous spatial predictions. All analyses were done in the statistical computing environment, R.

[20] To develop the LR models, we used an iterative procedure of building initial models, transforming data as needed, controlling collinearity, and then removing sites that were statistical outliers or had high influence. We used the R function *stepAIC* to select final LR models. *StepAIC* is an algorithm that combines both forward and backward stepwise selection to choose the model that minimizes the Akaike information criterion. This method produces models with predictive ability equal to that of models based on exhaustive variable selection [*Murtaugh*, 2009]. After developing an initial model, we used spread-level plots [*Fox*, 1997] to assess the residuals for heteroscedasticity and then applied the suggested power transformation to the response variable. This procedure both reduced the heteroscedasticity of residuals and increased the linearity of responses. An inspection of bivariate plots showed that only groundwater predictive variables needed to be transformed (log) to produce linear relationships. Collinearity was controlled by calculating the variance inflation factor (VIF) and iteratively removing predictors until all VIFs were less than 3 [*Zuur et al.*, 2009]. Sites that were statistical outliers in the initial models (tested using the Bonferroni outlier test) or influenced coefficient estimates by more than 20% were removed from the data set prior to developing the final model. Only variables that were significant at the  $p < 0.05$  level were retained in the final models.

[21] We used the same data sets used to create the final LR models (with outliers removed) to create random forest models based on 1500 trees (as implemented by the R function *randomForest*). The use of LR to identify outliers probably improved RF performance because RF does not have its own diagnostic tools to assess data quality. We optimized the number of predictors tried at each node using the *tuneRF* function. Although RF does provide estimates of each predictor’s importance, it uses all predictors without any selection as in LR. Modeling with multiple correlated predictors can bias importance estimates of predictors in RF models [*Strobl et al.*, 2008]. To create the most parsimonious models and reduce the number of correlated predictors, we modeled iteratively, removing correlated, or low importance predictors until a model’s out-of-bag-mean-square error began to increase. Prior to choosing the final RF model, we examined bivariate, partial-dependence plots for evidence of inconsistent relationships between response and predictors (i.e., three or more changes in direction of effect). Predictors with inconsistent relationships to the response indicate an indirect or spurious correlation, and these predictors were removed from the final model.

## 2.5. Model Evaluation, Validation, and Comparison

[22] We evaluated model fit with the coefficient of determination ( $R^2$ , also referred to as Nash-Sutcliffe model efficiency when applied to validation data), the absolute RMSE, and the nRMSE as a measure of relative accuracy. Fit was assessed for both training and validation data, although we used out-of-bag predictions (i.e., predictions from those trees not used in model training) to calculate pseudo  $R^2$  and RMSE for RF training data.

[23] We also used the equivalence testing strategy outlined by *Robinson et al.* [2005] to assess predictive accuracy, i.e., if the regression of observed-on-predicted values had an intercept equal to 0 and slope equal to 1. A more nuanced view of model performance is provided by separately assessing prediction bias (i.e., prediction mean is equivalent to observation mean, so regression intercept = 0) and similarity of individual predictions to their associated observations (i.e., regression slope = 1). Traditionally, tests of intercept and slope were made based on the null hypothesis of no difference between observed and modeled data (e.g.,  $\mu_{\text{obs}} = \mu_{\text{pred}}$ ). However, failure to reject this null hypothesis can be due to the test having insufficient power. Conversely, testing with large data sets might reject the null hypothesis even when the differences are not meaningful in an ecological or environmental management context. Equivalence testing avoids these problems by reversing the null hypothesis of agreement between predictions and observations to a null hypothesis of difference between the two (e.g.,  $\mu_{\text{obs}} \neq \mu_{\text{pred}}$ ). This switches the burden of proof on to the model [*Robinson et al.*, 2005] and results in concluding either that predictions are sufficiently similar to the observations (i.e., null hypothesis is rejected) or there is either insufficient evidence or a true difference between predictions and observations (i.e., null hypothesis is not rejected). A region of similarity is defined by the investigator to define what constitutes “sufficiently similar.” Our region of similarity was 25% of the estimate for both slope and intercept, and the probability level we used was  $\alpha = 0.05$ . We then performed a nonparametric bootstrap with the R function `equiv.boot` to produce 10,000 estimates of the intercept and slope, and reported the proportions that would fall in the region of equivalence. The null hypothesis of nonequivalence between observed and predicted would be rejected if <5% of the bootstrap estimates fell outside of the region of equivalence.

## 3. Results and Interpretation

### 3.1. Selected Models and Variable Importance

[24] The numbers of predictors retained in the LR models varied from 11 for the  $\text{SO}_4$  model to 16 for the ANC model (Table 6). The numbers of predictors retained in the RF models varied from seven for the  $\text{SO}_4$  model to 21 for the ANC model. All of the retained predictors had a consistent direction of effect for all models, except for atmospheric Cl and TN deposition, both of which had negative effects in the RF models and positive effects in the LR models.

[25] Most of the predictors included in the models had relative importance and directions of correlation consistent with expectations based on our understanding of the proc-

esses determining water chemistry. Among these was the dominant role of rock chemistry as a source for all constituents, the secondary effects of temperature on either or both evaporative concentration and weathering rates, and dilution effects of increasing precipitation. A few models (RF Ca, RF Mg, and RF  $\text{SO}_4$ ) were improved by using the rock chemistry grids weighted by precipitation, which accounted for the spatial interactions between rock composition and precipitation. Soil predictors were also included in most models, with soil bulk density being the most important soil predictor in seven of 10 models. Higher-density soils were associated with higher-constituent concentrations, likely due to their lower gas exchange rates and increased  $\text{pCO}_2$ , which increases carbonic acid concentrations and hence chemical weathering [*Ballard*, 2000]. Soil organic content was negatively correlated with ANC, probably a result of the additional organic acids or inhibition of calcite dissolution by organic compounds [*Morse and Arvidson*, 2002] associated with high soil organic content. Ca and Mg deposition was positively correlated with stream EC, ANC, Ca, and Mg, consistent with expectations associated with marine [*Evans et al.*, 2001] and dust inputs [*Likens et al.*, 1996]. Positive correlations between vegetation (EVI) and stream concentrations were expected because of the increase in physical weathering through root action and in chemical weathering via increased exposure to  $\text{CO}_2$ . Factors affecting rock/water contact had a complex relationship with constituent concentrations. Soil permeability was negatively correlated with concentrations, whereas concentrations were positively correlated with rock hydraulic conductivity and the base flow index. These relationships are in general agreement with the expectations of *Drever* [1997]. He noted that while high permeability in the vadose zone may reduce contact time resulting in reduced concentrations, low-permeability bedrock may reduce the amount of water in contact with rock also reducing concentrations. Topography and rock strength exhibited expected relationships, but were weak predictors that were selected in less than half of the models.

[26] Not all predictors performed as expected, or were clearly associated with a putative mechanism. The weak predictive ability of the percent of MgO relative to the percent of CaO in the Mg models was probably an artifact of our treating both dolomitic and calcareous clastic rock types the same and only characterizing the differences in CaO content within these rock types. Day of last freeze (DLF) was the strongest climatic predictor for LR  $\text{SO}_4$ , and was also included in the RF EC model, but was negatively correlated with both constituents. Because DLF was negatively correlated with mean temperature ( $r = -0.89$ ), we interpret DLF as a surrogate measure of both temperature and dilution due to snow melt. Greater DLFs were associated with lower constituent concentrations possibly resulting from cooler temperatures and greater dilution during summer months due to later snow melt. The importance of  $\text{SO}_4$  deposition relative to other atmospheric deposition was also unexpected.  $\text{SO}_4$  deposition occurred in seven models and was the most important atmospheric predictor in the Ca,  $\text{SO}_4$ , and LR ANC models. The positive correlation between ANC and atmospheric  $\text{SO}_4$  in the LR ANC model runs opposite to the expectation that increased acid deposition leads to decreased ANC. Other models of ANC

**Table 6.** Model Predictors in Rank Order of Importance and Direction of Association

Random Forest Model			Linear Regression Model			
Predictor	Direction	Importance <sup>a</sup>	Predictor	Direction	Importance <sup>b</sup>	Coefficient
<i>Electrical Conductivity</i>						
Percent CaO	+	63	Percent CaO	+	0.31	2.68E - 02
Percent S	+	42	Maximum temperature	+	0.28	3.90E - 03
Maximum temperature	+	41	Percent S	+	0.20	5.49E - 01
Mean wet days	-	37	Mean wet days	-	0.18	-2.30E - 03
Mean precipitation	-	35	Percent CaO CV	+	0.15	1.82E - 01
Soil bulk density	+	33	Soil bulk density	+	0.15	4.81E - 01
Soil permeability	-	33	Atmospheric Cl	+	0.12	3.72E - 01
Atmospheric Mg	+	32	Atmospheric SO <sub>4</sub>	+	0.12	3.05E - 01
Atmospheric Ca	+	32	Soil permeability	-	0.09	-1.17E - 02
Percent MgO	+	32	Log hydraulic cond	+	0.09	5.53E - 02
Atmospheric SO <sub>4</sub>	+	31	Base-flow index	+	0.05	6.29E - 01
Mean maximum EVI	+	30	Percent MgO CV	+	0.04	6.76E - 02
Compressive strength	-	30	Soil erodibility	+	0.04	3.86E - 01
Minimum precipitation	-	29	Percent MgO	+	0.04	7.09E - 03
Max wet days	-	28	Soil depth	-	0.04	-1.86E - 03
Soil erodibility	+	28	(Intercept)	+	0.00	7.33E - 01
Day last freeze	-	28				
Log hydraulic cond	+	27				
Mean summer precipitation	-	24				
<i>Acid Neutralization Capacity</i>						
Percent CaO	+	90	Percent CaO	+	0.38	1.96E - 02
Percent S	+	51	Maximum temperature	+	0.27	2.29E - 03
Maximum temperature	+	48	Soil organic content	-	0.16	-4.14E - 02
Mean precipitation	-	39	Soil bulk density	+	0.13	2.50E - 01
Atmospheric Cl	-	35	Percent S	+	0.12	2.09E - 01
Log hydraulic cond	+	35	Percent CaO CV	+	0.12	8.41E - 02
Mean wet days	-	34	Soil depth	-	0.11	-3.58E - 03
Soil bulk density	+	33	Maximum summer precipitation	-	0.11	-3.14E - 04
Atmospheric Ca	+	33	Soil permeability	-	0.11	-8.66E - 03
Percent MgO	+	32	Log hydraulic cond	+	0.10	3.91E - 02
Soil organic content	-	31	Mean summer precipitation	-	0.10	-4.39E - 06
Atm TN deposition	-	31	Mean aximum EVI	+	0.07	2.46E - 05
Atmospheric Mg	+	31	Percent MgO CV	+	0.06	5.28E - 02
Minimum precipitation	-	31	Atmospheric SO <sub>4</sub>	+	0.05	7.87E - 02
Mean summer precipitation	-	31	Water table depth	+	0.04	5.71E - 02
Soil permeability	-	30	Base-flow index	+	0.04	2.69E - 01
Mean temperature	+	30	(Intercept)	+	0.00	1.51E + 00
Soil erodibility	+	29				
Soil depth	-	26				
Compressive strength	-	25				
Mean maximum EVI	+	24				
<i>Calcium</i>						
Percent CaO/precipitation	+	85	Percent CaO	+	0.44	8.79E - 02
Maximum temperature	+	41	Maximum temperature	+	0.23	8.09E - 03
Mean maximum EVI	+	40	Percent S	+	0.21	1.27E + 00
Percent S/precipitation	+	40	Percent CaO CV	+	0.20	5.93E - 01
Mean wet days	-	38	Soil bulk density	+	0.19	1.84E + 00
Mean summer precipitation	-	37	Minimum precipitation	-	0.15	-1.18E - 02
Compressive strength	-	30	Atmospheric SO <sub>4</sub>	+	0.15	8.76E - 01
Soil bulk density	+	29	Soil permeability	-	0.11	-4.03E - 02
Atmospheric SO <sub>4</sub>	+	27	Mean maximum EVI	+	0.07	1.09E - 04
Atmospheric Ca	+	25	Soil depth	-	0.07	-9.43E - 03
			Atmospheric Cl	+	0.06	5.29E - 01
			(Intercept)	+	0.00	-5.68E - 01
<i>Magnesium</i>						
Percent CaO/precipitation	+	59	Percent CaO	+	0.30	1.09E - 02
Percent MgO/precipitation	+	39	Maximum temperature	+	0.26	1.71E - 03
Maximum temperature	+	36	Percent S/precipitation	+	0.20	1.53E + 02
Percent S	+	35	Percent MgO	+	0.18	1.70E - 02
Mean wet days	-	30	Mean EVI	+	0.15	4.87E - 05
Atmospheric Mg	+	28	Mean precipitation	-	0.14	-5.78E - 05
Mean summer precipitation	-	27	Percent CaO CV	+	0.13	7.24E - 02
Mean temperature	+	26	Soil permeability	-	0.12	-8.21E + 03
Mean maximum EVI	+	24	Soil bulk density	+	0.11	1.98E - 01
Percent MgO CV	+	19	Percent MgO CV	+	0.11	8.42E - 02

Table 6. (continued)

Random Forest Model			Linear Regression Model			
Predictor	Direction	Importance <sup>a</sup>	Predictor	Direction	Importance <sup>b</sup>	Coefficient
			Atmospheric Mg	+	0.10	2.23E + 00
			Log hydraulic cond	+	0.10	2.91E - 02
			Soil organic content	-	0.07	-1.69E - 02
			Mean summer precipitation	-	0.06	-2.05E - 06
			(Intercept)	+	0.00	9.06E - 01
			<i>Sulfate</i>			
Mean summer precipitation	-	28	Percent S	+	0.34	6.13E - 02
Mean wet days	-	23	Day last freeze	-	0.29	-3.66E - 04
Percent S/precipitation	+	22	Percent CaO/precipitation	+	0.21	9.73E - 01
Compressive strength	-	17	Atmospheric SO <sub>4</sub>	+	0.19	3.27E - 02
Soil bulk density	+	15	Soil bulk density	+	0.18	5.20E - 02
Atmospheric SO <sub>4</sub>	+	12	Percent CaO CV	+	0.13	1.16E - 02
Percent CaO	+	8	Soil permeability	-	0.12	-1.33E - 03
			Maximum mean EVI	+	0.11	5.29E - 06
			Atm TN deposition	+	0.10	1.01E - 02
			Soil depth	-	0.10	-4.01E - 04
			Catchment shape	+	0.06	2.56E - 02
			(Intercept)	+	0.00	1.05E + 00

<sup>a</sup>Random forest (RF) model importance is calculated as percent increase in mean squared error when predictor is removed.

<sup>b</sup>Linear regression (LR) model importance is calculated as the absolute value of the standardized coefficients.

in the western U.S. have not shown SO<sub>4</sub> deposition to be a significant predictor [Clow *et al.*, 2010; Nanus, 2008]. Although this relationship is possibly caused by an anion exchange of SO<sub>4</sub><sup>2-</sup> for OH<sup>-</sup> [Evans *et al.*, 2001], it is also possible that the relationship is not directly causal at all. Instead, the relationship might be produced by correlations of SO<sub>4</sub> deposition with other confounding environmental factors. Marine deposition is one possible confounding factor, a possibility supported by the correlation of SO<sub>4</sub> deposition with Cl deposition ( $r = 0.45$ ) in marine influenced areas west of the Sierra/Cascade Range. Other confounding factors are also possible (i.e., dust deposition), but we lack data to assess these relationships.

[27] We controlled for the alteration of stream chemistry by land use by selecting minimally altered sites, but we could not control for atmospheric inputs of anthropogenic sources of SO<sub>4</sub> or TN. Because our measured response for ANC and SO<sub>4</sub> includes some amount of anthropogenic inputs, our empirical models of these constituents is of a natural background plus anthropogenic inputs and include SO<sub>4</sub> and TN deposition as predictors. Although anthropogenic deposition is widespread, its effects on stream chemistry compared with that associated with land use are small.

### 3.2. Model Fit and Validation

[28] The models explained 60%–78% of the variation in the training data (Table 7 and Figure 3), with nRMSEs that were all <10%. The RF models had slightly better fits to the training data than the LR models, both in terms of  $R^2$  and RMSE. Direct comparison of RF and LR performance based on training data penalizes RF because RF  $R^2$  and RMSE values were calculated from out-of-bag predictions. A fairer comparison of the relative performance of the two model techniques is given by the independent validation data. In these comparisons, RF models had notably better model efficiencies and RMSEs than LR models for all constituents except SO<sub>4</sub>. The nRMSEs for RF models ranged from 3% to 11%. Model efficiencies calculated from the

independent validation data set showed that all models had good predictive ability when applied to other sites in the western U.S., except for the LR models for ANC and Mg. RMSEs were higher for the validation than the training data in all cases except the RF Ca and SO<sub>4</sub> models, but all validation nRMSEs were <15%.

[29] Model assessments based on equivalence tests showed even more striking differences between the RF and LR models. Three of the RF models showed no evidence of bias, i.e., the null hypothesis that the mean of predicted and observed values were not equivalent was rejected. For these models, >97.5% of the bootstrap sample estimates fell within the region of equivalence for the intercept. For the RF Mg model, the null hypothesis of  $\mu_{\text{obs}} \neq \mu_{\text{pred}}$  was not rejected, but there was little sign of consistent bias, with 87% of the bootstrapped sample estimates falling within the region of equivalence. The RF SO<sub>4</sub> model showed an underprediction bias, with 38% of the bootstrap sample estimates being above the region of equivalence. All of the LR models exhibited minor to severe underprediction bias, with 15%–99% of bootstrap sample estimates falling above the region of equivalence. The SO<sub>4</sub> models were the most biased of any of the LR or RF models.

[30] Although the plots of observed versus predicted concentrations do not show a clear tendency to underpredict, the null hypothesis of the slopes being not equivalent to 1 was not rejected for any model based on validation data. RF models for all constituents except SO<sub>4</sub> had 48%–71% of the bootstrap estimates of slope fall within the region of equivalence, indicating that these models failed to meet the specification of having a slope within 25% of 1. In all models except LR ANC, LR Mg, and RF SO<sub>4</sub>, the estimates of slope fell above the region of equivalence, indicating they tended to underpredict concentrations at higher levels. This test may be somewhat misleading because at least a portion of the decrease in slope from the 1:1 line is probably caused by the effect of regression toward the mean. Regression toward the mean always occurs

**Table 7.** Assessment of Model Performance

Model	Data <sup>a</sup>	<i>n</i>	<i>R</i> <sup>2b</sup>	RMSE	nRMSE	<i>r</i> <sup>2c</sup>	Equivalent Intercept <sup>d</sup>	Equivalent Slope <sup>e</sup>
<i>Electrical Conductivity</i>								
RF	Tng	1390	0.78	67.3	5.8	0.79	100	100
	Val	73	0.71	84.2	7.2	0.73	99.0	52.3
LR	Tng	1390	0.67	80.1	6.9	0.70	100	100
	Val	73	0.65	91.0	7.8	0.70	81.5	37.3
<i>Acid Neutralization Capacity</i>								
RF	Tng	1323	0.73	643.2	8.7	0.74	100	100
	Val	71	0.61	797.6	10.8	0.63	99.8	49.8
LR	Tng	1323	0.62	764.2	10.3	0.64	100	100
	Val	71	0.32	1046.3	14.2	0.33	85.0	41.2
<i>Calcium</i>								
RF	Tng	795	0.77	501.3	7.0	0.77	100	100
	Val	41	0.92	330.9	4.6	0.94	100.0	71.1
LR	Tng	795	0.67	629.1	8.8	0.65	100	99.6
	Val	41	0.61	720.7	10.1	0.76	12.4	4.4
<i>Magnesium</i>								
RF	Tng	754	0.73	368.0	5.2	0.73	100	99.3
	Val	41	0.58	437.6	6.2	0.58	86.5	48.9
LR	Tng	754	0.70	434.2	6.1	0.63	98.8	99.9
	Val	41	0.38	532.2	7.5	0.49	68.3	23.9
<i>Sulfate</i>								
RF	Tng	449	0.77	476.4	5.1	0.77	99.8	95.8
	Val	29	0.74	334.1	3.6	0.88	61.9	0.9
LR	Tng	449	0.60	883.2	9.5	0.38	36.5	22.3
	Val	29	0.79	303.0	5.8	0.79	0.4	0.3

<sup>a</sup>Tng, training data; Val, independent validation data.

<sup>b</sup>For training data,  $R^2$  was calculated as the coefficient of determination using transformed training data for LR and untransformed training data for RF. For validation data,  $R^2$  was calculated as Nash-Sutcliffe model efficiency using back transformed (LR) or untransformed (RF) validation data.

<sup>c</sup>Squared Pearson correlation between observations and associated model predictions.

<sup>d</sup>Percentage of 10,000 bootstrap simulations falling within the region of equivalence ( $Eq_0 = \hat{Y} \pm 25\%$ ) for the intercept = 0.

<sup>e</sup>Percentage of 10,000 bootstrap simulations falling within the region of equivalence ( $Eq_1 = m \pm 25\%$ ) for the slope = 1.

whenever two variables are less than perfectly correlated. When this happens, individual cases that are large for the observed value will be relatively less large for the predicted value, resulting in systematic disagreement between the two. *Copas* [1997] demonstrated how regression toward the mean causes validation data not to plot near their predicted values, but to regress toward the mean of the training data set. Although equivalence tests provide an objective basis for understanding a model's potential weaknesses, they must be interpreted with caution, given that a portion of the deviance of slope is due to regression toward the mean. An estimate of what proportion of the slope's deviance is due to regression toward the mean and what portion is due to model inadequacies would allow more informed decisions on the validity of a model.

## 4. Discussion

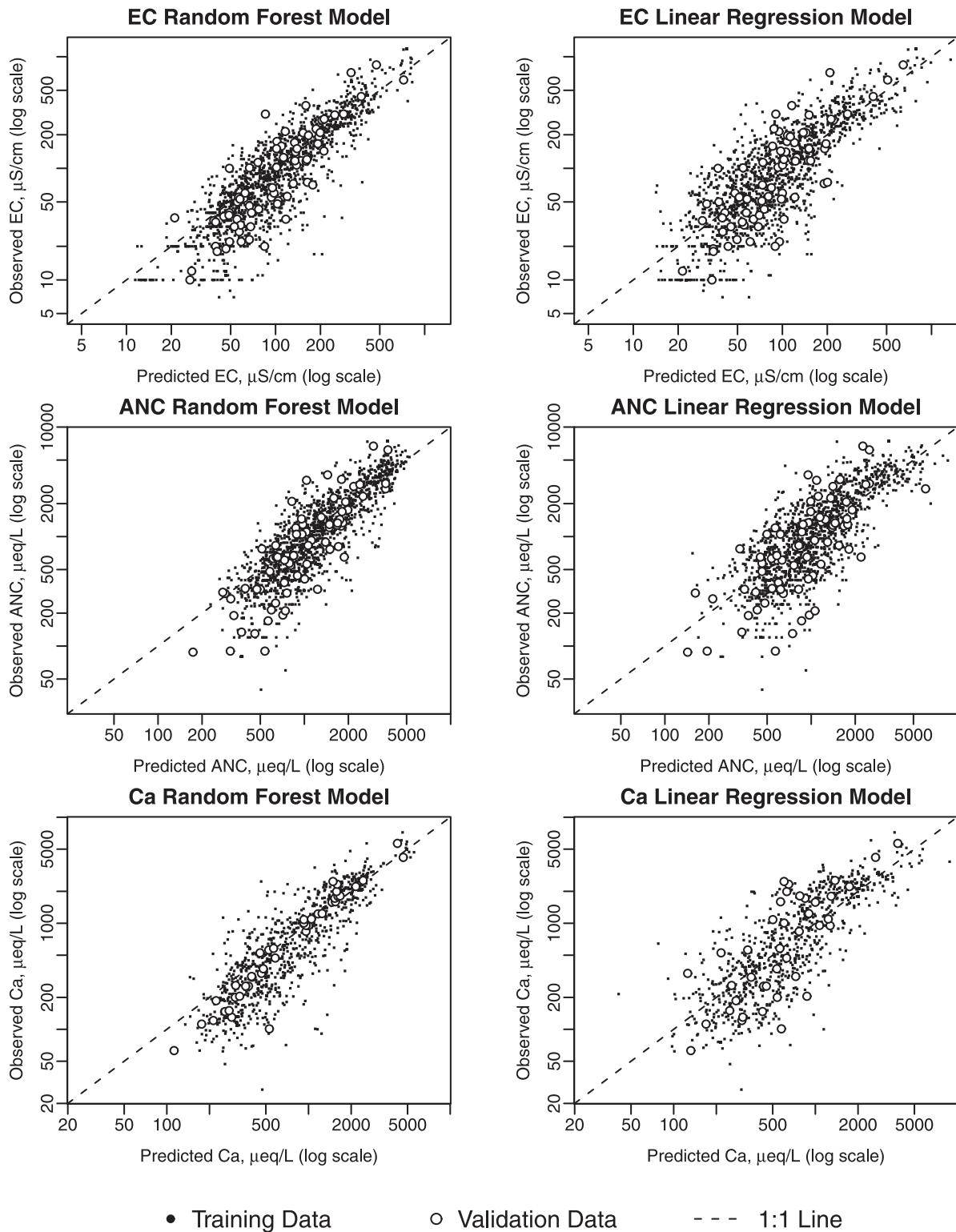
### 4.1. Comparison of Models Based on Continuous Geology With Previous Work

[31] The best assessment of the utility of our continuous characterization of geology is to compare the performance of our models with earlier empirical models (Table 8). Comparisons of this nature have received limited discussion in previous studies [although, see *Peterson et al.*, 2006], but are necessary to understand which modeling techniques and data provide the best predictions. We do not compare our results with those from process-based models

because they focus on temporal dynamics instead of spatial variation.

[32] Previously developed empirical models based on land use generally have weak predictive power. Our models based on landscape attributes accounted for substantially more variation in EC than models developed by *Baker et al.* [2005] and *Zheng et al.* [2008], and in ANC and  $SO_4$  than the model developed by *Peterson et al.* [2006]. Only the *Peterson et al.* [2006] EC model performed similarly to ours. We expect that models that parse spatial variation based solely on land use would tend to make weak predictions of natural background water chemistry because of the generally weak correlation between land use and underlying natural variation. The strong influence of anthropogenic land uses on water chemistry relative to natural variation might also obscure catchment response to natural variation in models based on data from both altered and unaltered sites. *Peterson et al.* [2006] also developed geostatistical models that included information from the spatial correlation patterns of neighboring sites, resulting in considerable improvement in model fit compared to their linear models (EC  $r^2 = 0.96$ , ANC  $r^2 = 0.90$ , and  $SO_4$   $r^2 = 0.40$ ). However, *Peterson et al.* noted that this approach is only practical when sites are located closer than their autocorrelation distances, providing limited ability to predict natural conditions across landscapes.

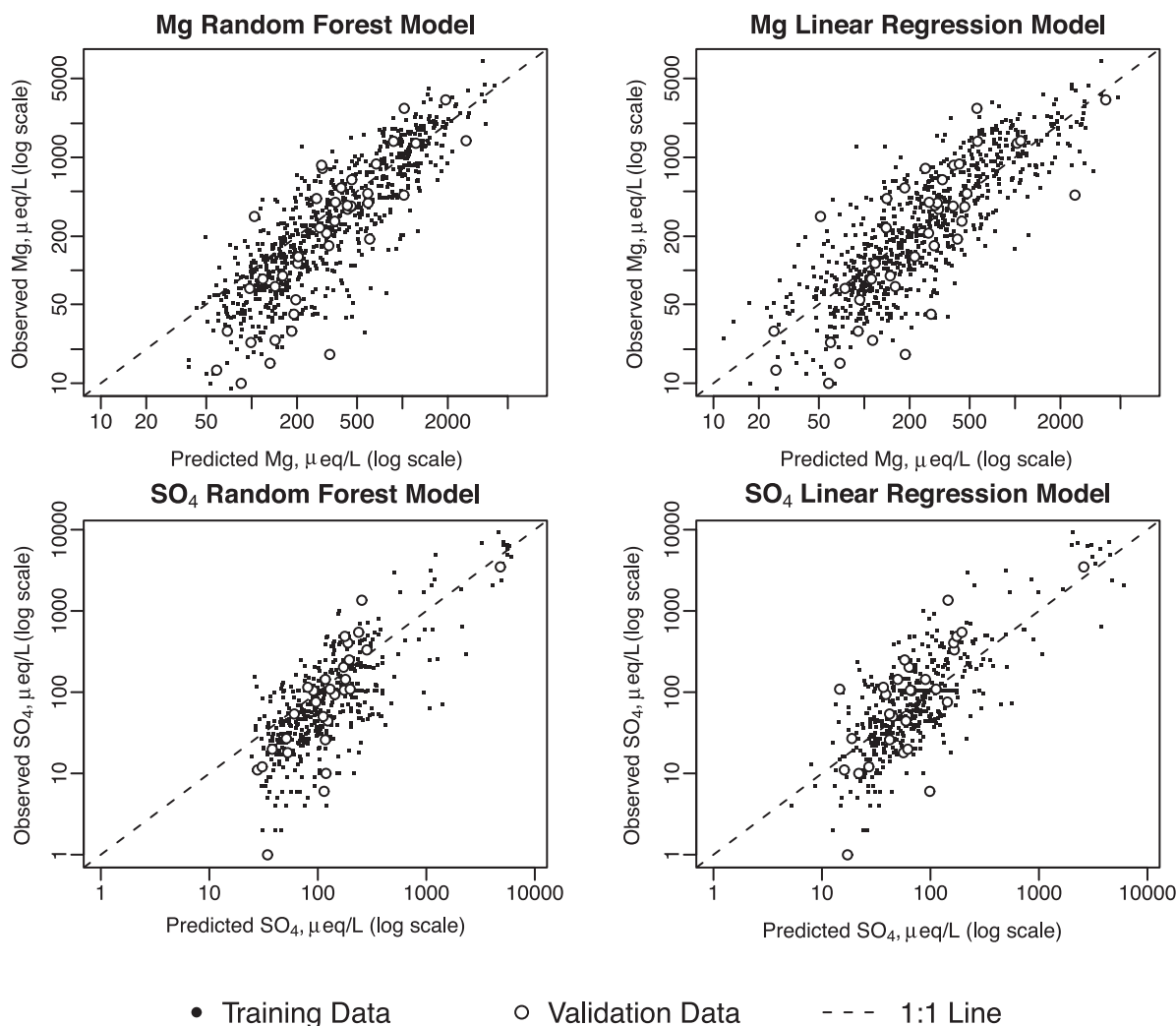
[33] Geologic classifications better characterize natural environmental variation than land use and often result in



**Figure 3a.** Plots of predicted versus observed values for both training and validation data by constituent and modeling technique. Linear regression (LR) predictions are back transformed. Plots are presented in log-log form to improve readability with the acid neutralization capacity (ANC) plots adjusted to make all values positive.

empirical models with better predictive ability. However, predictive ability of these models can vary widely when applied to different portions of the landscape. Models predicting ANC by *Berg et al.* [2005] and models predicting

ANC, Ca, and Mg by *Nedeltcheva et al.* [2006a, 2006b] showed wide variation in their  $R^2$  values when applied to areas differing in size or geology, respectively. In both cases, models for some portions of the landscape had



**Figure 3.** (continued)

performance similar to ours, but models of other areas were much weaker. *Clow et al.* [2010] developed a robust ANC model that is appreciably better than our ANC model. However, the ability of classified geology to successfully partition natural variation in the Clow et al. model may be partially due to their focus on an area three orders of magnitude smaller than ours containing less geologic heterogeneity. One of the few examples of geologic classifications applied at scales similar to ours are the models of annual mean dissolved  $\text{SiO}_2$  yields developed by *Jansen et al.* [2010] for 142 minimally disturbed catchments across the continental U.S. Their predictions based on nine rock classes and an estimate of runoff produced a squared Pearson correlation coefficient ( $r^2$ ) between observations and predictions of 0.89 for their training data, which is slightly higher than the precision of most of our models. Although both their empirical approach and predictors were similar to ours, it is difficult to directly compare their results with ours because of differences in the constituents examined. So, although geologic classifications can be used to make effective predictions for small areas or for  $\text{SiO}_2$  yield, using discrete geologic classes to characterize natural variation appears to lack sufficient information to make

predictions of biologically relevant constituents across large regions.

[34] All of these studies describing variations in lithology via classification are subject to the dilemma noted by *Jansen et al.* [2010], of either lumping lithologies too coarsely and oversimplifying the differences between them, or splitting lithologies too finely and creating a classification that is too complex to be practical. This dilemma becomes especially acute when trying to describe lithologies across large regions. This balance between resolution of how lithology is portrayed and the complexity of that portrayal is inherent in any classification, mandating at least some loss of information as different rock types are grouped together to make a usable classification. Because geologic map units often represent different rock types that are colocated (e.g., interbedded siliceous sandstone and limestone), any classification system will struggle with how to best represent these units [*Sullivan et al.*, 2007]. Also, any classification that optimally partitions variation in rocks by one attribute (e.g., rock chemical content) will necessarily partition other uncorrelated attributes such as those related to physical weathering (e.g., rock hardness) less well. Converting geologic units

**Table 8.** Summary of Previous Empirical Surface Water Chemistry Models

Study	Response	Predictors	Study Area (Extent $\times 10^6$ km <sup>2</sup> )	Train <i>n</i>	Model Type	Valid <i>n</i>	<i>R</i> <sup>2a,b</sup>	<i>r</i> <sup>2a,c</sup>
<i>Baker et al.</i> [2005]	EC	Land use and surficial geology	Great Lakes (0.181)	94	LR	0	0.27	–
<i>Peterson et al.</i> [2006]	EC	Land use, date, and coordinate	Maryland (0.032)	874	GLM	100	–	0.71
<i>Zheng et al.</i> [2008]	EC	Land use	W. Virginia (0.004)	56	LR	0	0.23	–
<i>Berg et al.</i> [2005]	ANC	Geology class, vegetation, and lake morphology	Sierra Nevada Mountains (0.090)	130	GLM	95	0.07–0.51	–
<i>Cresser et al.</i> [2000] <sup>d</sup>	Alkalinity	Continuous geology	R. Dee, Scotland (0.002)	18	LR		0.82	–
<i>Cresser et al.</i> [2006]	Alkalinity	Continuous geology and precipitation	N. Great Britain (0.09)	29	LR	0	0.85	–
<i>Nedeltcheva et al.</i> [2006b]	ANC	Geology class, precipitation, and catchment area	Vosges Mountains, France (0.003)	95	LR	0	0.30–0.81	–
<i>Nedeltcheva et al.</i> [2006a]	ANC	Geology class and precipitation	Vosges, France (0.003)	95	LR	0	0.65	–
<i>Peterson et al.</i> [2006]	ANC	Land use and date	Maryland (0.032)	874	GLM	100	–	0.41
<i>Clow et al.</i> [2010]	ANC	Geology class, catchment area, vegetation, and N deposition	Yosemite, California (0.003)	52	LR	0	0.87	–
<i>Cresser et al.</i> [2000] <sup>d</sup>	Ca	Continuous geology	R. Dee, Scotland (0.002)	18	LR		0.82	–
<i>Cresser et al.</i> [2006]	Ca	Continuous geology and precipitation	N. Great Britain (0.09)	29	LR	0	0.85	–
<i>Nedeltcheva et al.</i> [2006b]	Ca	Geology class, slope, catchment area, vegetation, and precipitation	Vosges Mountains, France (0.003)	95	LR	0	0.48–0.79	–
<i>Nedeltcheva et al.</i> [2006a]	Ca	Geology class, precipitation, and catchment area	Vosges Mountains, France (0.003)	95	LR	0	0.59	–
<i>Nedeltcheva et al.</i> [2006b]	Mg	Geology class, precipitation, catchment Area, and vegetation	Vosges Mountains France (0.003)	95	LR	0	0.70–0.79	–
<i>Nedeltcheva et al.</i> [2006a]	Mg	Geology class	Vosges, France (0.003)	95	LR	0	0.48	–
<i>Peterson et al.</i> [2006]	SO <sub>4</sub>	Land use, ecoregion, and coordinate	Maryland (0.032)	870	GLM	100	–	0.19

GLM, generalized linear model; other acronyms same as in Table 5.

<sup>a</sup>Assessment of fit was based on validation data, unless Valid *n* = 0, in which case fit was assessed for training data.

<sup>b</sup>*R*<sup>2</sup> is the coefficient of determination for the multiple regression models. Ranges represent *R*<sup>2</sup> for models developed for different portions of landscape.

<sup>c</sup>*r*<sup>2</sup> was reported as the squared Pearson correlation coefficient between observed and predicted data.

<sup>d</sup>Only results of upland base-flow models were reported.

into continuous measures of multiple chemical and physical characteristics of the rocks avoids unnecessarily grouping rocks together to make a useable classification, and also provides a better way to describe how different chemical and physical properties of rock interact with each other and with other factors to create different environments. Describing the environment as a continuum of various geologic properties instead of discrete classes should increase the precision of our estimates of chemical and physical attributes and thus improve our prediction of chemical weathering rates and resulting stream chemistries. This increased precision should also allow for greater understanding of how geology influences the distribution and diversity of biota at regional scales as seen by *Ander-son and Ferree* [2010].

[35] A comparison of our results with the earlier G-BASH models based on continuous characterizations of geology demonstrates both the advantages of the G-BASH approach, and its limitations. The G-BASH model performed well when applied to subcatchments within the River Dee basin [*Cresser et al.*, 2000; *Smart et al.*, 2001], but application to another basin by *Cresser et al.* [2006] produced systematic overpredictions. Once differences in dilution due to runoff were accounted for and the model reparameterized with data from both locations, the model predicted Ca and Gran alkalinity with slightly more preci-

sion than our models. Although our models and the G-BASH models both characterize geology continuously, they differ in their taxonomic and spatial resolution. G-BASH models were based on the measured CaO or MgO content of each formation mapped at 1:50,000, whereas our models used average lithology values for map units often consisting of multiple formations mapped at 1:250,000 or greater. This difference in approach occurred partly because *Cresser et al.* [2006] had access to high-resolution geologic data and partly because of the practical limitations of applying that resolution to an area 20 times larger than the one used by *Cresser et al.* The other key difference in approaches is our explicit inclusion of other geologic and environmental factors in our models as opposed to the post hoc correction for differences in precipitation applied by *Cresser et al.* [2006]. The limited amount of climatic variation within the study area of *Cresser et al.* also reduced the need to account for variations in temperature or vegetation. Although the G-BASH approach accounts for geologic variation better than geologic classification schemes, our model demonstrates the importance of incorporating other geologic and environmental influences in addition to rock CaO and MgO content. Accounting for these additional influences allowed us to predict how water chemistry varies across large landscapes, and also how it might vary with changes in temperature and precipitation expected from climate change.



#### 4.2. Model Applicability

[36] Model performance measures ( $R^2$ , RMSE, and equivalence tests) showed that our predictions of natural base flow water chemistry at independent validation sites were sufficiently precise and accurate to inform many stream bioassessments and restoration efforts. The precision of our models is probably near what is possible given the coarse spatial resolution of available data, the partially subjective nature of geologic maps, and the lack of predictors of temporal variation. The nRMSE of the best model for each constituent was <11% of the observed range of values. This level of precision met our objective and indicates these predictions should be useful in establishing reference-condition water chemistry values [sensu *Hawkins et al.*, 2010], which in turn should allow for more accurate ecological assessments. For example, we have improved predictions of the species composition expected under reference conditions across streams in Wyoming [*Hargrett et al.*, 2007], Idaho [*Cao et al.*, 2007], and Utah [J. Ostermiller, Utah DEQ, personal communication, 2008] by incorporating the predictions from our initial water chemistry models into biological niche models. Currently, most models developed for biological assessments do not include water chemistry as a predictor even though it is known to influence the abundance and distribution of stream biota [*Hawkins et al.*, 2010]. Improving biological models by incorporating water chemistry predictions will thus allow a more refined assessment of the degree to which the species composition observed at an assessed site differs from that expected under reference conditions. The models presented here should aid in improving the accuracy of biological assessments across the entire western United States. Comparing measured water quality with expected background conditions should also aid in diagnosing potential sources of biological impairment (e.g., a site with altered biology and markedly higher EC than predicted implies that the altered biology may be caused by stress associated with elevated conductivity). Understanding the expected natural background condition is also critical to establishing realistic ecosystem restoration goals [*Hobbs and Norton*, 1996]. Although these models only predict mean expected conditions, an upper prediction interval could be calculated to incorporate prediction uncertainty in these assessments. Models like these that incorporate the effects of temperature on water chemistry will be useful in predicting how water chemistry might change at site and regional scales with changing climate and how these changes in water chemistry might affect stream biota. Transformations, coefficients, and intercepts for the LR models are listed in Tables 5 and 6, and R objects for the RF models are available from the authors.

#### 4.3. Model Limitations

[37] Although the precision of our models was satisfactory for many purposes, they are not sufficient for all (e.g., acidic deposition sensitivity). Our models also tend to underpredict at high levels, with slopes of observations versus predictions greater than one. This tendency to underpredict was also seen in the model of dissolved  $\text{SiO}_2$  by *Jansen et al.* [2010]. This pattern of underprediction is also commonly seen in other applications of equivalence testing of slopes [e.g., *Pokharel and Froese*, 2008; *Eitel et al.*,

2008], and we suspect it is at least partly caused by the regression process itself. We conclude that, although we have less confidence in our predictions at high levels, the majority of our predictions provide an unbiased estimate of background base flow stream chemistry.

[38] The remaining error in our predictions results from some combination of measurement error (both predictor and response variables), unaccounted for processes, and temporal variation. Unfortunately, our current data set did not allow us to assess the magnitude of these sources of error. Although increased accuracy in measuring predictor variables should generally improve water chemistry predictions, the results of *Cresser et al.* [2000] do not suggest that increased resolution of geochemical data will necessarily yield significant improvements. In spite of rock chemistry's importance in determining stream chemistry, increasing resolution of two dimensional rock chemistry data may yield only small improvements in representing processes that occur within the three-dimensional geologic strata underlying watersheds. Because of the importance of dilution on constituent concentrations, we suspect that incorporating improved temporal and spatial estimates of stream discharge will improve model performance once those estimates become available.

[39] Although the LR and RF  $\text{SO}_4$  models were reasonably precise, they both exhibited more bias than the models of other constituents, according to the equivalence tests of the slope and the intercept of the observations versus predictions. Poor performance of  $\text{SO}_4$  models relative to other constituents was also seen in other studies [*Chen and Driscoll*, 2005; *Peterson et al.*, 2006] whose authors suggest that their models lacked important sources, such as  $\text{SO}_4$  deposition, or sinks such as retention of  $\text{SO}_4$  in wetlands. We suspect that three factors may be associated with the relatively poor performance of our  $\text{SO}_4$  models. First, the resolution of the geologic data for formations composed of discontinuous beds or lenses of easily erodible gypsum is very coarse. Although the resolution of state geologic maps is sufficient for representing spatial variation in sources of Ca and Mg, it may not be for very erodible rocks such as gypsum. Characterizing very spatially heterogeneous deposits of such a highly reactive rock as homogenous within a unit would likely lead to both over- and underpredictions. Second, our models do not account for bacterially mediated sulfate reduction that can result in losses of sulfur either by precipitation as sulfides or degassing as  $\text{H}_2\text{S}$ . This process can lower  $\text{SO}_4$  concentrations below what is delivered by deposition and has been observed in formations in our study area such as the Fort Union Formation [*Hem*, 1985], and may account for much of the unexplained variation in the portions of our study area with significant amounts of wetlands. Third, uptake of  $\text{SO}_4$  by either plants in terrestrial environments [*Likens et al.*, 2002] or phytoplankton in lakes or large pools [*Lehman and Branstrator*, 1994], or via adsorption by soils [*Sokolova and Aledseeva*, 2008] could influence stream water  $\text{SO}_4$  concentrations.

#### 4.4. Relative Importance of Environmental Factors on Stream Chemistry

[40] Across the multiple constituents that we modeled, we saw clear differences in the relative importance of different environmental factors on stream chemistry. In general, the order of importance of factors was: rock chemistry >

temperature > precipitation > soil = atmospheric deposition > vegetation > rock/water contact > topography. However, we cannot assess the relative importance of specific predictors (e.g., the importance of the percentage of CaO versus the percentage of S), because individual predictors within these categories were correlated with one another. The dominant effect of rock chemistry on stream chemistry is not surprising, especially the importance of whole rock percentages of CaO indicative of carbonate weathering. Ca in rocks is the ultimate source of Ca in streams (and makes up a large portion of both EC and ANC); and carbonate weathering is the most important contributor of solutes [Drever, 1997]. The importance of whole rock percentages of S in predicting all constituents probably reflects the contributions from high-solubility evaporites like CaSO<sub>4</sub> and MgSO<sub>4</sub> to EC, ANC, Ca, and Mg concentrations. Similar associations between SO<sub>4</sub> and both Ca and Mg were seen by Brenot *et al.* [2007].

[41] The importance of temperature relative to precipitation was unexpected, however. Although temperature is known to positively affect SiO<sub>2</sub> weathering [Gaillardet *et al.*, 1999; Kump *et al.*, 2000] and it affects mineral dissolution rates in the laboratory, previous field-based studies have not shown a clear relationship between temperature and Ca, Mg, ANC, or EC [Drever, 1997; White and Blum, 1995]. The effect of temperature is probably obscured by its covariation with other factors that affect weathering, namely precipitation, evaporation, vegetation cover, and soil development. To understand the effect of temperature one must either control for these other factors statistically, or select sites such that variation in these other factors is limited [Kump *et al.*, 2000]. Our modeling approach may have been better able to separate the effects of temperature from other factors than the work of White and Blum [1995] because of its larger sample size and inclusion of arid sites. Although part of the effect of temperature on chemical concentrations is almost certainly due to evaporative concentration [White and Blum, 1995], we conclude that evaporation explained only part of the temperature effect observed because relative humidity also directly affects evaporation and was not selected as a predictor.

[42] The relatively weak relationships between stream chemistry and soils, atmospheric deposition, and vegetation were expected. Base flow stream chemistry is closely controlled by groundwater sources [Soulsby *et al.*, 1998], so we expected that lithology data would better explain base flow chemistry than soil data. Nonetheless, we may be underestimating the role of soils on stream chemistry because we did not have spatially complete soil chemistry to include as a predictor. Atmospheric deposition can be an important source of solutes in areas with limited chemical weathering [Likens *et al.*, 1996; Driscoll *et al.*, 2001] or near sources of marine or anthropogenic deposition [Evans *et al.*, 2001; Chae *et al.*, 2004]. Ca deposition concentrations of 30  $\mu\text{eq L}^{-1}$  or greater commonly occur in the desert southwest and this concentration by itself would account for 20% of the stream Ca concentration at over 10% of our sites. However, because acid deposition in the western U.S is generally both lower and more localized than in the eastern U.S. [Wisniewsk and Keitz, 1983], we expected atmospheric deposition to have limited influence in our models. Our results show a clear association between stream water chemistry and both natural and anthropogenic

atmospheric deposition, but these associations were substantially smaller than the associations with chemical weathering and climate. However, we probably underestimated the effects of atmospheric deposition because we used only wet deposition data. Until spatially extensive dry deposition data are available, we cannot assess how important it might be in determining stream water chemistry. Studies comparing chemical weathering in vegetated and unvegetated catchments show that the presence of vegetation increases fluxes of Ca and Mg from basalts [Moulton *et al.*, 2000] and SiO<sub>2</sub> and Na from granites [Asano *et al.*, 2004]. Other authors examining the effect of vegetation at larger scales have shown either minor or mixed effects of vegetation [Drever, 1997; Jansen *et al.*, 2010], leading us to similar expectations.

[43] We found that the amount of rock/water contact and topographic measures had the least influence on water chemistry. Topography is generally correlated with temperature and soil development [Drever, 1997; Vitousek, 1977], so incorporating these influences into our model directly probably minimized the association of a surrogate variable like topography. Topographic effects on water chemistry have been most clearly observed in small catchments [Johnson *et al.*, 2000; Vitousek, 1977], whereas effects have not been observed in studies of larger catchments [White and Blum, 1995]. Wolock *et al.* [1997] observed that ANC and base cation concentration varied with subsurface contact time, but variation in subsurface contact time dampened in catchments >3 km<sup>2</sup>. Only 5% of our catchments were <3 km<sup>2</sup>, which may explain the limited importance of variables associated with rock/water contact and topography in our models.

[44] Although a strictly empirical approach to modeling cannot establish causation, it can identify those factors that may have the most influence on water chemistry. Our development of multiple regression models based on data from a wide variety of environmental conditions allowed us to separate the influence of factors like temperature, precipitation, vegetation, and soils that often confound one another and also assess the relative importance of these factors. As increasingly accurate spatial estimates of factors that can potentially influence water chemistry become available (e.g., lithology and climate), it will become possible to incorporate them into process models. Such information should improve model predicative power and allow for increased understanding of how past land use development and future climate change may affect stream chemistry.

[45] **Acknowledgments.** This research was supported by grants R-828637-01 and R-830594-01 from the National Center for Environmental Research (NCER) Science to Achieve Results (STAR) Program of the U.S. EPA. We thank Pete Kolesar for geology guidance and suggestions, Ryan Hill for GIS support, and Richard Cutler and John Van Sickle for statistical advice. Constructive suggestions for improving the manuscript were provided by Matt Baker, Michelle Baker, Ryan Hill, Helga Van Miegroet, John Van Sickle, Jacob VanderLaan, and Ellen Wakeley, as well as reviews by Malcolm Cresser and two other referees.

## References

- Anderson, M. G., and C. E. Ferree (2010), Conserving the Stage: Climate change and the geophysical underpinnings of species diversity, *PLoS One*, 5(7), e11554.
- Asano, Y., N. Ohte, and T. Uchida (2004), Sources of weathering-derived solutes in two granitic catchments with contrasting forest growth, *Hydrol. Processes*, 18, 651–666.

- Baker, E. A., K. E. Wehrly, P. W. Seelbach, L. Wang, M. J. Wiley, and T. Simon (2005), A multimetric assessment of stream condition in the northern lakes and forests ecoregion using spatially explicit statistical modeling and regional normalization, *Trans. Am. Fish. Soc.*, *134*, 697–710.
- Baker, M. E., M. J. Wiley, M. L. Carlson, and P. W. Seelbach (2003), A GIS model of subsurface water potential for aquatic resource inventory, assessment, and environmental management, *Environ. Manage.*, *32*, 706–719.
- Bales, R. C., M. Conklin, J. Morrill, K. Mangin, C. Henzel, and J. Henzel (2002), GLOBE QA/QC Rep., Univ. Arizona GLOBE Hydrology Team, Tucson, Ariz.
- Ballard, T. M. (2000), Impacts of forest management on northern forest soils, *For. Ecol. Manage.*, *133*, 37–42.
- Ballester, M. V. R., D. D. Victoria, A. V. Krusche, R. Coburn, R. L. Victoria, J. E. Richey, M. G. Logsdon, E. Mayorga, and E. Matricardi (2003), A remote sensing/GIS-based physical template to understand the biogeochemistry of the Ji-Parana river basin (Western Amazonia), *Remote Sens. Environ.*, *87*, 4285–4405.
- Berg, N. H., A. Gallegos, T. Dell, J. Frazier, T. Procter, J. Sickman, S. Grant, T. Blett, and M. Arbaugh (2005), A screening procedure for identifying acid-sensitive lakes from catchment characteristics, *Environ. Monit. Assess.*, *105*, 285–307.
- Billett, M. F., and M. S. Cresser (1992), Predicting stream-water quality using catchment and soil chemical characteristics, *Environ. Pollut.*, *77*, 263–268.
- Billett, M. F., and M. S. Cresser (1996), Evaluation of the use of soil ion exchange properties for predicting streamwater chemistry in upland catchments, *J. Hydrol.*, *186*, 375–394.
- Breiman, L. (2001), Random forests, *Machine Learning*, *45*, 5–32.
- Breiman, L., J. Friedman, R. Olshen, and C. Stone (1984), *Classification and Regression Trees*, 358 pp., Wadsworth, Belmont, Calif.
- Brenot, A., J. Carignan, C. France-Lanord, and M. Benoit (2007), Geological and land use control on delta S-34 and delta O-18 of river dissolved sulfate: The Moselle river basin, France, *Chem. Geol.*, *244*, 25–41.
- Bricker, O. P., and K. C. Rice (1989), Acidic deposition to streams: A geology-based method predicts their sensitivity, *Environ. Sci. Technol.*, *23*, 379–385.
- Cao, Y., C. P. Hawkins, J. Olson, and M. A. Kosterman (2007), Modeling natural environmental gradients improves the accuracy and precision of diatom-based indicators, *J. North Am. Benthol. Soc.*, *26*, 566–585.
- Chae, G. T., S. T. Yun, K. H. Kim, P. K. Lee, and B. Y. Choi (2004), Atmospheric versus lithogenic contribution to the composition of first- and second-order stream waters in Seoul and its vicinity, *Environ. Int.*, *30*, 73–85.
- Chen, L. M., and C. T. Driscoll (2005), Regional assessment of the response of the acid-base status of lake watersheds in the Adirondack region of New York to changes in atmospheric deposition using PnET-BGC, *Environ. Sci. Technol.*, *39*, 787–794.
- Chinnayakanahalli, K. (2006), *The Multi-Watershed Delineation Tool Software Users Manual*, Utah State Univ., Logan, Utah. (Available at <http://hydrology.usu.edu/mwdtool/>.)
- Clow, D. W., L. Nanus, and B. Huggett (2010), Use of regression-based models to map sensitivity of aquatic resources to atmospheric deposition in Yosemite National Park, USA, *Water Resour. Res.*, *46*, W09529, doi:10.1029/2009WR008316.
- Commission for Environmental Cooperation (CEC) (2006), Ecological regions of North America: Toward a common perspective, Montreal, Quebec, 71 pp.
- Copas, J. B. (1997), Using regression models for prediction: Shrinkage and regression to the mean, *Statist. Methods Med. Res.*, *6*, 167–183.
- Cosby, B. J., G. M. Hornberger, J. N. Galloway, and R. F. Wright (1985), Modeling the effects of acid deposition-assessment of a lumped parameter model of soil-water and streamwater chemistry, *Water Resour. Res.*, *21*, 51–63.
- Cramer, D., and D. Howitt (2004), *The Sage Dictionary of Statistics: A Practical Resource for Students in the Social Sciences*, 188 pp., Sage, London, U. K.
- Cresser, M. S., R. Smart, M. F. Billett, C. Soulsby, C. Neal, A. Wade, S. Langan, and A. C. Edwards (2000), Modelling water chemistry for a major Scottish river from catchment attributes, *J. Appl. Ecol.*, *37*, 171–184.
- Cresser, M. S., N. Ahmed, R. P. Smart, T. Arowolo, L. J. Calver, and P. J. Chapman (2006), Predicting gran alkalinity and calcium concentrations in river waters over a national scale using a novel modification to the G-BASH model, *Environ. Pollut.*, *143*, 361–366.
- Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, and K. T. Hess (2007), Random forests for classification in ecology, *Ecology*, *88*, 2783–2792.
- Daly, C., R. P. Neilson, and D. L. Phillips (1994), A statistical topographic model for mapping climatological precipitation over mountainous terrain, *J. Appl. Meteorol.*, *33*, 140–158.
- Davy-Bowker, J., R. T. Clarke, R. K. Johnson, J. Kokes, J. F. Murphy, and S. Zahradkova (2006), A comparison of the European Water Framework Directive physical typology and RIVPACS-type models as alternative methods of establishing reference conditions for benthic macroinvertebrates, *Hydrobiologia*, *566*, 91–105.
- Drever, J. I. (1997), *The Geochemistry of Natural Waters*, 3rd ed., 437 pp., Prentice Hall, Englewood Cliffs, N.J.
- Driscoll, C. T., G. B. Lawrence, A. J. Bulger, T. J. Butler, C. S. Cronan, C. Eagar, K. F. Lambert, G. E. Likens, J. L. Stoddard, and K. C. Weathers (2001), Acidic deposition in the northeastern United States: Sources and inputs, ecosystem effects, and management strategies, *BioScience*, *51*, 180–198.
- Eitel, J. U. H., D. S. Long, P. E. Gessler, and E. R. Hunt (2008), Combined spectral index to improve ground-based estimates of nitrogen status in dryland wheat, *Agron. J.*, *100*, 1694–1702.
- Evans, C. D., D. T. Monteith, and R. Harriman (2001), Long-term variability in the deposition of marine ions at west coast sites in the UK acid waters monitoring network: Impacts on surface water chemistry and significance for trend determination, *Sci. Total Environ.*, *265*, 11–5129.
- Fox, J. (1997), *Applied Regression Analysis, Linear Models, and Related Methods*, 597 pp., Sage, Thousand Oaks, Calif.
- Gaillardet, J., B. Dupre, P. Louvat, and C. J. Allegre (1999), Global silicate weathering and CO<sub>2</sub> consumption rates deduced from the chemistry of large rivers, *Chem. Geol.*, *159*, 3–30.
- Gherini, S. A., L. Mok, R. J. M. Hudson, G. F. Davis, C. W. Chen, and R. A. Goldstein (1985), The ILWAS Model: Formulation and application, *Water Air Soil Pollut.*, *26*, 425–459.
- Goldstein, R. A., S. A. Gherini, C. W. Chen, L. Mok, and R. J. M. Hudson (1984), Integrated acidification study (ILWAS)—a mechanistic ecosystem analysis, *Philos. Trans. R. Soc., B*, *305*, 409–425.
- Hargett, E. G., J. R. ZumBerge, C. P. Hawkins, and J. R. Olson (2007), Development of a RIVPACS-type predictive model for bioassessment of Wadeable streams in Wyoming, *Ecol. Indic.*, *7*, 807–826.
- Hawkins, C. P., J. R. Olson, and R. A. Hill (2010), The reference condition: predicting benchmarks for ecological and water-quality assessments, *J. North Am. Benthol. Soc.*, *29*, 312–343.
- Hem, J. D. (1985), Study and interpretation of the chemical characteristics of natural water, *Water-Supply Paper Rep.*, *2254*, 263 pp., U.S. Geological Survey, Reston, Va.
- Hendershot, W. H., S. Savoie, and F. Courchesne (1992), Simulation of stream-water chemistry with soil solution and groundwater-flow contributions, *J. Hydrol.*, *136*, 237–252.
- Herlihy, A. T., and J. C. Sifneos (2008), Developing nutrient criteria and classification schemes for Wadeable streams in the conterminous US, *J. North Am. Benthol. Soc.*, *27*, 932–948.
- Herlihy, A. T., S. G. Paulsen, J. Van Sickle, J. L. Stoddard, C. P. Hawkins, and L. L. Yuan (2008), Striving for consistency in a national assessment: the challenges of applying a reference-condition approach at a continental scale, *J. North Am. Benthol. Soc.*, *27*, 860–877.
- Hobbs, D. W. (1964), The tensile strength of rocks, *Int. J. Rock Mech. Min. Sci.*, *1*, 389–396.
- Hobbs, R. J., and D. A. Norton (1996), Towards a conceptual framework for restoration ecology, *Restoration Ecol.*, *4*, 93–110.
- Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira (2002), Overview of the radiometric and biophysical performance of the MODIS vegetation indices, *Remote Sens. Environ.*, *83*, 195–213.
- Jansen, N., J. Hartmann, R. Lauerwald, H. H. Durr, S. Kempe, S. Loos, and H. Middelkoop (2010), Dissolved silica mobilization in the conterminous USA, *Chem. Geol.*, *270*, 90–109.
- Johnson, C. E., C. T. Driscoll, T. G. Siccama, and G. E. Likens (2000), Element fluxes and landscape position in a northern hardwood forest watershed ecosystem, *Ecosystems*, *3*, 159–184.
- Kump, L. R., S. L. Brantley, and M. A. Arthur (2000), Chemical, weathering, atmospheric CO<sub>2</sub>, and climate, *Annu. Rev. Earth Planet. Sci.*, *28*, 611–667.
- Lehman, J. T., and D. K. Branstrator (1994), Nutrient dynamics and turnover rates of phosphate and sulfate in Lake Victoria, East Africa, *Limnol. Oceanogr.*, *39*, 227–233.
- Leland, H. V., and S. V. Fend (1998), Benthic invertebrate distributions in the San Joaquin River, California, in relation to physical and chemical factors, *Can. J. Fish. Aquat. Sci.*, *55*, 1051–1067.

- Leland, H. V., and S. D. Porter (2000), Distribution of benthic algae in the upper Illinois River basin in relation to geology and land use, *Freshwater Biol.*, *44*, 279–301.
- Likens, G. E., C. T. Driscoll, and D. C. Buso (1996), Long-term effects of acid rain: Response and recovery of a forest ecosystem, *Science*, *272*, 244–246.
- Likens, G. E., C. T. Driscoll, D. C. Buso, M. J. Mitchell, G. M. Lovett, S. W. Bailey, T. G. Siccama, W. A. Reiners, and C. Alewell (2002), The biogeochemistry of sulfur at Hubbard Brook, *Biogeochemistry*, *60*, 235–316.
- McCartan, L., J. D. Peper, L. J. Bachman, and J. W. Horton (1998), Application of geologic map information to water quality issues in the southern part of the Chesapeake Bay watershed, Maryland and Virginia, eastern United States, *J. Geochem. Explor.*, *64*, 355–376.
- Minshall, G. W., and J. N. Minshall (1978), Further evidence on the role of chemical factors in determining the distribution of benthic invertebrates in the River Duddon, *Arch. Hydrobiol.*, *83*, 324–355.
- Morse, J. W., and R. S. Arvidson (2002), The dissolution kinetics of major sedimentary carbonate minerals, *Earth Sci. Rev.*, *58*, 51–84.
- Moulton, K. L., J. West, and R. A. Berner (2000), Solute flux and mineral mass balance approaches to the quantification of plant effects on silicate weathering, *Am. J. Sci.*, *300*, 539–570.
- Murtaugh, P. A. (2009), Performance of several variable-selection methods applied to real ecological data, *Ecol. Lett.*, *12*, 1061–1068.
- Nanus, L. (2008), Regional assessment of the sources and effects of acidic deposition on lake chemistry in alpine and subalpine watersheds of national parks in the Rocky Mountains, United States, Ph.D. thesis, Dep. of Geog., Univ. of Colorado, Boulder, Colo.
- Nedel'tcheva, T., C. Piedallu, J. C. Gegout, J. M. Stussi, J. P. Boudot, N. Angeli, and E. Dambrine (2006a), Influence of granite mineralogy, rainfall, vegetation and relief on stream water chemistry (Vosges Mountains, northeastern France), *Chem. Geol.*, *231*, 1–15.
- Nedel'tcheva, T., C. Piedallu, J. C. Gegout, J. P. Boudot, N. Angeli, and E. Dambrine (2006b), Environmental factors influencing streamwater composition on sandstone (Vosges Mountains), *Ann. For. Sci.*, *63*, 369–376.
- Peterson, E. E., A. A. Merton, D. M. Theobald, and N. S. Urquhart (2006), Patterns of spatial autocorrelation in stream water chemistry, *Environ. Monit. Assess.*, *121*, 571–596.
- Pokharel, B., and R. E. Froese (2008), Evaluating alternative implementations of the Lake States FVS diameter increment model, *For. Ecol. Manage.*, *255*, 1759–1771.
- Pond, G. J., M. E. Passmore, F. A. Borsuk, L. Reynolds, and C. J. Rose (2008), Downstream effects of mountaintop coal mining: Comparing biological conditions using family- and genus-level macroinvertebrate bioassessment tools, *J. North Am. Benthol. Soc.*, *27*, 717–737.
- Renner, R. (2009), Salt-loving algae wipe out fish in Appalachian stream, *Environ. Sci. Technol.*, *43*, 9046–9047.
- Robinson, A. P., R. A. Duursma, and J. D. Marshall (2005), A regression-based equivalence test for model validation: Shifting the burden of proof, *Tree Physiol.*, *25*, 903–913.
- Rykiel, E. J. (1996), Testing ecological models: The meaning of validation, *Ecol. Modell.*, *90*, 229–244.
- Siroky, D. S. (2009), Navigating random forests and related advances in algorithmic modeling, *Statistics Surv.*, *3*, 147–163.
- Smart, R. P., C. Soulsby, C. Neal, A. Wade, M. S. Cresser, M. F. Billett, S. J. Langan, A. C. Edwards, H. P. Jarvie, and R. Owen (1998), Factors regulating the spatial and temporal distribution of solute concentrations in a major river system in NE Scotland, *Sci. Total Environ.*, *221*, 93–110.
- Smart, R. P., C. Soulsby, M. S. Cresser, A. J. Wade, J. Townend, M. F. Billett, and S. Langan (2001), Riparian zone influence on stream water chemistry at different spatial scales: A GIS-based modelling approach, an example for the Dee, NE Scotland, *Sci. Total Environ.*, *280*, 173–193.
- Sokolova, T. A., and S. A. Alekseeva (2008), Adsorption of sulfate ions by soils (a review), *Eurasian Soil Sci.*, *41*, 140–148.
- Soulsby, C., M. Chen, R. C. Ferrier, R. C. Helliwell, A. Jenkins, and R. Harriman (1998), Hydrogeochemistry of shallow groundwater in an upland Scottish catchment, *Hydrol. Proc.*, *12*, 1111–1127.
- Strobl, C., A. L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis (2008), Conditional variable importance for random forests, *BMC Bioinf.*, *9*.
- Stutter, M. I., L. K. Deeks, and M. F. Billett (2004), Spatial variability in soil ion exchange chemistry in a granitic upland catchment, *Soil Sci. Soc. Am. J.*, *68*, 1304–1314.
- Sullivan, T. J., J. R. Webb, K. U. Snyder, A. T. Herlihy, and B. J. Cosby (2007), Spatial distribution of acid-sensitive and acid-impacted streams in relation to watershed features in the Southern Appalachian Mountains, *Water Air Soil Pollut.*, *182*, 57–71.
- Townsend, C. R., A. G. Hildrew, and J. Francis (1983), Community structure in some southern English streams - the influence of physicochemical factors, *Freshwater Biol.*, *13*, 521–544.
- U.S. Environmental Protection Agency (USEPA) (2004), Evaluation of impacts to underground sources of drinking water by hydraulic fracturing of coalbed methane reservoirs, *Rep. EPA 816-R-04-003*, Washington, DC.
- U.S. Geological Survey (USGS) (2006), FGDC digital cartographic standard for geologic map symbolization (post script implementation), *Survey Techniques and Methods Rep. 11-A2*, U.S. Geol. Surv., Reston, Va.
- Vitousek, P. M. (1977), Regulation of element concentrations in mountain streams in northeastern United States, *Ecol. Monogr.*, *47*, 65–87.
- White, A. F., and A. E. Blum (1995), Effects of climate on chemical-weathering in watersheds, *Geochim. Cosmochim. Acta*, *59*, 1729–1747.
- Williams, W. D. (1987), Salinization of rivers and streams—an important environmental hazard, *Ambio*, *16*, 180–185.
- Wisniewski, J., and E. L. Keitz (1983), Acid-rain deposition patterns in the continental United States, *Water Air Soil Pollut.*, *19*, 327–339.
- Wolock, D. M., J. Fan, and G. B. Lawrence (1997), Effects of basin size on low-flow stream chemistry and subsurface contact time in the Neversink River Watershed, New York, *Hydrol. Processes*, *11*, 1273–1286.
- Wu, H., J. S. Kimball, N. Mantua, and J. Stanford (2011), Automated upscaling of river networks for macroscale hydrological modeling, *Water Resour. Res.*, *47*, W03517, doi:10.1029/2009WR008871.
- Zheng, L., J. Gerritsen, J. Beckman, J. Ludwig, and S. Wilkes (2008), Land use, geology, enrichment, and stream biota in the eastern Ridge and Valley Ecoregion: Implications for nutrient criteria development, *J. Am. Water Resour. Assoc.*, *44*, 1521–1536.
- Zuur, A. F., E. N. Ieno, and C. S. Elphick (2009), A protocol for data exploration to avoid common statistical problems, *Methods Ecol. Evolut.*, *1*, 3–14.

---

C. P. Hawkins and J. R. Olson, Western Center For Monitoring and Assessment of Freshwater Ecosystems, Department of Watershed Sciences, and Ecology Center, Utah State University Logan, Utah 84322-5210, USA. (john.olson@usu.edu)