5-2012

The Influence of Geology and Other Environmental Factors on Stream Water Chemistry and Benthic Invertebrate Assemblages

John R. Olson
Utah State University

Recommended Citation

Olson, John R., "The Influence of Geology and Other Environmental Factors on Stream Water Chemistry and Benthic Invertebrate Assemblages" (2012). *All Graduate Theses and Dissertations*. 1327.
https://digitalcommons.usu.edu/etd/1327

THE INFLUENCE OF GEOLOGY AND OTHER ENVIRONMENTAL FACTORS ON STREAM WATER CHEMISTRY AND BENTHIC INVERTEBRATE ASSEMBLAGES

by

John R. Olson

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Watershed Science

Approved:

Charles P. Hawkins Major Professor	Peter T. Kolesar Committee Member
Matthew E. Baker Committee Member	Helga Van Miegroet Committee Member
Michelle A. Baker Committee Member	Mark R. McLellan Vice President for Research and Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2012

ABSTRACT

The Influence of Geology and Other Environmental Factors on Stream Water Chemistry and Benthic Invertebrate Assemblages

by

John R. Olson, Doctor of Philosophy

Utah State University, 2012

Major Professor: Dr. Charles P. Hawkins
Department: Watershed Sciences

Catchment geology is known to influence water chemistry, which can significantly affect both species composition and ecosystem processes in streams. However, current predictions of how stream water chemistry varies with geology are limited in both scope and precision, and we have not adequately tested the specific mechanisms by which water chemistry influences stream biota. My dissertation research goals were to (1) develop empirical models to predict natural base-flow water chemistry from catchment geology and other environmental factors, (2) extend these predictions to nutrients to establish more realistic criteria for evaluating water quality, and (3) test the hypothesis that catchment geology significantly influences the composition of stream invertebrate assemblages by restricting weak osmoregulators from streams with low total dissolved solids (TDS). To meet goal 1, I first mapped geologic chemical and physical influences by associating rock properties with geologic map units. I then used these maps and other environmental factors as predictors of electrical conductivity (EC, a measure of TDS), acid neutralization capacity, and calcium, magnesium, and sulfate concentrations. The models explained 58 – 92% of the variance in these five constituents. Rock chemistry was the best predictor of stream water chemistry, followed by temperature, precipitation and other factors. To meet goal 2, I developed empirical models predicting naturally occurring stream total nitrogen and total phosphorus concentrations. These models explained most of the spatial variation among sites in total nitrogen and phosphorus and produced better predictions than previous models. By determining upper prediction limits that incorporated model error, I demonstrated how predictions of nutrient concentrations could be used to set site-specific nutrient criteria and accounted for natural variation among sites better than regional criteria. To meet goal 3, I experimentally manipulated (high and low) EC in both stream-side and laboratory flow-through microcosms and measured survival, growth, and emergence of 19 invertebrate taxa. Observed variation among taxa in survival between treatments predicted taxon EC optima estimated from field observations ($r^2 = 0.60$). Taxa with the greatest differences in survival between treatments also had the highest EC optima, indicating that the inability to persist in low EC likely restricts the distributions of some taxa.

(161 pages)

PUBLIC ABSTRACT

The Influence of Geology and Other Environmental Factors on Stream Water Chemistry and Freshwater Invertebrates

Determining if a stream has been degraded by human activities requires knowing what that stream's natural water quality and freshwater species composition would likely be without any alteration. However stream natural conditions vary greatly from stream to stream, making predicting natural conditions difficult. To determine natural stream conditions, I developed models to predict natural stream water chemistry at individual streams across the western USA. Specifically, the models predict a stream's electrical conductivity (a measure of the amount of solids dissolved in water), acid neutralization capacity, and concentrations of calcium, magnesium, sulfate, total phosphorus, and total nitrogen. These models predict chemistry expected under natural conditions because they are based on measurements of watershed characteristics not influenced by human activities, such as geology, climate, soils and topography. Model predictions allow comparison of current water chemistry with the water chemistry expected under natural conditions. These comparisons can then used to determine if protection or restoration efforts are needed.

To better understand how natural differences in water chemistry could affect freshwater species, I also ran two experiments in which I exposed a range of animals to waters with different amounts of dissolved solids. I found that low amounts of dissolved solids in streams affect the survival of some invertebrates, but not others. These differences in survival occurred because some animals living in dilute freshwater are better at maintaining the required balance between water and salts (i.e., osmoregulating) than other species. Those animals with poorer survival when exposed to water with low dissolved solids in my experiments also did not occur in streams with low dissolved solids in nature.

Combining models and experimental results showed that streams underlain by granite or similar rocks have low dissolved solids, causing some invertebrates to be restricted from these streams. These combined results explain why invertebrate distributions in nature are related to geology and provides insight into the basic ecological question of why animals live where they do. This research increases our understanding of both how geology influences water chemistry and how different invertebrates respond to water chemistry, improving our ability to predict the chemical and biological conditions of streams.

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.

I would like to thank my advisor, Dr. Chuck Hawkins, for the many opportunities and guidance he provided me that have made me a better scientist. I would especially like to thank my committee members, Drs. Michelle Baker, Matt Baker, Pete Kolesar, Helga Van Miegroet, and Jon Harrison, for their support and assistance throughout the entire process. I thank my lab mates Yong Cao, Jeff Ostermiller, Trey Simmons, Brian Creutzburg, Ryan Hill, Nora Burbank, Ellen Wakeley, Jake VanderLaan, and Robin Jones for their advice, assistance, and moral support. Chris Krober, Ryan Zaugg, Nicole Wright, Seth Flanigan, Jonathan Felis, and Jake Gibbson did much of the long, hard work in field and laboratory that made this research possible. Thanks to Drs. Richard Cutler and John Van Sickle for their statistical advice. Drs. Mark Vinson, Scott Miller, and the staff of the Bug Lab and Dr. Laurie McNeill, Joan McLean, and the staff of USU Water Research Lab provided me invaluable expertise, lab space, and equipment. Thanks also to the Watershed Sciences Department faculty and staff for their camaraderie, support, and encouragement.

I give special thanks to my family for first allowing me to pursue this, and then providing the support to enable me to finish. For me this was a huge challenge, and I could not have done it without all of you.

John R. Olson

DEDICATION

To the memories of Major Stuart Wolfer, Sergeant First Class James Moudy, Specialist Charles Wilson, and Private Nathan Thacker. We all took the same risks, and I was just lucky.

To Tanya, Chris, and Kyle for their unconditional love and support.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The idea that the “valley rules the stream” (Hynes 1975) is a central tenet of stream ecology, and catchment geology has long been recognized as a major driver of stream characteristics that influence aquatic biota. However, little progress has been made in quantifying how geology influences stream environments, how geology interacts with other environmental factors to produce different water chemistries, and how these differences in water chemistry affect organisms. Quantifying these relationships should improve our understanding of both the mechanisms causing these patterns between geology, chemistry, and biota, and the relative influence of different environmental factors on water chemistry. Quantifying these relationships will also allow prediction of reference condition stream chemistry and improve our ability to predict stream biota for use in bioassessment. Predictions of reference condition stream chemistry can be used to directly assess water quality by comparing to current conditions, to develop water quality criteria for monitoring, or to set goals for stream restoration.

New threats to water quality and aquatic biota increase the need for predictions of both natural water chemistry and biotic responses to water chemistry changes. Some of the newest energy extraction processes have been linked to changes in water chemistry. Mountain top mining, hydraulic fracturing, and coal bed methane production have all been associated with increases in TDS (Pond et al. 2008, Renner 2009, U.S. Environmental Protection Agency (USEPA) 2004). Agriculture irrigation return flows or runoff can also increase TDS. These processes do not necessarily have toxic effects via changes in pH or increased metal concentrations, but instead can change the structure of the stream communities because different organisms are adapted to different TDS concentrations (Pond et al. 2008). Models predicting background concentrations of TDS and major ions allow us to determine if stream water chemistry has been altered, and inform managers of potential restoration goals. Understanding how different organisms respond to changes in TDS is important to both predict the effects of changes in TDS on communities and as an indicator of a potential mechanism causing an observed change in community structure. Many of the components of stream communities (i.e., fish, macroinvertebrates, and algae) have been shown to be sensitive to changes in TDS. I focus on macroinvertebrates only because they are the assemblage most commonly used for bioassessment.

Ecologists have long observed that catchment geology influences macroinvertebrate distributions. The earliest observation of this pattern was a survey of the benthic fauna of Scottish Highland streams by Egglishaw and Morgan (1965). They found that streams with total cation concentrations < 400 µeq/L underlain by granite or schist had lower richness and abundances than those streams with greater cation concentrations associated with other lithologies. Minshall and Kuehne (1969) saw similar distribution patterns in their study of the River Duddon. Streams in the upper part of the catchment had cation concentrations <245 µeq/L and lower taxonomic richness. Greater taxonomic richness occurred in streams in the lower portion of the catchment that had greater cation concentrations. The difference in assemblage structure between the upper and lower portions of the catchment were due to the absence of most Ephemeroptera taxa and *Gammarus* from the upper catchment. These same patterns continue to be seen in studies up to present. Neff and Jackson (2011) found that biota differed between streams on the granitic Canadian Precambrian Shield and nearby streams on sedimentary rocks. Shearer and Young (2011) also found geology to have a significant effect on the structure of the macroinvertebrate assemblage among streams in Motueka River catchment in New Zealand. Because of these known associations between stream biota and geology, geology has been incorporated into stream classification systems like the European Union's WFD System-A typology (Davy-Bowker et al. 2006) and the River Environment Classification system (Snelder et al. 2004).

Many potential mechanisms have been proposed to explain the relationship between geology and macroinvertebrates, but they can be grouped into four general types of causative mechanisms (Figure 1-1). The first proposed mechanism is that water chemistry indirectly affects macroinvertebrates via its effect on food availability and quality. Water chemistry affects detritus processing rates by increased Ca concentrations increasing conditioning rates (Egglishaw 1968), periphyton assemblage abundance and composition by differences in periphyton ion and nutrient concentration optima (Leland and Porter 2000), and the flocculation of dissolved organic matter controlled by pH (Krueger and Waters 1983). Bedrock is also the primary source for all P in a catchment and can contribute N to streams in some circumstances (Holloway et al. 1998). All of these effects could influence macroinvertebrates via its influence on quantity and quality of food resources, although food resources are also affected by other factors (e.g., amount of incoming radiation or allochthonous inputs). A second potential mechanism is a direct effect of water chemistry on macroinvertebrates via the osmoregulatory challenge posed by living in a dilute medium. Both Minshall and Minshall (1978) and Willoughby and Mappin (1988) concluded that low ion concentrations were having a direct and variable effect on survival of macroinvertebrate taxa in the River Duddon, and that differences in survival among taxa were at least partly responsible for the distribution patterns seen there. A third potential mechanism was suggested by Huryn et al. (1995, see also Wiley et al. 1997, Jin and Ward 2007) is an indirect effect of geology on macroinvertebrate growth rates via its influence on hydrology and stream

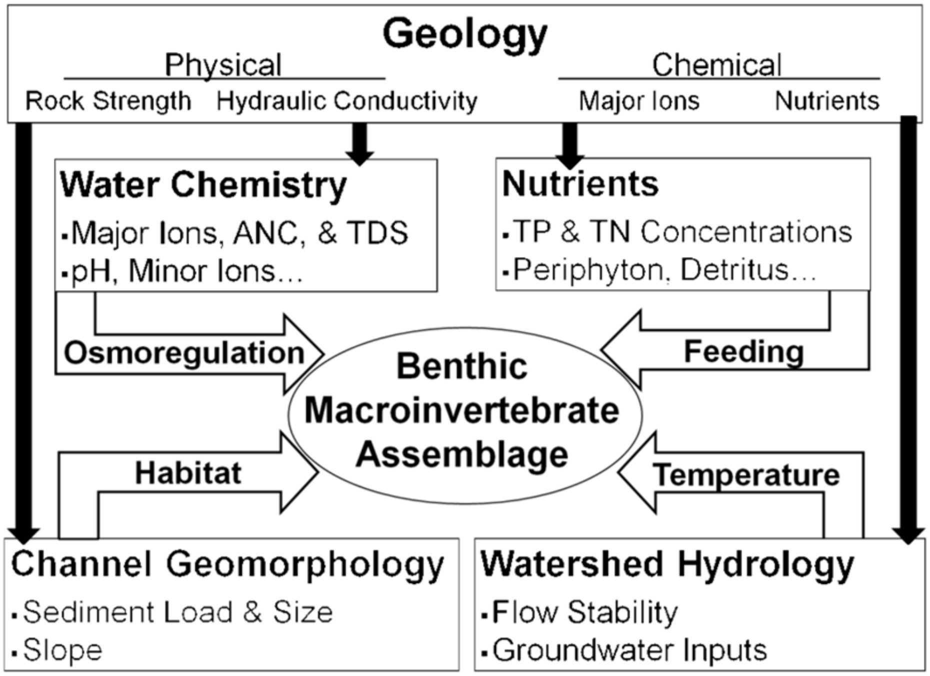


Fig. 1-1. Four potential mechanisms explaining the observed relationship between geology and stream macroinvertebrates.

temperatures. Streams with fractured or porous underlying geology have greater groundwater input than streams on less porous geology. Streams with greater groundwater inputs have more stable hydrologic and temperature regimes. These stable temperature regimes can lead to warmer winter temperatures and increased growth in the winter months (Huryn et al. 1995, Jin and Ward 2007) and to lower summer temperatures with associated higher dissolve oxygen concentrations (Wiley et al. 1997). A fourth potential mechanism is that lithology influences stream substrates (Sable and Wohl 2006), which affects stream macroinvertebrates. Some lithologies produce finer substrates than others (e.g., sandstone or shale produces finer substrates than granite or basalt) and weather at different rates, creating differences in embeddedness and channel morphology known to affect salmonid distributions (Nelson et al. 1992, Hicks and Hall 2003). Although a geology – substrate – macroinvertebrate pathway has not been directly demonstrated, substrate type is known to affect macroinvertebrates (Minshall 1984).

Although each of these four causal mechanisms probably account for some portion of geology's influence on macroinvertebrates, some of the proximal effects are more greatly influenced by factors other than geology. Food quality and quantity affects abundance and biomass of macroinvertebrates, but it may not have a strong effect on species richness or structure (Vinson and Hawkins 1998). Spatial variation in stream temperatures are mostly driven by atmospheric conditions (Caissie 2006), with differences in the amount of groundwater input accounting for a smaller proportion of the variation. Geology's effect on channel substrate size is less than the effects of catchment slope and approximately equal to the effects of precipitation (Snelder et al. 2011). The only causal path where geology is the dominant factor is via its effect on TDS and osmoregulation.

My three objectives for this research were to: 1) model how geology affects TDS and other major ions (Ca, Mg, SO_4, and Acid Neutralization Capacity - ANC), 2) leverage the data and methods developed to address objective one to predict natural background nutrient concentrations to support development of nutrient criteria, and 3) experimentally test if differences in TDS and ion concentrations affect macroinvertebrate fitness and hence their distributions. Objectives one and three examine the geology – water chemistry – macroinvertebrate potential causal path creating a link between geology and macroinvertebrates. Objective two meets a need of the regulatory community to establish criteria for water quality management that accounts for natural variation in water chemistry among streams. I address objective one (Chapter 2) by predicting naturally occurring concentrations of TDS (measured as Electrical Conductivity or EC) and other ions using empirical predictive models. I created these models by predicting reference site water chemistry from catchment measures of geology and other environmental factors known to be related to water chemistry. Measuring the physical and chemical characteristics of the underlying geology required that I first create maps of these characteristics from geologic maps, also presented in Chapter 2. I address objective two (Chapter 3) by developing I models to predict TP and TN concentrations in individual streams. To account for model errors and allow these predictions to be used in setting site-specific nutrient criteria, I also develop two methods for determining prediction. I address objective three (Chapter 4) using an experimental approach to determine if long-term exposure to different levels of EC differentially affect fitness of several macroinvertebrate taxa.

References

Caissie, D. 2006. The thermal regime of rivers: a review. Freshwater Biology 51:1389-1406.

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.

Egglishaw, H. J. 1968. The quantitative relationship between bottom fauna and plant detritus in streams of different calcium concentrations. Journal of Applied Ecology 5:731-740.

Egglishaw, H. J., and N. C. Morgan. 1965. A survey of the bottom fauna of streams in the Scottish Highlands. Part II. The relationship of the fauna to the chemical and geological conditions. Hydrobiologia 26:173-183.

Hicks, B. J., and J. D. Hall. 2003. Rock type and channel gradient structure salmonid populations in the Oregon Coast Range. Transactions of the American Fisheries Society 132:468-482.

Holloway, J. M., R. A. Dahlgren, B. Hansen, and W. H. Casey. 1998. Contribution of bedrock nitrogen to high nitrate concentrations in stream water. Nature 395:785-788.

Huryn, A. D., A. C. Benke, and G. M. Ward. 1995. Direct and indirect effects of geology on the distribution, biomass, and production of the freshwater snail Elimia. Journal of the North American Benthological Society 14:519-534.

Hynes, H. B. N. 1975. The stream and its valley. Verhandlungen des Internationalen Verein Limnologie 19:1-15.

Jin, H.-S., and G. M. Ward. 2007. Life history and secondary production of *Glossosoma nigrior* Banks (Trichoptera: Glossosomatidae) in two Alabama streams with different geology. Hydrobiologia 575:245-258.

Krueger, C. C., and T. F. Waters. 1983. Annual production of macroinvertebrates in 3 streams of different water-quality. Ecology 64:840-850.

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 Biology 44:279-301.

Minshall, G. W. 1984. Aquatic insect-substratum relationships. Pages 358-400 *in* V. H. Resh and D. M. Rosenberg (editors). The ecology of aquatic insects. Praeger Publishers, New York.

Minshall, G. W., and R. A. Kuehne. 1969. An ecological study of invertebrates of the Duddon, an English mountain stream. Archiv für Hydrobiologie 66:169-191.

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. Archiv für Hydrobiologie 83:324-355.

Neff, M. R., and D. A. Jackson. 2011. Effects of broad-scale geological changes on patterns in macroinvertebrate assemblages. Journal of the North American Benthological Society 30:459-473.

Nelson, R. L., W. S. Platts, D. P. Larsen, and S. E. Jensen. 1992. Trout distribution and habitat in relation to geology and geomorphology in the North Fork Humboldt River drainage, northeastern Nevada. Transactions of the American Fisheries Society 121:405-426.

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. Journal of the North American Benthological Society 27:717-737.

Renner, R. 2009. Salt-loving algae wipe out fish in Appalachian stream. Environmental Science & Technology 43:9046-9047.

Sable, K. A., and E. Wohl. 2006. The relationship of lithology and watershed characteristics to fine sediment deposition in streams of the Oregon Coast Range. Environmental Management 37:659-670.

Shearer, K. A., and R. G. Young. 2011. Influences of geology and land use on macroinvertebrate communities across the Motueka River catchment, New Zealand. New Zealand Journal of Marine and Freshwater Research 45:437-454.

Snelder, T. H., F. Cattaneo, A. M. Suren, and B. J. E. Biggs. 2004. Is the River Environment Classification an improved landscape-scale classification of rivers? Journal of the North American Benthological Society 23:580-598.

Snelder, T. H., N. Lamouroux, and H. Pella. 2011. Empirical modelling of large scale patterns in river bed surface grain size. Geomorphology 127:189-197.

USEPA. 2004. Evaluation of impacts to underground sources of drinking water by hydraulic fracturing of coalbed methane reservoirs. EPA 816-R-04-003. US Environmental Protection Agency, Washington, DC.

Vinson, M. R., and C. P. Hawkins. 1998. Biodiversity of stream insects: variation at local, basin, and regional scales. Annual Review of Entomology 43:271-293.

Wiley, M. J., S. L. Kohler, and P. W. Seelbach. 1997. Reconciling landscape and local views of aquatic communities: lessons from Michigan trout streams. Freshwater Biology 37:133-&.

Willoughby, L. G., and R. G. Mappin. 1988. The distribution of *Ephemerella-ignita* (Ephemeroptera) in streams - the role of pH and food resources. Freshwater Biology 19:145-155.

CHAPTER 2

PREDICTING NATURAL BASE-FLOW STREAM WATER CHEMISTRY IN THE WESTERN UNITED STATES*

Abstract

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 % CaO, % MgO, % S, uniaxial compressive strength, and hydraulic conductivity for western USA 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_4. 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_4 when assessed with independent observations. The Root Mean Square Error for predictions

---

* Coauthored by Charles P. Hawkins. Reproduced by permission of American Geophysical Union [*Olson and Hawkins,* 2012].

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.

Introduction

Statement of Problem

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 variation 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 1000s 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 variation in stream water chemistry [*Drever*, 1997 p. 283].

Background

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 (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 1000s 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., non-point sources associated with dispersed land use like livestock grazing or novel sources like 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.

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 U.S.; 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 [*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). 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.

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 lithogeochemical 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_2O, and Na_2O) 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 lithogeochemical 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).

Early water chemistry models predominantly focused on predicting concentrations of major cations and 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 [*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.

Objectives

Our general objective was to model natural base-flow water chemistry in the western U.S. streams from catchment geology and other environmental factors. We focused on developing models for Ca, Mg, SO_4, 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 storm-flow events and our understanding of the effects of storm-flow 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) less than 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.

Methods

Geology Characterization

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 2-1). The source geologic maps we used were the Preliminary Integrated Geologic Map Databases for the United States [*Ludington et al*., 2007; *Stoeser et al.*, 2007], 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.

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 % 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

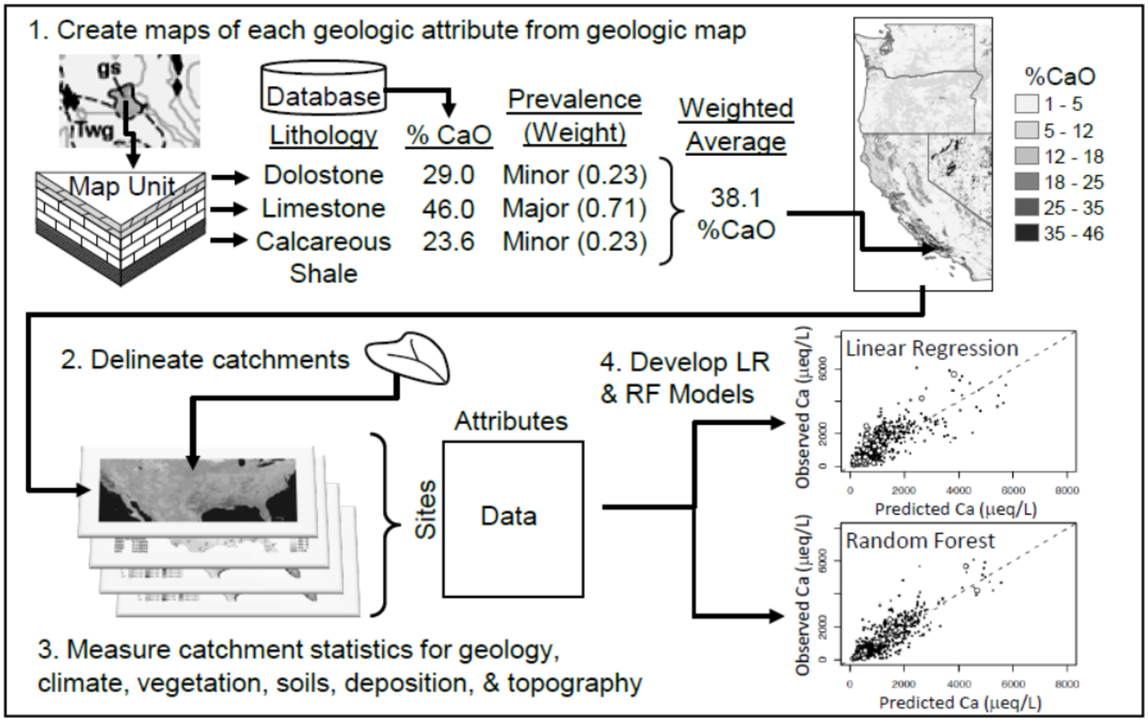


Figure 2-1. Diagram of work flow.

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].

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 descriptions to better parse physical or chemical variability within lithologies (see Table

2-1). 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 2-1 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.

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 non-calcareous, we used rock % CaO to partition samples into three

Table 2-1. Modifiers assigned to lithology by type (chemical or physical) and effect (only applicable lithologies are listed)

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)
non-calcareous (any clastic sedimentary)	till (any unsorted glacial deposit)
calcareous (any clastic sedimentary)	tuff (any volcanic)
carbonaceous (any coarse or fine detrital)	

groups representing non-calcareous, 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 non-calcareous 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 non-calcareous 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 +/- 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 classes were coded as no data so they would have no influence on the characterization of geologic map units.

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 2-2) 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 % CaO, % MgO, % S, UCS, and hydraulic conductivity.

Other Environmental Predictors of Water Chemistry

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 2-3). We

Table 2-2. 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

Table 2-3. Predictor variables used

Type	Variable	Units	Short Name
Geology [a]	Catchment mean whole rock CaO	%	% CaO
	Catchment mean whole rock MgO	%	% MgO
	Catchment mean whole rock S	%	% S
	Catchment mean unconfined compressive strength	MPa	Compressive Strength
	Catchment mean log geometric mean hydraulic conductivity	$x10^{-6}$ m/s	Log Hydraulic Cond
Climate [b]	Catchment mean of mean 1971-2000 annual precipitation	mm/year	Mean Precipitation
	Catchment mean of mean 1971-2000 annual min monthly precipitation	mm/month	Min Precipitation
	Catchment mean of mean 1971-2000 annual max monthly precipitation	mm/month	Max Precipitation
	Catchment mean of mean June-Sept 1971-2000 monthly precipitation	mm/month	Mean Summer Precip
	Catchment mean of mean 1971-2000 annual temperature	˚C	Mean Temperature
	Catchment mean of mean 1971-2000 annual min monthly temperature	˚C	Min Temperature
	Catchment mean of mean 1971-2000 annual max monthly temperature	˚C	Max Temperature
	Catchment mean of mean 1961-1990 first & last day of freeze	day of year	Day Last Freeze
	Catchment mean of mean 1961-1990 annual number of wet-days	days/year	Mean # Wet Days
	Catchment mean of mean 1961-1990 annual relative humidity	%	Relative Humidity
Atmospheric Deposition [c]	Catchment mean of mean 1994-2006 annual precipitation-weighted mean Ca concentration	mg/l	Atmospheric Ca
	Catchment mean of mean 1994-2006 annual precipitation-weighted mean Mg concentration	mg/l	Atmospheric Mg
	Catchment mean of mean 1994-2006 annual precipitation-weighted mean Na concentration	mg/l	Atmospheric Na
	Catchment mean of mean 1994-2006 annual precipitation-weighted mean Cl concentration	mg/l	Atmospheric Cl
	Catchment mean of mean 1994-2006 annual precipitation-weighted mean SO_4 concentration	mg/l	Atmospheric SO_4
	Catchment mean of mean 1994-2006 annual precipitation-weighted mean NO_3 concentration	mg/l	Atmospheric NO_3
	Catchment mean of mean 1994-2006 annual total inorganic nitrogen (TN) wet deposition	kg/ha	Atmospheric TN
Soil [d]	Catchment mean available water capacity	fraction	Soil Water Cap
	Catchment mean bulk density	g/cm^3	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/hr	Soil Permeability
	Catchment mean soil depth	m	Soil Depth
	Catchment mean water table depth	m	Water Table Depth

Table 2-3. Continued

Type	Variable	Units	Short Name
Topography [e]	Catchment elevation mean, min, max, and std deviation	m	MCE, MinCE, MaxCE, SDCE
	Catchment elevation relief ratio	dimensionless	Elevation Relief Ratio
	Catchment shape ratio (catchment area : length)	dimensionless	Catchment Shape
	Catchment area	km^2	Catchment Area
Vegetation [f]	Catchment mean of mean 2000-2009 annual Enhanced Vegetation Index	dimensionless	Mean EVI
	Catchment max of mean 2000-2009 annual Enhanced Vegetation Index	dimensionless	Max Mean EVI
	Catchment mean of mean 2000-2009 annual max Enhanced Vegetation Index	dimensionless	Mean Max EVI
Groundwater [g]	Catchment mean delivery velocity	m/day	Mean Delivery
	Catchment mean recharge velocity	m/day	Mean Recharge
	Catchment mean total flux	m/day	Mean Total Flux
	Catchment mean Base-Flow Index	dimensionless	Base-Flow Index
Rock/Water	Catchment mean % CaO / Mean Precipitation	dimensionless	% CaO/ Precipitation
Interactions [h]	Catchment mean % MgO / Mean Precipitation	dimensionless	% MgO/ Precipitation
	Catchment mean % S / Mean Precipitation	dimensionless	% S/ Precipitation

[a] Derived using method described in section 2.1 at a grid resolution of 90 x 90 m.

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

[c] National Atmospheric Deposition Program National Trends Network (NADP/NTN) 2.5 x 2.5 km resolution grids (obtained from the NADP website http://nadp.sws.uiuc.edu/ntn/).

[d] Natural Resource Conservation Service State Soil Geographic Database (NRCS STATSGO) 500 x 500 m resolution grids (obtained from the NRCS website http://soils.usda.gov/survey/geography/statsgo/).

[e] Calculated from National Elevation Database DEMs at 30 x 30 m resolution (obtained from the USGS website http://ned.usgs.gov/).

[f] MODIS satellite MOD13A1.V4 data collected every 16 days at 500 x 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 (http://lpdaac.usgs.gov).

[g] Velocity derived from MRI-Darcy model [*Baker et al*., 2003], at a 90 x 90 m resolution. Base-Flow Index values derived from interpolation of the ratio of annual max flow to minimum flow for all USGS gage data in the region.

[h] Derived by dividing each rock chemistry grid by the mean precipitation grid to account for spatial interactions.

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 (see Geology Characterization section) 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 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 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 Geology Characterization section) 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.

Water Chemistry Data and Catchment Assessments

We used base-flow water chemistry data collected at 1487 locations across the western U.S. (Figure 2-2) by multiple agencies (Table 2-4) to build empirical predictive models. The 13 western states (approximately 3.45 x10^6 km^2) from which we compiled data represent a wide diversity of climatic and geologic environments, ranging from boreal to sub-tropic 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 2-5) and sample concentrations reported as below detection limits were set to half of the reported detection limit. Some agencies measured

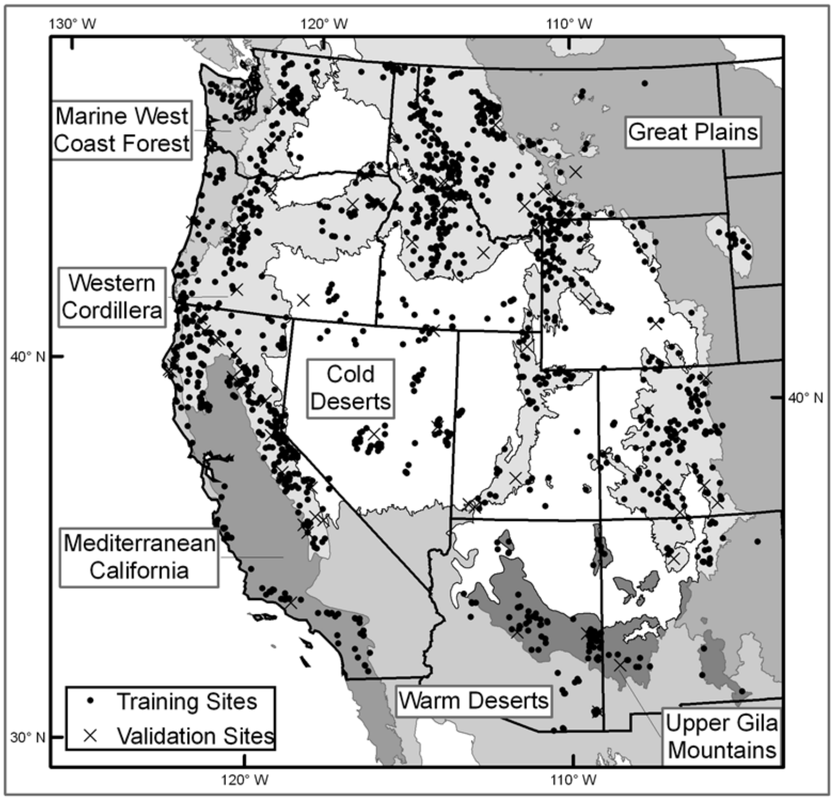


Figure 2-2. Map of 1414 training and 73 validation sites by ecoregion and state.

Table 2-4. Sources of water chemistry data

Data Source	# of sites	Years collected	Location/contact
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 & Environment	76	1992 - 2007	Chris Theel
Eastern Sierra Nevada Dataset	30	1999 - 2002	Dave Herbst
USEPA Environmental Monitoring & Assessment Program	339	2000 - 2004	http://www.epa.gov/emap2/
USGS National Water-Quality Assessment Program	60	1965 - 2008	http://water.usgs.gov/nawqa/
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 Lab, 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	http://waterdata.usgs.gov/nwis

Table 2-5. Summary of water chemistry training data

Constituent	Units	Min	Mean	Max	n[a]	Transform[b]
EC	μS/cm	7	133	1171	1391	0.20
ANC	μeq/L	-110	1271	7280	1324	0.14
Ca	μeq/L	27	998	7194	796	0.25
Mg	μeq/L	9	509	7108	755	0.16
SO_4	μeq/L	2	302	9279	450	0.51

[a] Number of sites used for model development after removal of outliers and sites with high influence.

[b] Exponent used for power transformations applied to data prior to LR modeling only.

ANC in the field, whereas others measured it in the lab. *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 μeq/L due to size of the titrant drop and inaccurate titrant concentrations. To assess whether the field and lab methods might show bias relative to each other, we compared lab and field ANC estimates by regressing each against lab-measured Ca concentrations. The intercept for field measured ANCs was 230 μeq/L greater than lab measured ANCs ($p < 0.00001$, on 342 field and 454 lab measurements of ANC). Slopes of the two regressions were similar (1.48 for field data and 1.41 for lab) 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.

We used the Multi-Watershed Delineation Tool [*Chinnayakanahalli,* 2006] to delineate catchment boundaries for each water chemistry site from DEMs (step 2, Figure 2-1). Catchment averages for all predictive variables were then calculated (step 3, Figure 2-1). We also calculated the coefficient of variation (CV) of each geologic variable as a measure of geologic heterogeneity within catchments.

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 µS/cm), Cl^- (>250 µeq/L), SO_4^{-2} (>250 µeq/L), TP (>90 µg/L), TN (>300 µg/L) or whose catchments contained > 5% agricultural or urban land use (assessed with the 2001 National Land Cover Dataset). These inspection criteria were based on both earlier reference site selection criteria used in the western U.S. [*Herlihy et al*., 2008; *Herlihy and Sefneos*, 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, clear-cuts, etc.). We removed sites from the dataset 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.

Modeling

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

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.

We used both multiple linear regressions (LR) and random forest (RF) regression [*Breiman*, 2001] to develop predictive models (step 4, Figure 2-1). We used both methods because we wanted to compare the performance of these two modeling approaches. RF is a non-parametric modeling approach and has been widely applied to a variety of classification and regression problems in genetics, bio-medical 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 sub-groups 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 non-parametric methods like RF, they can be easily used to make continuous spatial predictions. All analyses were done in the statistical computing environment, R.

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 Bonferroni outlier test) or influenced coefficient estimates by more than 20% were removed from the dataset prior to developing the final model. Only variables that were significant at the $p<0.05$ level were retained in the final models.

We used the same datasets 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.

Model Evaluation, Validation, and Comparison

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.

We also used the equivalence testing strategy outlined in *Robinson et al.* [2005] to assess predictive accuracy, i.e., if the regression of observed on predicted values had an intercept = 0 and slope = 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., $\mu_{obs} = \mu_{pred}$). 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., $\mu_{obs} \neq \mu_{pred}$). 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 non-parametric 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 less than 5% of the bootstrap estimates fell outside of the region of equivalence.

Results and Interpretation

Selected Models and Variable Importance

The numbers of predictors retained in LR models varied from 11 for the SO_4 model to 16 for the ANC model (Table 2-6). The numbers of predictors retained in RF models varied from 7 for the 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.

Most of the predictors included in the models had relative importance and directions of correlation consistent with expectations based on our understanding of the processes

Table 2-6. Model predictors in rank order of importance and direction of association

Random Forest Model			Linear Regression Model			
Predictor	Direction	Importance[a]	Predictor	Direction	Importance[b]	Coefficient
			Electrical Conductivity			
% CaO	+	63	% CaO	+	0.31	2.68E-02
% S	+	42	Max Temperature	+	0.28	3.90E-03
Max Temperature	+	41	% S	+	0.20	5.49E-01
Mean # Wet Days	-	37	Mean # Wet Days	-	0.18	-2.30E-03
Mean Precipitation	-	35	% 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_4	+	0.12	3.05E-01
Atmospheric Ca	+	32	Soil Permeability	-	0.09	-1.17E-02
% MgO	+	32	Log Hydraulic Cond	+	0.09	5.53E-02
Atmospheric SO_4	+	31	Base-Flow Index	+	0.05	6.29E-01
Mean Max EVI	+	30	% MgO CV	+	0.04	6.76E-02
Compressive Strength	-	30	Soil Erodibility	+	0.04	3.86E-01
Min Precipitation	-	29	% 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 Precip	-	24				
			ANC			
% CaO	+	90	% CaO	+	0.38	1.96E-02
% S	+	51	Max Temperature	+	0.27	2.29E-03
Max Temperature	+	48	Soil Organic Content	-	0.16	-4.14E-02
Mean Precipitation	-	39	Soil Bulk Density	+	0.13	2.50E-01
Atmospheric Cl	-	35	% S	+	0.12	2.09E-01
Log Hydraulic Cond	+	35	% CaO CV	+	0.12	8.41E-02
Mean # Wet Days	-	34	Soil Depth	-	0.11	-3.58E-03
Soil Bulk Density	+	33	Max Precipitation	-	0.11	-3.14E-04
Atmospheric Ca	+	33	Soil Permeability	-	0.11	-8.66E-03
% MgO	+	32	Log Hydraulic Cond	+	0.10	3.91E-02
Soil Organic Content	-	31	Mean Summer Precip	-	0.10	-4.39E-06
Atm TN Deposition	-	31	Mean Max EVI	+	0.07	2.46E-05
Atmospheric Mg	+	31	% MgO CV	+	0.06	5.28E-02
Min Precipitation	-	31	Atmospheric SO_4	+	0.05	7.87E-02
Mean Summer Precip	-	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 Max EVI	+	24				

[a.] RF model importance is calculated as % increase in mean squared error when predictor is removed.

[b.] LR model importance is calculated as the absolute value of the standardized coefficients.

Table 2-6. Continued

Random Forest Model			Linear Regression Model			
Predictor	Direction	Importance	Predictor	Direction	Importance	Coefficient
			Calcium			
% CaO / Precipitation	+	85	% CaO	+	0.44	8.79E-02
Max Temperature	+	41	Max Temperature	+	0.23	8.09E-03
Mean Max EVI	+	40	% S	+	0.21	1.27E+00
% S / Precipitation	+	40	% CaO CV	+	0.20	5.93E-01
Mean # Wet Days	-	38	Soil Bulk Density	+	0.19	1.84E+00
Mean Summer Precip	-	37	Min Precipitation	-	0.15	-1.18E-02
Compressive Strength	-	30	Atmospheric SO_4	+	0.15	8.76E-01
Soil Bulk Density	+	29	Soil Permeability	-	0.11	-4.03E-02
Atmospheric SO_4	+	27	Mean Max 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
			Magnesium			
% CaO / Precipitation	+	59	% CaO	+	0.30	1.09E-02
% MgO / Precipitation	+	39	Max Temperature	+	0.26	1.71E-03
Max Temperature	+	36	% S / Precipitation	+	0.20	1.53E+02
% S	+	35	% 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 Precip	-	27	% CaO CV	+	0.13	7.24E-02
Mean Temperature	+	26	Soil Permeability	-	0.12	-8.21E-03
Mean Max EVI	+	24	Soil Bulk Density	+	0.11	1.98E-01
% MgO CV	+	19	% MgO CV	+	0.11	8.42E-02
			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 Precip	-	0.06	-2.05E-06
			(Intercept)	+	0.00	9.06E-01
			Sulfate			
Mean Summer Precip	-	28	% S	+	0.34	6.13E-02
Mean # Wet Days	-	23	Day Last Freeze	-	0.29	-3.66E-04
% S / Precipitation	+	22	% CaO / Precipitation	+	0.21	9.73E-01
Compressive Strength	-	17	Atmospheric SO_4	+	0.19	3.27E-02
Soil Bulk Density	+	15	Soil Bulk Density	+	0.18	5.20E-02
Atmospheric SO_4	+	12	% CaO CV	+	0.13	1.16E-02
% CaO	+	8	Soil Permeability	-	0.12	-1.33E-03
			Max 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

determining water chemistry. Among these was the dominant role of rock chemistry as a source for all constituents, 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 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 ten models. Higher density soils were associated with higher constituent concentrations, likely due to their lower gas exchange rates and increased 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 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 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.

Not all predictors performed as expected, or were clearly associated with a putative mechanism. The weak predictive ability of % MgO relative to % 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 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 SO_4 deposition relative to other atmospheric deposition was also unexpected. SO_4 deposition occurred in 7 models and was the most important atmospheric predictor in the Ca, SO_4, and LR ANC models. The positive correlation between ANC and atmospheric SO_4 in the LR ANC model runs opposite to the expectation that increased acid deposition leads to decreased ANC. Other models of ANC in the western U.S. have not shown SO_4 deposition to be a significant predictor [*Clow et al*., 2010; *Nanus,* 2008]. Although this relationship is possibly caused by anion exchange of SO_4^{2-} for OH^- [*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_4 deposition with other confounding environmental factors. Marine deposition is one possible confounding factor, a possibility supported by the correlation of SO_4 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.

We controlled for 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_4 or TN. Because our measured response for ANC and SO_4 includes some amount of anthropogenic inputs, our empirical models of these constituents is of natural background plus anthropogenic inputs and include SO_4 and TN deposition as predictors. Although anthropogenic deposition is widespread, its effects on stream chemistry compared with that associated with land use are small.

Model Fit and Validation

The models explained 60 – 78% of the variation in the training data (Table 2-7 and Figure 2-3), with nRMSEs that were all less than 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_4. The nRMSEs for RF models ranged from 3 – 11%. Model efficiencies calculated from the independent validation dataset 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_4 models, but all validation nRMSEs were < 15%.

Table 2-7. Assessment of model performance

Model	Data	n	R^{2} [a]	RMSE	nRMSE	r^{2} [b]	Equivalent Intercept [c]	Equivalent Slope [d]
Electrical Conductivity								
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
ANC								
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
Calcium								
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
Magnesium								
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
Sulfate								
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

[a] 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.

[b] Squared Pearson correlation between observations and associated model predictions.

[c] Percentage of 10,000 bootstrap simulations falling within the region of equivalence ($Eq_0 = \hat{Y} \pm 25\%$) for the intercept = 0.

[d] Percentage of 10,000 bootstrap simulations falling within the region of equivalence ($Eq_1 = m \pm 25\%$) for the slope = 1.

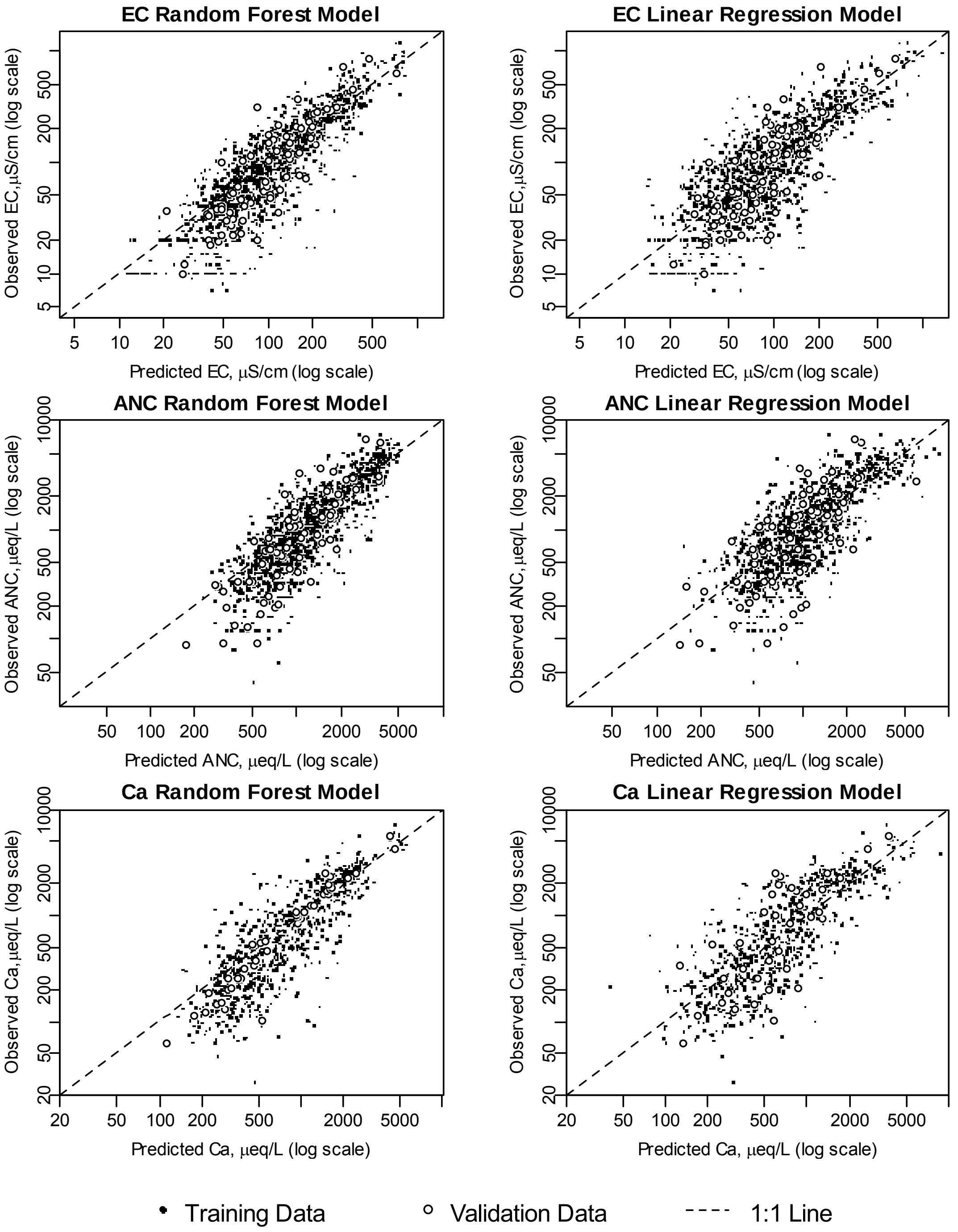


Figure 2-3. Plots of predicted vs. observed values for both training and validation data by constituent and modeling technique. LR predictions are back transformed. Plots are presented in log – log form to improve readability with the ANC plots adjusted to make all values positive.

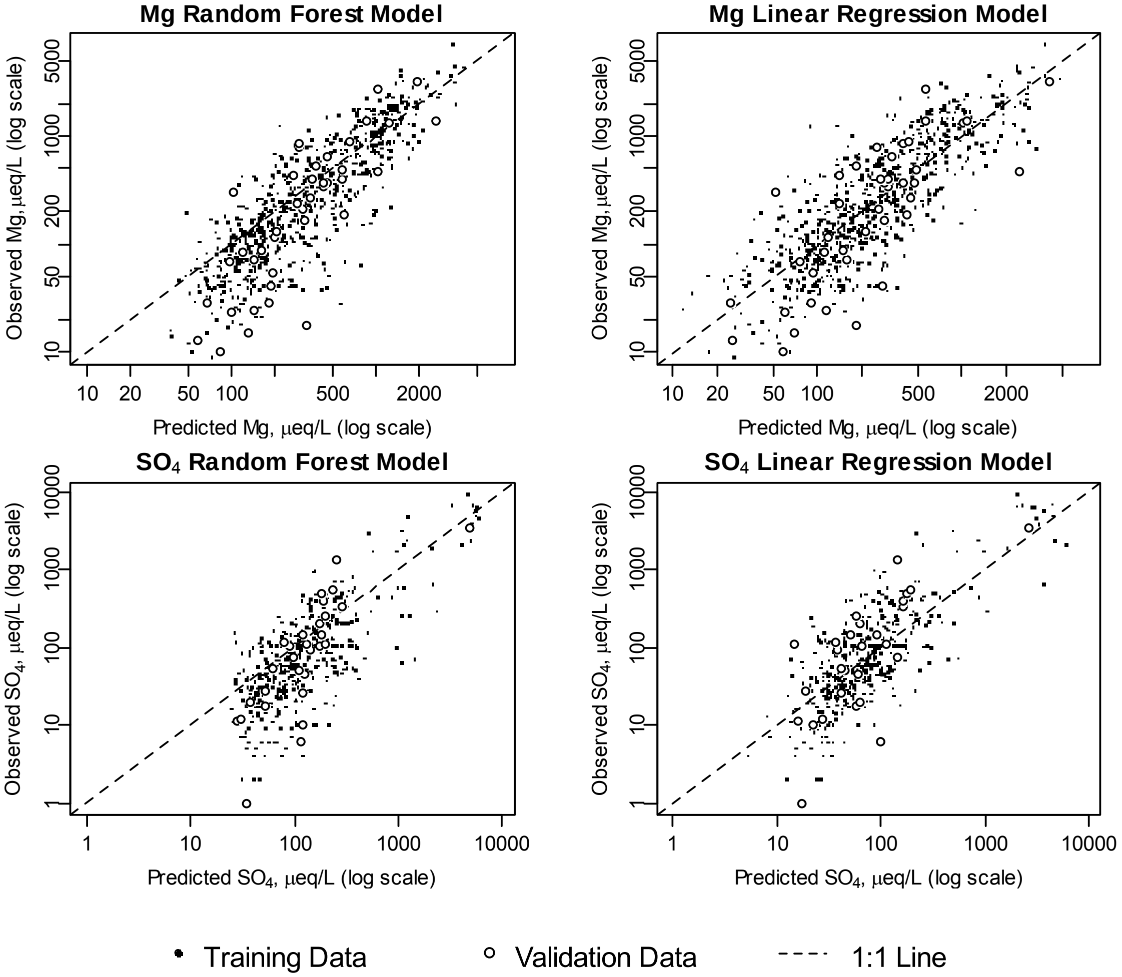


Figure 2-3. Continued.

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, more than 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 $\mu_{obs} \neq \mu_{pred}$ 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_4 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_4 models were the most biased of any of the LR or RF models.

Although the plots of observed vs. 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_4 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_4, 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 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 dataset. 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.

Discussion

Comparison of Models Based on Continuous Geology with Previous Work

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 2-8). 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.

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

Table 2-8. Summary of previous empirical surface water chemistry models

Study	Response	Predictors	Study Area (Extent x10^6 km^2)	Train n	Model Type	Valid n	$R^{2\ a,\ b}$	$r^{2\ a,c}$
Baker et al. [2005]	EC	Land Use & Surficial Geology	Great Lakes (0.181)	94	LR	0	0.27	-
Peterson et al. [2006]	EC	Land Use, Date, & Coordinate	Maryland (0.032)	874	GLM	100	-	0.7
Zheng et al. [2008]	EC	Land Use	W. Virginia (0.004)	56	LR	0	0.23	-
Berg et al. [2005]	ANC	Geology Class, Vegetation, & Lake Morphology	Sierra Nevada Mountains (0.090)	130	GLM	95	0.07 - 0.51	-
Cresser et al. [2000][d]	Alkalinity	Continuous Geology	R.Dee Scotland (0.002)	18	LR		0.82	-
Cresser et al. [2006]	Alkalinity	Continuous Geology & Precipitation	N. Great Britain (0.09)	29	LR	0	0.85	-
Nedeltcheva et al. [2006a]	ANC	Geology Class, Precipitation, & Catchment Area	Vosges Mountains France (0.003)	95	LR	0	0.30 - 0.81	-
Nedeltcheva et al. [2006b]	ANC	Geology Class & Precipitation	Vosges France (0.003)	95	LR	0	0.65	-
Peterson et al. [2006]	ANC	Land Use & Date	Maryland (0.032)	874	GLM	100	-	0.4
Clow et al. [2010]	ANC	Geology Class, Catchment Area, Vegetation, & N deposition	Yosemite California (0.003)	52	LR	0	0.87	-
Cresser et al. [2000][d]	Ca	Continuous Geology	R.Dee Scotland (0.002)	18	LR		0.82	-
Cresser et al. [2006]	Ca	Continuous Geology & Precipitation	N. Great Britain (0.09)	29	LR	0	0.85	-
Nedeltcheva et al. [2006a]	Ca	Geology Class, Slope, Catchment Area, Vegetation, & Precipitation	Vosges Mountains France (0.003)	95	LR	0	0.48 - 0.79	-
Nedeltcheva et al. [2006b]	Ca	Geology Class, Precipitation, & Catchment Area	Vosges Mountains France (0.003)	95	LR	0	0.59	-
Nedeltcheva et al. [2006a]	Mg	Geology Class, Precipitation, Catchment Area, & Vegetation	Vosges Mountains France (0.003)	95	LR	0	0.70 - 0.79	-
Nedeltcheva et al. [2006b]	Mg	Geology Class	Vosges France (0.003)	95	LR	0	0.48	-
Peterson et al. [2006]	SO_4	Land Use, Ecoregion, & Coordinate	Maryland (0.032)	870	GLM	100	-	0.2

[a]. Assessment of fit was based on validation data, unless Valid n = 0 in which case fit was assessed for training data.

[b]. R2 is the coefficient of determination for the multiple regression models. Ranges represent R2 for models developed for different portions of landscape.

[c]. r2 was reported as the squared Pearson correlation coefficient between observed and predicted data.

[d]. Only results of upland base-flow models were reported.

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.* 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.

Geologic classifications better characterize natural environmental variation than land use and often result in 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 and 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 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 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, 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 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.

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 co-located (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 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 *Anderson and Ferree* [2010].

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 over-predictions. Once differences in dilution due to runoff were accounted for and the model re-parameterized with data from both locations, the model predicted Ca and Gran alkalinity with slightly more precision 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.

Model Applicability

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 below 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 [*Hargett 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 U.S. 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 2-5 and 2-6, and R objects for the RF models are available from the authors.

Model Limitations

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 vs. predictions greater than one. This tendency to underpredict was also seen in the model of dissolved 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 & 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.

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 dataset 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 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.

Although the LR and RF 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 vs. predictions. Poor performance of 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 SO_4 deposition, or sinks such as retention of SO_4 in wetlands. We suspect that three factors may be associated with the relatively poor performance of our 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 like 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 H_2S. This process can lower 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 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 SO_4 concentrations.

Relative Importance of Environmental Factors on Stream Chemistry

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 %CaO vs. % 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 % 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 % S in predicting all constituents probably reflects the contributions from high solubility evaporites like $CaSO_4$ and $MgSO_4$ to EC, ANC, Ca, and Mg concentrations. Similar associations between SO_4 and both Ca and Mg were seen by *Brenot et al*. [2007].

The importance of temperature relative to precipitation was unexpected however. Although temperature is known to positively affect SiO_2 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 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.

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 µeq/L 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_2 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.

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 greater than 3 km^2. Only 5% of our catchments were < 3 km^2, which may explain the limited importance of variables associated with rock/water contact and topography in our models.

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.

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).

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 report, *Rep.*, University of Arizona GLOBE Hydrology Team, Tuscon, 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*, 429-445.

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. N. Am. Benthol. Soc.*, *26*, 566-585.

CEC (2006), Ecological regions of North America: Toward a common perspective, Commission for Environmental Cooperation, Montreal, Quebec.

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.

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.

Copas, J. B. (1997), Using regression models for prediction: Shrinkage and regression to the mean, *Statistical Methods in Medical Research*, *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 Publications, London.

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*, 115-129.

Fox, J. (1997), *Applied Regression Analysis, Linear Models, and Related Methods*, 597 pp., Sage Publications, Thousand Oaks, Calif.

Gaillardet, J., B. Dupre, P. Louvat, and C. J. Allegre (1999), Global silicate weathering and CO_2 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. N. 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. N. 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. N. Am. Benthol. Soc*., 27, 860-877.

Hobbs, D. W. (1964), The tensile strength of rocks, *International Journal of Rock Mechanics and Mining*, *1*, 389-396.

Hobbs, R. J., and D. A. Norton (1996), Towards a conceptual framework for restoration ecology, *Restoration Ecology*, *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_2, 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.

Ludington, S., B. C. Moring, R. J. Miller, P. A. Stone, A. A. Bookstrom, D. R. Bedford, J. G. Evans, G. A. Haxel, C. J. Nutt, K. S. Flyn, and M. J. Hopkin (2007), Preliminary integrated geologic map databases for the United States Western States: California, Nevada, Arizona, Washington, Oregon, Idaho, and Utah, *Open-File Rep. 2005-1305*, U.S. Geological Survey, Reston, Va.

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, *Ecology Letters*, *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.

Nedeltcheva, T., C. Piedallu, J. C. Gegout, J. P. Boudot, N. Angeli, and E. Dambrine (2006a), Environmental factors influencing streamwater composition on sandstone (Vosges Mountains), *Ann. For. Sci.*, *63*, 369-376.

Nedeltcheva, T., C. Piedallu, J. C. Gegout, J. M. Stussi, J. P. Boudot, N. Angeli, and E. Dambrine (2006b), Influence of granite mineralogy, rainfall, vegetation and relief on stream water chemistry (Vosges Mountains, north-eastern France), *Chem. Geol.*, *231*, 1-15.

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. Copyright 2012 American Geophysical Union.

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. N. 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 Surveys*, *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 Science*, 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. Processes*, *12*, 1111-1127.

Stoeser, D. B., 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*, U.S. Geological Survey, Reston, Va.

Strobl, C., A. L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis (2008), Conditional variable importance for random forests, *BMC Bioinf.*, *9,* 307.

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.

USEPA (2004), Evaluation of impacts to underground sources of drinking water by hydraulic fracturing of coalbed methane reservoirs, *Rep. EPA 816-R-04-003*, US Environmental Protection Agency, Washington, D.C.

USGS (2006), FGDC digital cartographic standard for geologic map symbolization (post script implementation), *Survey Techniques and Methods Rep. 11-A2*, U.S. Geological Survey, 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:10.1029/2009WR00887

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 in Ecology & Evolution*, *1*, 3-14.

CHAPTER 3

DEVELOPING SITE-SPECIFIC NUTRIENT CRITERIA FROM EMPIRICAL MODELS[*]

Abstract

Ecologically meaningful and scientifically defensible nutrient criteria are needed to protect the water quality of the Nation's streams. To protect aquatic life uses, these criteria should be based on our best understanding of naturally occurring nutrient concentrations. Previous approaches to predicting natural background nutrient concentrations have relied on some form of landscape categorization (e.g., nutrient ecoregions) to account for natural variability among waterbodies. However, the natural variation within these regions is still so high that use of a single criterion would under protect naturally occurring low-nutrient streams and overprotect naturally occurring high-nutrient steams. We developed Random Forest models to predict how baseflow concentrations of total P (TP) and total N (TN) vary among western U.S. streams in response to continuous spatial variation in nutrient sources, sinks, or other processes affecting nutrient concentrations. Both models were relatively accurate (Root Mean Squared Errors < 12% of the range of observations for independent validation sites) and made better predictions than previous models of natural nutrient concentrations. However, the models were not very precise ($r^2 = 0.46$ for the TP model, and $r^2 = 0.23$ for the TN model). An analysis of the sources of variation showed that our models accounted for a majority of the spatial variation in nutrient concentrations, and much of the imprecision was due to temporal or measurement variation. We applied two methods to determine upper prediction limits that incorporated model error and could be used as

---

[*] Coauthored by Charles P. Hawkins.

site-specific nutrient criteria. These site-specific nutrient criteria better accounted for natural variation among sites than did criteria based on regional average conditions, would increase protection for streams with naturally low nutrient concentrations, and specify more attainable conditions for those streams with naturally higher nutrient concentrations.

Introduction

Nutrient pollution of streams increases plant and microbial growth and shifts ecosystems toward a more eutrophic state, eventually affecting downstream lakes and estuaries. Nutrient pollution has increased dramatically over the last 50 years, with over 50% of stream and 78% of coastal waters now exhibiting eutrophication (USEPA 2011). To prevent further harm and set standards for restoration, the Clean Water Act requires that criteria be established to protect the designated uses of each waterbody. Criteria can be in either narrative or numeric form, but the USEPA has long recommended numeric nutrient criteria be used to identify the level of impairment, prioritize water bodies for management, and set remediation goals for individual water bodies (USEPA 2011). Where the designated use is to sustain naturally occurring biota, numeric nutrient criteria should be developed that protect the trophic states within which biota evolved (Dodds 2007). That is, nutrient criteria designed to protect biota should be based on naturally occurring nutrient concentrations. Even when the designated use is not the protection of aquatic life, understanding how a proposed criterion relates to the expected natural condition would inform decision makers how much a system has been altered. The challenge in establishing meaningful numeric nutrient criteria is in estimating the nutrient concentrations that should occur in streams under natural conditions, especially if those streams have been previously altered by human activities.

Several approaches have been developed to predict background nutrient conditions and define criteria. One approach is to base a criterion on some percentile value of the distribution of nutrient concentrations observed at reference sites within a region (e.g., 75% in USEPA 2000; 86% in Suplee et al. 2007). Another is to model background nutrient concentrations as a function of ecoregion, runoff, and atmospheric deposition (for N) or in-stream loss (for P) (Smith et al. 2003). In a third approach, Dodds and Oakes (2004) modeled nutrient concentrations as a function of land use disturbance within separate ecoregions, depending on the ecoregions to control for natural variation. Because disturbance was used as a predictor in the model, naturally occurring concentrations were predicted by running the model with disturbance set to zero at altered sites. All of these approaches control for natural variation in nutrient concentrations caused by differences in geology, climate, or vegetation by spatially classifying sites into nutrient ecoregions that separate sites into groups with similar environments. However, the ability of such regionalizations to sufficiently control for natural variation in water chemistry and other ecosystem attributes is questionable (Hawkins et al. 2010).

Even when landscape classifications are based on known environmental drivers, they often account for insufficient amounts of natural variation in nutrient conditions to allow the prediction of expected natural nutrient concentrations. Herlihy and Sifneos (2008) concluded that the 14 nutrient ecoregions covering the contiguous U.S. do not control natural variability well enough to allow establishment of regional criteria, specifically in the Pacific Northwest. Even within some of the finer resolution level III ecoregions (85 regions for the contiguous U.S.), TP and TN concentrations varied 3 fold or greater among reference sites (Fig. 5. in Herlihy and Sifneos 2008). Similarly, Cheruvelil et al. (2008) found that multiple regionalization schemes were ineffective in partitioning natural variation in TP and TN among minimally disturbed lakes in Michigan. Robertson et al. (2006) also noted several inherent problems in accounting for variation with ecoregions, including the difficulty of developing a single classification that adequately parses natural variation of multiple chemical constituents when each constituent responds to a different set of processes. They also noted that ecoregions are often confounded with land use because human development occurs disproportionately in ecoregions with favorable environmental attributes. For example, if the amount of agriculture is correlated with natural differences in soil and vegetation type, then regions delineated based on soils or vegetation are likely to differ in water chemistry because of both differences in land use as well as variation in natural features. Identifying appropriate background concentrations in streams that flow across multiple regions and assigning criteria to such streams is also problematic (Dodds and Oakes 2004).

Others have tried using typological or reach-level classification approaches to better control for natural variation in nutrient concentrations (Snelder et al. 2004, Robertson et al. 2006, Herlihy and Sefneos 2008). Although these typologies were more effective than ecoregions, nutrient concentrations still varied up to an order of magnitude within some classes. Because many of the environmental drivers important to water chemistry vary continuously (e.g., climate, topography, vegetation), any discrete classification imposed on these gradients must contain a certain amount of within-class variation.

If large amounts of unexplained natural variation occur within landscape or waterbody classes, it is difficult to establish criteria that are both attainable and protective across the range of expected conditions. Any criterion chosen from across a large range of possible natural conditions will be under-protective for some sites and over-protective for others. An example of under-protection would be a site with very low natural nutrient concentrations, but in a highly variable region with a criterion significantly higher than that site’s natural background condition. A site like this would have to be substantially altered before the nutrient concentrations violated the criterion and prompted action. Ice and Binkley (2003) describe an example of over protection in which the nutrient concentrations found in 3 streams draining undisturbed forest watersheds would exceed regional criteria, indicating that these criteria were set too low. They concluded that “Water quality standards will be acceptable only when they reflect what is physically achievable…” (Ice and Binkley 2003). Given the monetary and societal costs associated with developing TMDLs and restoring streams to meet them, it is critical that management decisions are guided by criteria that are achievable and reliable.

Nutrient criteria should be based on the best estimates of expected natural or near natural conditions, but making these estimates is difficult given the complex environmental processes that influence nutrient concentrations. Smith et al. (2003) developed regression models to predict natural background nutrient concentrations, but because they lacked access to information on vegetation, soils, or geology, they also relied on ecoregions to account for all of these environmental effects. Ice and Binkley (2003) noted that although ecoregions explain some variation in nutrient concentrations, they do not account for the influence of finer-scale factors such as geology or forest type. Dodds and Oakes (2004) called for the consideration of spatially variable characteristics such as geology, slope, and drainage area to better account for natural variation in water chemistry within ecoregions. Recently, new spatial data describing environmental factors that can influence water chemistry have been produced (see Chapter 2). Also, new modeling techniques that account for both nonlinear and interacting predictors have been developed (e.g., Random Forests and Artificial Neural Networks). These advancements in both data and modeling provide an opportunity to develop models in which stream nutrient concentrations are predicted as joint functions of potential nutrient sources and sinks instead of relying on spatial classifications like ecoregions.

Our main objectives were to develop models to predict baseflow nutrient concentrations for individual stream reaches and then to identify site-specific nutrient criteria based on these model predictions. We first describe how we modeled site-specific variation in naturally occurring TN and TP concentrations. We then describe two methods for estimating prediction error and demonstrate how these methods can be applied to estimate the highest probable naturally occurring nutrient concentration at a site, i.e., a site-specific nutrient criterion.

Methods

Nutrient concentration data

We assembled a dataset of TP and TN concentrations from samples collected during baseflow conditions by multiple agencies from 823 reference condition streams across the western U.S. (Figure 3-1 and Table 3-1). Sample TP and TN concentrations were measured from unfiltered grab samples by persulfate oxidation and colorimetry (TP and TN) or calculated as the sum of total Kjeldahl nitrogen plus nitrate and nitrite (TN). We used concentrations derived from individual grab samples instead of long-term averages or estimates of nutrient loads despite the noisiness of this type of data (Knowlton and Jones 2006), because most regulatory agencies use estimates from grab samples in their assessment programs. Also, the number of sites with grab sample data far exceeds the number of sites that have the frequent, multiple measurements needed to calculate loads. The data from many grab samples allowed us to develop models whose scope included a broad range of environments. Sites were originally identified as being in reference condition by the sampling agency, but to ensure consistency we also screened

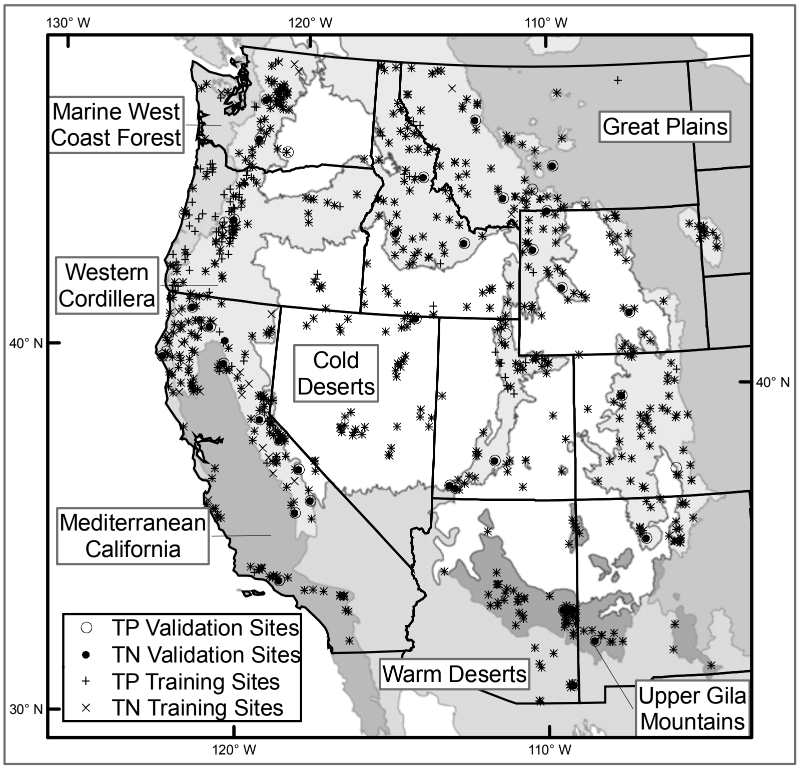


Figure 3-1. Map of 782 training and 41 validation sites by nutrient, ecoregion, and state.

sites to verify their catchments had little to no human disturbance except for atmospheric deposition (see Chapter 2 for details).

Table 3-1. Sources of water chemistry data

Data Source	# of sites	Years collected	Location/contact
Arizona Department of Environmental Quality	25	1994 - 2008	Patrice Spindler
California Department of Fish and Game	46	2003 - 2008	Andrew Rehn
Eastern Sierra Nevada Dataset	22	2000 - 2002	Dave Herbst
USEPA Environmental Monitoring & Assessment Program	337	2000 - 2004	http://www.epa.gov/emap2/
USGS National Water-Quality Assessment Program	41	1973 - 2008	http://water.usgs.gov/nawqa/
New Mexico Environment Department	25	1999 - 2007	Shann Stringer
Oregon Department of Environmental Quality	67	1992 - 2002	Shannon Hubler
Utah State University	255	2001 - 2003	John Olson
USGS National Water Information System	5	1981 - 1995	http://waterdata.usgs.gov/nwis

Environmental predictors

We used a GIS to measure spatial variation in factors potentially affecting nutrient concentrations among sites. These factors include direct effects associated with spatial variation in sources (e.g., rock P, N deposition) and sinks (e.g., P deposition in lakes, removal of N by denitrification). We also measured factors that could indirectly affect nutrient concentrations (e.g., factors associated with evaporation or aquatic and terrestrial nutrient processing rates). Temporal data describing seasonal changes in climate or vegetation were also measured. Our measurements of spatial data included both average upstream catchment conditions and the value of each variable at the sampling point. Catchments were delineated by applying the Multi-Watershed Delineation Tool (Chinnayakanahalli 2006) to 30 m Digital Elevation Models. In total, these measurements produced 182 potential predictor variables for each site. The major categories of predictors and the specific predictors selected for the final models are described below. The full list and descriptions of predictors is available in Appendix Table A.

Data on potential sources of P and N include descriptions of underlying geology, amounts of atmospheric deposition, and distributions of N-fixing plants. All geologic assessments were derived from the Preliminary Integrated Geologic Map Databases for the United States (Ludington et al. 2007, Stoeser et al. 2007). Because basalts can be sources of elevated stream P (Meybeck 1982), we measured the percent of each catchment underlain by volcanic rocks. We also measured each catchment's average bedrock composition of P_2O_5, N, CaO, MgO, and S (see Chapter 2 for details). Because bedrock N in the form of NH_4 is more easily weathered than organic forms (primarily kerogen, Holloway and Dahlgren 2002), we also estimated the amount of bedrock NH_4. Although NH_4 exists in other rock types, we based our estimates of NH_4 rock content on metamorphic rocks only because mineralization of N is associated with diagenesis and metamorphism (Holloway and Dahlgren 2002). We extracted bedrock N values from all geologic map units associated with metamorphic rocks and applied this value as our estimate of bedrock NH_4 concentration. Atmospheric deposition was measured as the long-term (1994-2006) average wet deposition concentrations of NO_3, Ca, Na, and SO_4 from the National Atmospheric Deposition Program National Trends Network. Because dry deposition can be a major source of N we also estimated catchment average annual dry + wet TN deposition. These estimates were obtained by applying the Watershed Deposition Tool to analyze Community Multiscale Air Quality model output (CMAQ, Schwede et al. 2009) and estimate long-term average deposition for available data (2002 – 2006). N-fixing plants can be the dominant source of N in some streams (e.g., Compton et al. 2003), so we developed several predictors describing the potential distribution of N-fixing woody plants identified by the USDA PLANTS Database as naturally occurring in the western U.S. These plants included *Alnus incana*, *Alnus rubra*, *Ceanothus velutinus*, and *Prosopis glandulosa*. To develop maps of the potential distributions of these species under natural conditions, we used the LANDFIRE Biophysical Settings Model descriptions and layers which together describe pre-settlement vegetation patterns. We first identified which LANDFIRE Biophysical Settings Model descriptions listed each species as either occurring or dominant (LANDFIRE 2011b). We then extracted those grid cells associated with the identified Biophysical Settings Model from the LANDFIRE Biophysical Settings layer (LANDFIRE 2011a) to create layers describing the expected locations where each species would be either present or dominant in our study area. We also calculated *Alnus rubra* percent cover for each catchment from estimates of current forest composition derived from Gradient

Nearest Neighbor imputation (Ohmann et al. 2007) of areas across the Pacific Northwest by the Landscape, Ecology, Modeling, Mapping, and Analysis project (LEMMA 2011).

Potential sinks for nutrients include uptake or retention by vegetation, soils, lakes or wetlands; denitrification; and chemical precipitation or adsorption. To characterize spatial differences in potential vegetative uptake we used long-term (2000-2009) average MODIS satellite Enhanced Vegetation Index (EVI) values (Huete et al. 2002) as a proxy for spatial variation in plant biomass. Because MODIS EVI data are available in weekly increments starting in 2000 we could potentially use it to characterize temporally specific differences in vegetative uptake also (i.e., EVI for the specific time of the sample or increase in EVI in the previous month). However 10% of our data was collected before MODIS became operational, so we relied on day of year of the sample to account for seasonal variations in vegetative uptake. We characterized major differences in vegetation composition with data from the 2001 National Land Cover Dataset (NLCD, Homer et al. 2004). We used maps of soil organic carbon (SOC) (Global Soil Data Task Group 2000) and soil organic matter (SOM) content (NRCS 2011) to characterize the potential release or immobilization of nutrients by soils caused by microbial uptake or chelation associated with SOC or SOM. To describe potential differences in nutrient retention by lakes and wetlands, we measured the percent of each catchment classified as lake, wetland, or both (i.e., water body) in both the NLCD and the National Hydrography Dataset (NHD, USGS 2006). We also assessed the size of the largest water bodies in each catchment and the amount of flow routed through these water bodies in the NHD data. We also measured environmental variables associated with differences in conditions favorable to denitrification, such as soil bulk density (lower pore connectivity with increased density creates more anaerobic sites) or the amount of surface–subsurface hydrologic exchange in streams (increased exchange brings more N in contact with hyporheic waters). Soil density was obtained from the U.S. General Soil Map (NRCS 2011). Surface-subsurface hydrologic exchange was characterized by both average catchment hydraulic conductivity and an index of groundwater velocity estimated with the MRI-Darcy model (Baker et al. 2003). The MRI-Darcy model applies Darcy's equation within a GIS environment (see Chapter 2 for details). We also measured other factors that could potentially influence chemical precipitation or adsorption of nutrients where spatial data were available. These variables included the amount of Ca available from either bedrock or atmospheric sources that could act as a co-precipitate with P, and soil pH which could influence adsorption or cation exchange.

We used long-term estimates (1971-2000) of average precipitation, number of wet days, air temperature, day of last freeze, and relative humidity produced by the Parameter-elevation Regression on Independent Slopes Model (PRISM, Daly et al. 1994) to estimate the effects of dilution and evaporative concentration. Because temporal variation in precipitation can influence nutrient concentrations, we also measured PRISM monthly mean precipitation for the month of the sample, mean precipitation for the month previous to the sample, and mean annual precipitation for the year previous to the sample.

We also measured other factors that could potentially affect processing rates or retention, or that could act as proxies for factors we could not measure. These variables included soil order and properties (e.g., available water content, erosion factor, and percent hydric soils), topography (e.g., elevation, relief, and catchment shape), catchment area, Level II ecoregion, and average channel slope. We also included measurements of other atmospheric deposition components not directly related to nutrient concentrations like Mg, Na, Cl, and SO_4.

Model development and evaluation

We used the nonparametric modeling technique Random Forest (RF, Breiman 2001) to develop empirical predictive models. RF models outperformed multiple linear regression models for other water chemistry constituents because of their ability to account for both interactions between variables and nonlinear relationships (see Chapter 2). RF models are ensembles of classification and regression trees (CART, Breiman et al. 1984), where observations are recursively split into groups, minimizing the remaining unexplained variance within each group. Splits are constructed as a series of binary rules based on one of the explanatory variables. However, CART models are sensitive to small changes in training data. RF overcomes this limitation by growing multiple individual trees using a bootstrap sample of the training data and a random sample of the predictors at each split. RF predications are then generated by averaging the predictions of all trees. RF estimates the predictive accuracy of the model from observations that were excluded from each bootstrap sample (the out of bag error) and the importance of each predictor by measuring how out of bag error changes when each predictor is permuted. We implemented RF using the R package randomForest (Liaw and Wiener 2009) creating 1500 trees for each model. To create the most parsimonious model and minimize the number of correlated predictors, we modeled iteratively, removing correlated or low importance predictors until a model's out of bag error began to increase. We used partial-dependence plots to visualize relationships between nutrient concentrations and predictors, and removed any predictors for which the direction of response in nutrient concentrations reversed directions more than three times because such patterns are likely spurious relationships. After predictor variables were selected, we used the tuneRF function to optimize the size of the random sample of the predictors tried at each split. To correct for a small bias inherent in RF regression models (Zhang and Lu 2012), we also applied the bias correction function internal to the randomForest package.

We used both the training (internal) data and an external validation dataset to evaluate model performance. External validation data were selected by randomly sampling 5% of sites, stratified by level II ecoregion (CEC 2006) to ensure the validation set represented all environments. Internal evaluations were based on out of bag observations (analogous to cross validation) allowing us to assess how well the models performed across the widest range of conditions. External validation allowed us to rigorously assess the applicability of these models to completely independent observations. We quantified model performance with the Nash-Sutcliffe Model Efficiency coefficient (NSE) and r^2 values associated with linear regressions of observed vs. predicted concentrations (Piñeiro et al. 2008). We assessed model bias (systematic over- or underprediction) and consistency (deviance between observations and predictions remains constant over their ranges) by testing if the regression of observed vs. predicted concentrations had an intercept = 0 and a slope = 1 using an equivalence test (Robinson et al. 2005). Intercepts ≠ 0 indicate model bias, whereas slopes ≠ 1 indicate that model predictions lack consistency across the range and model over- or underpredicts at the extremes. The equivalence test approach reverses the test from a null hypothesis of agreement between observations and predictions to a null hypothesis of having less than a given difference. This test shifts the burden of proof to the model, and rejection of the null hypothesis indicates predictions are sufficiently similar to the observations for that particular application. A failure to reject the null hypothesis indicates there is either insufficient evidence of a similarity between predictions and observations or a true difference exists. The amount of difference we considered significant (i.e., region of equivalence) was 25% for slope and intercept, assessed with α

= 0.05. Instead of applying the equivalence test once, we used a bootstrap analysis with 10,000 resamples of predictions and observations to estimate the proportion of results that would fall within the region of equivalence for both intercept and slope. We also used the Root Mean Squared Error (RMSE) to assess model accuracy. Finally, we compared the performance of our model with the only other model predicting background nutrient concentrations across the western U.S., the SPARROW model developed by Smith et al. (2003).

Because our predictors primarily describe static spatial variation among sites, we also wanted to assess how much variation in nutrient concentrations could potentially be attributed to temporal or measurement variation. We assessed the magnitude of temporal or measurement variation in concentrations by calculating the ratio of between-site variance (spatial signal) to within-site variance related to temporal and measurement noise, i.e. the signal to noise (S:N) ratio (Kaufmann et al. 1999). For example, if more variation existed among multiple sites than existed among all repeated samples from the same sites, then the S:N ratio would be high. We then used these S:N ratios to estimate the best possible r^2 that static predictors could produce. Variance among sites was calculated from observations in each training data set. Variance among within-site replicate samples was based on a subset of 41 EMAP and USU sites sampled multiple times for both TP and TN. These samples exhibited temporal variation comparable to that seen by Chételat and Pick (2001). We calculated pooled sample variance for these replicate samples. We then calculated the S:N ratio from these two variances and the maximum possible r^2 value as: $max(r^2) = S{:}N/(S{:}N + 1)$ (Van Sickle 2006 and illustrated in figure 2 of Stoddard et al. 2008). We calculated among-site variance with data from all sites instead of just sites with replicate samples because this larger data set provides a more representative estimate of the natural variation in stream nutrient concentrations across the western U.S.

Determining highest probable concentrations based on model predictions

Site-specific nutrient criteria should incorporate both the model prediction of nutrient concentrations and prediction uncertainty arising from unaccounted variation, imperfect model structure, and error in measuring predictor values and nutrient concentrations. Prediction uncertainty can be quantified by establishing a prediction interval describing the range of conditions expected at a site. The upper prediction limit (PL) of this interval would provide a value based on a site's most likely nutrient condition and would account for all uncertainties associated with that prediction arising from unexplained variation and model uncertainty. Distribution based statistical methods (e.g., linear regression) are able to produce prediction intervals from an assumed normal distribution, but non-distributional methods like RF cannot. Quantile Regression Forests have been proposed as a method for determining prediction intervals (Meinshausen 2006), but this approach has two shortcomings. RF models cannot extrapolate beyond the range of the data used to construct them, so quantiles based on RF models become constrained at the lower and upper ends of the data. Also, the quantiles produced by quantile random forest models do not account for the uncertainties associated with the estimates of a given quantile. To develop reliable prediction intervals for our RF models, we instead relied on two forms of empirically derived prediction intervals.

The first method, referred to as the Simple Empirical Error (SEE) method, empirically determines the amount of error for each prediction from a bootstrap sample of residuals from the training data (suggested by John Van Sickle, USEPA-Corvallis, OR, personal communication). For each prediction, we sampled all residuals 500 times with replacement and added each sampled residual to the prediction to create an empirical distribution of the prediction plus error. The 95th percentile of this distribution was then selected as the upper PL for that prediction.

The second method is a variation of the UNcertainty Estimation based on Local Errors and Clustering (UNEEC) method of Shrestha and Solomatine (2008). UNEEC is similar to the first method in that errors are determined from a bootstrap sample of residuals from the training data, but instead of using a sample of all residuals, UNEEC only uses residuals from those samples similar to the site we are trying to predict. Sample residuals for similar sites were derived by first clustering all training observations by their environmental properties and then bootstrap sampling the residuals of each cluster and selecting the 95th percentile as the error for that cluster. For each prediction, probability of membership in each cluster is determined and these probabilities are used to calculate a weighted average of the 95th percentile errors for all clusters. This weighted average error is then added to the prediction to determine the upper 95th percentile PL for that prediction. We created clusters based on those environmental variables selected for the RF model. These environmental data were first standardized to a common scale and then clustered (k-means clustering). We selected the number of clusters to both minimize the sum of squares and ensure the minimum number of samples included in each cluster was greater than 50. We then randomly sampled the residuals of the training data for each cluster 500 times with replacement and determined the 95th percentile value. Probability of cluster membership for new observations was determined by applying a separate RF model built with the same transformed environmental variables used in clustering. These probabilities of cluster membership were then used as weights when calculating the average 95th percentile error to be added to each prediction to determine the upper PL.

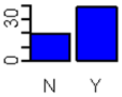
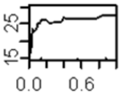
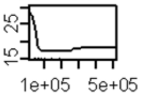
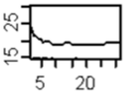
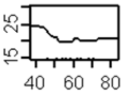
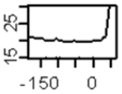
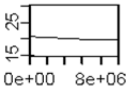
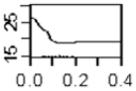
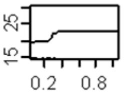
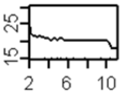
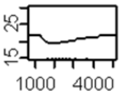
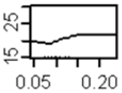
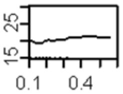
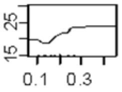
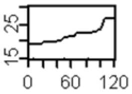
Results

Model structure and performance

Relationships between nutrient concentrations and most predictors were consistent with our understanding of how the natural environment influences nutrient concentrations (Tables 3-2 and 3-3). Both models included factors related to both sources and sinks, but the TP model had two predictors clearly more important than the others, both relating to geologic sources. The TN model did not include any clearly dominant predictors, and TN was almost equally influenced by predictors related to both sources and sinks. The TP predictors were also almost entirely static (with the exception of previous year's precipitation), whereas the TN model included temporal measures like day of year and the prior two months precipitation.

We tried eliminating correlated variables during variable selection, but in several cases removing correlated predictors degraded model performance. To maximize the model's ability to make predictions, correlated variables were retained if they improved model performance. The only predictors in our TN model that were strongly correlated were atmospheric SO_4 and NO_3 deposition (r=0.9). Correlated TP predictors included: relative humidity and soil organic carbon (r=0.8), relative humidity and atmospheric Ca deposition (r=0.64), relative humidity and previous year's precipitation (r=0.63), soil organic carbon and previous year's precipitation (r=0.67), local minimum temperature and EVI (r=0.63), % volcanic lithology and rock P concentration (r=0.69), and soil erosion factor and soil water capacity (r=0.61). RF models are robust to the effects of correlated predictors (Cutler et al. 2007). However, correlated predictors can cause variable importance measures to be unreliable (Strobl et al. 2008), so inferences

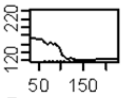
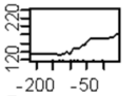
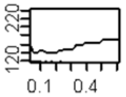
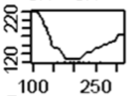
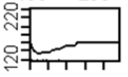
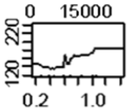
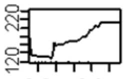
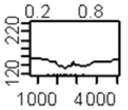
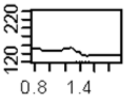
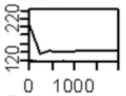
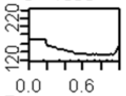
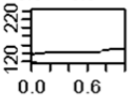
Table 3-2. Predictors, relative importance, direction of effect, and associated mechanisms for TP model

Predictor	Imp[a]	Effect[b]	Interpretation
Gila Mtns. Ecoregion	36 (0.8)		Unusual combination of steep topography and large amounts of young basalts in Upper Gila Mountains / Mogollon Rim result in increased P (Rampe et al. 1981)
% Volcanic Lithology	31 (1.8)		Volcanic rocks are large source of P (Dillon & Kirchner 1975), that chemically weather more quickly than other igneous rocks types (Gislason et al. 1996)
Previous Year's Precipitation	26 (1.6)		Captures both spatial variability in amount of dilution with increasing discharge and the wash out of retained P by previous year's flood events (House 2003)
Rock % CaO	24 (1.3)		Increased Ca availability results in co-precipitation with P in river bed (House 2003) although contrasts with increased rock weathering (Mulholland 1992, Cross & Schlesinger 2001)
Relative Humidity	24 (1.5)		Low humidity results in increased evaporative concentration of solutes, in addition to affecting TP by decreasing soil organic carbon
Local Minimum Temperature	22 (0.9)		Lack of freezing results in increased water flow and less nutrient uptake (Green & Finlay 2010), more shading and less uptake by periphyton, or less P sorption by wetlands (Wang et al. 2007)
Area largest water body	21 (1.0)		Presence of lakes or wetlands acts as a sink for P in catchment, with larger water bodies retaining more P (Smith et al. 2003)
Average Channel Slope	21 (0.9)		Low slopes have greater hydrologic flushing of P from saturated surface soils (D'Arcy & Carignan 1997). Lower P retention on high slopes (Hill et al. 2010) may result in less P export at baseflow.
Atmospheric Ca Deposition	21 (1.5)		Ca deposition acts as surrogate for dust deposition, the major source for atmospheric P (Reynolds 2001)
Soil Organic Carbon	20 (2.8)		High SOC increases biotic P uptake (Kirschbaum 2000) and immobilization in organic form (Stevenson & Cole 1999), but also correlated with old leached soils (Walker & Syers 1976)
Enhanced Vegetation Index	19 (1.1)		Initially increasing vegetation retains P resulting in less P in streams, but at higher levels additional vegetation results increases chemical weathering and release of P
Soil Water Capacity	19 (1.0)		Higher water capacities associated with fine soils are correlated with higher surface runoff and erosion rates (Panagopoulos et al. 2007)
Soil Erosion K Factor	18 (1.3)		Highly erodible soils transport greater amounts of P with suspended sediment
Rock % P	16 (1.1)		High rock P acts as source of P within catchment, % rock P captures variation in P among non-volcanic rocks
% Alfisols	15 (1.1)		Alfisols may increase P export by providing Fe allowing P to bind to dissolved humic matter (Dillon & Molot 1997) or reduce P retention in upper horizons due to low clay content (Bhadha & Jawitz 2010).

[a] Importance, listed as % increase in mean squared error when predictor is removed, with standard error of the mean in parentheses calculated from 50 separate models.

[b] Effect illustrated as partial dependence plots of each predictor with all other predictors held constant. Change in predictor is displayed on the X axis and change in TP is displayed on the Y axis.

Table 3-3. Predictors, relative importance, direction of effect, and associated mechanisms for TN model

Predictor	Imp[a]	Effect[b]	Interpretation
Mean Number of Wet Days	27 (0.9)		Wet soils resulting from more wet days increases denitrification (Bollmann & Conrad 1998), more precipitation increases dilution, dry conditions favor N build up and flushing (Van Miegroet et al. 2001)
Minimum Temperature	25 (0.7)		Higher temperatures increase N-fixation in soil (Cleveland et al. 1999) and streams (Marcarelli & Wurtsbaugh 2006) and also litter decomposition (Park et al. 2003) and soil nitrification
Atmospheric Na Deposition	24 (0.9)		Mechanism appears to be increased release of NH_4 associated with Na (and then nitrification of the released NH_4), not direct exchange with Cl (Jana Compton, USEPA, personal comm.)
Day of Year	24 (0.8)		Phenology related uptake of N by plants, reaching its maximum in summer (Also note higher inputs in spring than fall from snowmelt)
Prior 2 Months Precipitation	23 (0.9)		Precipitation favors N fixation, litter decomposition (Lewis et al. 1999), and flushing (Kane et al. 2008) in mesic areas, whereas in xeric areas it may increase plant uptake and/or denitrification
Atmospheric NO_3 Deposition	21 (0.8)		Source of N from anthropogenic and natural sources
Atmospheric SO_4 Deposition	21 (0.7)		No known mechanism, but Likens et al. (2002) and Cai et al. (2011) saw same pattern, may be surrogate for other source (like dry N deposition), or perhaps SO_4 displaces NO_3 adsorbed to soils
Enhanced Vegetation Index	20 (1.1)		Increasing biomass results in N uptake to a point, where forest maturation results in decreased N retention
Soil Bulk Density	18 (1.0)		Increasing soil density creates more anaerobic sites, therefore increases denitrification in soils (Torbert & Wood 1992)
Ground Water Index	16 (0.9)		Higher ground water index could reflect either up welling of ground water with little soil contact or increasing hyporheic zone contact increasing denitrification (Grimaldi & Chaplot 2000)
% Evergreen	15 (0.8)		Kane et al. (2008) saw same pattern, perhaps due to slower processing or lower N content of evergreen litter (Lopez et al. 2001, Washburn & Arthur 2003), or as surrogate for a soil attribute
% *Alnus rubra* dominated	10 (0.8)		*Alnus rubra* (where present) is a major source of N (Compton et al. 2003)

[a] Importance, listed as % increase in mean squared error when predictor is removed, with standard error of the mean in parentheses calculated from 50 separate models.

[b] Effect illustrated as partial dependence plots of each predictor with all other predictors held constant. Change in predictor is displayed on the X axis and change in TN is displayed on the Y axis.

regarding the relative importance of different processes in Tables 3-2 and 3-3 should be made with caution.

Both models predicted nutrient concentrations without significant bias, but were relatively imprecise (Table 3-4 and Figure 3-2). The TP model accounted for less than half of the variation in TP concentrations, and the TN model accounted for less than a third of TN concentrations. However, both models did have positive, if modest, Nash Sutcliffe Model Efficiencies indicating some predictive power. RMSEs of both models were less than 12% of the range of observed values (TP range: 1 – 192 µg/L, TN range: 5 – 960 µg/L). Only the TP model showed any evidence of bias, which was only slight (-2.3 µg/L) with 16% of the bootstrapped validation samples having an intercept less than the specified region of equivalence. Both models had slopes equivalent to 1 when assessed with training data, but did not when assessed with validation data, indicating predictions were not always consistent with observed values at new locations. For validation data, 51% of the bootstrap slope estimates for the TP model fell above the region of equivalence and the slope of all predictions together was 1.3, although this

Table 3-4. Assessment of model performance and comparison with predictions of the SPARROW model

Nutrient	Model	Data	n	r^2 [a]	NSE [b]	RMSE	Equivalent Intercept [c]	Equivalent Slope [d]
TP	RF	Tng	752	0.40	0.40	16.2	100.0	100.0
		Val	40	0.46	0.43	20.5	83.8	22.2
	SPARROW	Tng	752	0.02	-0.40	24.7	20.6	0
		Val	40	0.04	-0.10	28.5	56.1	16.4
TN	RF	Tng	665	0.32	0.32	113.9	100.0	99.6
		Val	35	0.23	0.16	80.1	96.8	34.6
	SPARROW	Tng	665	0.04	-0.40	163.8	100.0	0
		Val	35	0	-0.58	109.6	75.7	0.4

[a] Squared Pearson correlation between observations and associated model predictions.

[b] Nash-Sutcliffe Model Efficiency.

[c] Percentage of 10,000 bootstrap simulations falling within the region of equivalence ($Eq_0 = \hat{Y} \pm 25\%$) for the intercept = 0.

[d] Percentage of 10,000 bootstrap simulations falling within the region of equivalence ($Eq_1 = m \pm 25\%$) for the slope = 1.

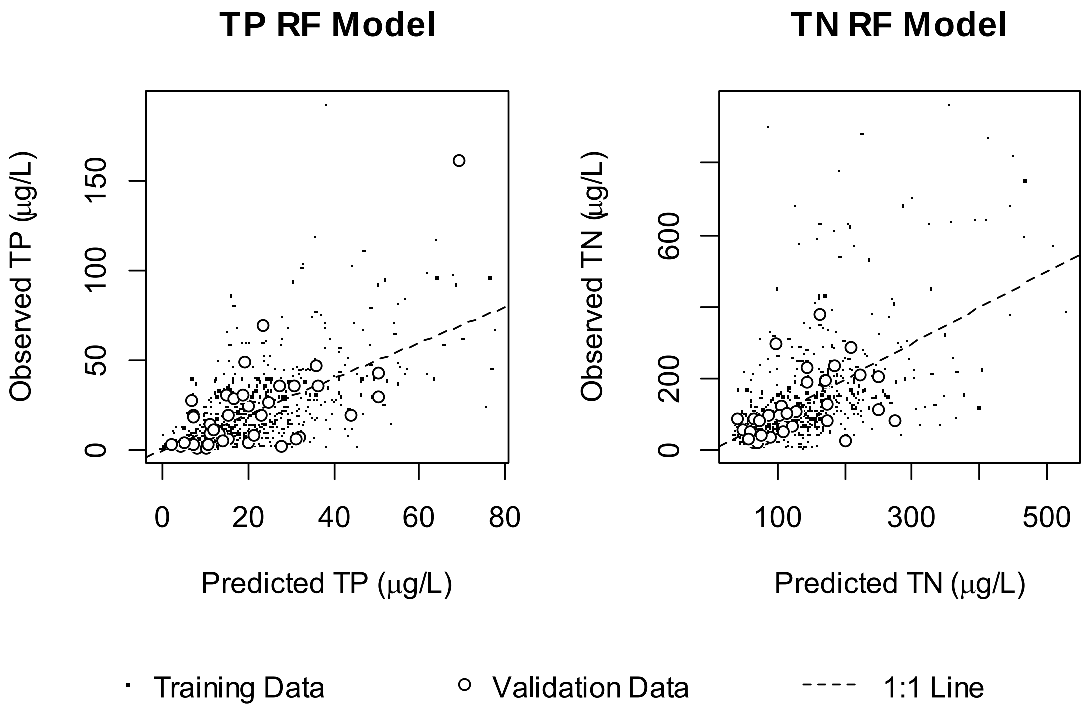


Figure 3-2. Plots of TP and TN observed vs. predicted values for both training and validation data.

result is heavily influenced by the single validation observation above 100 µg/L. This slope > 1 indicates that the model increasingly overpredicted with increasing TP concentrations. The equivalence test for slope showed the opposite pattern for the TN model, with 64% of the bootstrap estimates of slope falling below the region of equivalence and a smaller slope (0.66), indicating under-predictions at higher concentrations. Both models explained much more variance than did predictions based on the SPARROW model (Table 3-4) with RMSEs 25% lower than those for the SPARROW model.

Although our models had relatively low r^2 values, the results of our S:N analysis indicated that both models explained a large proportion of the static spatial variation

Table 3-5. Assessment of Signal to Noise (S:N) Ratio

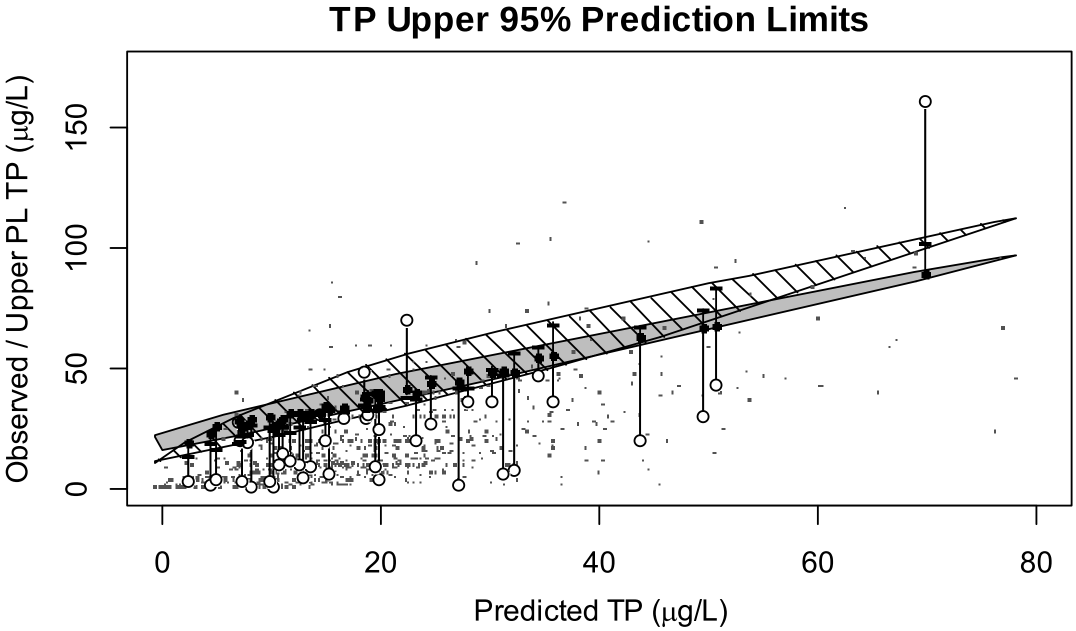
Model	Var_{sites} (Signal)	Var_{reps} (Noise)	S:N	Max $r^{2\ a}$
TP	438	520	0.84	0.46
TN	19175	12155	1.58	0.61

[a.] Highest possible r^2 value for a given S:N ratio calculated as S:N/(S:N + 1).

(Table 3-5). The TP model accounted for 87% of the static spatial variation in concentrations, i.e., the model explained 40% of the observed variation compared to a maximum possible of 46%. The TN model accounted for 53% of the spatial variation. The remaining unexplained variation is either due to temporal variation or measurement error.

Determining the highest probable concentration based on model predictions

The SEE and UNEEC methods produced similar upper PLs (Figure 3-3). Each method produced site-specific upper PLs, as opposed to a single line produced by distribution-based methods. For visual clarity, we plotted the envelopes containing individual upper PLs of training sites instead of the cloud of individual upper PLs themselves. Both methods identified identical numbers of training and validation sites to be greater than their upper PL (Table 3-6). Prediction interval coverage probabilities (PICPs, the probability that all observed values fit within their prediction limits) calculated from validation data indicated that 90% and 94% of predictions were within the prediction limits for both TP and TN, respectively, for both methods. Ideally the PICP would equal the selected prediction limit of 95%. The TN model identified approximately the correct number of sites as above the upper PL, but upper PLs for the TP model were conservative, identifying more sites above the limit than expected.



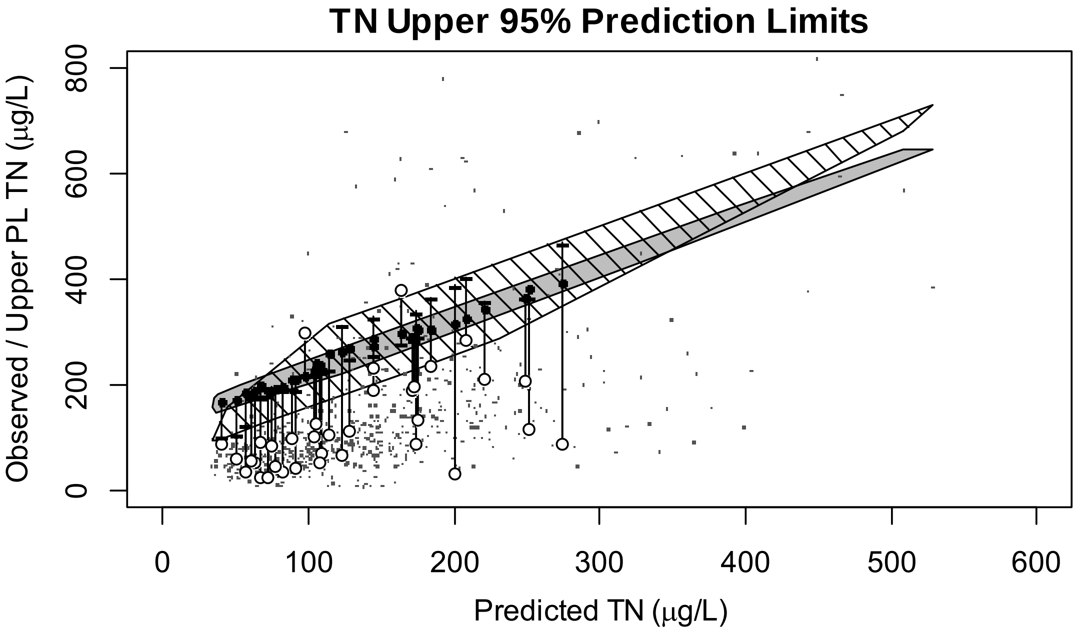


Figure 3-3. Plots of TP and TN observed vs. predicted values and upper prediction limits for both training and validation data. Observations are plotted as grey dots (training data) or open circles (validation data). Regions containing upper PLs for training data are plotted as filled grey (SEE method) or cross-hatch (UNEEC method). Site-specific upper PLs for validation data are plotted as filled circles (SEE method) or bars (UNEEC method).

Table 3-6. Performance of upper PLs

	SEE Method				UNEEC Method			
	Training Data		Validation Data		Training Data		Validation Data	
Model	# over upper PL[a]	PICP[b]	# over upper PL[c]	PICP	# over upper PL[a]	PICP[b]	# over upper PL[c]	PICP
TP	68	91%	4	90%	68	91%	4	90%
TN	61	91%	2	94%	61	91%	2	94%

[a] n for TP training data is 752 and for TN training data is 665.

[b] PICP = Prediction Interval Coverage Probability (Sherstha and Solomatine 2008).

[c] n for TP validation data is 40 and for TN validation data is 35.

Although both SEE and UNEEC identified the same number of sites as having concentrations greater than the upper PL, the specific sites identified as being over their PL varied between methods. For predicted high concentrations, the UNEEC method's upper PLs were larger than PLs produced by the SEE method, and the reverse was true for smaller predicted concentrations. This pattern occurred because of the heteroscedasticity in model errors (seen in Figure 3-2), where larger predictions were made with larger errors. The SEE method applies the same error to all predictions, so therefore does not account for heteroscedasticity in model errors.

Discussion

Model performance

Our results showed that spatial variation in natural background TP and TN concentrations can be accurately predicted from geographic data, albeit not as precisely as we would like. We consider our models to be accurate because the TN model exhibited no consistent bias and the bias of the TP model was less than 2% of the range of natural variation in TP concentration among our sites. Model predictions are generally applicable across the study area, as demonstrated by the low RMSEs at validation sites for both models. Because geothermal inputs can greatly affect nutrient concentrations, streams with significant geothermal inputs are the major exception to the generality of our predictions. The concordance of the observed relationships between predictors and nutrient concentrations with known mechanisms influencing TP and TN concentrations in streams further increases our confidence in the robustness of model predictions. The fact that the models accounted for a majority (87% for TP, 53% for TN) of the spatial variation in TP and TN concentrations indicates that the models were highly successful in capturing site-specific differences in reference conditions. We consider these models to be primarily spatial because the one or two predictors with temporal components (i.e., previous year’s precipitation in TP model, and day of year and prior 2 months precipitation in TN model) were of only moderate importance in either model.

Model predictions based on measures of continuously varying environmental factors also clearly outperformed the SPARROW model predictions that are based on regional predictors, runoff, and in-steam losses (Smith et al. 2003). Although the SPARROW model we used relies on ecoregions to control for spatial variation in nutrient sources, newer versions of the SPARROW model (Wise and Johnson 2011, Garcia et al. 2011) have begun to directly account for variation in natural sources of nutrients. These new SPARROW models include P concentrations in stream sediment, a proxy for P concentrations of underlying geology, and distributions of N-fixing *Alnus rubra* as natural sources of nutrients. These later models may predict natural background concentrations better than the Smith et al. (2003) model, but because they only predict annual yields we could not compare their predictions with our results.

Predictors

Most of the relationships between environmental factors and nutrient concentrations matched expectations based on previous studies, but relationships between nutrient concentrations and relative humidity, Ca deposition, EVI, precipitation, and SO_4 deposition were not as clearly related to known mechanisms. Increasing TP concentrations with decreasing humidity could be caused by evaporative concentration (Reddy et al. 1999). However, there is no reason to expect that atmospheric Ca deposition is directly linked to TP. Instead it is likely that the NADP measure of wet Ca deposition is correlated with dust deposition (Brahney 2012) and that this variable may be acting as a surrogate for the deposition of P in dust (Reynolds et al. 2001). Decreasing TP and TN concentrations with increasing EVI was expected due to increasing nutrient retention with increasing vegetation cover. However, this pattern only held for areas with lower EVI values associated with grasslands and scrub, and the opposite pattern occurred in areas with higher EVI values associated with forests (i.e., nutrient concentrations increased with increasing EVI). These increasing nutrient concentrations in forested areas could be attributed to lower nutrient retention by mature forest (Vitousek and Reiners 1975), built up litter fall from decades of fire suppression acting as a source of nutrients (Miller et al. 2005), or decreased microbial biomass resulting in lower P retention (Chen et al. 2003). Additional vegetation could also lead to increased rock weathering (as seen for other elements, see Chapter 2) which would release additional P.

The relationship between TN concentrations and precipitation also showed different directions of effect in different environments. TN concentrations declined with additional precipitation in xeric areas, but increased with additional precipitation in mesic ones. Although TN concentrations have been observed to be positively correlated with precipitation in mesic areas (e.g., Hill 1986, Vanderbilt et al. 2003) and negatively correlated with precipitation in xeric areas (e.g., Lewis and Grant 1979, Alvarez-Cobelas et al. 2010), these two patterns have not been observed in the same data set before. Increasing precipitation in mesic areas can lead to increased TN concentrations due to increased N fixation in wet soils (Cleveland et al. 1999), litter decomposition (Lewis et al. 1999), and flushing caused by greater stream/hill slope connectivity (Kane et al. 2008). Howarth et al. (2006) also proposed that increased precipitation results in shorter water residence times that limit the amount of contact between runoff and denitrifying organisms in the streambed. We suspect the negative relationship we observed between precipitation and TN concentrations in xeric areas is caused by water-dependent plant uptake. Greater precipitation in xeric areas may also create more anoxic zones in soils and thus increase denitrification (Bollmann and Conrad 1998). The relationship that is the least interpretable was the positive association between TN and atmospheric SO_4 deposition. This relationship is similar to the relationship seen by Cai et al. (2011) between stream NO_3 and atmospheric SO_4 deposition in streams in Great Smoky Mountains National Park. Although SO_4 deposition could have a direct effect on stream TN by suppressing plant growth and hence N uptake, it is also likely that SO_4 deposition is a surrogate for another process or N source such as dry deposition.

Volcanic rocks are a known source of P, but we were surprised at how important they were in predicting stream TP relative to measures of percent rock P. During model development, we created models without percent volcanic lithology as a predictor to assess its importance relative to measures of percent rock P. That model performed nearly as well as our TP model with volcanic lithology (r^2 of 0.37 vs. 0.40) and percent rock P became the most important predictor, indicating that most of the explanatory power of volcanic rocks is related to their P content. Faster weathering rates of volcanic rocks could explain the remaining difference in the importance between these two predictors. Another reason volcanic rock could have been a better predictor of TP relative to rock phosphorous is that our estimates of basalt P content may have been biased by applying median P rock content derived from global databases, which may not accurately reflect P values for basalts in the USA. However, the importance of the Gila Mountains/Mogollon Rim Ecoregion in predicting stream TP concentrations suggests a different explanation. Streams in the upper Gila Mountains/Mogollon Rim Ecoregion had an average TP concentration more than double the concentration seen in the rest of our study area (48 μg/L vs. 18 μg/L). The high TP concentrations in this region are likely due to the occurrence of large, recently active (within 1000-3000 years) basalt flows, which weather faster than older basalts (Gislason et al. 1996). The importance of volcanic rocks in predicting TP in western U.S. streams may be related to their relatively young age and fast weathering relative to other rock types.

Several environmental factors associated with nutrient concentrations in other studies were not selected as predictors in our models. Rock N and dry N deposition have both been shown to be sources of N (Holloway and Dahlgren 2002, Fenn et al. 2003), which increases TN concentrations in streams and lakes. Rock N content was positively related to stream TN in our data as observed elsewhere (Williard et al. 2005, Gardner and McGlynn 2009), indicating that rock N is a source. However, this relationship was weak and including it as a predictor did not improve model fit. Rock N may act as a significant source of stream TN only in specific circumstances where rock N content is high and readily weathered (e.g., Gardner and McGlynn 2009), such as in carbonaceous or oil shales. We also included estimates of dry N deposition derived from the CMAQ model in the TN model, but including these estimates slightly decreased model performance compared with models that included only wet N deposition (i.e., NADP data). This decrease in model performance with inclusion of dry N deposition estimates does not imply that dry deposition is not influencing stream TN, but rather any potential model improvement associated with the inclusion of dry deposition was swamped by errors in deposition estimates. CMAQ dry deposition estimates are based on emissions data instead of measured deposition as in the NADP data. Errors in deposition estimates could be caused by inaccurate emissions data, errors in the model estimating the distribution and amount of deposition, or both.

Factors associated with downstream nutrient losses and nutrient colimitation, both of which could potentially modify the amount of nutrients exported from catchments, were also not included in our models. Including catchment area in our models, which is related to travel time and stream size and is associated with nutrient loss (Prairie and Kaiff 1986, Smith et al. 2003), decreased performance of both the TP and TN models. The lack of a relationship with catchment area in our study area probably occurred for several reasons. First, previous estimates of in-stream loss rates are mostly from agricultural catchments (e.g., Alexander et al. 2000), which have larger loss rates than reference catchments (Prairie and Kaiff 1986, Mulholland et al. 2008). Greater uptake in streams flowing through agricultural catchments is probably caused by their higher nutrient concentrations, despite their lower uptake efficiencies (Mulholland et al. 2008). Second, although NH_4 uptake is positively related to stream size, the relationship between NO_3 uptake and stream size is much nosier (Tank et al. 2008). The noisy NO_3 – stream size relationship may obscure any effect that uptake of NH_4 by algae might have on TN concentrations because NO_3 concentrations are much higher than NH_4 concentrations. Third, surrogates for denitrification (i.e., ground water index) or streambed P adsorption or precipitation (i.e., Ca availability or channel slope) might have been more strongly associated with N and P removal because they are more direct surrogates of nutrient sinks than stream size. We also examined the possibility that P and N might be colimiting in streams as they are in lakes (Dodds et al. 2002). If N and P are colimiting, we would expect concentrations of one to be associated with concentrations of the other. For example, a P limited system would have lower N uptake and higher N export (and TN concentrations) at low P than at high P due to stoichiometric constraints on a stream's ability to use excess N. We assessed if potential interactions between TP and TN improved predictions of each nutrient by including each nutrient as a predictor of the other. TP (either measured or predicted) had no effect on the performance of the TN model, but including measured TN slightly improved the r^2 of the TP model (0.40 to 0.42). However, because the use of predicted TN did not improve the models and including measured TN as a predictor would prevent the application of these models to unmeasured locations, we elected not to include TN as a predictor in the final TP model.

Model shortcomings and possible improvements

Although the models made unbiased predictions of stream TP and TN concentrations in the western USA, these predictions could be potentially improved by addressing two model shortcomings. The first shortcoming of our models is their reliance on some predictors that can be altered by land use, which could potentially bias predictions of nutrient concentrations expected under natural conditions at altered sites. Vegetation predictors (e.g., EVI and % evergreen) may be especially problematic in this regard, but land use alteration could also alter soil properties (bulk density and SOC). Because these predictors had relatively low importance, these predictors could simply be dropped from the models. A better approach would be to replace these predictors with estimates of potential vegetation (e.g., Landfire Biophysiscal Settings Layer) or predicted natural soil properties (e.g., Malone et al. 2011). We did not pursue these options in this study because it was not clear a priori which vegetation and soil attributes would be important.

Another shortcoming of our models is their relatively coarse precision. The effect of model imprecision is to increase upper prediction limits, making criteria based upon these upper limits less protective than they would be if models were more precise. We attribute most of the poor model precision to temporal and measurement variation in grab sample concentrations that was unaccounted for by our models. A comparison of the variation explained by our models with that potentially associated with spatial differences among streams indicates the majority of unexplained variation is some combination of temporal and measurement error. Much of the unexplained temporal variation was probably associated with seasonal and yearly differences in runoff, flushing, freezing, or snowmelt. As models that characterize natural runoff and hydrologic regimes become available (e.g., Li et al. 2010), temporally and spatially explicit predictions of flow should enable better nutrient predictions (Helton et al. 2011). Also, some of the unexplained variation in nutrient concentrations may be due to differences in methods used to determine nutrient concentrations that occurred over time or between agencies. TN measurements before 1999 were almost 4-fold higher on average than measurements taken after 1999, resulting in a positive relationship between year of sample and TN model residuals. This decrease in measured TN concentrations might be partially due to the change from the Kjeldahl digestion method to persulfate oxidation and colorimetry method that occurred around this time. Patton and Kryskalla (2003) analyzed samples with both methods and observed that TN values obtained with persulfate oxidation and colorimetry were on average 15% lower than concentrations obtained with the Kjeldahl digestion method. It should be possible to improve model performance by limiting data to observations measured with a single method or adjusting concentrations to account for the method used if that information is known. We chose to retain these earlier samples in our data to maximize the number of environments represented in our model, but recommend that future work be based on TN estimates derived from a single method. Developing models based on long-term average concentrations or loads should eliminate much of the residual error associated with temporal variation in grab sample concentrations. However using long-term averages to establish criteria for all of the streams that need to be assessed is not practical because of costs associated with such long-term measurements. A better approach would be to focus on predicting temporal variation in the nutrient concentrations observed from grab samples. Models that could predict both spatial and temporal variation would provide a better basis for establishing criteria and can provide potentially important ecological information on the location and timing of natural nutrient fluctuations that influence primary producers (e.g., Butzler and Chase 2009).

Much of the remaining unexplained spatial variation is likely associated with some combination of natural and anthropogenic factors not included in our models. Natural factors that we did not consider include inputs from migrating fish (either excreted or from carcasses), the effect of flow modification by beaver dams, variation in uptake with spatial or temporal changes in stream metabolism, and natural disturbances that affect catchment or riparian vegetation (e.g., Houlton et al. 2003, Eshleman et al. 2004). MODIS-derived EVI could be used to detect vegetation disturbances, but model development and application would then be restricted to the last 10 years, the period for which MODIS observations are available. Development of models of stream gross primary production and respiration (e.g., Bernot et al. 2010) would allow us to incorporate these metabolic factors that control nutrient uptake and denitrification rates (Mulholland et al. 2008). Potential anthropogenic sources of unexplained spatial variation include either historical (e.g., logging) or highly localized land use (e.g., cabins with septic systems near creeks), that was not caught by our screening. Dry N deposition is another potentially important anthropogenic source, as is nutrient inputs delivered by dust (Ballantyne et al. 2011). As the measurement or prediction of dry N deposition and dust improves it should be possible to account for these inputs from national datasets like the NADP.

Developing nutrient criteria

Both the SEE and UNEEC methods appear suitable for establishing upper prediction limits. PLs produced by both methods were conservative, finding 1-5% more sites above their PL than expected from the chosen prediction interval (e.g., Prediction Interval Coverage Probabilities were 1 to 5% < the chosen prediction interval of 95%). However, complete agreement may be difficult to achieve given that other applications of the UNEEC method resulted in PICPs that deviated from desired prediction levels by 4-9% (Solomatine and Shrestha 2009, Malone et al. 2011). The UNEEC method better accounted for data heteroscedasticity, but this modest improvement required a much more complicated approach. UNEEC's more complicated method may make it more difficult for managers and stakeholders to understand. The UNEEC method also assumes that prediction error is different under the different natural environmental conditions identified in the clustering step (Shrestha and Solomatine 2008). Although this assumption may be a reasonable, it has not been rigorously tested. Choice of method will involve a tradeoff between the ability to potentially account for heteroscedasticity in prediction errors and ease in understanding how criteria are identified.

Concluding Remarks

Model-derived, site-specific criteria should better account for natural variation in nutrient concentrations than do regional criteria based on average regional conditions. As seen in other studies, observed nutrient concentrations for minimally altered reference sites varied over an order of magnitude within ecoregions (Figure 3-4). Comparing this variation with proposed regional criteria (horizontal lines in Figure 3-4) highlights the difficulty of establishing a single criterion protective of most streams without overprotecting some significant minority of streams. For example, the criteria proposed by Herlihy and Sifneos (2008) and Smith et al. (2003) for TP in nutrient ecoregion II (Western Forested Mountains, Figure 3-4A) would protect the majority of sites, but be overprotective of 25% of sites with naturally high TP concentrations. The site-specific criteria identified for TP in this ecoregion by our approach are generally higher than these regional criteria, but avoid being overprotective. Also, in approximately 15% of cases, the site-specific criteria would be more protective than the regional criteria. This same pattern of model-based upper PLs being higher than the Herlihy and Sifneos (2008) regional criteria also occurred for TN in nutrient ecoregion II. In nutrient ecoregion III (Xeric West), our site-specific criteria were generally higher than the Smith et al. (2003) regional criteria for TP and TN. However, our PL based site-specific criteria were generally lower than criteria developed from models by Dodds and Oakes (2004). The higher expected nutrient concentrations identified by Dodds and Oakes could have resulted from prediction error that occur when effects of land use are not fully captured in land use - nutrient models. Hill and Hawkins (in review) noted that stream temperature models developed from only reference site data predicted lower temperatures than did models built from data collected at both reference and non-reference sites that

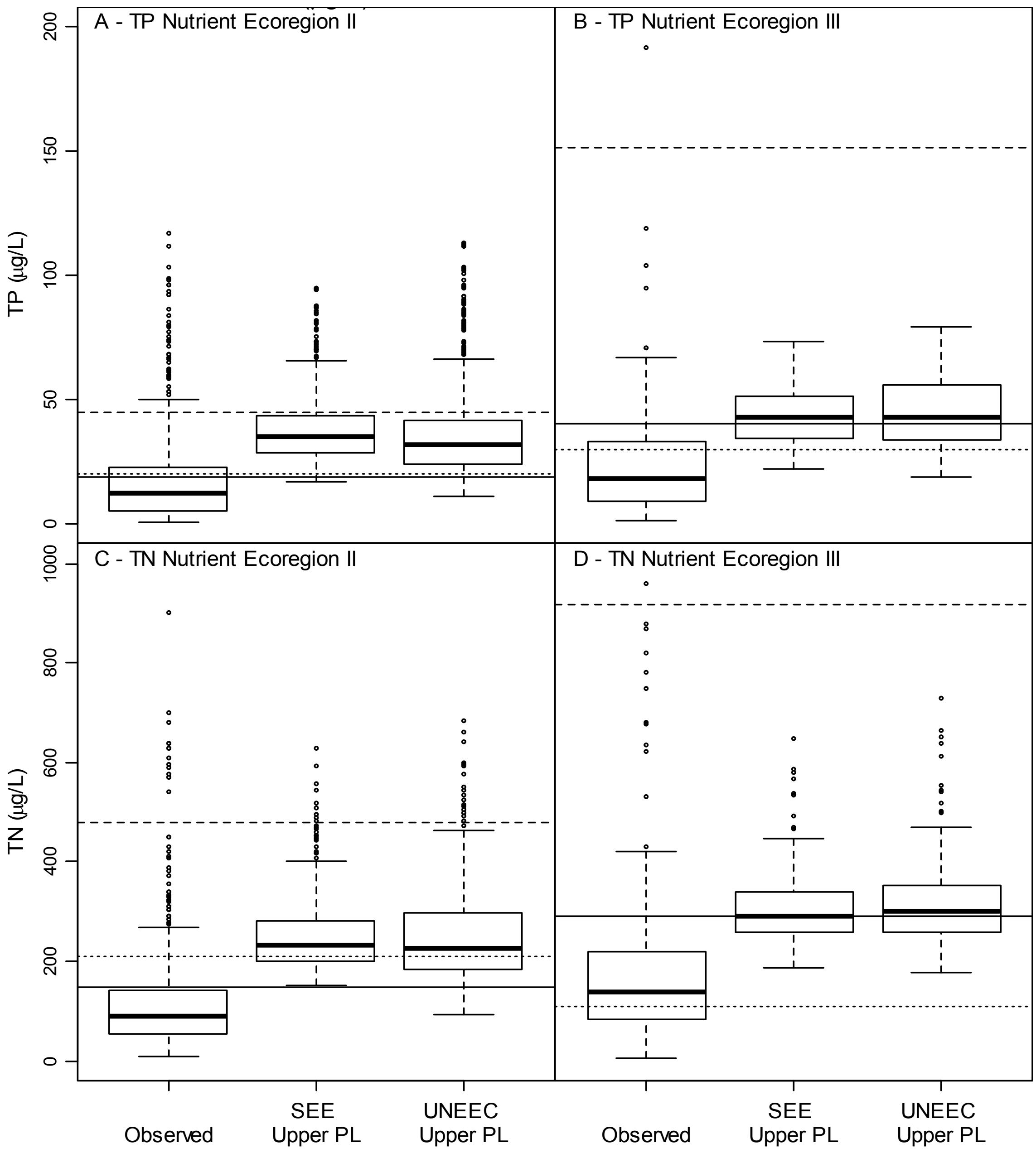


Figure 3-4. Comparison of observed concentrations and upper PLs for TP in Nutrient Ecoregion II Western Forested Mountains (A) and III Xeric West (B) and TN in Nutrient Ecoregions II (C) and III (D) with regional criteria from Herlihy and Sifneos (2008, solid lines), Dodds and Oakes (2004, dashed lines), and Smith et al. (2003, dotted lines). In all four cases, significant variation occurs within each region making any criterion identified over or under protective in many instances. Site-specific criteria based on upper PLs, although often higher than the regional criteria, better account for this observed variation.

statistically controlled for the effects of land use. In some cases model-based upper PLs agreed on average with proposed regional criteria (i.e., the Herlihy and Sifneos criterion in Figures 3-4B and D or the Smith et al. criterion in Figures 3-4C), but use of site-specific criteria would result in lower thresholds in about half the cases and a higher thresholds in the other half.

Establishing meaningful nutrient criteria for individual streams is challenging, but necessary for development and application of scientifically defensible and ecologically meaningful water quality standards. Model-based, site-specific criteria will protect streams with naturally low nutrient concentrations from eutrophication better than regional criteria that are based, in part, on data from streams with naturally high concentrations. Conversely, streams with naturally higher nutrient concentrations should not be held to a standard that is impossible to achieve. Making site-specific predictions across large regions might appear challenging, but models based on readily available geographic predictors can now be easily developed and applied within a GIS framework to produce spatially explicit maps of expected nutrient conditions. Similar site-specific predictions have been made of stream bed surface grain sizes across France (Snelder et al. 2011). As additional data describing the spatial and temporal factors affecting nutrient concentrations become available, models can be improved resulting in nutrient criteria that are even more reliable and protective.

References

Alexander, R. B., R. A.Smith, and G. E. Schwarz. 2000. Effect of stream channel size on the delivery of nitrogen to the Gulf of Mexico. Nature 403: 758-761.

Alvarez-Cobelas, M., R. Sanchez-Andres, S. Sanchez-Carrillo, and D. G. Angeler. 2010. Nutrient contents and export from streams in semiarid catchments of central Spain. Journal of Arid Environments 74:933-945.

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. Environmental Management 32:706-719.

Ballantyne, A. P., J. Brahney, D. Fernandez, C. L. Lawrence, J. Saros, and J. C. Neff. 2011. Biogeochemical response of alpine lakes to a recent increase in dust deposition in the Southwestern, US. Biogeosciences 8: 2689-2706.

Bernot, M. J., D. J. Sobota, R. O. Hall, P. J. Mulholland, W. K. Dodds, J. R. Webster, J. L. Tank, L. R. Ashkenas, L. W. Cooper, C. N. Dahm, S. V. Gregory, N. B. Grimm, S. K. Hamilton, S. L. Johnson, W. H. McDowell, J. L. Meyer, B. Peterson, G. C. Poole, H. M. Valett, C. Arango, J. J. Beaulieu, A. J. Burgin, C. a. Crenshaw, A. M. Helton, L. Johnson, J. Merriam, B. R. Niederlehner, J. M. O'Brien, J. D. Potter, R. W. Sheibley, S. M. Thomas, and K. Wilson. 2010. Inter-regional comparison of land-use effects on stream metabolism. Freshwater Biology 55:1874-1890.

Bhadha, J. H., and J. W. Jawitz. 2010. Characterizing deep soils from an impacted subtropical isolated wetland: implications for phosphorus storage. Journal of Soils and Sediments 10:514-525.

Bollmann, A., and R. Conrad. 1998. Influence of O_2 availability on NO and N_2O release by nitrification and denitrification in soils. Global Change Biology 4:387-396.

Brahney, J. 2012. The biogoechemical response of alpine lakes to changes in nutrient and dust deposition. Dissertation. University of Colorado, Boulder, Colorado.

Breiman, L. 2001. Random forests. Machine Learning 45:5-32.

Breiman, L., J. Friedman, R. Olshen, and C. Stone. 1984. Classification and regression trees. Wadsworth, Belmont, CA.

Butzler, J. M. and J. M. Chase. 2009. The effects of variable nutrient additions on a pond mesocosm community. Hydrobiologia 617:65-73.

Cai, M., J. S. Schwartz, R. B. Robinson, S. E. Moore, and M. A. Kulp. 2011. Long-term annual and seasonal patterns of acidic deposition and stream water quality in a Great Smoky Mountains high-elevation watershed. Water Air and Soil Pollution 219:547-562.

Chen, C. R., L. M. Condron, M. R. Davis, and R. R. Sherlock. 2003. Seasonal changes in soil phosphorus and associated microbial properties under adjacent grassland and forest in New Zealand. Forest Ecology and Management 177: 539-557.

Cheruvelil, K. S., P. A. Soranno, M. T. Bremigan, T. Wagner, and S. L. Martin. 2008. Grouping lakes for water quality assessment and monitoring: The roles of regionalization and spatial scale. Environmental Management 41:425-440.

Chetelat, J., and F. R. Pick. 2001. Temporal variability of water chemistry in flowing waters of the northeastern United States: does river size matter? Journal of the North American Benthological Society 20:331-346.

Chinnayakanahalli, K. 2006. The multi-watershed delineation tool. (Available from: http://hydrology.usu.edu/mwdtool/).

Chinnayakanahalli , K. J., C. P. Hawkins, D. G. Tarboton, and R. A. Hill. 2011. Natural flow regime, temperature and the composition and richness of invertebrate assemblages in streams of the western United States. Freshwater Biology 56:1248-1265.

Cleveland, C. C., A. R. Townsend, D. S. Schimel, H. Fisher, R. W. Howarth, L. O. Hedin, S. S. Perakis, E. F. Latty, J. C. Von Fischer, A. Elseroad, and M. F. Wasson. 1999. Global patterns of terrestrial biological nitrogen (N-2) fixation in natural ecosystems. Global Biogeochemical Cycles 13:623-645.

Commission for Environmental Cooperation (CEC). 2006. Ecological regions of North America: toward a common perspective. (Available from: http://www.epa.gov/wed/pages/ecoregions/na_eco.htm).

Compton, J. E., M. R. Church, S. T. Larned, and W. E. Hogsett. 2003. Nitrogen export from forested watersheds in the Oregon Coast Range: The role of N_2-fixing red alder. Ecosystems 6:773-785.

Cross, A. F., and W. H. Schlesinger. 2001. Biological and geochemical controls on phosphorus fractions in semiarid soils. Biogeochemistry 52:155-172.

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. Journal of Applied Meteorology 33:140-158.

D'Arcy, P., and R. Carignan. 1997. Influence of catchment topography on water chemistry in southeastern Quebec Shield lakes. Canadian Journal of Fisheries and Aquatic Sciences 54: 2215-2227.

Dillon, P. J., and W. B. Kirchner. 1975. Effects of geology and land-use on export of phosphorus from watersheds. Water Research 9:135-148.

Dillon, P. J., and L. A. Molot. 1997. Effect of landscape form on export of dissolved organic carbon, iron, and phosphorus from forested stream catchments. Water Resources Research 33: 2591-2600.

Dodds, W. K. 2007. Trophic state, eutrophication and nutrient criteria in streams. Trends in Ecology & Evolution 22:669-676.

Dodds, W. K., and R. M. Oakes. 2004. A technique for establishing reference nutrient concentrations across watersheds affected by humans. Limnology and Oceanography-Methods 2:333-341.

Dodds, W. K., V. H. Smith, and K. Lohman. 2002. Nitrogen and phosphorus relationships to benthic algal biomass in temperate streams. Canadian Journal of Fisheries and Aquatic Sciences 59:865-874.

Eshleman, K. N., D. A. Fiscus, N. M. Castro, J. R. Webb, and A. T. Herlihy. 2004. Regionalization of disturbance-induced nitrogen leakage from mid-Appalachian forests using a linear systems model. Hydrological Processes 18:2713-2725.

Fenn, M. E., J. S. Baron, E. B. Allen, H. M. Rueth, K. R. Nydick, L. Geiser, W. D. Bowman, J. O. Sickman, T. Meixner, D. W. Johnson, and P. Neitlich. 2003. Ecological effects of nitrogen deposition in the western United States. Bioscience 53:404-420.

Garcia, A. M., A. B. Hoos, and S. Terziotti. 2011. A regional modeling framework of phosphorus sources and transport in streams of the Southeastern United States. Journal of the American Water Resources Association 47:991-1010.

Gardner, K. K., and B. L. McGlynn. 2009. Seasonality in spatial variability and influence of land use/land cover and watershed characteristics on stream water nitrate concentrations in a developing watershed in the Rocky Mountain West. Water Resources Research 45:W08411.

Gislason, S. R., S. Arnorsson, and H. Armannsson. 1996. Chemical weathering of basalt in southwest Iceland: Effects of runoff, age of rocks and vegetative/glacial cover. American Journal of Science 296:837-907.

Global Soil Data Task Group. 2000. Global Gridded Surfaces of Selected Soil Characteristics (International Geosphere-Biosphere Programme - Data and Information System). Oak Ridge National Laboratory Distributed Active Archive Center. (Available from: http://www.daac.ornl.gov).

Green, M. B., and J. C. Finlay. 2010. Patterns of hydrologic control over stream water total nitrogen to total phosphorus ratios. Biogeochemistry 99:15-30.

Grimaldi, C., and V. Chaplot. 2000. Nitrate depletion during within-stream transport: Effects of exchange processes between streamwater, the hyporheic and riparian zones. Water Air and Soil Pollution 124:95-112.

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

Helton, A. M., G. C. Poole, J. L. Meyer, W. M. Wollheim, B. J. Peterson, P. J. Mulholland, E. S. Bernhardt, J. A. Stanford, C. Arango, L. R. Ashkenas, L. W. Cooper, W. K. Dodds, S. V. Gregory, R. O. Hall, S. K. Hamilton, S. L. Johnson, W. H.

McDowell, J. D. Potter, J. L. Tank, S. M. Thomas, H. M. Valett, J. R. Webster, and L. Zeglin. 2011. Thinking outside the channel: modeling nitrogen cycling in networked river ecosystems. Frontiers in Ecology and the Environment 9:229-238.

Herlihy, A. T., and J. C. Sifneos. 2008. Developing nutrient criteria and classification schemes for wadeable streams in the conterminous US. Journal of the North American Benthological Society 27:932-948.

Hill, A. R. 1986. Stream nitrate-N loads in relation to variations in annual and seasonal runoff regimes. Water Resources Bulletin 22:829-839.

Hill, B. H., F. H. McCormick, B. C. Harvey, S. L. Johnson, M. L. Warren, and C. M. Elonen. 2010. Microbial enzyme activity, nutrient uptake and nutrient limitation in forested streams. Freshwater Biology 55:1005-1019.

Hill, R. A., and C. P. Hawkins. In review. Predicting thermal reference conditions in western USA streams. Freshwater Science.

Holloway, J. M., and R. A. Dahlgren. 2002. Nitrogen in rock: Occurrences and biogeochemical implications. Global Biogeochemical Cycles 16:1118.

Homer, C., C. Q. Huang, L. M. Yang, B. Wylie, and M. Coan. 2004. Development of a 2001 National Land-Cover Database for the United States. Photogrammetric Engineering and Remote Sensing 70:829-840.

Houlton, B. Z., C. T. Driscoll, T. J. Fahey, G. E. Likens, P. M. Groffman, E. S. Bernhardt, and D. C. Buso. 2003. Nitrogen dynamics in ice storm-damaged forest ecosystems: Implications for nitrogen limitation theory. Ecosystems 6:431-443.

House, W. A. 2003. Geochemical cycling of phosphorus in rivers. Applied Geochemistry 18:739-748.

Howarth, R. W., D. P. Swaney, E. W. Boyer, R. Marino, N. Jaworski, and C. Goodale. 2006. The influence of climate on average nitrogen export from large watersheds in the Northeastern United States. Biogeochemistry 79:163-186.

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 Sensing of Environment 83:195-213.

Ice, G., and D. Binkley. 2003. Forest streamwater concentrations of nitrogen and phosphorus - A comparison with EPA's proposed water quality criteria. Journal of Forestry 101:21-28.

Kane, E. S., E. F. Betts, A. J. Burgin, H. M. Clilverd, C. L. Crenshaw, J. B. Fellman, I. H. Myers-Smith, J. A. O'Donnell, D. J. Sobota, W. J. Van Verseveld, and J. B. Jones. 2008. Precipitation control over inorganic nitrogen import-export budgets across watersheds: a synthesis of long-term ecological research. Ecohydrology 1:105-117.

Kaufmann, P. R., P. Levine, E. G. Robison, C. Seeliger, and D. V. Peck. 1999. Quantifying physical habitat in wadeable streams. EPA/620/R-99/003. USEPA Western Ecology Division, Corvallis, OR.

Kirschbaum, M. U. F. 2000. Will changes in soil organic carbon act as a positive or negative feedback on global warming? Biogeochemistry 48:21-51.

Knowlton, M. F., and J. R. Jones. 2006. Natural variability in lakes and reservoirs should be recognized in setting nutrient criteria. Lake and Reservoir Management 22:161-166.

LANDFIRE. 2011a. LANDFIRE 1.0.5 Biophysical settings layer. (Available from: http://landfire.cr.usgs.gov/viewer/).

LANDFIRE. 2011b. LANDFIRE 1.0.5 Vegetation Dynamics Model Descriptions. (Available from: http://www.landfire.gov/NationalProductDescriptions24.php).

LEMMA. 2011. Landscape, Ecology, Modeling, Mapping, and Analysis Project. (Available from: http://www.fsl.orst.edu/lemma/main.php?project=nwfp15&id= studyAreas).

Lewis, W. M., and M. C. Grant. 1979. Relationships between stream discharge and yield of dissolved substances from a Colorado mountain watershed. Soil Science 128:353-363.

Lewis, W. M., J. M. Melack, W. H. McDowell, M. McClain, and J. E. Richey. 1999. Nitrogen yields from undisturbed watersheds in the Americas. Biogeochemistry 46:149-162.

Li, M., Q. Shao, L. Zhang, and F. H. S. Chiew. 2010. A new regionalization approach and its application to predict flow duration curve in ungauged basins. Journal of Hydrology 389:137-145.

Liaw, A., and M. Wiener. 2009. Package randomForests. (Available from: http://cran.r-project.org/web/packages/randomForest/index.html).

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.

Lopez, E. S., I. Pardo, and N. Felpeto. 2001. Seasonal differences in green leaf breakdown and nutrient content of deciduous and evergreen tree species and grass in a granitic headwater stream. Hydrobiologia 464:51-61.

Ludington, S., B. C. Moring, R. J. Miller, P. A. Stone, A. A. Bookstrom, D. R. Bedford, J. G. Evans, G. A. Haxel, C. J. Nutt, K. S. Flyn, and M. J. Hopkin. 2007. Preliminary integrated geologic map databases for the United States Western States: California, Nevada, Arizona, Washington, Oregon, Idaho, and Utah. U.S. Geological Survey Open-File Report 2005-1305. (Available from: http://pubs.usgs.gov/of/2005/1305/).

Malone, B. P., A. B. McBratney, and B. Minasny. 2011. Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. Geoderma 160:614-626.

Marcarelli, A. M., and W. A. Wurtsbaugh. 2006. Temperature and nutrient supply interact to control nitrogen fixation in oligotrophic streams: An experimental examination. Limnology and Oceanography 51:2278-2289.

Meinshausen, N. 2006. Quantile regression forests. Journal of Machine Learning Research 7:983-999.

Meybeck, M. 1982. Carbon, nitrogen, and phosphorus transport by world rivers. American Journal of Science 282:401-450.

Miller, W. W., D. W. Johnson, C. Denton, P. S. J. Verburg, G. L. Dana, and R. F. Walker. 2005. Inconspicuous nutrient laden surface runoff from mature forest Sierran watersheds. Water Air and Soil Pollution 163:3-17.

Mulholland, P. J. 1992. Regulation of nutrient concentrations in a temperate forest stream - roles of upland, riparian, and instream processes. Limnology and Oceanography 37:1512-1526.

Mulholland, P. J., A. M. Helton, G. C. Poole, R. O. Hall, S. K. Hamilton, B. J. Peterson, J. L. Tank, L. R. Ashkenas, L. W. Cooper, C. N. Dahm, W. K. Dodds, S. E. G. Findlay, S. V. Gregory, N. B. Grimm, S. L. Johnson, W. H. McDowell, J. L. Meyer, H. M. Valett, J. R. Webster, C. P. Arango, J. J. Beaulieu, M. J. Bernot, A. J. Burgin, C. L. Crenshaw, L. T. Johnson, B. R. Niederlehner, J. M. O'Brien, J. D. Potter, R. W. Sheibley, D. J. Sobota, and S. M. Thomas. 2008. Stream denitrification across biomes and its response to anthropogenic nitrate loading. Nature 452:202-U246.

Natural Resources Conservation Service (NRCS). 2011. U.S. General Soil Map (STATSGO2). (Available from: http://soildatamart.nrcs.usda.gov).

Ohmann, J. L., M. J. Gregory, and T. A. Spies. 2007. Influence of environment, disturbance, and ownership on forest vegetation of Coastal Oregon. Ecological Applications 17:18-33.

Panagopoulos, I., M. Mimikou, and M. Kapetanaki. 2007. Estimation of nitrogen and phosphorus losses to surface water and groundwater through the implementation of the SWAT model for Norwegian soils. Journal of Soils and Sediments 7:223-231.

Park, J. H., M. J. Mitchell, P. J. McHale, S. F. Christopher, and T. P. Meyers. 2003. Impacts of changing climate and atmospheric deposition on N and S drainage losses from a forested watershed of the Adirondack Mountains, New York State. Global Change Biology 9:1602-1619.

Patton, C. J., and J. R. Kryskalla. 2003. Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory - Evaluation of alkaline persulfate digestion as an alternative to Kjeldahl digestion for determination of total and dissolved nitrogen and phosphorus in water. Water-Resources Investigations Report 03-4174. U. S. Geological Survey, Denver, Colorado.

Piñeiro, G., S. Perelman, J. P. Guerschman, and J. M. Paruelo. 2008. How to evaluate models: Observed vs. predicted or predicted vs. observed? Ecological Modelling 216:316-322.

Prairie, Y. T., and J. Kalff. 1986. Effect of catchment size on phosphorus export. Water Resources Bulletin 22:465-470.

Rampe, J. J., R. D. Jackson, and M. R. Sommerfeld. 1981. Physicochemistry of the upper Gila River watershed: I. San Francisco River and Clifton Hot Springs. Arizona-Nevada Academy of Science, Journal 16:1-6.

Reddy, K. R., R. H. Kadlec, E. Flaig, and P. M. Gale. 1999. Phosphorus retention in streams and wetlands: A review. Critical Reviews in Environmental Science and Technology 29:83-146.

Reynolds, R., J. Belnap, M. Reheis, P. Lamothe, and F. Luiszer. 2001. Aeolian dust in Colorado Plateau soils: Nutrient inputs and recent change in source. Proceedings of the National Academy of Sciences of the United States of America 98:7123-7127.

Robertson, D. M., D. A. Saad, and D. M. Heisey. 2006. A regional classification scheme for estimating reference water quality in streams using land-use-adjusted spatial regression-tree analysis. Environmental Management 37:209-229.

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 Physiology 25:903-913.

Schwede, D. B., R. L. Dennis, and M. A. Bitz. 2009. The watershed deposition tool: A tool for incorporating atmospheric deposition in water-quality analyses. Journal of the American Water Resources Association 45:973-985.

Shrestha, D. L., and D. P. Solomatine. 2008. Data-driven approaches for estimating uncertainty in rainfall-runoff modelling. International Journal of River Basin Management 6:109-122.

Smith, R. A., R. B. Alexander, and G. E. Schwarz. 2003. Natural background concentrations of nutrients in streams and rivers of the conterminous United States. Environmental Science & Technology 37:3039-3047.

Snelder, T. H., B. J. F. Biggs, and M. A. Weatherhead. 2004. Nutrient concentration criteria and characterization of patterns in trophic state for rivers in heterogeneous landscapes. Journal of the American Water Resources Association 40:1-13.

Snelder, T. H., N. Lamouroux, and H. Pella. 2011. Empirical modelling of large scale patterns in river bed surface grain size. Geomorphology 127:189-197.

Solomatine, D. P., and D. L. Shrestha. 2009. A novel method to estimate model uncertainty using machine learning techniques. Water Resources Research 45:W00B11.

Stevenson, F. J., and M. A. Cole. 1999. Cycles of soil: carbon, nitrogen, phosphorus, sulfur, micronutrients. 2nd edition. John Wiley and Sons, New York.

Stoddard, J. L., A. T. Herlihy, D. V. Peck, R. M. Hughes, T. R. Whittier, and E. Tarquinio. 2008. A process for creating multimetric indices for large-scale aquatic surveys. Journal of the North American Benthological Society 27:878-891.

Stoeser, D. B., 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. U.S. Geological Survey Open-File Report 2005-1351. (Available from: http://pubs.usgs.gov/of/2005/1351/).

Strobl, C., A. L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis. 2008. Conditional variable importance for random forests. BMC Bioinformatics 9:307.

Suplee, M. W., A. Varghese, and J. Cleland. 2007. Developing nutrient criteria for streams: An evaluation of the frequency distribution method. Journal of the American Water Resources Association 43:453-472.

Tank, J. L., E. J. Rosi-Marshall, M. A. Baker, and R. O. Hall. 2008. Are rivers just big streams? A pulse method to quantify nitrogen demand in a large river. Ecology 89:2935-2945.

Torbert, H. A., and C. W. Wood. 1992. Effects of soil compaction and water-filled pore-space on soil microbial activity and N losses. Communications in Soil Science and Plant Analysis 23:1321-1331.

US Environmental Protection Agency (USEPA). 2000. Nutrient criteria technical guidance manual, rivers and streams. EPA-822-B-00-002. USEPA Office of Water, Washington D.C.

US Environmental Protection Agency (USEPA). 2011. Working in partnership with states to address phosphorus and nitrogen pollution through use of a framework for state nutrient reductions. USEPA Office of Water, Washington D.C.

US Geologic Survey (USGS). 2006. National Hydrography Dataset (NHD) Medium Resolution. (Available from: http://nhd.usgs.gov/data.html).

Van Miegroet, H., I. F. Creed, N. S. Nicholas, D. G. Tarboton, K. L. Webster, J. Shubzda, B. Robinson, J. Smoot, D. W. Johnson, S. E. Lindberg, G. Lovett, S. Nodvin, and S. Moore. 2001. Is there synchronicity in nitrogen input and output fluxes at the Noland Divide Watershed, a small N-saturated forested catchment in the Great Smoky Mountains National Park? The Scientific World 1: 480-492.

Van Sickle, J. 2006. Relation between S/N ratio of a metric Y, and the ability of a stressor to explain that metric via linear regression. US Environmental Protection Agency.

Vanderbilt, K. L., K. Lajtha, and F. J. Swanson. 2003. Biogeochemistry of unpolluted forested watersheds in the Oregon Cascades: temporal patterns of precipitation and stream nitrogen fluxes. Biogeochemistry 62:87-117.

Vitousek, P. M., and W. A. Reiners. 1975. Ecosystem succession and nutrient retention - a hypothesis. Bioscience 25:376-381.

Walker, T. W., and J. K. Syers. 1976. Fate of phosphorus during pedogenesis. Geoderma 15:1-19.

Wang, G. P., J. S. Liu, H. Y. Zhao, J. D. Wang, and J. B. Yu. 2007. Phosphorus sorption by freeze-thaw treated wetland soils derived from a winter-cold zone (Sanjiang Plain, Northeast China). Geoderma 138:153-161.

Washburn, C. S. M., and M. A. Arthur. 2003. Spatial variability in soil nutrient availability in an oak-pine forest: potential effects of tree species. Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere 33:2321-2330.

Williard, K. W. J., D. R. Dewalle, and P. J. Edwards. 2005. Influence of bedrock geology and tree species composition on stream nitrate concentrations in mid-Appalachian forested watersheds. Water Air and Soil Pollution 160:55-76.

Wise, D. R., and H. M. Johnson. 2011. Surface-water nutrient conditions and sources in the United States Pacific Northwest. Journal of the American Water Resources Association 47:1110-1135.

Zhang, G., and Y. Lu. 2012. Bias-corrected random forests in regression. Journal of Applied Statistics 39:151-160.

CHAPTER 4

AN EXPERIMENTAL ASSESSMENT OF THE EFFECTS OF LOW TOTAL DISSOLVED SOLIDS ON THE SURVIVAL AND DISTRIBUTION OF STREAM MACROINVERTEBRATES*

Summary

1. Freshwater taxa must osmoregulate to maintain water and ion balances. If taxa differ in their ability to osmoregulate, variation among streams in total dissolved solids could influence the spatial distribution of taxa. Previous studies have largely focused on the effects of high total dissolved solids (TDS) on freshwater taxa, but the effect of low TDS on taxa distributions has been rarely investigated.

2. We used stream-side and laboratory flow-through microcosm experiments to assess the effects of low TDS (measured as electrical conductivity - EC) on three indices of fitness (survival, growth, and emergence) for 19 stream invertebrate taxa. We then tested the hypothesis that one or more fitness indices would predict the observed distribution of these taxa in nature.

3. In the stream-side experiment, we exposed 13 taxa to stream water with naturally low (< 25 μS/cm) and high (> 125μS/cm) EC for 83 days. In the laboratory experiment we exposed 16 taxa (10 of which were the same taxa used in the stream-side experiment) to low (<30 μS/cm) and high (>300 μS/cm) EC treatments for 55 days. Both experiments controlled for differences in habitat, temperature, food availability, and pH. We measured how survival, growth, and adult emergence responded to treatments.

4. Taxa survival varied from significantly higher survival in high EC (3 taxa), to no

---

* Coauthored by Charles P. Hawkins.

difference between treatments (15 taxa), to significantly higher survival in low EC treatments (1 taxon). Emergence was higher in low EC for 1 taxon, higher in high EC for one taxon, and similar in the two treatments for the remainder of the taxa. Growth rates differed between treatments for only one taxon.

5. The difference in survival between treatments predicted taxon EC optima derived from a previous field study ($r^2 = 0.60$, $p < 0.003$). Taxa with the greatest difference in survival between high and low treatments all had the highest EC optima, indicating that the inability to persist in low conductivity environments likely restricts the distributions of some stream macroinvertebrate taxa.

Introduction

The amount of solutes in stream water plays an important role in determining distributions of aquatic invertebrate taxa. For instance, Egglishaw and Morgan (1965) showed that streams in Scotland with relatively low total dissolved solids (TDS, <400 μeq/L of cations) had significantly lower abundances and taxa richness than streams with high TDS, with some taxa apparently restricted from lower TDS streams. Similar patterns of taxa distributions were seen among the tributaries of the River Duddon (Minshall & Kuehne 1969; Minshall & Minshall 1978). In the Duddon catchment, streams with low TDS (<245 μeq/L of cations) were dominated by Plecoptera with nearly no Ephemeroptera or *Gammarus*, whereas streams with higher TDS were dominated by Ephemeroptera, and *Gammarus* were common. Although early work examining the mechanism responsible for this pattern suggested these differences might be due to nutrient availability or pH (Egglishaw 1968; Sutcliffe & Carrick 1973), later work demonstrated that TDS was directly responsible for the pattern (Minshall & Minshall 1978; Willoughby & Mappin 1988), most likely because of osmoregulatory challenges posed by low TDS water. Because freshwater invertebrates must maintain higher internal concentrations of solutes than the medium they live in, they face physiological challenges in maintaining internal solute concentrations and volume against an osmotic gradient (Bradley 2009). As water becomes increasingly dilute, this challenge increases.

Most experimental studies examining the effect of TDS on aquatic biota have primarily focused on the effects of high TDS conditions (e.g., Kefford et al. 2004; 2007). However, some of these same studies show that low salinity also differentially affects survival, growth, or reproduction of some taxa (Hassell, Kefford & Nugegoda 2006; Kefford et al. 2007). These studies show inverted U shaped responses of multiple measures of organism performance to increasing TDS, as measured by electrical conductivity (EC). This pattern was seen for several taxa, including *Physa acuta* (Kefford & Nugegoda 2005), *Cloeon* spp., *Centroptilum* spp., *Chironomus* spp. (Hassell, Kefford & Nugegoda 2006), *Aedes aegypti* (Clark, Flis & Remold 2004), *Glyptophysa alicine*, *Glacidorbis* spp. (Kefford et al. 2007). With the exception of *Centroptilum*, survival, growth, and reproductive success increased with increasing EC over the natural range of EC found in most temperate streams (i.e., EC <1000 µS/cm), and began declining at much higher EC levels than found in most temperate streams (ranging from 1000 – 4000 µS/cm). The mayfly *Centroptilum* is an exception, increasing in survival and emergence up to 500 µS/cm and declining at levels higher than that. Some taxa were not affected by low ECs, including *Paragnetina media* (Kapoor 1979), *Dinotoperla thwaitesi*, *Anisocentropus* spp., and *Plectrocnemia* sp. (Kefford et al. 2007). These studies indicate that taxa differ in their ability to tolerate low EC conditions. Such differences may play an important role in structuring aquatic communities. We expect that some taxa specialize in inhabiting very dilute environments and have developed adaptations that increase their osmoregulatory abilities. Adaptations to dilute environments include a relatively impermeable waxy epicuticle, dilute urine production, and ion uptake structures (including chloride cells, chloride epithelia, rectal gills, and papillae). Not all taxa have developed these structures to the same level, hence taxa would be expected to exhibit a range of abilities to cope with low EC environments.

Differences in osmoregulatory ability could help explain observed distribution patterns, with some taxa excluded from low EC water and others inhabiting both high and low EC conditions. Taxa without strong osmoregulatory adaptations (i.e., soft bodied species or those with minimal ability to uptake ions such as some Chironomidae or Tipulidae) should be poor osmoregulators and should thus be restricted to higher EC environments. Other taxa with more impermeable integuments and ion uptake structures (e.g., Plecoptera) would better maintain ion and water balances, and be better able to live in extremely dilute environments. Differences in osmoregulatory abilities among taxa should be expected because of the trade-offs involved with the cost of developing strong osmoregulatory abilities. One such trade-off is the development of impermeable integuments that minimize water uptake and ion loss, but also reduce respiratory ability by restricting dissolved oxygen diffusion (Charmantier, Charmantier-Daures & Towle 2009). Another trade-off is the allocation of energy to ion uptake, at both the animal's surface and internally as part of the process of producing dilute urine, at the cost of other functions like growth and reproduction (Fiance 1978).

Our understanding of how aquatic taxa vary in their responses to osmotic conditions is limited, and this limitation makes it difficult to predict how taxa osmotic ability might interact with stream TDS conditions to impact taxa distributions. Most previous research on osmoregulation has focused on terrestrial insects, crustaceans, and mosquitoes (Bradley 2009). These intensive studies of specific taxa have greatly increased our understanding of the variety of osmoregulatory mechanisms. Much work has also been done on determining the upper salinity tolerances of a wide array of taxa, especially hyporegulators specializing in saline conditions. However, the response of only a few hyperregulators to low EC conditions has been examined, and these studies have often compared relatively extreme EC conditions (i.e., high TDS versus de-ionized water) in artificial settings (i.e., unfed animals in containers without flow). Low EC conditions (i.e., <100 μS/cm) are wide-spread among head-water streams (most commonly in mountains, but also in some low-land streams), and represent an important habitat. To establish how well different taxa tolerate low TDS conditions, and how these tolerances might relate to the distributions of aquatic macroinvertebrates, we need to quantify the responses of more taxa to low EC conditions while controlling for the effects of other factors (e.g., temperature, resources, and flow) that can also influence distributions.

Understanding taxa responses to different osmotic environments is important in increasing our understanding of the biology of freshwater taxa, but we also need to predict how taxa will respond to human caused changes in TDS. Examples of biota being directly threatened by alteration of TDS/EC include the effects of 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 2004). Because of the evolutionary tradeoffs associated with osmoregulatory adaptations, taxa that have specialized in living in dilute environments may be at a competitive disadvantage when EC increases and other taxa can then invade. Also, a better understanding of how taxa respond to spatial variation in water chemistry could enhance the accuracy of bioassessments by improving the predictions of taxa occurrences on which these assessments are based (Hawkins, Olson & Hill 2010).

Our goal was to address two related questions. Do taxa vary in their response to naturally occurring low TDS conditions? Does this variation help explain taxa distribution patterns we see in nature? Our approach was to measure how life-history end points important to taxa persistence (i.e., survival, growth, and adult emergence) respond to ecologically relevant differences in TDS/EC. The restriction from low TDS environments seen in some taxa distributions, the results of previous studies, and the differences in osmoregulatory ability that we discussed earlier all led us to three expectations. The first is that some taxa will exhibit lower survival, growth and emergence in low EC conditions than in high, presumably due to weaker osmoregulatory systems. Second, we expect that other taxa with stronger osmoregulatory systems will show equal survival, growth and emergence in low and high EC conditions. Third, we expect that observed differences in these responses to EC conditions among taxa will help explain distribution patterns observed for these taxa, with taxa exhibiting poorer survival, growth, or emergence in dilute environments being restricted from those environments. Although previous work has established a direct effect of low TDS on some aquatic invertebrates, additional work is needed to expand our understanding of how TDS affects distributions. We especially need to determine if the patterns seen over large EC ranges in Australia (i.e., Hassell, Kefford & Nugegoda 2006; Kefford et al. 2007) still hold for smaller ranges more relevant in temperate North America. Examining a broad array of taxa will allow us to better understand the diversity of taxa responses to EC. Also, by examining the effect of TDS under close to natural conditions, while controlling for other indirect effects associated with EC, we can test if correlations between distribution patterns and EC are, at least in part, directly caused by TDS.

Materials and Methods

General Approach

We used two different experimental designs that both used 600 mL flow-through microcosms, but differed in their realism. The first was a steam-side experiment with two naturally occurring EC conditions (<25 µS/cm vs. > 120 µS/cm) as treatments. The second was a laboratory experiment in which we manipulated EC to produce two EC treatments (<30 µS/cm vs. > 300 µS/cm). In both settings we controlled for temperature, food and habitat availability, and flow. Taxa were chosen to represent different apparent levels of adaptation to dilute environments. To assess whether differences in observed low TDS tolerances were related to taxa distributions, we then compared the results of these experiments with published EC optima.

Experimental Animals

We selected 19 experimental taxa (Table 4-1) based on their availability (collectible within a 2 hour drive of either the stream-side or laboratory experiment) and suitability to experimental conditions. We also considered the range of apparent EC preferences of each taxon based on survey data collected across the western USA to ensure taxa likely differed in their sensitivity to EC. We identified animals at the collection sites and transported them back to the experiment site in stream water. Animals were held in stream water (< 10˚ C) and then measured and placed in microcosms within 24 hours. For the stream-side experiment, we tracked growth by measuring body length (from labrum to end of abdomen) to the nearest 0.1 mm and then converting length to mass using published length to mass relationships. For the laboratory experiment, we measured wet weights to the nearest 0.1 mg after briefly blotting animals with filter paper. To track individual growth and minimize the potential for animal size to bias our

Table 4-1. Experimental animals, collection sites, and numbers

Taxa	Experiment	Collection Site*	Collection Site EC (μS/cm)	Organisms per microcosm	Total Number of Organisms
Callibaetis Eaton, 1881	Lab	Spring Hollow	280	7	84
Chloroperlidae	Stream-side	Timber	22	4-5	52
	Lab	Upper Blacksmith	307	5-6	66
Drunella coloradensis Dodds, 1923	Lab	Temple	297	6-7	80
Drunella doddsii Needham, 1927	Stream-side	Piermont	17	4	48
Drunella grandis Eaton, 1884	Stream-side	Negro	276	5-6	58
Hyalella azteca Saussure, 1858	Stream-side	Muncy	345	5	60
	Lab	Unnamed Spring	277	7	84
Hesperoperla pacifica Banks, 1900	Stream-side	Bassett	21	4-5	49
	Lab	Logan	308	4-6	57
Hydropsyche Pictet, 1834	Stream-side	Negro	276	5-6	63
	Lab	Logan	308	5-7	68
Hydroptila Dalman, 1819	Stream-side	Muncy	345	5-6	69
Isoperla Banks, 1906	Stream-side	Piermont	17	5	60
	Lab	Blue Pond	313	4	48
Leptophlebiidae	Stream-side	Timber	22	4-5	52
	Lab	Temple	297	4	48
Malenka Ricker, 1952	Stream-side	Negro	276	6	72
	Lab	Temple	297	3-4	42
Micrasema McLachlan, 1876	Stream-side	Kalamazoo	306	6	72
	Lab	Upper Blacksmith	307	5	60
Pagastia Oliver, 1959	Stream-side	Kalamazoo	306	3-4	42
	Lab	Blue Pond	313	5-6	65
Pteronarcella Banks, 1900	Lab	Upper Blacksmith	307	2-3	34
Pteronarcys Newman, 1838	Lab	Blacksmith	390	4-5	52
Rhyacophila Pictet, 1834	Stream-side	Piermont	17	4	48
	Lab	Logan	308	4-5	41
Skwala Ricker, 1943	Lab	Logan	308	4-5	47
Zapada Ricker, 1952	Lab	Upper Blacksmith	307	3	18

*Collection site coordinates : Bassett Ck - 39.442 N, 114.532 W, Blacksmith R - 41.624 N, 111.796 W, Blue Pond Spring- 42.104 N, 111.497 W, Kalamazoo Ck - 39.567 N, 114.589 W, Logan R - 41.746 N, 111.742 W, Muncy Ck - 39.603 N, 114.569 W, Negro Ck - 39.273 N, 114.310 W, Piermont Ck - 39.478 N, 114.586 W, Spring Hollow - 41.748 N, 111.715 W, Temple Fork - 41.829 N, 111.579 W, Timber Ck - 39.402 N, 114.612 W, Unnamed Spring - 41.758 N, 111.804 W, Upper Blacksmith R. - 41.609 N, 111.586 W results, we distributed animals among experimental units equally by size (i.e., placing equal number of large and small organisms in each microcosm). Organism densities ranged from 3 to 7 per microcosm (Table 4-1). All animals were provided biofilm conditioned rocks collected from the stream closest to each experimental site and fish food flakes (ad libitum) as food sources. This diet was supplemented with macroalgae (*Monostroma*) for herbivorous taxa and live *Tubifex* worms for predatory taxa. We inspected each microcosm daily and removed dead larva or emerged adults. Every 10-14 days, we removed microcosms from the experiment and replaced rocks with freshly collected conditioned rocks. At this time we also recorded the status of each animal. In the stream-side experiment we also re-measured body lengths at these times. Animals that pupated were left undisturbed until they emerged or the experiment was over. If pupae were attached to rocks, the rocks were not replaced. If dead bodies were discovered intact, they were remeasured.

Steam-Side Microcosm Experiment

We conducted stream-side microcosm experiments at the confluence of the two first-order tributaries of Piermont Creek located in an undeveloped portion of the Humboldt-Toiyabe National Forest in the Schell Creek Range of eastern Nevada (Fig. 4-1). We chose this location because the two tributaries are underlain by different geologies that produce a natural source of both low (<25 µS/cm) and high (> 120 µS/cm) EC water at their confluence. Although these streams differ greatly in EC and alkalinity, they have similar pH (Table 4-2). Water from both streams was diverted above the experimental site, filtered with a 100 µm screen to minimize colonization by other organisms, and then gravity fed to the experiment site through ¾-inch polyethylene pipe. Because the low EC stream was approximately 5˚ C cooler than the high EC stream, we equalized

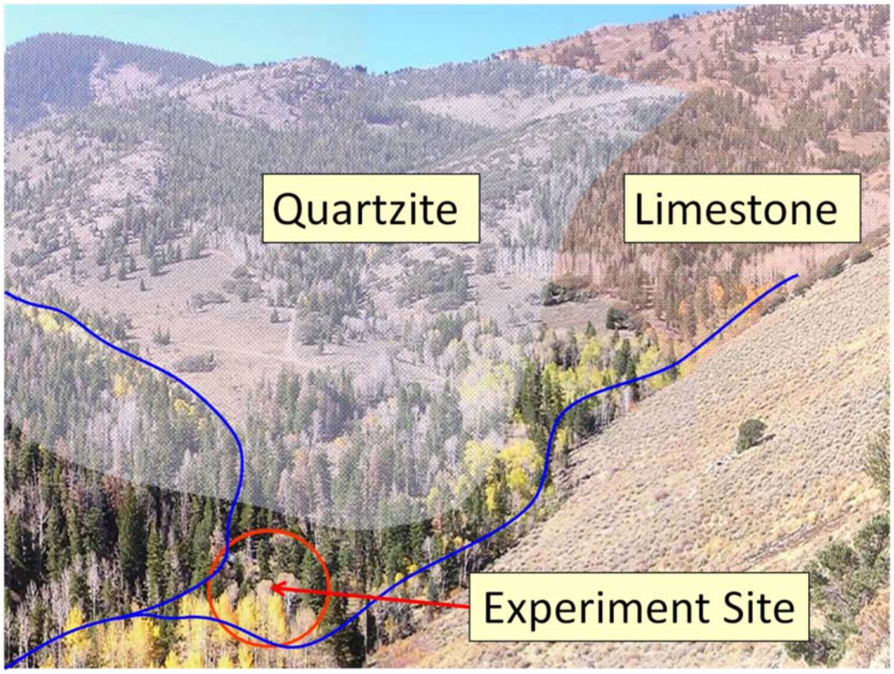


Fig. 4-1. Photograph of location of the stream-side experiment at Piermont Creek in the Humboldt-Toiyabe National Forest. The two first-order tributaries of Piermont Creek with their contrasting lithologies producing both low (<25 µS/cm) and high (> 120 µS/cm) EC water is illustrated. Photo by JRO.

temperatures between treatments by heating some of the cooler water from the low EC stream. Heating was done by creating a second diversion on the low EC tributary 300 m above the main diversion and solar heating this water by passing it through ¼-inch polyethylene hoses laid out on an exposed slope. This heated water was added to water from the main diversion in a header tank to raise its temperature to equal that of the northern tributary. We adjusted the flow of heated water daily as needed to keep the

Table 4-2. Lithology and water chemistry of tributaries of Piermont Creek, NV

	Southern Tributary	Northern Tributary
Dominant Lithology	Quartzite	Limestone
Specific Conductivity	< 25 µS/cm^2	> 120µS/cm^2
Alkalinity	20 mg/L $CaCO_3$	120 mg/L $CaCO_3$
pH	7.8	8.0

temperature of the two treatments equal. Experimental units were shaded to minimize heating of water as it passed through the experimental units. Temperatures were thus allowed to fluctuate naturally on a daily and seasonal basis (mean = 8˚ C, range: 1.1˚ - 17.4˚ C). Flow from the header tanks was set at 15 mL/s and monitored daily.

We used a flow through microcosm design to minimize changes in water chemistry and water quality resulting from animal feeding or excretion and to approximate natural conditions within the microcosms. We constructed microcosms from 600 mL plastic food storage containers, with the top of each container cut open and sealed with a screen to allow air flow into the microcosm. 1-mm screen was used, except for taxa with small adults for which we used 500 µm screens. We placed two stones (64 to 90 mm wide) in each microcosm as both a food source and substrate. Thirteen microcosms (one per taxon) were connected in series as a single experimental unit with each microcosm separated by a 500 µm screen (Fig. 4-2a). Except for a <1˚C increase in temperature along the length of a series, all other factors remained constant. We used twelve experimental units, grouped into six blocks of two experimental units each, with EC treatments assigned randomly to experimental units within each block (Fig. 4-2b). We arranged taxa in the same order in both experimental units within a block, and then systematically changed the order between blocks to achieve maximal interspersion of taxa. This interspersion ensured that across blocks all taxa were located equally often at the top and bottom of the series of microcosms. The stream-side experiment ran for 83 days (28 July to 20 October 2004).

Laboratory Microcosm Experiment

We used the same microcosms in the laboratory experiment as we used in the steam-side experiment, but changed how we created the two EC treatments and how microcosms were arranged (Fig. 4-3). For the high EC treatment, we used unchlorinated well water pumped on the Utah State University campus with water chemistry similar to the nearby Logan River (EC > 300 µS/cm and alkalinity > 4000 µeq/L $CaCO_3$). For the low EC treatment, we diluted well water with distilled water until it had an EC< 30 µS/cm. Water was circulated from a header tank, through all of the microcosms, into a 19L glass aquarium, and then back into the header tank. Water was lifted to the header tank by bubbling compressed air into the bottom of small tubes connecting the aquarium and the header tank, which also aerated the water. We adjusted flow rates by controlling the amount of compressed air released so that flow matched the same 15mL/s rate used in the stream-side experiment. Water temperatures were maintained at a constant 10˚ C by placing the aquaria in flow-through baths of 10˚ C well water. We monitored EC weekly, and added additional distilled or well water to the aquaria to maintain the EC difference. We also used twelve experimental units in this experiment, grouped into six blocks of two experimental units each, with treatments assigned randomly within each block. In this experiment, we arranged microcosms in parallel instead of in series (Fig. 4-3b) and kept taxa order the same within each block, but systematically changed the order between blocks to maximize spatial interspersion of taxa in the experiment. All experimental units were exposed to the same 16:8·h light:dark photoperiod. The laboratory experiment ran for 55 days (3 September to 21 November 2005).

Relationship between Taxa Response to EC Conditions and Distributions

To assess whether differences in response to EC conditions are related to distributions, we compared observed differences in survival between treatments with taxa EC optima derived from a field survey. We quantified taxa response to EC

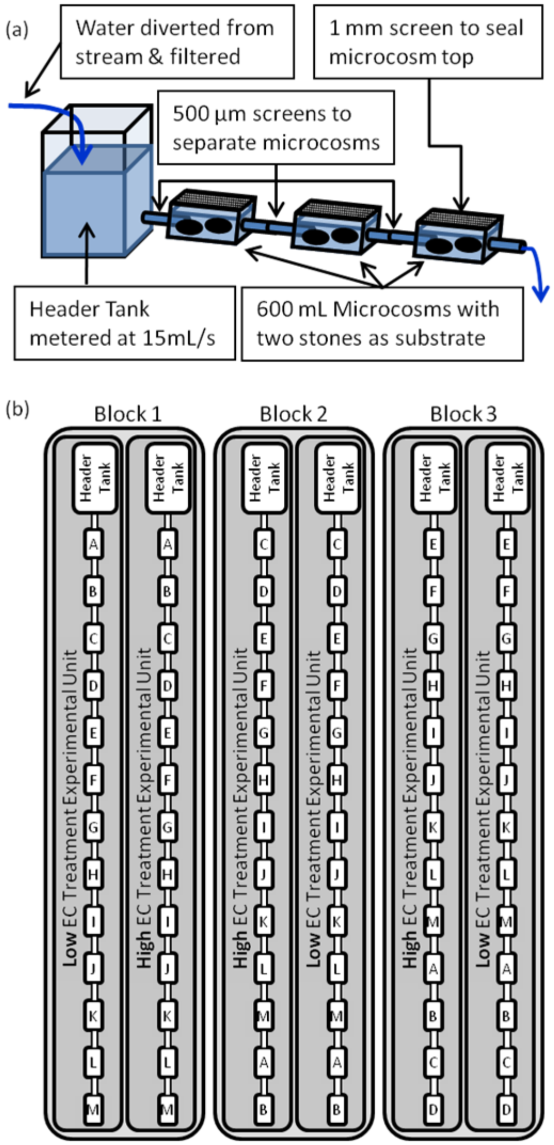


Fig. 4-2. Diagram of stream-side experiment design. (a) side view of single experimental unit with details and (b) plane view showing 6 of 12 experimental units. Letters A-M indicate micorcosms occupied by different taxa. Taxa order is rotated systematically between blocks and high and low EC treatments are assigned randomly within each block.

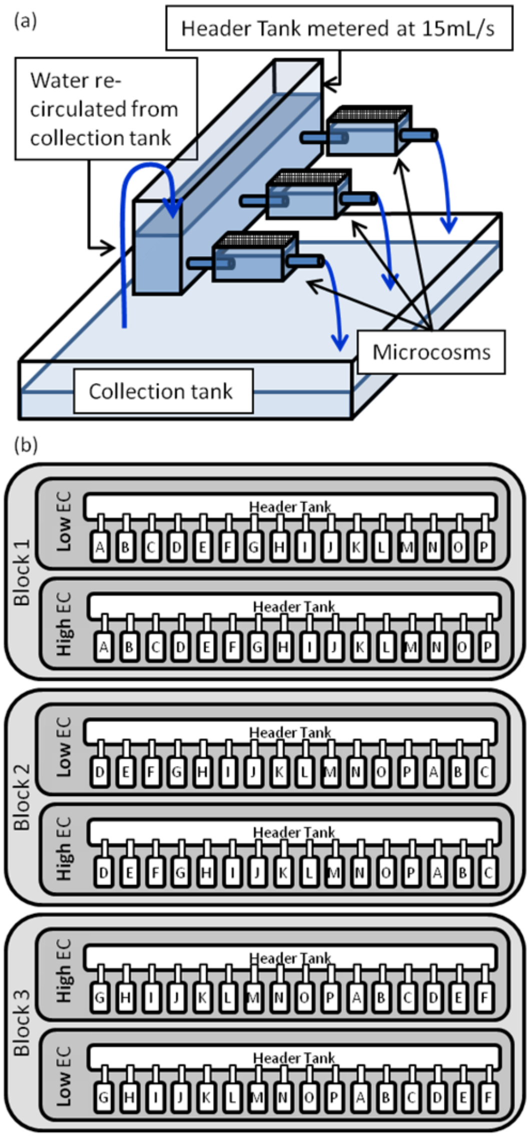


Fig. 4-3. Diagram of laboratory experiment design. (a) side view of single experimental unit with details and (b) plane view showing 6 of 12 experimental units. Letters A-P indicate micorcosms occupied by different taxa. Taxa order is rotated systematically between blocks and high and low EC treatments are assigned randomly within each block.

conditions as the difference in mean survival times in the high and low EC treatments. Optima calculated from field data provide a way of quantifying how taxa distributions vary with EC conditions (Blocksom & Winters 2006). We used the EC optima expressed as weighted average abundances provided by Black, Munn & Plotnikoff (2004). Black, Munn & Plotnikoff (2004) provided optima for taxa that best matched our experiment in terms of taxonomy and location. In instances where Black et al. did not list an optima matching a taxon we used, we either applied the optima listed for a coarser taxonomic resolution (i.e., family or higher), or used the optima of a closely related taxon. We were able to match optima with 17 of the taxa used in our experiments (all except *Pteronarcys* and *Pteronarcella*). We determined the direction and strength of this relationship by regressing EC optima against survival differences, i.e., we developed a model that predicted EC optima (distribution) from experimentally determined survivorship under low and high EC conditions.

Statistical Analysis

We used the area under Kaplan-Meier survival curves to determine mean survival times in days and then used this data to calculate differences in survival between high and low EC treatments. Because some taxa did not experience 50% mortality, medians could not be used. We used Mantel-Haenszel tests to test for significance in survival between treatments. We used the R package “Survival” to conduct survival analyses. We tested for differences in the percentage of each taxon that successfully emerge as adults in each treatment by applying the Fisher exact test to a two by two contingency table. We calculated growth rates both as the change in body length (mm/day, stream-side only) or mass (mg/day, both stream-side and laboratory) with time, and as specific growth (G, mg/(mg*day)) calculated as:

$$G = \frac{\ln(\mathrm{Mass}_{\mathrm{final}}/\mathrm{Mass}_{\mathrm{initial}})}{\text{time interval}} * 100$$

(Hawkins 1986). Where appropriate, data were $\log_{10}$ transformed to improve the assumptions of normality and homogeneity of variance. We used ANCOVA to test for significant differences in growth rates between treatments, with initial mass as the covariate to control for the effect of size on growth rates. We used the R package “stats” to perform emergence and growth analyses.

Results

Survival

The survival results met our expectations that some taxa would be sensitive to EC and others would be insensitive. The majority of taxa did not have significant differences in survival between treatments (Table 4-3), with 9 of these having less than 5 day difference in survival. Only 4 taxa had significantly different survival between the two treatments. Except for three taxa in the laboratory experiment, mean survival times ranged from 3 to 9 weeks in the streamside experiment and from 3 to 7 weeks in the laboratory experiment. In the laboratory experiment, *Drunella coloradensis* and *Pagastia* had mean survival times of less than a week because many individuals emerged as adults shortly after the start of the experiment. *Isoperla* had a mean survival of 10 days in the laboratory experiment, and all animals died within 19 days. *Drunella grandis*, *Hyalella azteca*, and *Malenka* all had significantly better survival in high EC conditions than in low EC conditions. *Hesperoperla pacifica* had significantly longer survival in low EC conditions, but only in the stream-side experiment. The difference in survival for *H. pacifica* in the laboratory experiment was less than 1 day. Of the 10 taxa that were used in both the stream-side and laboratory experiments, *H. pacifica* was also the only taxon with significantly different results between experiments.

Table 4-3. Mean survival (in days) in high and low EC treatments. Standard errors are in parentheses, and data shown in bold where significant at a 0.05 level

Taxa	Experiment	High EC Mean Survival (SE)	Low EC Mean Survival (SE)	High - Low Survival Diff	p-value
Callibaetis	Lab	43.3 (1.9)	40.3 (2.4)	3.0	0.41
Chloroperlidae	Stream-side	48.3 (5.0)	46.9 (5.8)	1.5	0.73
	Lab	28.2 (2.7)	29.7 (2.9)	-1.5	0.53
Drunella coloradensis	Lab	5.8 (0.4)	5.7 (0.3)	0.1	0.73
Drunella doddsii	Stream-side	34.8 (2.6)	32.8 (2.7)	2.0	0.58
Drunella grandis	**Stream-side**	**35.2 (4.7)**	**20.9 (1.8)**	**14.3**	**0.03**
Hyalella azteca	**Stream-side**	**62.0 (4.3)**	**40.0 (3.8)**	**22.0**	**0.00**
	Lab	**35.5 (2.1)**	**17.5 (1.8)**	**18.0**	**0.00**
Hesperoperla pacifica	**Stream-side**	**26.2 (4.1)**	**42.9 (6.3)**	**-16.7**	**0.04**
	Lab	30.4 (2.6)	29.7 (2.8)	0.7	0.88
Hydropsyche	Stream-side	31.5 (4.6)	29.8 (3.6)	1.7	0.53
	Lab	21.0 (2.6)	27.2 (2.6)	-6.2	0.11
Hydroptila	Stream-side	24.7 (2.1)	28.2 (2.4)	-3.5	0.56
Isoperla	Stream-side	26.3 (2.8)	22.8 (2.5)	3.5	0.53
	Lab	10.0 (1.2)	10.4 (1.3)	-0.3	0.86
Leptophlebiidae	Stream-side	44.0 (6.3)	48.3 (5.7)	-4.3	0.80
	Lab	22.3 (3.0)	25.3 (2.8)	-3.0	0.49
Malenka	**Stream-side**	**62.2 (4.3)**	**52.1 (4.6)**	**10.1**	**0.05**
	Lab	**23.3 (3.2)**	**14.9 (2.6)**	**8.5**	**0.05**
Micrasema	Stream-side	45.5 (4.4)	37.2 (4.0)	8.2	0.15
	Lab	35.9 (2.6)	38.9 (2.2)	-3.1	0.23
Pagastia	Stream-side	46.3 (5.8)	38.7 (5.9)	7.6	0.23
	Lab	14.6 (1.2)	14.3 (1.1)	0.3	0.92
Pteronarcella	Lab	41.8 (2.4)	43.7 (2.5)	-1.9	0.29
Pteronarcys	Lab	49.0 (3.1)	42.7 (3.8)	6.3	0.21
Rhyacophila	Stream-side	39.6 (5.4)	47.0 (4.7)	-7.4	0.21
	Lab	23.3 (4.0)	22.1 (3.4)	1.2	0.80
Skwala	Lab	27.0 (3.3)	23.5 (3.3)	3.5	0.24
Zapada	Lab	41.4 (1.5)	30.8 (5.9)	10.7	0.21

Adult Emergence

The emergence results also agreed with our expectations that emergence success of taxa would vary in response to the treatments. The only significant differences in emergence were seen in the stream-side experiment. More *Micrasema* emerged in the high EC treatment and more *H. pacifica* emerged in the low EC treatment (Table 4-4). However, differences in emergence for these two taxa were not significantly different between treatments in the laboratory experiment. Emergence of the other 16 taxa with terrestrial adult stages was not significantly different between treatments. Emergence varied from a high of 66% (*D. coloradensis*) to no emergence, with < 2% of individuals emerging in 6 taxa (i.e., *Callibaetis,* Chloroperlidae*, Hydropsyche, Pteronarcella, Pteronarcys*, and *Skwala*). Across treatments, we observed almost twice the proportion of emergence in the stream-side experiment (17%) than in the laboratory experiment (9%), probably as a consequence of the length and timing of the experiments.

Growth

None of the taxa tested had significantly greater growth in the high EC treatment than the low EC treatment, and only *Drunella doddsii* had significantly greater growth in the low EC treatment than the high EC treatment (Table 4-5). Contrary to our expectations that less efficient osmoregulators would have slower growth in low EC treatments, 10 of the 19 taxa showed greater growth in low EC conditions, although these differences were not significant at $p < 0.05$. These 10 taxa with greater growth in low EC conditions included 3 of the 4 taxa that had greater survival or emergence in high EC conditions (*D. grandis, H. azteca*, and *Micrasema*).

Table 4-4. Emergence in high and low EC treatments. Data shown in bold where significant at a 0.1 level

Taxa	Experiment	High EC % Emergence	Low EC % Emergence	p-value
Callibaetis	Lab	0	2	1.00
Chloroperlidae	Stream-side	0	0	1.00
	Lab	0	3	1.00
Drunella coloradensis	Lab	66	62	0.82
Drunella doddsii	Stream-side	46	42	1.00
Drunella grandis	Stream-side	28	17	0.53
Hesperoperla pacifica	**Stream-side**	**0**	**20**	**0.05**
	Lab	0	0	1.00
Hydropsyche	Stream-side	0	0	1.00
	Lab	0	0	1.00
Hydroptila	Stream-side	12	29	0.13
Isoperla	Stream-side	27	20	0.76
	Lab	4	8	1.00
Leptophlebiidae	Stream-side	4	11	0.61
	Lab	0	04	1.00
Malenka	Stream-side	39	19	0.12
	Lab	0	5	1.00
Micrasema	**Stream-side**	**47**	**25**	**0.08**
	Lab	0	3	1.00
Pagastia	Stream-side	24	10	0.41
	Lab	41	33	0.61
Pteronarcella	Lab	0	0	1.00
Pteronarcys	Lab	0	0	1.00
Rhyacophila	Stream-side	0	0	1.00
	Lab	0	0	1.00
Skwala	Lab	0	0	1.00
Zapada	Lab	22	44	0.62

Relationship between Taxa Response to EC Conditions and Distributions

EC optima were significantly related to survival differences between high and low EC conditions for both the stream-side experiment (p=0.0028) and the laboratory experiment

Table 4-5. Growth rates in high and low EC treatments. Data shown in bold where significant at a 0.05 level. Growth measured as change in length (mm/day), mass (mg/day), and specific growth (mg/mg*day)

		Mean Growth Rate (Length)			Mean Growth Rate (Mass)			Mean Specific Growth Rate		
Taxa	Experiment	p-value	High EC	Low EC	p-value	High EC	Low EC	p-value	High EC	Low EC
Callibaetis	Lab				0.18	0.1	0.1	0.7	1.75	1.8
Chloroperlidae	Stream-side	0.13	0.02	0	0.26	0.01	0	0.09	0.82	1.3
	Lab				0.89	0.05	0.1	0.91	1.67	1.61
Drunella coloradensis	Lab				0.79	1.1	1.4	0.85	2.15	2.58
Drunella doddsii	**Stream-side**	**0.04**	**0.01**	**0**	**0.05**	**0.03**	**0.1**	**0.04**	**0.51**	**1.29**
Drunella grandis	Stream-side	0.12	0.02	0	0.08	0.04	0.1	0.15	0.46	0.93
Hyalella azteca	Stream-side	0.54	0.02	0	0.74	0.01	0	0.44	1.32	1.51
	Lab				0.52	0.06	0.1	0.45	1.23	1.67
Hesperoperla pacifica	Stream-side	0.11	0.09	0	0.41	0.28	0.2	0.11	1.97	0.66
	Lab				0.36	1.32	0.2	0.51	0.66	0.38
Hydropsyche	Stream-side	0.37	0.02	0	0.45	0.03	0	0.32	0.65	0.46
	Lab				0.18	1.01	0.3	0.19	1.36	0.29
Hydroptila	Stream-side	0.24	0	0	0.31	0	0	0.2	0.15	-0.1
Isoperla	Stream-side	0.76	0.04	0.1	0.87	0.04	0.1	0.69	0.88	1.38
	Lab				0.68	0.84	0.7	0.56	3.14	2.56
Leptophlebiidae	Stream-side	0.45	0.03	0	0.29	0.01	0	0.61	2	1.84
	Lab				0.84	0.04	0	0.54	1.27	1.75
Malenka	Stream-side	0.96	0.03	0	0.73	0.01	0	0.78	2.07	2.01
	Lab				0.12	0.05	0	0.56	2.15	1.45
Micrasema	Stream-side	0.61	0.01	0	0.5	0	0	0.72	0.53	0.81
	Lab				0.15	0.03	0	0.06	0.94	1.62
Pagastia	Stream-side	0.39	0.01	0	0.54	0	0	0.34	0.08	0.49
	Lab				0.93	0.43	0.4	0.8	2.96	3.2
Pteronarcella	Lab				0.57	0.24	0.2	0.56	1.4	1.6
Pteronarcys	Lab				0.76	1.48	1.9	0.87	1.01	1.12
Rhyacophila	Stream-side	0.14	-0	0.1	0.1	-0	0.1	0.21	-0.1	1.39
	Lab				0.39	0.34	0.1	0.45	2.05	0.8
Skwala	Lab				0.35	0.24	0.1	0.11	0.79	0
Zapada	Lab				0.73	0.03	0	0.7	0.57	0.97

(p=0.0017) (Fig. 4-4). This relationship was positive, with taxa having longer survival times in high EC conditions also having higher EC optima. Survival differences from the stream-side experiment explained 60% of the variability in EC optima, whereas the survival differences from the laboratory experiment explained 48%. We omitted *D. grandis* from this analysis as an outlier because the EC optima used for this taxon was one developed for the family Ephemerellidae, which is unlikely to adequately describe the distribution of this species given the extensive ecological diversity within this family (Hawkins 1984; 1985; 1986). Taxa that had significant differences in survival (e.g.,

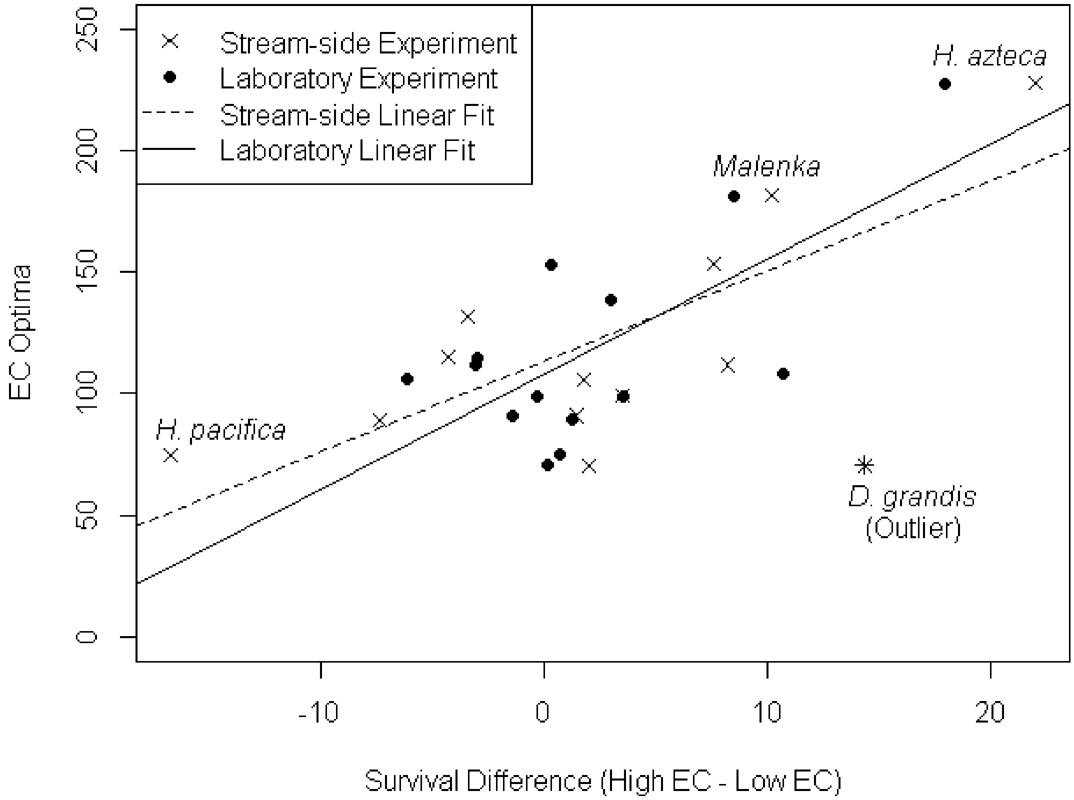


Fig. 4-4. Relationship between survival differences in high and low EC conditions and EC optima derived from field surveys by Black, Munn & Plotnikoff (2004).

Malenka and *H. azteca*) also had the greatest EC optima, indicating that these taxa seldom occur in low EC environments.

Discussion

The weight of evidence supports the hypothesis that the fitness of stream invertebrate taxa is affected by TDS and that these differences in fitness affect spatial distributions. Survival experiments by Willoughby & Mappin (1988) found that *Ephemerella ignita* and *Amphinemura sulcicollis* were tolerant of low TDS water, but *Baetis muticus* and *Beatis rhodani* were not. Kefford et al. (2007) reviewed previous studies examining salinity effects on invertebrates and found that of the 11 taxa that had been tested for sensitivity to low TDS conditions, 3 showed detrimental effects of low TDS but the remainder were unaffected. Our tests on 19 additional taxa are consistent with these earlier studies, and 25% of the taxa we tested exhibited significant differences in either survival or emergence between our two treatments. Because we controlled for differences in temperature, habitat, and food sources, we conclude that these differences in survival and emergence were due to some direct effect of differences in EC. By doubling the number of taxa investigated and examining responses over commonly observed EC differences, our tests show it is highly likely that ecologically significant variation in fitness occurs among taxa exposed to low to moderate levels of TDS.

Does this variation in taxa response to low TDS conditions help explain taxa distribution patterns we see in nature? Experimental results are often interpreted as explaining distributions, but rarely are taxa responses to TDS directly related to their distributions. Willoughby & Mappin (1988) did select test taxa and experimental conditions to directly test whether the observed distributions of these taxa in the River

Duddon corresponded to their survival when exposed to different TDS conditions found in the catchment. They found that the responses to TDS were consistent with their observations from the field for 3 taxa (i.e., taxa not found in low TDS conditions had poorer survival in low TDS conditions and taxa found in low TDS conditions had equal or better survival in low TDS conditions). One of their taxa not found in low TDS conditions (*E. ignita*) had lower survival in low TDS conditions (consistent with expectations), but the authors concluded the differences were not significant. Our comparison of observed survival differences in high and low EC conditions and field derived EC optimum revealed a fairly strong correlation between them. Taxa that were sensitive to low EC treatments in our experiments had higher EC optima indicating they are found primarily in high EC conditions. This correlation supports the notion that differences in the ability of taxa to persist under low TDC conditions partly determine distributions of taxa. Some of the unexplained variation is probably associated with the level of taxonomic resolution we had to use. It is unlikely that all of the species used in our experiments matched those collected by Black, Munn & Plotnikoff (2004), and differences in optima among species within the same genus would have contributed unexplained variance in our analyses.

In contrast with survivorship and emergence, none of the taxa we tested had faster growth in high EC conditions than low. These results are consistent with previous work that examined differences in growth between animals exposed to low and high EC treatments (Eggert & Burton 1994; Hassell, Kefford & Nugegoda 2006). This agreement among separate studies supports the conclusion that EC conditions do not affect growth, although two alternative explanations should also be considered. First, the precision of mass estimates made on aquatic invertebrates may not be sufficient to detect ecologically significant differences in growth rates. Error in mass estimated by wet weight has been shown to range from 4-15 % of the mean (Marcus, Sutcliffe & Willoughby 1978) and by up to 20% for estimates from body lengths (Benke et al. 1999). Second, estimates of growth based on wet weights or lengths are susceptible to bias from increased water uptake in taxa unable to adequately control water flow and regulate their volume. Additional water uptake by weak osmoregulators could explain why we observed taxa with decreased survival and emergence in low EC conditions to have greater apparent growth in the same conditions. Future work examining how differing EC conditions affect growth should adopt a paired cohort approach that would allow mass to be measured directly as ash free dry mass and control for both of these effects.

Differences in taxa response to variation in TDS conditions are usually attributed to variation in osmoregulatory ability among taxa (Willoughby & Mappin 1988; Hassell, Kefford & Nugegoda 2006; Kefford et al. 2007). Only two of the taxa we tested have had their osmotic abilities quantified, and their survivorship and emergence was consistent with measurements of osmotic performance. Colby (1972) concluded that *Pteronarcys* was a strong osmoregulator relative to other taxa, and as expected *Pteronacrys* showed no difference in survival in our experiments. Buchwalter, Jenkins & Curtis (2002) showed *Callibaetis* to be a moderate osmoregulator and it showed only minor differences in survival. Unfortunately, direct measurements of osmotic abilities are rare in the literature and are focused mostly on various mosquitoes. Demonstrating that differences in survivorship and emergence in different TDS/EC conditions are caused by osmoregulatory differences will be difficult until the osmoregulatory abilities of more taxa are measured.

Weak osmoregulatory ability could explain the better performance of some taxa in high than low TDS conditions, but not the better performance in low TDS conditions we observed for *H. pacifica* (Table 4-3 and Fig. 4-4). Being a strong osmoregulator should not decrease performance at high ECs, until the environmental EC increases past an animal's hemolymph concentration and the osmotic gradient switches direction, which occurs at much higher EC conditions than we used in either experiment. In natural streams with high EC, poor survival and emergence of strong osmoregulators could be the result of increased competition with taxa that do not have to allocate energy into osmoregulatory structures and processes or differences in resource availability in low and high EC environments. However, our control of these factors in our experiments excludes these mechanisms. The fact that we only observed poor survival and emergence in high EC conditions in the stream-side experiment, and not in the laboratory experiment, indicates the difference is likely related to one of the natural water sources. Water chemistry analysis did not indicate any contamination by heavy metals or nutrients. The presence of taxon-specific pathogens in the high EC water source is a possible explanation. For example, some nematodes and fungi are known to specialize in parasitizing certain genera of the family Perlidae, and they are also negatively affected by low EC conditions (Micieli et al. 2012; Wood-Eggenschwiler & Barlocher 1983). As the ecology of pathogens affecting aquatic invertebrates becomes better understood, this possibility can be more fully assessed.

Previous studies have shown that taxa responses to low TDS conditions can vary from none to significant in terms of survival and emergence, albeit for small number of taxa. Our work confirms these findings across a larger range of taxa, even when tested against a much smaller but more ecologically relevant range of TDS. We also show that differences in survival under different EC conditions can predict observed variation in EC optima, consistent with the mechanistic hypothesis that a taxon's distribution is related to its ability to withstand osmotic stress. The taxa responses to TDS we observed also agree with our limited understanding of osmoregulatory variation among taxa, but to assess if osmoregulation causes these taxa responses will require a much better understanding of variation in osmoregulatory ability among taxa. Similar experiments on additional taxa are also needed to broaden our understanding of which taxa are sensitive to low TDS conditions.

Understanding how environmental conditions influence habitat suitability for different taxa is a primary goal of ecology and a cornerstone of bioassessment (Hawkins 2006). Measures of organism habitat preference, like EC optima, allow field survey data to be used to assess potential causes of impairment or to establish water quality criteria. As an example of the latter, the USEPA has recently established benchmarks for allowable stream EC in the Appalachian region based on the response of multiple taxa to EC conditions (USEPA 2011). These applications of optima assume a causal relationship between optima and an organism's response to its environment. This study provides experimental evidence that some taxa are directly affected by stream EC conditions, and this effect is related to the EC optima observed for these taxa. Improving our understanding of how aquatic biota respond to different osmotic challenges will allow for stronger causal inferences of the impacts of modifying EC, and enable predictions of how future changes in EC caused by climate or land use changes might influence distributions of individual aquatic taxa and entire communities.

References

Benke, A.C., Huryn, A.D., Smock, L.A. & Wallace, J.B. (1999) Length-mass relationships for freshwater macroinvertebrates in North America with particular reference to the southeastern United States. *Journal of the North American Benthological Society*, **18**, 308-343.

Black, R.W., Munn, M.D. & Plotnikoff, R.W. (2004) Using macroinvertebrates to identify biota-land cover optima at multiple scales in the Pacific Northwest, USA. *Journal of the North American Benthological Society*, **23**, 340-362.

Blocksom, K.A. & Winters, L. (2006) The evaluation of methods for creating defensible, repeatable, objective and accurate tolerance values for aquatic taxa. Office of Research and Development, USEPA, Cincinnati, OH. EPA 600/R-06/045.

Bradley, T.J. (2009) *Animal Osmoregulation*, Oxford University Press, New York, NY.

Buchwalter, D.B., Jenkins, J.J. & Curtis, L.R. (2002) Respiratory strategy is a major determinant of [3H] water and [14C] chlorpyrifos uptake in aquatic insects*. Canadian Journal of Fisheries and Aquatic Sciences*, **59**, 1315-1322.

Charmantier, G., M. Charmantier-Daures & Towle (2009) Osmotic and ionic regulation in aquatic arthropods. *Osmotic and Ionic Regulation, Cells and Animals* (ed. D. H. Evans), pp. 165-230, CRC Press, Boco Raton, FL.

Clark, T.M., Flis, B.J. & Remold, S.K. (2004) Differences in the effects of salinity on larval growth and developmental programs of a freshwater and a euryhaline mosquito species (Insecta : Diptera, Culicidae). *Journal of Experimental Biology*, **207**, 2289-2295.

Colby, C. (1972) Salt and water balance in stonefly naiads, *Pteronarcys californica* Newport. *Comparative Biochemistry and Physiology Part A*, **41**, 851-860.

Eggert, S.L. & Burton, T.M. (1994) A comparison of *Acroneuria-lycorias* (Plecoptera) production and growth in northern Michigan hard-water and soft-water streams. *Freshwater Biology*, **32**, 21-31.

Egglishaw, H. & Morgan, N. (1965) A survey of the bottom fauna of streams in the Scottish Highlands. Part II. The relationship of the fauna to the chemical and geological conditions. *Hydrobiologia*, **26**, 173-183.

Egglishaw, H.J. (1968) The quantitative relationship between bottom fauna and plant detritus in streams of different calcium concentrations. *Journal of Applied Ecology*, **5,** 731-740.

Fiance, S.B. (1978) Effects of pH on the Biology and Distribution of *Ephemerella funeralis* (Ephemeroptera). *Oikos*, **31**, 332-339.

Hassell, K.L., Kefford, B.J. & Nugegoda, D. (2006) Sub-lethal and chronic salinity tolerances of three freshwater insects: *Cloeon* sp and *Centroptilum* sp (Ephemeroptera : Baetidae) and *Chironomus* sp (Diptera : Chironomidae). *Journal of Experimental Biology*, **209**, 4024-4032.

Hawkins, C.P. (1984) Substrate associations and longitudinal distributions in species of Ephemerellidae (Ephemeroptera:Insecta) from Western Oregon. *Freshwater Invertebrate Biology*, **3**, 181-188.

Hawkins, C.P. (1985) Food habits of species of Ephemerellid mayflies (Ephemeroptera: Insecta) in streams of Oregon. *American Midland Naturalist*, **113**, 343-352.

Hawkins, C.P. (1986) Variation in individual growth-rates and population-densities of Ephemerellid mayflies. *Ecology*, **67**, 1384-1395.

Hawkins, C. P. (2006) Quantifying biological integrity by taxonomic completeness: Its utility in regional and global assessments. Ecological Applications, 16, 1277-1294.

Hawkins, C.P., Olson, J.R. & Hill, R.A. (2010) The reference condition: predicting benchmarks for ecological and water-quality assessments. Jo*urnal of the North American Benthological Society*, **29**, 312-343.

Kapoor, N.N. (1979) Osmotic regulation and salinity tolerance of the stonefly nymph, *Paragnetina media*. *Journal of Insect Physiology*, **25**, 17-20.

Kefford, B.J. & Nugegoda, D. (2005) No evidence for a critical salinity threshold for growth and reproduction in the freshwater snail *Physa acuta*. *Environmental Pollution*, **134**, 377-383.

Kefford, B.J., Nugegoda, D., Zalizniak, L., Fields, E.J. & Hassell, K.L. (2007) The salinity tolerance of freshwater macroinvertebrate eggs and hatchlings in comparison to their older life-stages: a diversity of responses - The salinity tolerance of freshwater macroinvertebrate eggs and hatchlings. *Aquatic Ecology*, **41**, 335-348.

Kefford, B.J., Palmer, C.G., Pakhomova, L. & Nugegoda, D. (2004) Comparing test systems to measure the salinity tolerance of freshwater invertebrates. *Water SA*, **30**, 499-506.

Marcus, J.H., Sutcliffe, D.W. & Willoughby, L.G. (1978) Feeding and growth of *Asellus-aquaticus* (Isopoda) on food items from the littoral of Windermere, including green leaves of *Elodea-canadensis*. *Freshwater Biology*, **8**, 505-519.

Micieli, M.V., Risso, P., Achinelly, M.F., Villar, M.M. & Muttis, E. (2012) Population dynamics between the mermithid *Strelkovimermis spiculatus* (Nematoda: Mermithidae) and the floodwater mosquito *Ochlerotatus albifasciatus* (Diptera: Culicidae) over time. *Biological Control*, **61**, 55-63.

Minshall, G.W. & Kuehne, R.A. (1969) An ecological study of invertebrates of the Duddon, an English mountain stream. *Archiv für Hydrobiologie*, **66**, 169-191.

Minshall, G.W. & Minshall, J.N. (1978) Further evidence on the role of chemical factors in determining the distribution of benthic invertebrates in the River Duddon. *Archiv für Hydrobiologie*, **83**, 324-355.

Pond, G.J., Passmore, M.E., Borsuk, F.A., Reynolds, L. & Rose, C.J. (2008) Downstream effects of mountaintop coal mining: comparing biological conditions using family- and genus-level macroinvertebrate bioassessment tools. *Journal of the North American Benthological Society*, **27**, 717-737.

Renner, R. (2009) Salt-loving algae wipe out fish in Appalachian stream. *Environmental Science & Technology*, **43**, 9046-9047.

Sutcliffe, D.W. & Carrick, T.R. (1973) Studies on mountain streams in the English Lake District I. pH, calcium and the distribution of invertebrates in the River Duddon. *Freshwater Biology*, **3**, 437-462.

USEPA (2004) Evaluation of impacts to underground sources of drinking water by hydraulic fracturing of coalbed methane reservoirs. US Environmental Protection Agency, Washington, DC. EPA 816-R-04-003.

USEPA (2011) A field-based aquatic life benchmark for conductivity in central Appalachian streams. Office of Research and Development, National Center for Environmental Assessment, Washington, DC. EPA/600/R-10/023.F.

Williams, W.D. (1987) Salinization of rivers and streams - an important environmental-hazard. *Ambio*, **16**, 180-185.

Willoughby, L.G. & Mappin, R.G. (1988) The distribution of *Ephemerella-ignita* (Ephemeroptera) in streams - the role of pH and food resources. *Freshwater Biology*, **19**, 145-155.

Wood-Eggenschwiler, S. & Barlocher, F. (1983) Aquatic hyphomycetes in sixteen streams in France, Germany and Switzerland. *Transactions of the British Mycological Society*, **81**, 371-379.

CHAPTER 5

CONCLUSION

My research has advanced our understanding of how catchment geology influences streams and their biota in two ways. The first is the development of models predicting natural water chemistry from geology and other environmental factors. The second is showing how differences in EC caused by geology and other factors influences the distribution of aquatic macroinvertebrates. In addition to increasing our understanding of how geology influences water chemistry, these models also have direct application to assessing stream ecologic health.

The development of models predicting natural water chemistry help quantify how geologic and environmental predictors interact to produce spatial differences in water chemistry. Geology is known to strongly influence many of the constituents of water chemistry, but the development of empirical predictive models allowed me to quantify the influence of geology relative to other environmental factors (Table 2-6). My analyses showed that, at a regional scale, geology has a greater influence on major ion concentrations and EC than climate, soils, vegetation, or topography. By relating a combination of spatial and temporal variables to stream nutrient concentrations, I was able to determine which factors likely have the greatest influence on TP and TN concentrations (Tables 3-2 and 3-3). Inferences of the relative strength of different predictors in these models should be made cautiously, because the majority of variation in nutrient concentrations remains unexplained. However, TP was most heavily influenced by the P content of the underlying rock, as expected. Contrary to the conclusions of Holloway et al. (1998), rock N was not related to stream TN indicating limited influence of rock N on stream TN concentrations.

The empirical approach I used to model stream chemistry as a function of geology and other environmental factors indicated some processes may be more important than previously understood. For example, temperature was the third most important predictor of the concentration of major ions in stream water, presumably due to the effects of increasing temperature on evapo-concentration and weathering rates. This finding implies that as climates become warmer in the future, stream chemistries will change. The effects of atmospheric deposition in wet and dry forms and as dust are also not generally included in most process based models, but my empirical models clearly showed effects of deposition on all constituents of water chemistry examined. The data on sources and sinks used in my empirical models, such as atmospheric deposition or rock chemistry, can also be used in process based models to account for spatial differences among catchments, which should reduce their dependence on local calibration and increase their transferability among catchments.

I also developed predictive models that showed that geologically driven differences in stream EC accounted for the majority of the variation in the EC optima of 19 macroinvertebrate taxa (Figure 4-3). EC optima are a measure of how macroinvertebrate distributions respond to differences in stream EC. The observed strength of this relationship is evidence of geology's influence on macroinvertebrate distributions. This relationship between macroinvertebrates and geologically driven differences in water chemistry provides insight into the relative importance of basic ecological processes that influence macroinvertebrate distributions. Because of geology's role in creating diverse chemical habitats, streams across a range of geologies should be conserved to maximize the number of taxa protected.

Predictive models not only increase our understanding of how geology and other environmental factors interact to produce different water chemistries and macroinvertebrate distributions, but can also be used to assess water quality. Predictions of water chemistry expected under natural conditions can be compared directly with current water chemistry to assess if water quality has been altered. Comparing stream EC with predicted EC could help determine how much land uses like mountain top removal / valley fill operations have changed stream chemistry. Water chemistry predictions can also be used to improve bioassessments. Because metrics used in bioassessments depend on predicting some benchmark biological condition to compare with current conditions, increasing the accuracy of these predicted benchmarks will increase the accuracy of the resulting biologic inferences. Current bioassessments largely ignore water chemistry in establishing benchmarks because there has not been any method to accurately predict background water chemistry. My predictive models address this need, and should lead to more accurate bioassessments in the future. Predictions of water chemistry can also improve the selection of reference sites used in bioassessment. Currently, reference sites are chosen based on regional thresholds for various water chemistry components (i.e., SO_4, Cl, TN, and TP). Site specific predictions of background concentrations can be used instead as the benchmarks used to assess if a site is reference quality. By accounting for unexplained variance, predictions of natural nutrient concentrations should result in more appropriate site-specific nutrient criteria. Accounting for natural spatial variation in nutrient concentrations should produce criteria that are both more attainable and better protective than current criteria that only account for natural variation using regional classifications.

The maps of rock chemical and physical characteristics I created should also be useful in a wide range of ecological applications beyond aquatic ecology. Few studies have used geologic data to predict diversity and distributions, even though geology is recognized as one of the abiotic factors controlling taxonomic diversity and distributions of many terrestrial plants and animals (Anderson and Ferree 2010). One of the reasons geologic data has not been used is the nature of geologic maps. Geologic maps primarily characterize lithology in terms of its age, structure, and formative process instead of current chemical and physical properties needed for ecologic prediction. My maps of geologic chemical and physical properties should be much more relevant predictors of the abundance and distributions of terrestrial plants than many of the coarse surrogates used in existing models (sensu Elith and Leathwick 2009).

Future work on modeling water chemistry should focus both on improving predictions and on expanding the number of water chemistry constituents covered. Including temporally and spatially specific estimates of catchment discharge has the greatest potential for improving model performance because stream discharge greatly influences solute concentrations in streams and is not directly accounted for in my models. These estimates require a method for estimating discharge in ungaged catchments, which has been the focus of hydrologists for the last decade (Sivapalan et al. 2003). As estimates of discharge become available, including them will both improve water chemistry predictions and also allow for predictions to be made at other than base-flow conditions. The poor predictive power of my TN model indicates that additional predictors are needed to better account for both sources (i.e., a reliable measure of dry deposition) and sinks (i.e., better estimates of denitrification and uptake). My nutrient predictive models might also be improved by accounting for the spatial arrangements of sources and sinks of nutrients relative to each other in time and space. Sinks, such as soils with greater potential for denitrification, would have a greater effect on the amount of nutrients entering streams if they are located between source areas and the stream channel. Other constituents of water chemistry affecting stream biota or being used as indicators of water quality should be modeled using these same approaches (e.g., Si, Al, Fe, Na, Cl, and K). Si, K, and Fe can be limiting resources for biota (e.g., Si is needed by diatoms, K is needed by aquatic macrophytes and fungi, and many anaerobic microbes use Fe as an electron acceptor). Na, Cl, and Al also can vary with natural sources (either geologic or marine) and can be significantly increased by human activities leading to toxic effects on stream biota and reduced water quality for human use. Modeling these additional chemical constituents could then be used to predict spatial variation in aquatic assemblages other than macroinvertebrates and support a more complete assessment of water quality.

References

Anderson, M. G., and C. E. Ferree. 2010. Conserving the stage: climate change and the geophysical underpinnings of species diversity. Plos One:5.

Elith, J., and J. R. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology Evolution and Systematics 40:677-697.

Holloway, J. M., R. A. Dahlgren, B. Hansen, and W. H. Casey. 1998. Contribution of bedrock nitrogen to high nitrate concentrations in stream water. Nature 395:785-788.

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

APPENDICES

Appendix A - Potential Predictors Evaluated for Nutrient Model

Table A. Potential Predictors Evaluated for Nutrient Model

Type	Variable	Units	Data Source	TP[a]	TN[a]
Geology	Catchment mean & point whole rock CaO	%	Olson & Hawkins 2012[b]	C	
	Catchment mean & point whole rock MgO	%	Olson & Hawkins 2012		
	Catchment mean & point whole rock S	%	Olson & Hawkins 2012		
	Catchment mean & point unconfined compressive strength	MPa	Olson & Hawkins 2012		
	Catchment mean & point log geometric mean hydraulic conductivity	$x10^{-6}$ m/s	Olson & Hawkins 2012		
	Catchment mean geometric mean hydraulic conductivity	$x10^{-6}$ m/s	Olson & Hawkins 2012		
	Catchment mean & point whole rock P_2O_5	%	Olson & Hawkins 2012	C	
	Catchment mean & point whole rock N	%	Olson & Hawkins 2012		
	Catchment mean whole rock NH_4	%	This study		
	Catchment Coefficient of Variation of rock CaO	%	Olson & Hawkins 2012		
	Catchment Coefficient of Variation of rock MgO	%	Olson & Hawkins 2012		
	Catchment Coefficient of Variation of rock SO	%	Olson & Hawkins 2012		
	Catchment Coefficient of Variation of rock P_2O_5	%	Olson & Hawkins 2012		
	Catchment Coefficient of Variation of rock N	%	Olson & Hawkins 2012		
	Catchment areal percent underlain by mafic volcanic rocks	%	Integrated geologic map[c.]		
	Catchment areal percent underlain by volcanic rocks	%	Integrated geologic map	C	
Temporal	Year of sample	year	Water Chem Data[d]		
	Day of year sample collected	day of year	Water Chem Data		C

[a.] Indicates if variable was selected for final model. “C” indicates catchment level variable selected, “P” indicates point level variable selected.

Table A. Continued.

Type	Variable	Units	Data Source	TP	TN
Climate	Catchment mean & point of mean 1971-2000 annual precipitation	mm/year	PRISM[e]		
	Catchment mean of mean 1971-2000 annual min monthly precipitation	mm/ month	PRISM		
	Catchment mean of mean 1971-2000 annual max monthly precipitation	mm/ month	PRISM		
	Catchment mean of mean June-Sept 1971-2000 monthly precipitation	mm/ month	PRISM		
	Catchment mean of mean 1971-2000 annual temperature	˚C	PRISM		
	Catchment mean of mean 1971-2000 annual min monthly temperature	˚C	PRISM	P	C
	Catchment mean of mean 1971-2000 annual max monthly temperature	˚C	PRISM		
	Catchment mean of mean 1961-1990 first & last day of freeze	day of year	PRISM		
	Catchment mean of mean 1961-1990 annual number of wet-days	days/year	PRISM		C
	Catchment mean of mean 1961-1990 annual relative humidity	%	PRISM	C	
	Catchment mean of mean 1961-1990 annual max number of wet-days	days/year	PRISM		
	Catchment mean of mean 1961-1990 annual min number of wet-days	days/year	PRISM		
	Catchment mean of mean precipitation over two months prior to the sample	mm/ month	PRISM		C
	Catchment mean of mean precipitation over the year prior to the season sample was taken	mm/ year	PRISM	C	
	Catchment mean of mean precipitation over the month of the sample	mm/ month	PRISM		
	Catchment mean of mean precipitation over the month prior to the sample month	mm/ month	PRISM		
	Catchment mean of mean precipitation occurring as snow over the year prior to the sample	mm/year	PRISM		
Geography	Latitude	degrees	Water Chem Data		
	Longitude	degrees	Water Chem Data		
	Level II Ecoregion	Name	CEC Map[f]	C[g]	

Table A. Continued.

Type	Variable	Units	Data Source	TP	TN
Atmospheric Deposition	Catchment mean of mean 1994-2006 annual precipitation-weighted mean Ca concentration	mg/l	NADP^{h}	C	
	Catchment mean of mean 1994-2006 annual precipitation-weighted mean Mg concentration	mg/l	NADP		
	Catchment mean of mean 1994-2006 annual precipitation-weighted mean Na concentration	mg/l	NADP		C
	Catchment mean of mean 1994-2006 annual precipitation-weighted mean Cl concentration	mg/l	NADP		
	Catchment mean of mean 1994-2006 annual precipitation-weighted mean SO_4 concentration	mg/l	NADP		C
	Catchment mean of mean 1994-2006 annual precipitation-weighted mean NO_3 concentration	mg/l	NADP		C
	Catchment mean of mean 1994-2006 annual total inorganic nitrogen (TN) wet deposition	kg/ha	NADP		
	Catchment mean of mean 2002-2006 annual total inorganic nitrogen (TN) wet & dry deposition calculated from CMAQ model using Watershed Deposition Tool	lbs/ acre	CMAQ^{i}		
Soil	Catchment mean & point available water capacity	fraction	STATSGO^{j}	C	
	Catchment mean & point bulk density	g/cm^{3}	STATSGO		C
	Catchment mean & point soil erodibility (K factor)	dimensionless	STATSGO	C	
	Catchment mean & point organic matter content	% weight	STATSGO		
	Catchment mean & point soil permeability	inches/ hr	STATSGO		
	Catchment mean & point soil depth	m	STATSGO		
	Catchment mean & point soil pH	pH	STATSGO		
	Catchment mean water table depth	m	STATSGO		
	% Catchment area in each of 8 soil orders (Alfisol, Aridisol, Andisol, Entisol, Inceptsol, Mollisol, Spodosol, or Ultisol)	%	STATSGO	C^{k}	
	Catchment mean soil organic carbon to 1 m depth	kg-C/m^{2}	IGBP-DIS Soil Data^{l}	C	

Table A. Continued.

Type	Variable	Units	Data Source	TP	TN
Lakes & Wetlands	% Catchment area covered by lakes	%	NHD[m]		
	% Catchment area covered by lakes	%	NHD		
	% Catchment area covered by open water	%	NLCD[n]		
	Area of largest lake in catchment	m^2	NHD		
	Flow weighted (using flow accumulation) lake area index	dimensionless	NHD		
	Flow weighted (using flow accumulation) largest lake area index	dimensionless	NHD		
	% Catchment area covered by wetland	%	NHD		
	% Catchment area covered by wetland	%	NLCD		
	% Catchment area covered by wooded wetland	%	NLCD		
	% Catchment area covered by herbaceous wetland	%	NLCD		
	Area of largest wetland in catchment	m^2	NHD		
	Flow weighted (using flow accumulation) wetland area index	dimensionless	NHD		
	Flow weighted (using flow accumulation) largest wetland area index	dimensionless	NHD		
	% Catchment area covered by lakes & wetlands	%	NHD		
	% Catchment area covered by open water or wetlands	%	NLCD		
	Area of largest lake or wetland in catchment	m^2	NHD	C	
	Flow weighted (using flow accumulation) lake & wetland area index	dimensionless	NHD		
	Flow weighted (using flow accumulation) largest lake or wetland area index	dimensionless	NHD		
Topography	Catchment elevation mean, min, max, and std deviation	m	NED[o]		
	Catchment elevation relief ratio	dimensionless	NED		
	Catchment shape ratio (catchment area : length)	dimensionless	NED		
	Catchment area	km^2	NED		
	Catchment mean channel slope	%	NED	C	

Table A. Continued.

Type	Variable	Units	Data Source	TP	TN
N-fixing Plants	Catchment areal coverage of LANDFIRE Biophysical Settings where *Alnus rubra* is predicted to be dominant	%	LANDFIRE[p]		C
	Catchment areal coverage of LANDFIRE Biophysical Settings where *Alnus rubra* is predicted to occurs	%	LANDFIRE		
	Catchment areal *Alnus rubra* coverage from LEMMA	%	LEMMA[q]		
	Catchment areal coverage of LANDFIRE Biophysical Settings where any moderate N-fixing plant is predicted to occur	%	LANDFIRE		
	Catchment areal coverage of LANDFIRE Biophysical Settings where *Alnus incana* is predicted to occur	%	LANDFIRE		
	Catchment areal coverage of LANDFIRE Biophysical Settings where *Ceanothus velutinus* is predicted to occur	%	LANDFIRE		
	Catchment areal coverage of LANDFIRE Biophysical Settings where *Prosopis glandulosa* is predicted to occur	%	LANDFIRE		
	Occurrence of *Alnus rubra* at sample point	Y/N	LANDFIRE		
	Occurrence of any moderate N-fixing plant at sample point	Y/N	LANDFIRE		
Ground-water	Catchment mean & point groundwater delivery velocity	m/day	MRI-Darcy Model[r]		C[s]
	Catchment mean & point groundwater recharge velocity	m/day	MRI-Darcy Model		
	Ratio of catchment mean delivery: recharge	dimensionless	MRI-Darcy Model		
	Catchment mean and maximum precipitation weighted ground water delivery Index	dimensionless	MRI-Darcy Model		
	Log_{10} Catchment mean and maximum precipitation weighted ground water delivery Index	dimensionless	MRI-Darcy Model		
	Catchment mean Base-Flow Index	dimensionless	USGS Gage Data[t]		

Table A. Continued.

Type	Variable	Units	Data Source	TP	TN
Vegetation	Catchment mean of mean 2000-2009 annual Enhanced Vegetation Index	dimensionless	MODIS[u]	C	C
	Catchment max of mean 2000-2009 annual Enhanced Vegetation Index	dimensionless	MODIS		
	Catchment mean of mean 2000-2009 annual max Enhanced Vegetation Index	dimensionless	MODIS		
	Catchment mean evergreen land cover	%	NLCD		C
	Catchment mean deciduous land cover	%	NLCD		
	Catchment mean mixed forest land cover	%	NLCD		
Water Chemistry	Predicted Electric Conductivity	µS/cm	Olson & Hawkins 2012		
	Predicted Acid Neutralization Capacity	µeq/L	Olson & Hawkins 2012		
	Predicted TP	µg/L	This study		
	Predicted TN	µg/L	This study		
	Measured TP	µg/L	This study		
	Measured TN	µg/L	This study		

[b] Derived using method described in section 2.1 of Olson & Hawkins (2012) at a grid resolution of 90 x 90 m. See Olson, J. R. and C. P. Hawkins (2012), Predicting natural base-flow stream water chemistry in the western United States, Water Resources Research, 48: WR011088.

[c] Preliminary integrated geologic map databases for the United States (obtained from http://pubs.usgs.gov/of/2005/1351/index_map.htm).

[d] See table 3-1 for sources of water chemistry data.

[e] PRISM climate data. 2 x 2 km resolution grids were used for the 1961–1990 data, and 800 x 800 m resolution grids were used for the 1971–2000 data. See Daly, C., R. P. Neilson, and D. L. Phillips (1994), A statistical topographic model for mapping climatological precipitation over mountainous terrain, Journal of Applied Meteorology, 33, 140-158.

[f] CEC (2006), Ecological regions of North America: toward a common perspective, Commission for Environmental Cooperation, Montreal, Quebec. Obtained from http://www.epa.gov/wed/pages/ecoregions.htm.

[g] Only Ecoregion 13 was selected for TP model, see text for details.

[h] National Atmospheric Deposition Program National Trends Network (NADP/NTN) 2.5 x 2.5 km resolution grids (obtained from the NADP website available at http://nadp.sws.uiuc.edu/ntn/).

[i] Community Multiscale Air Quality (CMAQ) model output analyzed using the Watershed Deposition Tool (available at http://www.epa.gov/AMD/EcoExposure/deposition Mapping.html). See Schwede, D. B., R. L. Dennis, and M. A. Bitz (2009), The watershed deposition tool: A tool for incorporating atmospheric deposition in water-quality analyses, Journal of the American Water Resources Association, 45:973-985.

[j] Natural Resource Conservation Service State Soil Geographic Database (NRCS STATSGO) 500 x 500 m resolution grids (obtained from the NRCS website available at http://soils.usda.gov/survey/geography/statsgo/).

[k] Only % Alfisol was selected.

[l] IGBP-DIS (1998) SoilData(V.0) A program for creating global soil-property databases, IGBP Global Soils Data Task, France. Obtained from The Atlas of the Biosphere: http://atlas.sage.wisc.edu/.

[m] National Hydrography Dataset, NHDWaterbody features identified as natural (lakes or wetlands), obtained from http://nhd.usgs.gov.

[n] National Land Cover Dataset, 2001, 30 x 30 m resolution grids, obtained from http://www.mrlc.gov/.

[o] Calculated from National Elevation Database DEMs at 30 x 30 m resolution (obtained from the USGS website available at http://ned.usgs.gov/).

[p] LANDFIRE Refresh 2008 (lf_1.1.0) Biophysical Settings (lf_110bps), 30 x 30 m resolution grids (obtained from http://landfire.cr.usgs.gov).

[q] Landscape Ecology, Modeling, Mapping, and Analysis (LEMMA) Modeling Region 200 March 2010, 30 x 30 m grid resolution (obtained from http://www.fsl.orst.edu/lemma/main.php?project=common &id=mr&model_region=200&ref=nwfp15).

[r] Groundwater flow velocity derived from MRI-Darcy model (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, Environmental Management, 32, 706-719), at a 90x 90 m resolution.

[s] Catchment maximum selected.

[t] 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.

[u] MODIS satellite MOD13A1.V4 data collected every 16 d at 500 x 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).

Appendix B - Permission-to-Reprint Letter

We are pleased to grant permission for the use of the material requested for inclusion in your thesis. The following non-exclusive rights are granted to AGU authors:

All proprietary rights other than copyright (such as patent rights).
The right to present the material orally.
The right to reproduce figures, tables, and extracts, appropriately cited.
The right to make hard paper copies of all or part of the paper for classroom use.
The right to deny subsequent commercial use of the paper.

Further reproduction or distribution is not permitted beyond that stipulated. The copyright credit line should appear on the first page of the article or book chapter. The following must also be included, “Reproduced by permission of American Geophysical Union.” To ensure that credit is given to the original source(s) and that authors receive full credit through appropriate citation to their papers, we recommend that the full bibliographic reference be cited in the reference list. The standard credit line for journal articles is: "Author(s), title of work, publication title, volume number, issue number, citation number (or page number(s) prior to 2002), year. Copyright [year] American Geophysical Union."

If an article was placed in the public domain, in which case the words “Not subject to U.S. copyright” appear on the bottom of the first page or screen of the article, please substitute “published” for the word “copyright” in the credit line mentioned above.

Copyright information is provided on the inside cover of our journals. For permission for any other use, please contact the AGU Publications Office at AGU, 2000 Florida Ave., N.W., Washington, DC 20009.

Michael

Michael Connolly
Program Manager, Journals
American Geophysical Union
+1.202.777.7365
MConnolly@agu.org
www.agu.org

AGU galvanizes a community of Earth and space scientists that collaboratively advances and communicates science and its power to ensure a sustainable future.

CURRICULUM VITAE

John Robert Olson
July 2012

Utah State University
Department of Watershed Sciences &
Western Center for Monitoring and Assessment of Freshwater Ecosystems
5210 Old Main Hill, Logan, UT 84322
(435) 770-4533 FAX: (435) 797-1871 email: john.olson@usu.edu

Research Interests

- Applications of spatial & macro ecology to aquatic systems
- Affects of geology on aquatic systems
- Aquatic ecosystem ecology
- Bioassessment

Education

Ph.D. (2012)	Utah State University: Watershed Science Dissertation Title: The Influence of Geology and Other Environmental Factors on Stream Water Chemistry and Benthic Invertebrate Assemblages.
M.S. (2002)	Columbus State University: Environmental Science Thesis Title: Using GIS and land use data to select candidate reference sites for stream bioassessment.
B.A. (1986)	University of California, Santa Barbara: Geography

Professional Experience

Postdoctoral Fellow (2012- present)	Utah State University – Department of Watershed Sciences: Creating nation-wide maps of geologic characteristics, and empirically modeling natural water chemistry for establishing baselines and nutrient criteria.
Assistant Research Engineering Officer (2008 - Present)	U.S. Army Research Laboratory - Battlefield Environment Division, Computational and Information Sciences Directorate: Level I Science and Technology Manager, responsible for coordinating research efforts between the Army Research Laboratory, supporting universities, and users groups.
Research Assistant (2002- 2012)	Utah State University – Department of Watershed Sciences: Improve understanding of the linkages between geology, water chemistry, and stream invertebrate assemblages and apply this to predict natural conditions to improve both water quality and biological assessments.

Research Assistant (2000-2001)	Columbus State University – Department of Environmental Science: Ecoregion Reference Site Project, Phase II - Identification and characterization of stream bioassessment reference sites in Georgia.
Acquisition Officer (1996-1999)	U.S. Army Infantry Center: Project management & requirement establishment for body armor, eye protection, & parachute programs.
Infantry Officer (1985-1995) (2007-2008)	Various assignments, including: Commander responsible for planning & leading 220 man company through 2 deployments; Personnel & Logistics Officer responsible for all human resource & logistics planning and execution for 550 man battalion; Senior Advisor to Iraqi Army in northern Iraq advising the Iraqi Army on training, administration, and logistics, and oversaw a $21 million construction and supply budget.

Teaching Experience

Instructor (2002 – 2006)	Utah State University – Department of Military Science: Advanced Tactics and Operations (MS 3020), Command and Staff Functions (MS 4010), and Officer Perspectives (MS 4020)
Teaching Assistant (1999-2000)	Columbus State University – Department of Biology: Principles of Biology (BIOL 1215)
Teaching Assistant (1999)	Columbus State University – Department of Environmental Science: Ecological Methods (ENVS 6207)
Instructor (1999-2001)	Columbus State University – Department of Military Science: Wilderness Survival (ROTC 1215) & Leadership Values and Skills (ROTC 1216)

Publications

Journal Articles:

Olson, J.R., and C.P. Hawkins. 2012. Predicting natural base-flow stream water chemistry in the western United States. Water Resources Research WR011088.

Bennett, S.N., **J.R. Olson**, J.L. Kershner, and P. Corbett. 2010. Influence of propagule pressure and stream characteristics on introgression between native westslope cutthroat trout and introduced rainbow trout in British Columbia. Ecological Applications 20(1), 263–277.

Hawkins, C.P., **J.R. Olson**, and R.A. Hill. 2010. The reference condition: predicting benchmarks for ecological and water-quality assessments. Journal of the North American Benthological Society 29(1): 312–343.

Cao, Y., C.P. Hawkins, **J. Olson**, M.A. Kosterman. 2007. Modeling natural environmental gradients improves the accuracy and precision of diatom-based indicators. Journal of the North American Benthological Society 26(3): 566-585.

Hargett, E.G., J.R. Zumberge, C.P. Hawkins, and **J.R. Olson**. 2007. Development of a RIVPACS-type predictive model for bioassessments of wadeable streams in Wyoming. Ecological Indicators 7(4): 807-826.

Book Chapters:

Olson, J.R., D.L. Hughes, and M.P. Brossett. 2010. Comparison of bioassessment methods. Chapter 2 in Hughes, D.L., M.P. Brossett, J.A. Gore, and **J.R. Olson** (editors). Rapid bioassessment of stream health. CRC Press, Boca Raton, Florida.

Olson, J.R., D.L. Hughes, J.A. Gore, and M.P. Brossett. 2010. Candidate reference conditions. Chapter 4 in Hughes, D.L., M.P. Brossett, J.A. Gore, and **J.R. Olson** (editors). Rapid bioassessment of stream health. CRC Press, Boca Raton, Florida.

Reports & Manuals:

Chinnayakanahalli, K., R. Hill, **J. Olson**, C. Kroeber, D.G. Tarboton, and C.P. Hawkins. 2006. The multi-watershed delineation tool: GIS software in support of regional watershed analyses, users manual. Department of Civil and Environmental Engineering and Department of Aquatic, Watershed, & Earth Resources, Utah State University.

Gore, J.A., **J.R. Olson**, D.L. Hughes, M. Brossett. 2004. Reference conditions for wadeable streams in Georgia with a multimetric index for the bioassessment and discrimination of reference and impaired streams. Georgia Department of Natural Resources, Atlanta, GA.

Hawkins, C.P., J.D. Ostermiller, M.R. Vinson, R.J. Stevenson, and **J.R. Olson**. 2003. Stream algae, invertebrate, and environmental sampling associated with biological water quality assessments field protocols. Department of Aquatic, Watershed, & Earth Resources, Utah State University.

Manuscripts in Review or Revision:

Olson, J.R., and C.P. Hawkins. *In review*. Developing site-specific nutrient criteria from empirical models. Submitted to Freshwater Science.

Olson, J.R., and C.P. Hawkins. *In revision*. An experimental assessment of the effects of low total dissolved solids on the survival and distribution of stream macroinvertebrates. Submitted to Functional Ecology.

Awards

University Graduate Research Assistant of the Year (2010) – Utah State University Robins Award.

Collage Graduate Research Assistant of the Year (2010) – Collage of Natural Resources, Utah State University.

Best Student Methods Presentation (2006) - Annual Meeting of the North American Benthological Society, Anchorage, AK.