Machine learning and hurdle models for improving regional predictions of stream water acid neutralizing capacity

Nicholas A. Povak,[1] Paul F. Hessburg,[1] Keith M. Reynolds,[2] Timothy J. Sullivan,[3] Todd C. McDonnell,[3] and R. Brion Salter[1]

Received 2 April 2012; revised 9 May 2013; accepted 12 May 2013.

[1] In many industrialized regions of the world, atmospherically deposited sulfur derived from industrial, nonpoint air pollution sources reduces stream water quality and results in acidic conditions that threaten aquatic resources. Accurate maps of predicted stream water acidity are an essential aid to managers who must identify acid-sensitive streams, potentially affected biota, and create resource protection strategies. In this study, we developed correlative models to predict the acid neutralizing capacity (ANC) of streams across the southern Appalachian Mountain region, USA. Models were developed using stream water chemistry data from 933 sampled locations and continuous maps of pertinent environmental and climatic predictors. Environmental predictors were averaged across the upslope contributing area for each sampled stream location and submitted to both statistical and machine-learning regression models. Predictor variables represented key aspects of the contributing geology, soils, climate, topography, and acidic deposition. To reduce model error rates, we employed hurdle modeling to screen out well-buffered sites and predict continuous ANC for the remainder of the stream network. Models predicted acid-sensitive streams in forested watersheds with small contributing areas, siliceous lithologies, cool and moist environments, low clay content soils, and moderate or higher dry sulfur deposition. Our results confirmed findings from other studies and further identified several influential climatic variables and variable interactions. Model predictions indicated that one quarter of the total stream network was sensitive to additional sulfur inputs (i.e., ANC < 100 μeq L^{-1}), while <10% displayed much lower ANC (<50 μeq L^{-1}). These methods may be readily adapted in other regions to assess stream water quality and potential biotic sensitivity to acidic inputs.

Citation: Povak, N. A., P. F. Hessburg, K. M. Reynolds, T. J. Sullivan, T. C. McDonnell, and R. B. Salter (2013), Machine learning and hurdle models for improving regional predictions of stream water acid neutralizing capacity, *Water Resour. Res.*, *49*, doi:10.1002/wrcr.20308.

1. Introduction

[2] Since the industrial revolution, industrially derived atmospherically deposited sulfur (S) has acidified streams across the eastern United States [*United States Environmental Protection Agency*, 2009], Europe [*Christophersen et al.*, 1990; *Hruška et al.*, 2002; *Schöpp et al.*, 2003], China [*Galloway et al.*, 1987], southeast Asia, and other industrialized regions [*Galloway*, 2001; *Menz and Seip*, 2004]. Within these areas, increased acidification has depleted base cations from soils through sulfate (SO_4^{2-}) leaching and exchange of associated acidity with base cations in the soil solution [*Hendershot et al.*, 1991], mobilized inorganic aluminum (Al_i) to streams [*Sullivan*, 2000], and reduced richness of fish and aquatic invertebrates [*Guerold et al.*, 2000; *Rago and Wiener*, 1986; *T. J. Sullivan et al.*, 2007; *United States Environmental Protection Agency*, 2009].

[3] Recent reductions in industrial emissions, particularly in the United States and Europe, have reduced atmospheric S, but lagged effects of prior deposition are still apparent, particularly in geologically sensitive headwater streams [*Driscoll et al.*, 2003; *Guerold et al.*, 2000; *United States Environmental Protection Agency*, 2009]. Results from recent eastern U.S. lake evaluations suggest that recovery from chronic exposure to atmospheric S may take decades [*Driscoll et al.*, 2003].

[4] Acid neutralizing capacity (ANC) is one measure of stream water acid-base status, which is reasonably well correlated with biological health and species richness in acid-sensitive systems [*Lien et al.*, 1992; *T. J. Sullivan et al.*, 2007; *United States Environmental Protection Agency*,

Additional supporting information may be found in the online version of this article.

[1]Wenatchee Forestry Sciences Laboratory, U.S. Forest Service, Pacific Northwest Research Station, Wenatchee, Washington, USA.
[2]Corvallis Forestry Sciences Laboratory, U.S. Forest Service, Pacific Northwest Research Station, Corvallis, Oregon, USA.
[3]E&S Environmental Chemistry Inc., Corvallis, Oregon, USA.

Corresponding author: N. A. Povak, Wenatchee Forestry Sciences Laboratory, U.S. Forest Service, Pacific Northwest Research Station, 1133 N. Western Ave, Wenatchee, WA 98801, USA. (npovak@fs.fed.us)

©2013. American Geophysical Union. All Rights Reserved.
0043-1397/13/10.1002/wrcr.20308

2009]. ANC is the sum of the concentrations of all major base cations, minus concentrations of anionic sulfate (SO_4^{2-}), nitrate (NO_3^-), and chloride (Cl^-), reported in $\mu eq\ L^{-1}$. As rates of acidic deposition increase in watersheds, especially those with shallow acid-sensitive soils, surface water ANC generally decreases, but in proportion to the natural resupply of base cations. Consequently, rebalancing long-term acid-base chemistry in acid-impacted watersheds partially depends on reducing atmospheric S to levels below the natural resupply of base cations [*Sullivan*, 2000]. Base cation resupply comes from soil mineral base cation weathering (BC_w) or exogenous inputs [*Cosby et al*., 1985; *Henriksen and Posch*, 2001; *McDonnell et al*., 2010]. The inherent capacity for a given watershed to buffer against acidic inputs over the long term is therefore related to the set of environmental conditions that influence the amount and transport of base cations within the watershed [*McDonnell et al*., 2012; *Sullivan et al*., 2007].

[5] When monitoring stream water acid-base status, ANC is often chosen over other metrics, such as pH, due to its relative insensitivity to changes in concentrations of CO_2, aluminum reactions, and presence of organic acids [*Neal et al*., 1999]. ANC is widely used in studies of regional critical loads (CLs) [*Clark et al*., 2012; *Duan et al*., 2000; *Henriksen et al*., 1995; *United States Environmental Protection Agency*, 2009] because it is the predominant chemical criterion used in the determination of CLs of acidity. Various ANC thresholds are associated with biological effects [*United States Environmental Protection Agency*, 2009]. Negative effects on macroinvertebrate and fish species richness have been associated with ANC concentrations between ~50 and 100 $\mu eq\,L^{-1}$ [*Cosby et al*., 2006; *Sullivan et al*., 2007], and more substantial effects are observed at lower levels [*Cosby et al*., 2006; *Sullivan et al*., 2007; *United States Environmental Protection Agency*, 2009].

[6] Stream water ANC status is used in the calculation of regional estimates of steady-state CLs to identify acid-sensitive stream reaches (T. C. McDonnell et al., Critical loads of sulfur deposition for aquatic resource protection in the southern Appalachian Mountains, submitted to *Water Resources Research*, 2013, hereinafter referred to as McDonnell et al., submitted manuscript, 2013). The CL is a quantitative estimate of the level of sustained S deposition above which harmful ecosystem effects are likely [*Nilsson and Grennfelt*, 1988]. Taken together, accurately estimated ANC and identifying CL exceedances within individual stream reaches can inform decisions about where best to mitigate S deposition in aquatic habitats. ANC modeling reported here was used in conjunction with BC_w and CL estimation for the study region (McDonnell et al., submitted manuscript, 2013). All three estimates are used in a decision support modeling framework to guide resource management and policy decisions regarding S emissions [*Reynolds et al*., 2012].

[7] The southern Appalachian Mountain region has a long history of atmospheric S deposition and contains threatened aquatic resources. The region exhibits complex land use patterns superimposed on steep climatic and topographic gradients. As such, the diverse environmental settings associated with this study area make it well suited to ANC estimation.

[8] Developing regression models that explain the contributions of biogeochemical and climatic variables to acid neutralization can be difficult, particularly when modeling these relations at large spatial scales. Difficulties arise from inherent geographic variability in interactions among biological, geochemical, and climatic variables that can influence the susceptibility of a stream to S deposition [*Levin*, 1992; *Turner*, 1989]. These interactions can be nonlinear and temporally nonstationary and therefore are poorly addressed by traditional modeling frameworks [*Elith et al*., 2008].

[9] Machine-learning techniques have recently been introduced to mainstream ecological research in an effort to address these factors. The main advantages of machine learning over statistical regression techniques are that resulting models (a) are robust against multicollinearity and outliers; (b) include methods to reduce model overfitting; (c) better identify important predictor variables, nonlinear relationships, and complex interactions among predictors; (d) are unaffected by data transformations; and (e) can incorporate categorical, ordinal, or continuous numeric predictors [*Elith et al*., 2008; *Franklin and Miller*, 2009; *Olden et al*., 2008]. A disadvantage of machine-learning techniques is that most are nonparametric and do not produce model coefficients associated with traditional statistical models. Techniques include ensemble decision trees, neural networks, support vector machines, and Bayesian belief networks [*Hastie et al*., 2005].

[10] Here we employ machine learning to predict the biogeochemical, climatic, vegetative, and acidic deposition that are associated with low-ANC streams in the southern Appalachian Mountain region. To accomplish this, we

[11] (1) gathered available stream water ANC data sets within the region;

[12] (2) used available remotely sensed, surveyed, or process-modeled climate, land cover, atmospheric deposition, geologic, edaphic, and topographic data;

[13] (3) compared traditional and machine-learning approaches to select the best performing model; and

[14] (4) used hurdle modeling and data resampling to address data imbalances.

[15] Our objectives were to develop and evaluate models that best explained observed ANC, identify key explanatory variables, predict ANC for a continuous stream network, and identify acid-sensitive streams and focal areas for future sampling, monitoring, and potential remediation.

[16] This research advances the work of *Sullivan et al*. [2007], who used logistic regression (logR) over a portion of the same region. Here we compare machine-learning and traditional model performance, address known imbalances in the ANC data distribution by incorporating hurdle modeling, evaluate a much broader set of potential predictors, average predictor variables to the upslope-contributing area of each stream water pour point across the study region, and provide continuous estimates of ANC across a larger region.

2. Methods

2.1. Study Area and Background

[17] The study area is the southern Appalachian Mountain region (14.3×10^6 ha), which extends from northern

Georgia to southern Pennsylvania, and from eastern Kentucky to central Virginia (Figure 1). The region is primarily composed of the Blue Ridge, Ridge and Valley, and Central Appalachian ecoregions [*Omernik*, 1987]. The Blue Ridge ecoregion is dominated by metamorphic and igneous parent materials, whereas the Ridge and Valley and Central Appalachian ecoregions are primarily sedimentary, with northeast to southwest trending sandstone ridges and limestone valleys (Figure 1). Elevations range from about 300 to 2000 m. The dominant land cover of the area consists of oak, hickory, pine, spruce, and hemlock forests, interspersed with crop and pasture lands, and urban areas.

2.2. ANC Data

[18] Water chemistry data were used to calculate stream water ANC. Data were obtained from national and regional databases, including the National Stream Survey, Environmental Monitoring and Assessment Program, Virginia Trout Stream Sensitivity Study (VTSSS), and others [see also *Sullivan et al*., 2007; *Sullivan et al*., 2004]. A total of 933 sampled sites were included in this study.

[19] Water chemistry data were collected mainly during the spring season between 1986 and 2009. Approximately 43% of data were collected during the 2000 VTSSS survey, 34% were collected between 1986 and 1996, and 23% were from 2003 to 2009. All water chemistry samples were georeferenced to a synthetic stream network created using a hydrologically conditioned 30 m digital elevation model (DEM) [*United States Environmental Protection Agency and United States Geological Survey*, 2005] within a geographical information system.

[20] ANC was calculated as the sum of the charge balance of Ca^{2+}, Mg^{2+}, K^{+}, Na^{+}, Cl^{-}, NO_3^{-}, NH_4^{+}, and

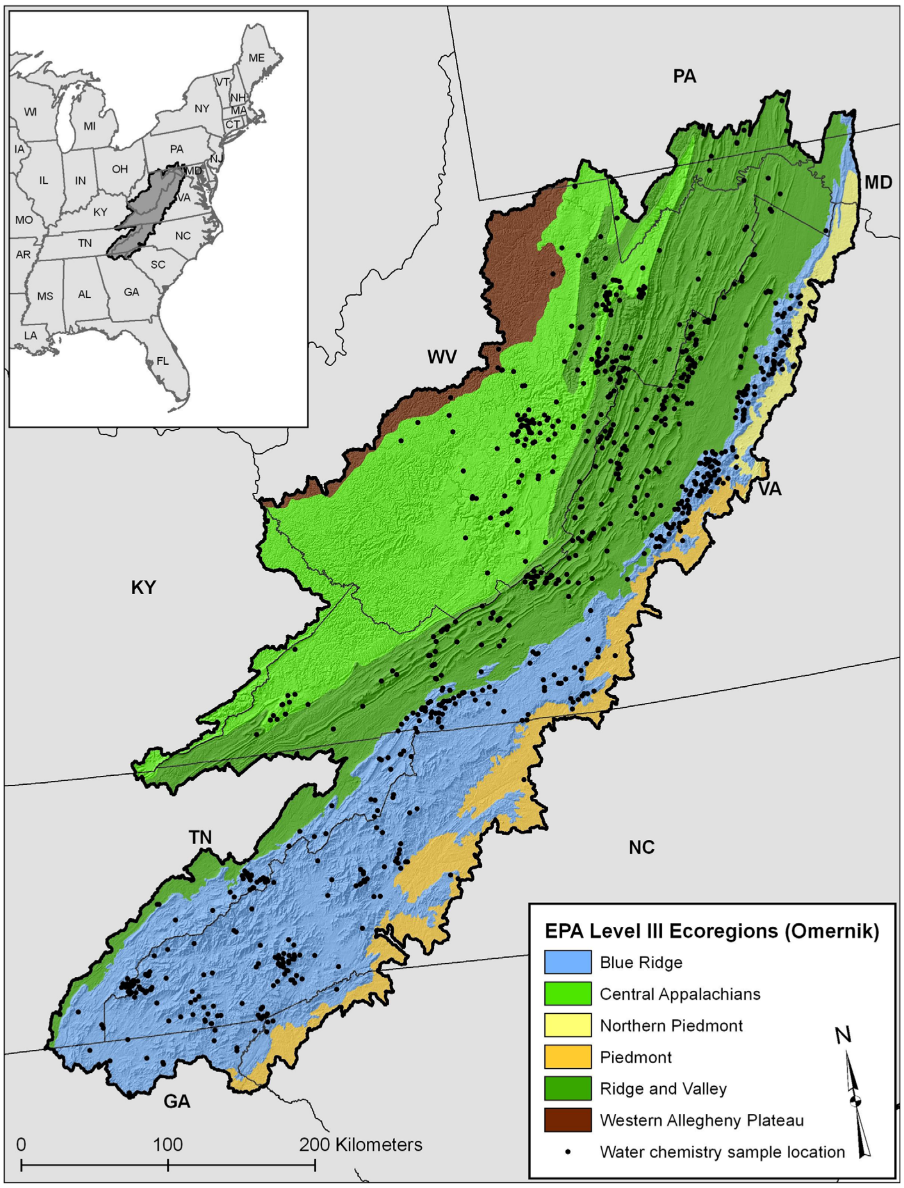

Figure 1. Distribution of water chemistry samples within *Omernik* [1987] ecoregions for the southern Appalachian Mountain region.

SO_4^{2-}. Calculated ANC values across the study area ranged from −109 to 3889 μeq L^{-1} (mean = 188 ± 414 (standard deviation, SD) μeq L^{-1}, median = 72), and the data were right skewed (Figure 2). We address this later in sections 2.6 and 2.7.

2.3. Predictor Variables

[21] An initial set of environmental predictors were chosen to represent broad- to fine-scale climatic, lithologic, geomorphic, topographic, edaphic, vegetation, land ownership, and S deposition conditions that were potentially influential to ANC (see Table S1 for complete details). All data layers were resampled to 30 m raster grids (Table S1), and data values were averaged across the upslope contributing area of each 30 m grid cell, using the methods described by *McDonnell et al.* [2012]. The equation for upslope-averaging is as follows:

$$\overline{P} = \frac{P_i + \sum_j^N P_j}{N+1} \tag{1}$$

where $\overline{P}$ is the upslope averaged value for the candidate cell (P_i), $\sum_j^N P_j$ is the summation of all cell values upslope of P_i, and N is the total number of upslope cells. Upslope-averaging enabled us to attribute the average of each predictor variable across the landscape draining into each individual cell. Environmental data were obtained from several sources, as described later.

[22] Eighteen climate variables representing 1961–1990 climate normals were taken from the 1 km resolution Ameriflux data set developed by *Hargrove and Hoffman* [2004] for the conterminous United States. Climate variables represented various aspects of the temperature, precipitation, and insolation regimes, each conditioned by local growing and nongrowing seasons. Five soil, one topography, two vegetation, and ten productivity variables were also provided by this data set (Table S1).

[23] The National Land Cover Dataset [*Homer et al.*, 2007] was used to quantify the percent areal coverage of major land cover types, including the cover of coniferous, hardwood, and of all forest types combined. Two additional classes were derived that combined mixed coniferous and hardwood forest by weighted averaging. A landownership variable was also included to represent the percentage of catchment area in federal versus nonfederal ownership [*National Atlas of the United States*, 2006]. Due to different management histories on federal versus privately owned lands in the eastern United States (more intensive logging on nonfederal lands), this layer was intended to provide proxy information on degree of past logging disturbance, a known influence on ANC [*Sullivan et al.*, 1999].

[24] Using a 30 m DEM, we derived three topographic variables: a steady-state topographic wetness index (TWI) [*Wood et al.*, 1990], surface area ratio (SAR), and flow accumulation (FAC). TWI was computed as the log of the catchment size divided by the catchment slope (radians), to represent the propensity of each grid cell to accumulate water [*Moore et al.*, 1993]. SAR measured terrain roughness as the ratio of sloped to flat surface area of a grid cell [*Jenness*, 2004]. FAC represented the total area contributing overland flow to a grid cell [*Jenson and Domingue*, 1988]. These variables represented watershed characteristics and were not upslope averaged.

[25] Additional soils data were obtained from the Soil Survey Geographic [*NRCS Soil Survey Staff*, 2010a] and the U.S. General Soil Map [*NRCS Soil Survey Staff*, 2010b] databases. Soil variables included percent clay, soil pH, and soil depth. A broad-scale lithology classification provided by *Sullivan et al.* [2007] was used to capture the percent composition of parent materials across the study area. Classes included siliceous, argillic, felsic, mafic, and carbonate substrates. Mapped surface lithologies were composites from State geologic maps [*United States Geological Survey*, 2005a, 2005b].

[26] Total wet and dry S deposition were calculated based on 3 year averages of the NADP (National Atmospheric Deposition Program) [*Grimm and Lynch*, 2004] interpolated wet (375 m resolution) and CMAQ (Community Multiscale Air Quality) [*Byun and Schere*, 2006] modeled dry deposition (12 km resolution), centered on the 2002 weather year.

[27] The initial set of predictors included 57 variables; only those with Pearson's correlation scores <0.7 were retained leading to a modeling set of 33 variables. Among correlated variables, those with the highest Pearson's correlation with ANC were retained. Other correlation cutoff values were considered and evaluated, but none improved the final models.

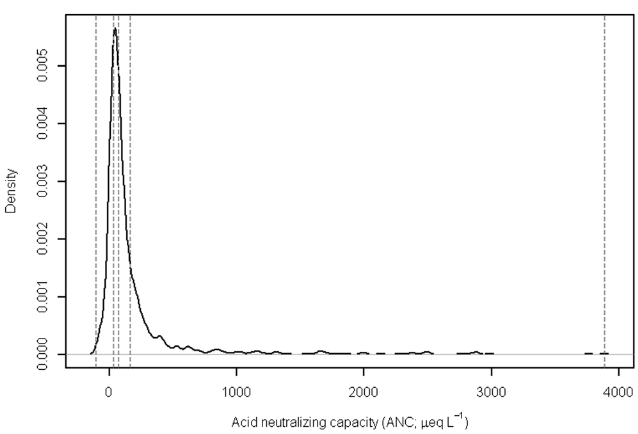

Figure 2. A plot of the kernel density function estimated for the 933 sampled ANC values across the southern Appalachian Mountain region. Gray vertical lines indicate minimum, first quantile, median, second quantile, and maximum ANC values.

2.4. Model Development

[28] To identify best modeling approaches, we compared traditional regression and machine-learning techniques using a common set of performance metrics (described later). Statistical models included linear models (LMs) and logR, random forest (RF), and boosted classification and regression trees (BCT and BRT).

[29] Exploratory ecological analyses like this one are opportunistic, involving combined data sets, which can yield imbalanced sampling designs and skewed data distributions [*Barandela et al.*, 2003; *Chawla et al.*, 2002]. While most statistical models, including advanced machine-learning

algorithms [*Elith and Leathwick*, 2009; *Elith et al*., 2008], assume balanced data designs, some produce robust predictions with moderate imbalances but few perform well with large imbalances. We explored a variety of techniques to ameliorate the influence of imbalanced data in our study and increase the overall accuracy of model predictions.

[30] We began by predicting ANC as a continuous response using LM, RF, and BRT models. Next, we used two-stage hurdle modeling, in combination with the regression techniques, to screen out sites with a high probability of being well buffered, and then predicted a continuous ANC value for the remainder. Finally, we tested two data resampling techniques that imposed a balanced data distribution within the hurdle modeling framework.

2.5. Machine-Learning Algorithms

[31] We compared the performance of BCT, BRT (*gbm* package in R v2.12.2) [*Ridgeway*, 2006], and RF (*randomForest* package in R v2.12.2) [*Liaw and Wiener*, 2002] models against LM and logR models (*stats* package in R) [*R Development Core Team*, 2011] to determine the best approach, considering the available data and the prediction goals. Models were evaluated using a 75%/25% training/testing set, over 50 iterations to ensure accuracy of predicted model error rates. Model error metrics are discussed later in section 2.6.

[32] Machine-learning algorithms are relatively new to ecological research [*Hastie et al*., 2005; *Olden et al*., 2008]. Many are data driven, meaning that models do not produce a parameterized statistical model but identify patterns in the data with few assumptions regarding the underlying probability distribution of the training data [*Breiman*, 2001b]. Accordingly, the main advantage of machine-learning algorithms is their potential to model complex and nonlinear relationships without having to satisfy the assumptions associated with parametric models (e.g., normally distributed residuals and linear relationships among dependent and independent variables). This is accomplished in a variety of ways depending on the type of learning used [*Franklin and Miller*, 2009; *Gahegan*, 2003; *Olden et al*., 2008].

[33] Boosted and RF methods represent ensemble versions of traditional classification and regression tree (CART) analysis that output a majority vote or mean value from a series of trees [*De'ath and Fabricius*, 2000; *Elith et al*., 2008; *Prasad et al*., 2006]. In lieu of a parameterized statistical model, CART splits the data into successively smaller and more homogenous groups until some stopping criterion is met. This is done by iteratively sorting each predictor variable and then splitting the data into two mutually exclusive groups at each iteration. This is repeated for every value of the predictor and for all predictors individually. The predictor and splitting value are chosen for each split that minimizes the within-group heterogeneity of the response variable [*Breiman*, 1984; *De'ath and Fabricius*, 2000].

[34] RF and boosted methods build tens to thousands of CART trees; predictions made from the models are based on a majority vote (classification) or averaged value (regression). Boosting works by using the entire set of predictors and data to build each individual tree, and the algorithm is sequential, using information about model residuals of past trees to guide development of subsequent trees [*De'ath*, 2007; *Elith et al*., 2008]. RF models use a bagging algorithm, in which individual trees are built using random samples of the predictors and of the data, each in a predefined proportion. Model performance is assessed using an independent test set, *n*-fold cross validation (BCT, BRT), or out-of-bag estimates (RF). Out-of-bag samples refer to the training data left while building individual regression trees. Mean out-of-bag error rates are estimated by models for each out-of-bag sample by entering them into the tree they were omitted from, calculating error estimates for each tree, and averaging error estimates across all trees.

[35] Variable importance measures are also calculated by these algorithms. For BRT/BCT, variable importance is based on the number of times the variable was chosen as a predictor in the individual trees and weighted by the deviance the variable explained across all trees [*Elith et al*., 2008]. For RF, variable importance is calculated as the difference between the error rate of an individual tree and the error rate of the tree calculated using randomly assigned values for the specific predictor, averaged across all trees in the RF model [*Breiman*, 2001a].

2.6. Hurdle Modeling

[36] Initially, we found that single regression models exhibited high error rates, particularly for streams with ANC values >150 $\mu\text{eq}\,\text{L}^{-1}$. A two-stage hurdle model was used to minimize RMSE of predicted ANC values (Figure 3). In a first stage, a binomial model (e.g., logR, BCT, and RF) predicted the probability that ANC values for any 30 m grid cell were below a specified threshold (e.g., ANC < 200 $\mu\text{eq}\,\text{L}^{-1}$). If a cell exhibited a high probability (e.g., probability > 0.5) of a low ANC value, that cell was entered into a second regression model (e.g., LM, BRT, or RF), where continuous ANC values were predicted. If a cell exhibited a low probability (e.g., <0.5) of a low ANC value, it was considered well buffered, assigned an arbitrarily high ANC value, and not considered further by the continuous model. Tested ANC threshold values were 150, 200, 250, and 300 $\mu\text{eq}\,\text{L}^{-1}$; tested probability cutoff values were (0.4, 0.5, 0.6, and 0.7).

[37] Threshold and continuous models were trained separately to identify the optimal statistical model, predictors, and parameters. Models were constructed using 3, 5, 7, 10, 15, 20, 25, and 33 of the most influential predictors. Performance was evaluated for each of these models to optimize model parsimony and prediction accuracy. Performance metrics were calculated for threshold and continuous models using a random 25% draw on the data set. Threshold models were compared using misclassification rate, κ statistic [*Maclure and Willet*, 1987], G mean, and area under the receiver operating curve (AUC) [*He and Garcia*, 2009]. Continuous models were compared using model RMSE (lower numbers indicate higher prediction accuracy) and coefficient of determination (R^2).

[38] Within the hurdle model, data resampling techniques were used to help reduce the influence of imbalanced data on model performance. These included both (1) randomly oversampling the "high-ANC" stream sample sites (those with values greater than the specified threshold, see this section above) until the sample number equaled that of

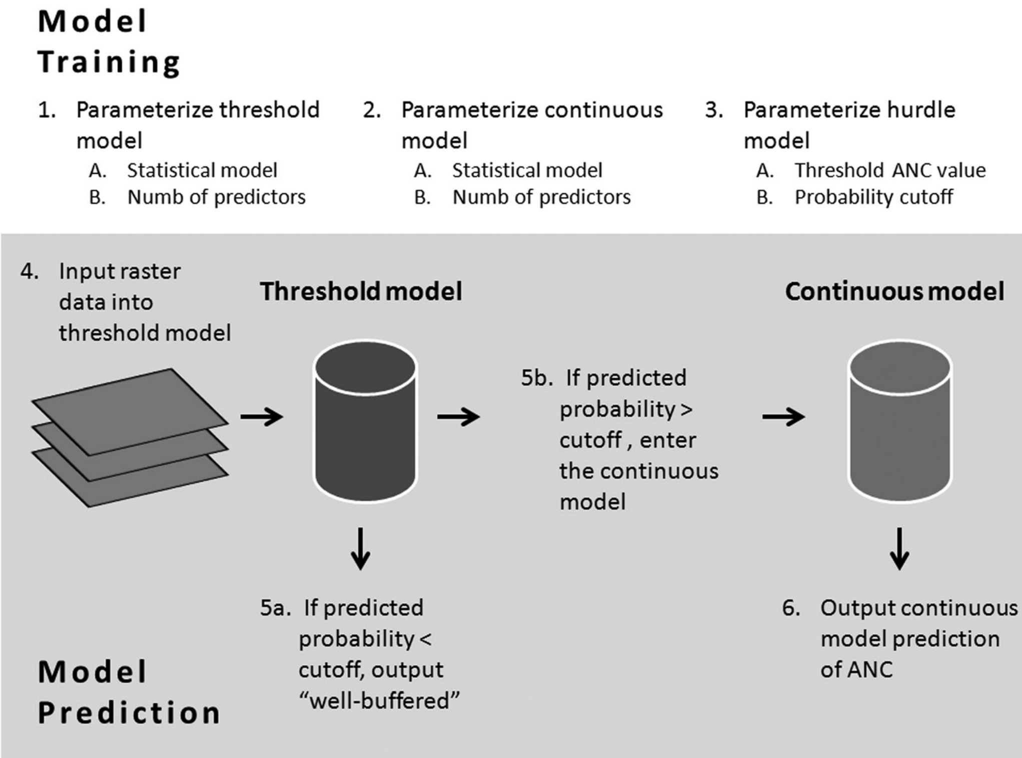

Figure 3. Conceptual diagram of the hurdle modeling framework for predicting ANC values in the southern Appalachian Mountain region.

the low-ANC sites, and (2) randomly undersampling the low-ANC sites until the sample sizes were equal.

2.7. Validating the Complete Hurdle Model

[39] Once both the threshold and continuous models were trained and the optimal models identified, the complete hurdle model (threshold + continuous models) was trained and evaluated using a randomly drawn training set consisting of 75% of the sampled water chemistry sites. During model training, three parameters were adjusted to optimize hurdle model parameterization: the threshold ANC value (150, 200, 250, or 300 μeq L^{-1}), the probability cutoff value (0.4, 0.5, 0.6, or 0.7), and the data resampling technique (no resample, oversample high ANC, undersample low ANC), resulting in 48 unique model parameterizations. All combinations of these parameterizations were run 50 times each, using unique random draws on the data to create the training and testing sets, and model statistics were averaged to identify optimal hurdle model parameterization. All model parameterizations were compared by plotting mean ($\pm$SD) RMSEs for the continuous models against mean ($\pm$SD) false-positive rates for the threshold models. The false-positive rate, as opposed to overall model error rate, was used to minimize instances in which a truly low-ANC site was erroneously predicted as high ANC. The "best" model parameterization exhibited the lowest values for the two error rates.

2.8. Spatial Autocorrelation

[40] Spatial autocorrelation in the final ANC model was assessed with a Moran's I correlogram on model residuals using the *spdep* package within R [*Bivand et al.*, 2011]. This statistic varies between -1 (perfect negative correlation) and 1 (perfect positive correlation); values approaching zero indicate complete spatial randomness. The correlogram also displays the results of null hypothesis testing at varying geographic distances.

3. Results

3.1. LM Regression and Machine Learning

[41] We initially predicted continuous surfaces of ANC using RF, BRT, and LM. RMSE of predictions (based on independent test sets withheld from model training over 50 iterations) were 7%–14% lower for machine-learning algorithms (RF: 258.1 $\pm$ 36.2, BRT: 277.2 $\pm$ 38.7) compared with LM regression (298.3 $\pm$ 37.4), but error rates were relatively high for all models in relation to the scale of the ANC values. Error rates for all models were highest for sites with ANC $>$150 μeq L^{-1} (Figure 4).

3.2. Hurdle Modeling Training

3.2.1. Threshold and Continuous Model Selection

[42] Compared to the initial continuous models, hurdle models reduced overall RMSE rates and identified key predictors contributing to low ANC. Within the hurdle modeling framework, RF models displayed the lowest error rates of all tested (Figure 5). Across all models, performance declined when fewer than 10 predictors were entered into the model (Figure 5).

[43] RF threshold models exhibited AUC scores $>$0.9, κ scores $>$0.7, and error rates $<$8% when models included between 7 and 20 predictors (Figure 5). Over this range of predictors, LM exhibited $\sim$10% error rates, AUC scores

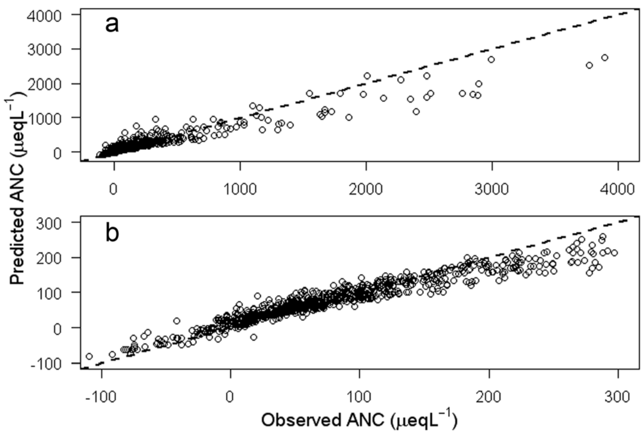

Figure 4. Predicted versus observed ANC values resulting from (a) the RF-only model (i.e., single RF regression model) and (b) the final hurdle model. Unlike Figure 6, where models were developed using a testing/training set, models here were trained and predicted using all 933 sample sites. Note the difference in x- and y-axis scaling between Figures 4a and 4b.

<0.9, and κ scores ~ 0.5. Similar results were found for the continuous models where R^2 for RF models were ≥ 0.5 compared to 0.40–0.48 for LM, and RMSE rates were consistently high for RF.

[44] The best final hurdle model included RF threshold and continuous models. Both models included 10 predictors and were derived from 1000 individual regression trees (*ntree* parameter) with a bagging fraction of 30% (*mtry* parameter).

3.2.2. Variable Importance

3.2.2.1. Threshold Model

[45] From the threshold model, low ANC occurred in small forested catchment areas with noncarbonate lithology, cool and moist climates, with relatively low soil pH and clay levels (Table 1 and Figure 6). With few exceptions, predictors influenced ANC in a nonlinear manner, and nonlinearities occurred at the ends of the response curves. For example, the probability of a low value of ANC increased nonlinearly with average percent forested cover, but the majority of samples (Figure 6, sixth panel, hash marks on the x axis) had between 90% and 100% forested cover.

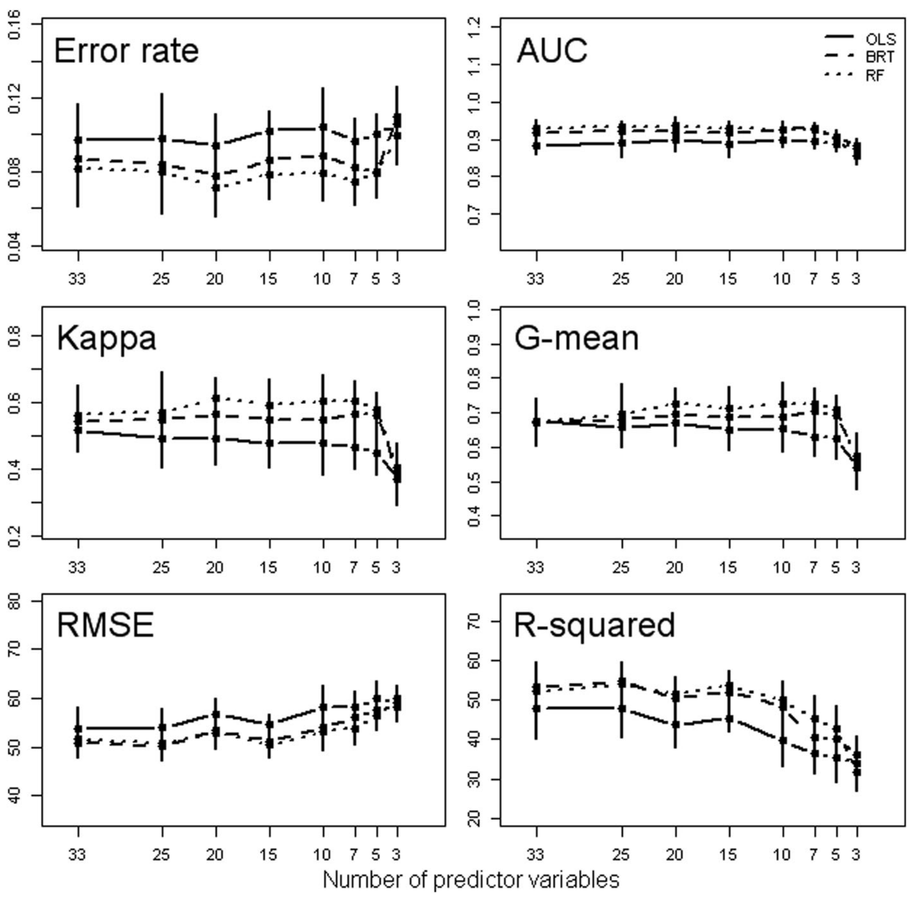

Figure 5. Validation results for (top two rows) threshold and (lower row) continuous models of the hurdle modeling framework. Comparisons are shown for the ordinary least squares (LM; continuous model only), BRT, and RF models. *Abbreviations*: AUC, area under the receiver operator curve; RMSE, root-mean-square error. Each unique combination of models was run 15 times to obtain estimates of variability due to the collection of training data unique to each run. Error bars represent the $\pm$SD of the error estimates. Variables included in each model were chosen using the top predictor variables based on the average variable importance measures from the BRT/BCT and RF models. Error rate, AUC, κ, G mean, and R^2 statistics are unitless, and RMSE has units of μeq L^{-1}.

Table 1. Predictor Variables Included in the (Left) Threshold Model and (Right) Continuous Model Within the Hurdle Modeling Framework[a]

Threshold Model Variables	Short Name	Relative Importance	Continuous Model Variables	Short Name	Relative Importance
Percent carbonate lithology	LITH_CAR	20.23	Percent siliceous lithology	LITH_SIL	18.64
Soil pH	SOIL_PH	12.20	Mean penultimate maximum days without precipitation while $\geq$10°C	NPDAYMAX	11.98
Mean penultimate maximum days VPD <1000 Pa while $\geq$10°C	VWDAYMAX	12.08	Mean number of GS days above 32.2°C	AB90GROW	11.51
Mean 95% of maximum GS temperature difference	DIFF95GR	10.68	Mean penultimate maximum days vapor pressure deficit <1000 Pa while $\geq$10°C	VWDAYMAX	9.24
Percent land in public ownership	PUBLIC	10.60	Dry sulfur deposition	S_DRY	9.13
Percent forest cover	FOREST	8.86	Percent forest cover	FOREST	8.79
Percent argillic lithology	LITH_ARG	7.14	Percent soil clay	SOIL_CLAY	8.42
Mean precipitation sum during the local non-GS	PRECIPNG	6.83	Soil pH	SOIL_PH	7.99
Mean penultimate maximum consecutive days VPD >750 Pa while $\geq$10°C	VDCONTDAY	6.17	Topographic wetness index	TWI	7.38
Mean non-GS gross primary productivity	GPPNG	5.20	Flow accumulation	FAC	6.92

[a]GS, growing season. Relative importance for a single predictor variable is based on the mean increase in mean-square error (regression) or model accuracy (classification) for each decision tree when the values of the predictor variable are randomized during model calibration. The values have been standardized to sum to 100%. See Table S1 for description of the predictor variables.

[46] Significant interactions among variables were identified by the RF model for percent calcareous, percent public land, the maximum number of very wet days (i.e., vapor pressure deficit <1000 Pa, VWDAYMAX), and maximum number of continuous very dry days (i.e., VPD >750 Pa; VDCONTDAY; Figure S1). Areas with <10% calcareous lithology in fairly moist environments had the highest probabilities of being low ANC; probabilities decreased most with slight increase in calcareous lithology and in the driest environments. Other important interactions included the amount of catchment area in public lands, and the highest probabilities of low ANC occurred on public land and in moist environments. However, the interaction between public lands and carbonaceous lithology suggested that probabilities were highest outside of public lands and on noncalcareous lithology. Slight increases in public lands reduced the probability of low ANC, but probabilities increased slightly with increased amounts of public land; increases in calcareous lithology reduced the probability of low ANC only for catchments with negligible amounts of public land.

3.2.2.2. Continuous Model

[47] From the continuous model, low ANC occurred in areas with siliceous lithologies, a relatively moist, cool, and short growing season, in conditions with low soil pH and clay levels, and in small, forested catchment areas (Table 1 and Figure 7). Like the threshold model, most variables had nonlinear response functions, and nonlinearities generally occurred at the extremes of the response curves.

[48] Important interactions were identified for the amount of siliceous lithology, precipitation, temperature, forest cover, and topography variables (Figure S2). TWI had a highly nonlinear influence on ANC levels (Figure 7), but when interacting with the level of siliceous lithology this was less apparent and ANC appeared most influenced by lithology. Interactions among climate variables clearly showed that low ANC was associated with low temperatures and high precipitation. Catchments with warm climates appeared to reduce the influence of siliceous lithology on ANC; catchments with 80% of their area in siliceous lithology and >10 days with temperatures >32.2°C had a predicted ANC of ~140 $\mu eq\,L^{-1}$ compared to only ~50 $\mu eq\,L^{-1}$ in the coldest environments. The amount of forested area had a similar influence, and catchments with <90% forest cover and <80% siliceous lithology appeared well buffered.

[49] To better understand differences in important drivers of ANC at high compared to low elevations, we stratified low (<500 m, first quartile) from high-elevation (>850 m, third quartile) sites and developed continuous RF models for observations with ANC $\leq$300 $\mu eq\,L^{-1}$ separately for each elevation setting. Drivers of low ANC for the low-elevation model were similar to those of the full model (Figure S3), but in the high-elevation model, low ANC was found in areas with high coniferous and low deciduous cover (Figure S4). Wet S deposition was also a leading predictor in the high-elevation model, but not in the low-elevation model or hurdle model, and ANC was lowest for intermediate levels of deposition.

3.2.3. Hurdle Model Performance

[50] Based on model performance results presented in Figure 8, the final model parameterization included RF threshold and continuous models, imbalanced data (no resampling), and a 0.7 probability cutoff, with a 300 $\mu eq\,L^{-1}$ ANC threshold (model 46, Figure 8, Table S2). This model displayed a 5.6% omission error rate, 9.5% overall error rate of classification, and an RMSE of 107.5 $\mu eq\,L^{-1}$ based on 50 model iterations. Overall, hurdle modeling reduced model RMSE by 140%–177%. Model RMSE was lowest for stream segments with ANC <150 $\mu eq\,L^{-1}$ (Figure 4); but this model consistently underpredicted ANC for values >150 $\mu eq\,L^{-1}$. Underprediction was likely associated with few sampled water chemistry data sites in streams with ANC >150 $\mu eq\,L^{-1}$. When only sites with

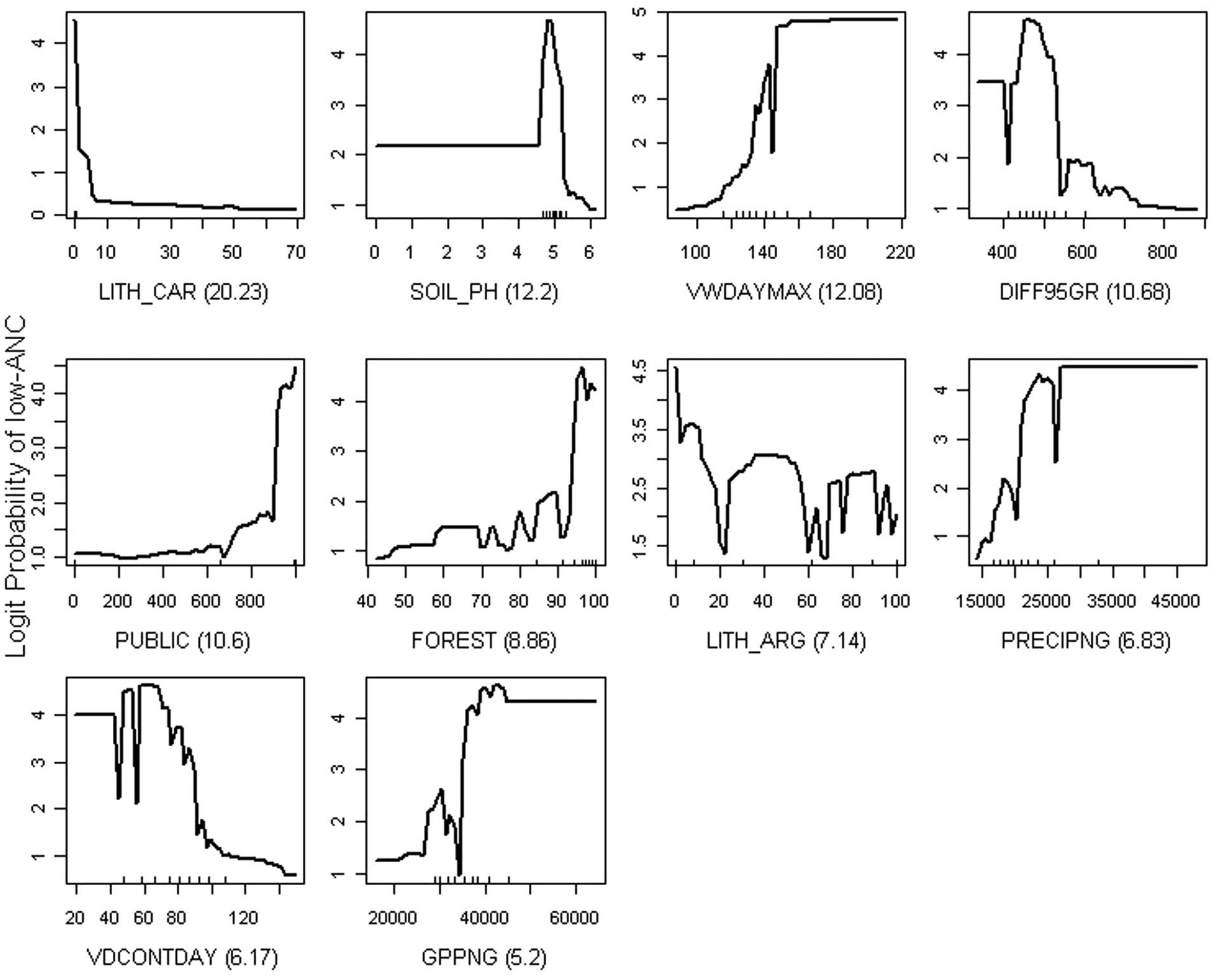

Figure 6. Response curves showing relations between the predicted ANC and individual predictor variables included in the threshold model within the hurdle modeling framework. Black tick marks on the *x* axis indicate decile classes for the predictors. The *y* axis indicates the relative effect of the predictor on ANC on a logit scale. In general, higher *y*-axis values indicate a higher probability of predicting a low ANC value. See Table S1 for a description of predictors.

ANC <150 $\mu eq\,L^{-1}$ were included in the model, RMSE was 36.6 $\mu eq\,L^{-1}$.

[51] The majority of the RMSE for well-buffered sites was related to threshold model misclassifications of high-ANC sites as low; a behavior consistent across all models tested. When misclassifications were removed from the error analysis, the resultant RMSE was 49.4 $\mu eq\,L^{-1}$ for the full model. Results of the spatial autocorrelation analysis indicated that spatial autocorrelation among model residuals was negligible (Table 2).

3.2.4. ANC Model Predictions

[52] Stream water chemistry often associated with biological harm (e.g., ANC < 100 $\mu eq\,L^{-1}$) occurred throughout the study area (Figure 9). Approximately 8% of the stream network displayed ANC values <50 $\mu eq\,L^{-1}$, and 24.5% with values <100 $\mu eq\,L^{-1}$ (Figure 10). Lithologic patterns clearly influenced ANC predictions, with high-ANC areas occurring mainly in limestone valley bottoms, and low ANC occurring on siliceous bedrock.

[53] We compared predictions from RF hurdle models to those from a linear regression model (LM). Differences in the patterns of predictions made by the LM and RF hurdle model were apparent. The LM predicted >84% of the stream network (as measured by length) had ANC values >150 $\mu eq\,L^{-1}$, compared to only 58% for the RF hurdle model (Figure 10). The RF model predicted 35% of streams had ANC between values 50 and 150 $\mu eq\,L^{-1}$ compared to only 8.9% predicted by LM (Figure 10). For the most acidic class (i.e., ANC < 0 $\mu eq\,L^{-1}$), the LM predicted a slightly higher percentage of streams (3.9% LM versus 0.9% RF hurdle).

[54] LM tended to predict higher ANC levels particularly in the central Appalachian region (Figures S5 and S6). However, in the southern Blue Ridge and in pockets of eastern and central West Virginia, where the RF hurdle model predicted low ANC rates (e.g., ANC < 50 $\mu eq\,L^{-1}$), the LM generally predicted very low ANC rates (Figures S5 and S6). Low ANC predictions from the LM were fairly isolated geographically and generally corresponded to areas of highest sampling intensity, potentially indicating that the LM had poor predictive ability outside the limits of the training data.

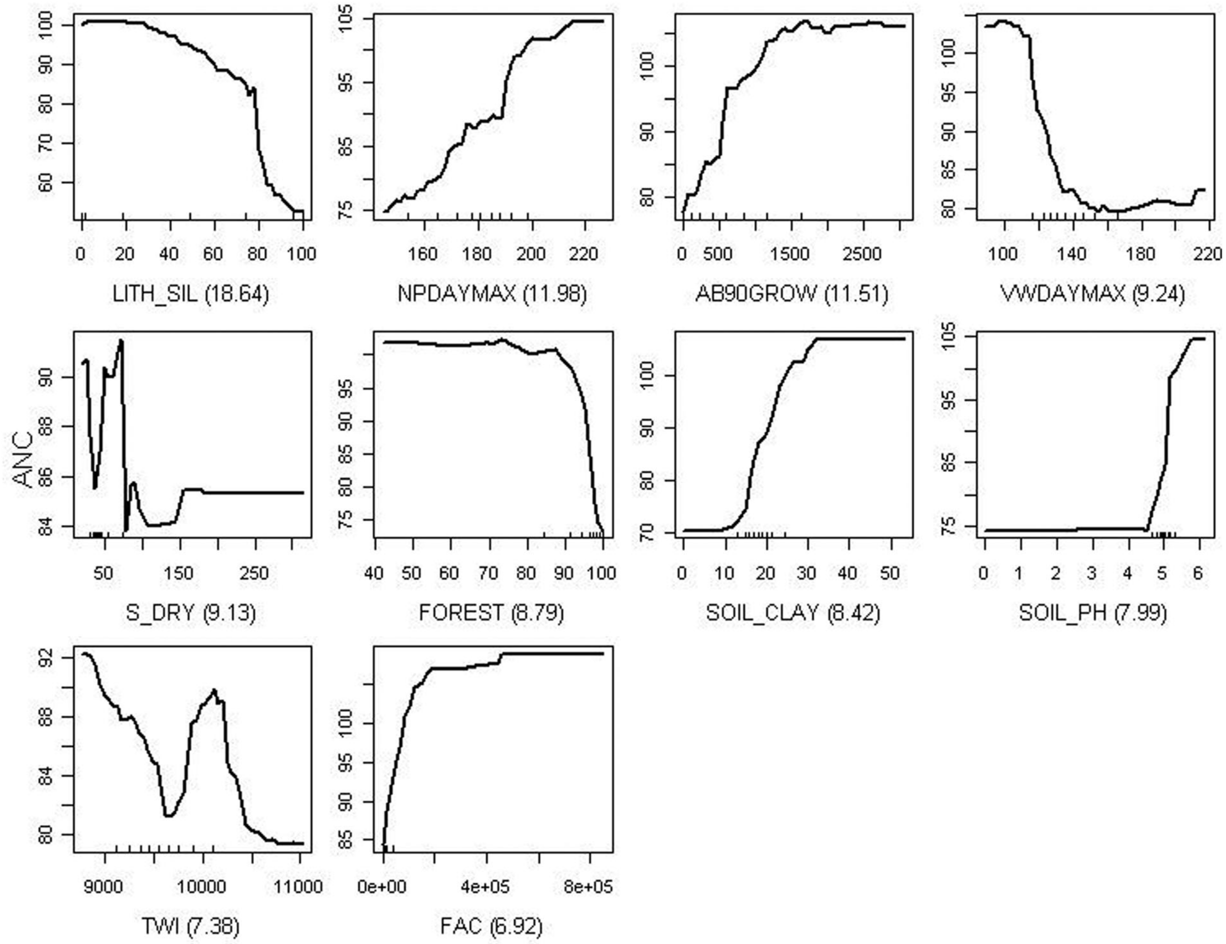

Figure 7. Response curves showing relations between the predicted ANC and individual predictor variables included in the continuous model within the hurdle modeling framework. Black tick marks on *x* axis indicate decile classes for the predictors. The *y* axis indicates the relative effect of the predictors on ANC. In general, lower *y*-axis values indicate lower ANC values. Note that the majority of climate variables have been multiplied by a constant, which had no effect on predictions or model performance. See Table S1 for a description of the predictor variables.

[55] Spatial patterning of RF model uncertainty was also apparent across the study area. Model uncertainty was expressed as the SD of continuous predictions made among the ensemble of individual trees comprising the RF model. As might be expected, model uncertainty was highest in the western portion of the study domain where few samples occurred (Figure 9) and lowest in the areas with the most samples.

4. Discussion

4.1. Modeling Low ANC

[56] Biogeochemical and climatic influences on ANC almost certainly involve many interactions among environmental and climatic processes that together influence the ultimate susceptibility of a stream to the acidifying effects of S deposition. Some of these factors are as yet unknown, and others are known, but complex, nonlinear, nonstationary over space and time, and difficult to measure [*Driscoll et al.*, 1987; *United States Environmental Protection Agency*, 2009]. Known processes include (1) nutrient uptake by plants, transport of litter fall, and delivery to neighboring streams; (2) mobilization of Al^{3+}; (3) erosion and sedimentation of cations and anions from upslope catchments; (4) deposition of S from exogenous sources; and (5) physical and chemical weathering of parent materials proximal and distal to the stream channel [*Christophersen and Neal*, 1990; *Driscoll et al.*, 1987; *Johnson and Host*, 2010; *Reuss et al.*, 1987; *Sullivan*, 2000]. Modeling the susceptibility of aquatic systems to acidic deposition at a regional scale requires an understanding of these processes, access to adequately resolved data layers that accurately portray the complexity of environmental conditions, and reliable statistical modeling.

[57] Our aim was to predict how integrated biological, geological, chemical, and climatic processes within catchments can be used to predict stream water chemistry, which has linear flow properties that may concentrate or dilute local buffering capacity within stream networks, and which result in a unique correlation structure among observed responses and the environments driving the response. Results from this study indicated that predictions of ANC

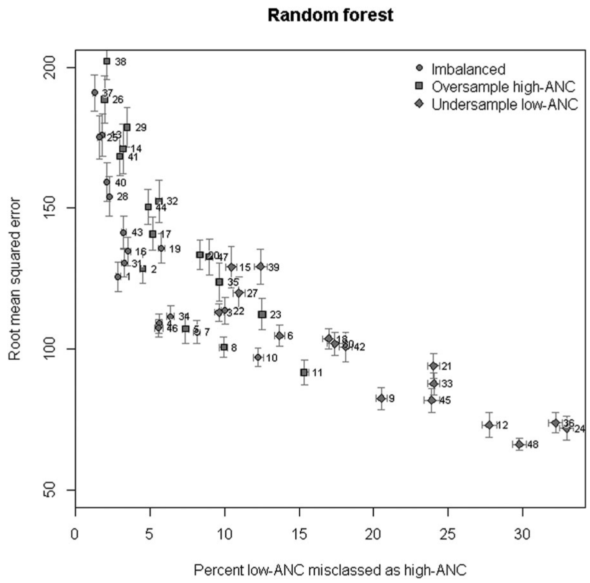

Figure 8. Scatterplot of the performance of 48 RF continuous models varying by combination of data resampling method (3), ANC threshold (4), and probability threshold (4). Validation statistics are based on predictions to a randomly drawn 25% subset of the data. A complete list of model results is included in Table S2. Model number 46, which showed the lowest RMSE and percent misclassified, was used as the final model.

within a region can be made using a combination of remotely sensed, spatially explicit environmental data, a representative sample of water chemistry across the region, and a modeling framework appropriate for the analysis.

[58] We addressed several aspects of the environmental and stream water chemistry during the modeling process including the spatial scale of the environmental predictors, the timespan over which water chemistry was sampled, and the spatial distribution of water sample sites.

[59] Stream water chemistry is governed by biological and physical processes that are both endogenous and exogenous to the contributing watersheds [*Sullivan et al*., 2007]. Water chemistry may be sampled at points along streams, but the environmental and edaphic influences on that chemistry must be related to the contributing space assignable to those points [*Johnson and Host*, 2010; *Steel et al*., 2010]. We accounted for this by using an upslope-averaging technique [*McDonnell et al*., 2012] that generated values for predictor variables that reflected conditions over the contributing area to each point along the stream network. Point estimates of the biogeoclimatic setting at the location of water chemistry sample points would not adequately represent catchment level influences on water chemistry.

Table 2. Results From a Moran's I Correlogram on Model Residuals From the Hurdle Model[a]

Lag	Estimate	Expected	Variance	SD	p
1	−0.001	−0.001	0.000	−0.014	0.989
2	0.004	−0.001	0.000	0.815	0.415
3	−0.007	−0.001	0.000	−1.058	0.290
4	−0.002	−0.001	0.000	−0.130	0.897
5	**−0.016**	**−0.001**	**0.000**	**−2.116**	**0.034**
6	−0.001	−0.001	0.000	0.099	0.921

[a]Values represented in bold indicate significant autocorrelation.

[60] The water chemistry data used in this study were obtained from a series of sampling efforts undertaken intermittently over a span of years (1986–2009). These data were used because they were readily available, sampled a fairly large spatial extent, were collected at fairly high density, and used a consistent method for calculating stream water acid-base status. All sample data used for modeling were collected during spring flow. As such, the model represents a description of the environment expected to influence spring season ANC concentrations, which are often the lowest that occur in these streams [*Herlihy et al*., 1993]. Although S deposition declined substantially between 1986 and 2009, stream water ANC has shown little recovery, mainly because of base cation depletion and continued S adsorption on soils [*Sullivan et al*., 2011, 2008, 2004]. As a result, we expect that data collected over a period of two decades generally reflect ambient stream acid-base chemistry.

[61] The lack of an a priori sampling design necessitated an assessment of potential sampling bias. The effects of data resampling on model prediction were mixed. Models that undersampled low-ANC sites (the majority sample) reduced RMSE but increased threshold model error rates. Models that oversampled high-ANC sites (the minority sample) produced negligible differences in model performance (Figure 8). Hence, rebalancing the sample was unnecessary, and RF models performed reasonably well, even with unbalanced sampling. Our results suggest that the multivariate signature of low ANC was unique enough to be readily detected against the background noise of other ANC levels. Moreover, the RF and boosted models consistently outperformed traditional regression approaches, and two-stage hurdle modeling adequately minimized effects of sampling bias on predicted ANC.

[62] Comparison of mean RMSE values for RF models with and without hurdle modeling showed that the hurdle model provided a more than twofold decrease in model RMSE (258.1, RF-only, and 107.5, hurdle model). The RMSE for ANC predictions <150 $\mu eq\, L^{-1}$ was 36.6 $\mu eq\, L^{-1}$. This observation is key: we can accept higher error rates in predictions of high ANC but require more accurate predictions where ANC values are within the range of known ecological degradation. Model performance was substantially improved for the ANC conditions that are most biologically relevant (≤ 100 $\mu eq\, L^{-1}$).

[63] A major aim of ecological models is often to predict various phenomena or responses over large geographic areas. When models produce sufficiently accurate predictions, they provide valuable insight into pattern and process interactions, and variability of those interactions. Our model allowed us to (1) identify areas of potential concern for ecological remediation and management; (2) identify discrete areas where exogenous factors such as the broad-scale climatic setting or large areas of heavy acidic S deposition may trump local factors such as topography or soils (e.g., the western portion of the region where ANC predictions were low despite deeply dissected topography; presumably due to precipitation and S deposition patterns); (3) identify areas where spatial clustering of environmental degradation is predicted; and (4) locate areas of high

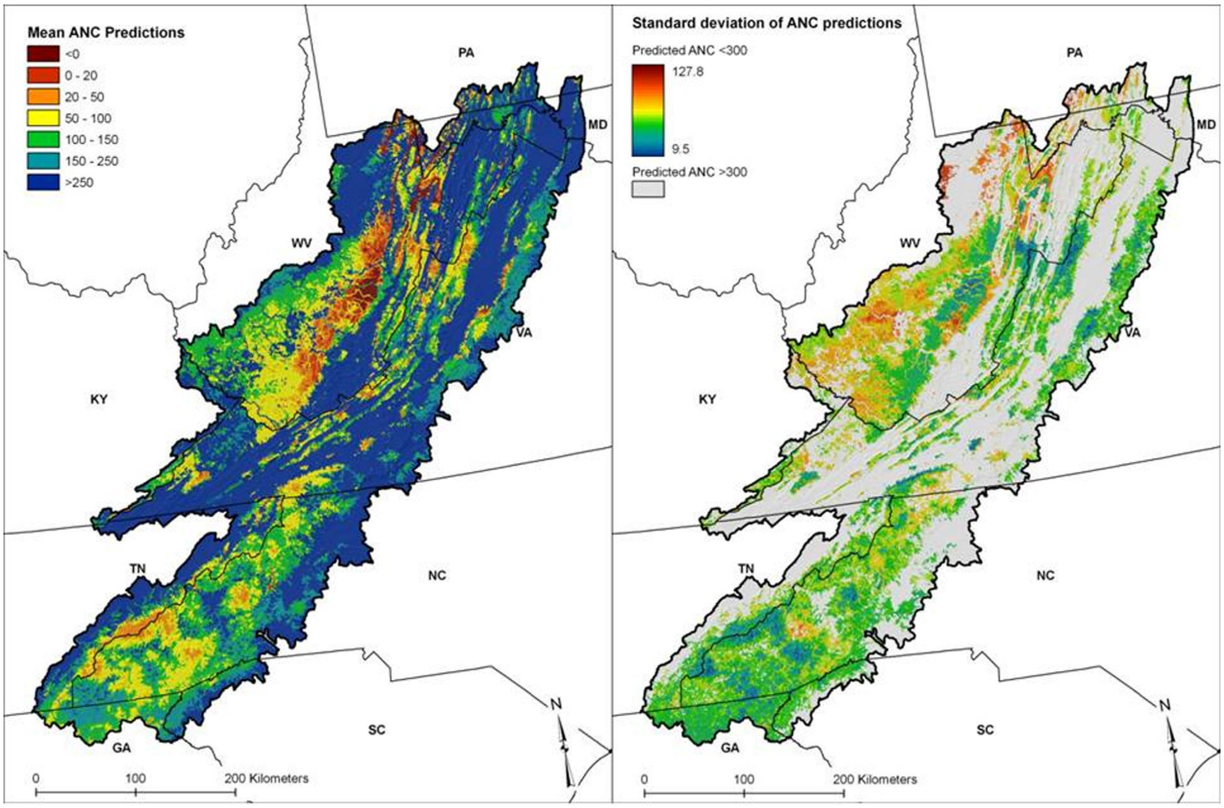

Figure 9. (left) ANC predictions from the best hurdle model, number 46 (see also Figure 6) and (right) the SD of predictions made by the continuous model within the final hurdle model. SDs were calculated from the predictions made from the ensemble of 1000 individual regression trees that made up the continuous RF model. Areas shown in white (filtered out by the threshold model) were predicted to exhibit ANC values >300 μeq L^{-1} and were not submitted to the continuous model).

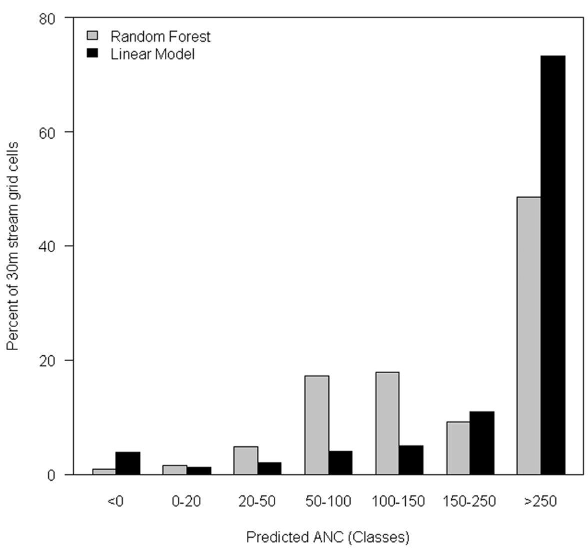

Figure 10. Histogram of the percentage of 30 m stream network grid cells predicted to be within seven ANC (μeq L^{-1}) classes. Dark shaded bars represent the predictions made from the linear regression model, and the lighter bars are predictions made from the RF hurdle model.

uncertainty, which may indicate inadequate sampling or variable interactions among the predictors.

[64] Knowledge of the uncertainty in model predictions is essential to evaluating model performance across geographic areas. Where model uncertainty is high, there may be deficiencies in sampling, unique environments not captured by the training data, lack of phenomenological understanding, or errors in model parameterization. Model validation results indicated that higher model uncertainty occurred in areas where observed ANC values were between 100 and 300 μeq L^{-1}; mapped SD values (Figure 9) confirmed this observation. These areas would benefit from additional sampling and monitoring.

[65] Predictions from our model indicate that streams with ANC <100 μeq L^{-1} make up approximately one quarter of the total stream length in the study area. ANC levels <100 μeq L^{-1} are associated with potential reductions in fish and macroinvertebrate taxonomic richness [*Cosby et al.*, 2006]. Approximately 10% of the total stream length in the study area was predicted to exhibit ANC <50 μeq L^{-1}. Streams with ANC <50 μeq L^{-1} are sensitive to episodic acidification, and fish species richness is greatly reduced [*Schindler*, 1988]. Low-ANC streams were generally located in the mountainous regions of the Central Appalachian and Blue Ridge Mountain provinces and in the central Ridge and

Valley Province (Figure 9). These areas are characterized by rugged topography, relatively steep temperature and precipitation gradients, high percentage of forest cover, and a variety of geologic parent materials.

[66] In lieu of using a statistical modeling approach, previous estimates of regional stream water ANC in our study area have been made using stratified random sampling designed to evaluate the proportion of surveyed streams contained within specified ANC classes [*Herlihy et al.*, 1993]. Estimates from these surveys generally corroborate our findings when individual ecoregions are assessed, although direct comparison is not possible. The authors concluded that about 13%–25% of the sample streams displayed ANC values <50 μeq L^{-1}; our area prediction was lower ($<10\%$).

4.2. Endogenous and Exogenous Drivers of Stream Water Acidity

[67] Results from the ANC modeling suggested low ANC of streams in the southern Appalachian Mountain region derived from lithologies, land-surface forms, temperature and precipitation regimes, exogenous S deposition, and endogenous patterns of physiognomies, soils, and topographies within catchments.

[68] Streams situated on siliceous lithologies had relatively low ANC, and this variable was the strongest predictor in the continuous model. Other studies have also found strong empirical relationships between lithology and stream water ANC [*Herlihy et al.*, 1993; *Puckett and Bricker*, 1992; *Sullivan et al.*, 2007]. Soils created from siliceous materials tend to be shallow, acidic [*Krug and Frink*, 1983], and generally exhibit low productivity, conductivity, and BC_w [*Herlihy et al.*, 1993]. Despite the known importance of lithology on governing stream water quality and other ecological processes, regionwide data layers describing mineralogy [*NRCS Soil Survey Staff*, 2010a] are of inconsistent quality and resolution [*Sullivan et al.*, 2007]. For instance, areas of West Virginia and Georgia delineate mineralogy at different resolutions, which likely contributes to reduced model performance in these areas (Figure 9). Because lithology and soil characteristics were important predictors in both the threshold and continuous models (Table 1), we suspect that improved quality, resolution, and consistency in soil and geologic data would improve model prediction and spatial accuracy.

[69] Low ANC was also associated with a high percentage of forest cover (Figures 6 and 7). *Sullivan et al.* [2007] and *Herlihy et al.* [1998] found a similar relationship between low ANC and high forest cover for streams within our study area. Afforested areas in Europe have also been found to consistently incite high levels of acidity in soils and stream water [*Gee and Stoner*, 1989; *Jenkins et al.*, 1990; *Miles*, 1986; *Whitehead et al.*, 1988]. Forests can have a variable influence on water chemistry depending on the prevailing climate, underlying geology, catchment size, successional stage of forest development, forest composition, and legacy influences of past disturbances, among others [*Harriman and Morrison*, 1982; *Sullivan*, 2000]. Forests generally scavenge wet and dry deposition of atmospheric pollutants (including, but not limited to S) and translate these pollutants to watershed soils and the associated stream reach. Furthermore, forests can take up base cations and reduce the capacity of soils to buffer against acidic inputs. However, forested areas in the Appalachian region are generally located at higher elevations and on ridges where catchments are small and reside on siliceous bedrocks [*Herlihy et al.*, 1993]. Therefore, the geographic locations of forested areas may be coincident with areas predisposed to having low ANC due to the edaphic and lithologic setting.

[70] When high-elevation sites were modeled alone, percent coniferous forest cover was the leading variable in the model, followed by deciduous forest cover. Moreover, these cover types displayed opposite influences on ANC (i.e., high coniferous cover was associated with low ANC, and high deciduous cover was related to high ANC; Figure S4). This observation is generally consistent with other works [*Cronan et al.*, 1978; *Nihlgård*, 1970]. Indeed, high-elevation catchments dominated by deciduous forest ($\geq 60\%$ cover; threshold value, Figure S4) had mean ANC of 83.8 μeq L^{-1}, compared to -19.8 μeq L^{-1} for catchments with a lesser deciduous component. In our data set high-elevation deciduous forests were generally located in larger catchments, with a lower percentage of siliceous parent material, less dry S deposition, and warmer and drier climates than coniferous forests. Whether coniferous forest types contributed to low ANC at higher elevations through acidic foliage deposition, pollutant sequestration, and subsequent acidic throughfall and stemflow [*Dunford et al.*, 2012; *Miles*, 1986], or if the environmental setting of coniferous forests influenced low ANC remains unclear. It is plausible that the influences of catchment size, mineral substrate, weathering rates, and dominant vegetation are highly interactive and change across the study region. The continuous model in our study identified important interactions among forest cover and lithology, and ANC was higher in catchments dominated by siliceous lithology when forest cover was low, compared to well-forested catchments with similar lithologies (Figure S2).

[71] Although elevation is generally considered a surrogate for climate at broad scales, elevation was not explicitly included as a predictor variable because remotely sensed climate data are now regionally available at a scale and accuracy sufficient for our application. Predictive modeling is most robust when predictor variables expected to be directly linked to the modeled response are used in place of those that may have indirect effects [*Austin*, 2002]. For example, elevation is often included as a predictor in species distribution modeling [*Austin*, 2002], but rarely do species respond to elevation alone. Rather, more direct predictors such as temperature and precipitation, which often are correlated with elevation, are more directly related to a species' distributional limits [*Austin*, 2002; *Elith and Leathwick*, 2009].

[72] Elevation is often used as a predictor in the ANC literature, where increases in elevation are often associated with (1) increased S deposition [*Lawrence et al.*, 1999], (2) higher precipitation [*Sullivan et al.*, 1999], (3) cooler temperatures, (4) higher percent siliceous parent material, (5) thinner and coarser soils with lower cation exchange capacity, (6) smaller catchments with steep slopes, and (7) higher percent forest cover [*Herlihy et al.*, 1998, 1993; *Lynch and Dise*, 1985; *Sullivan et al.*, 2007], all of which lead to decreased surface water ANC [*Herlihy et al.*, 1993;

Sullivan et al., 2007; *Sullivan et al*., 2011]. However, it is not clear how these factors rank in importance in driving ANC at regional scales.

[73] Our modeling results indicated that elevation was either uncorrelated or loosely correlated with stream water ANC ($r = -0.21$), total S deposition ($r = 0.11$), percent siliceous lithology ($r = -0.04$), soil depth ($r = 0.06$), soil clay ($r = -0.26$), catchment size ($r = -0.09$), and forest cover ($r = -0.01$). However, elevation was correlated with some, but not all, climate variables. These variables included number of days >32°C (AB90GROW, $r = -0.70$), nongrowing season precipitation levels ($r = 0.63$), number of days without precipitation (NPDAYMAX, $r = -0.63$), number of humid days (VWDAYMAX, $r = 0.70$), nongrowing season solar insolation (SOLARTOTNG, $r = 0.61$), and nongrowing season gross primary productivity (GPPNG, $r = 0.60$). Three of these climate variables (NPDAYMAX, VWDAYMAX, and AB90GROW) were among the top five variables in the continuous model in our study and indicated that low-ANC streams coincided with areas that experience few precipitation-free days, few dry days, and few days >32°C. These variables had higher variable importance scores than other edaphic, topographic, vegetation, and sulfur deposition variables in the model, suggesting that climate may drive regional stream water acidity.

[74] The importance of identifying climatic correlates of stream water ANC with our models has implications for monitoring and continued modeling of future stream water acid-base chemistry [*Benčoková et al*., 2011; *Evans*, 2005; *Wright et al*., 2006]. Climate projections for the southern Appalachian Mountain region indicate both warmer and drier (south) and warmer and wetter (north) futures (year 2100 projections) [*Hayhoe et al*., 2008; *Karl et al*., 2009; *Solomon*, 2007]. ANC modeling results reported here indicate that areas with relatively low precipitation and warm temperatures have relatively high stream water ANC. A warmer climate would increase BC_w and therefore stream ANC. Drier conditions would decrease wet S deposition and leaching losses of base cations from soils. Interactions from the continuous model (Figure S2) suggest that higher temperatures and lower precipitation can moderate the influence of siliceous lithology on reducing stream water ANC levels. Monitoring ANC over time will be necessary to better understand these interactions.

[75] Ecosystem responses to longer-term climatic change are largely unknown due to the potential complexity of interactions among the changes. Accompanying changes in temperature, precipitation, and insolation regimes, changes in overall productivity, vegetative communities, plant cover, carbon and nutrient uptake by plants, and influences of disturbances may all affect the acid-base chemistry of streams within the study domain. Thus, it will be important to continue to monitor and model ANC. Moreover, new combinations or changes in the importance of predictors will likely necessitate water chemistry sampling that more evenly samples the variability of conditions represented by the predictors.

[76] Since enactment of the Clean Air Act Amendments of 1990, levels of S deposition in the eastern United States have decreased by more than 40% [*Baumgardner et al*., 2002]. Ongoing monitoring suggests that reduced deposition improved stream and lake acid-base chemistry at some locations [*United States Environmental Protection Agency*, 2009]. In the current assessment, the amount of S deposited was an important predictor of ANC, but S deposition was only included in one model (the continuous model) and was not included in any other of the models in the analysis. Instead, other variables related to the surrounding geology, soils, climate, and vegetation were generally more influential than the amount of S deposition in the catchment.

[77] Furthermore, the continuous model did not indicate a monotonic decrease in ANC with increasing levels of S deposition; rather, nonlinearities in this relation were apparent. Modeled ANC was lowest for intermediate levels of dry S deposition, pointing to interactions among S and other predictors. For example, intermediate levels of dry S deposition typically occurred in montane environments, where wet deposition was highest, catchment areas were small, and climate was cool and moist with frequent cloud cover. These interactions were nonstationary across the study domain and could not be accounted for using a single global model.

[78] Nonlinearities may also derive from the rather coarse resolution of CMAQ modeled dry and wet S deposition or from CMAQ prediction errors. Considerable uncertainty exists in these layers, but it is difficult to quantify the extent of these uncertainties as no "true estimates" of deposition exist to test model errors [*United States Environmental Protection Agency*, 2009].

[79] Another possible explanation for the predicted nonlinear relationship between S deposition and ANC may be related to the temporal discontinuity between a portion of the observed ANC data and the modeled CMAQ data. Because the water chemistry data were taken over a 24 year period and the CMAQ data used to represent dry S deposition were taken from 2002, the modeled dry S deposition may not accurately represent conditions at the time of the water chemistry sampling. However, CMAQ data for 2002 likely represent the relative extent to which each catchment has been exposed to elevated S deposition. Furthermore, base cation depletion of soils has limited the ANC recovery of the most acid-sensitive streams [*Sullivan et al*., 2004, 2011]. Nevertheless, results here indicate that reductions in S inputs alone may not lead directly to increases in stream water ANC because interactions with other driving variables, such as temperature and precipitation, confound simple monotonic interpretation of relations between acidic S inputs and stream water ANC.

5. Conclusion

[80] Models developed here predict stream water ANC across the southern Appalachian Mountains. Results suggest that aquatic biota may be at risk from the deleterious effects of low ANC (ANC < 100 μeq L^{-1}) [*Sullivan*, 2000] over approximately one fourth of the stream network. Acid-sensitive areas have siliceous lithology, cool and moist climates, and forests with low soil clay and pH levels, and relatively small contributing areas. Our findings suggest that predicting future ANC will require incorporation of data from ongoing stream water chemistry monitoring into a modeling framework capable of quantifying the inherently nonlinear interactions among relevant biogeochemical and climatic variables. Continued water chemistry monitoring

should include previously sampled locations and be expanded to include undersampled geographic areas and environments (e.g., areas with high model uncertainty). Although correlation does not imply causation, the identification of several climate variables as key drivers of ANC suggests that future climate change may influence future stream acid-base chemistry in unique and complex ways. The analysis approach taken here can be readily applied in other regions where adequate coverage of high-resolution, spatially explicit environmental data is available, and where stream water quality surveys have been conducted at a reasonable sampling intensity.

References

Austin, M. (2002), Spatial prediction of species distribution: An interface between ecological theory and statistical modelling, *Ecol. Model.*, *157*(2–3), 101–118.

Barandela, R., E. Rangel, J. Sánchez, and F. Ferri (2003), Restricted decontamination for the imbalanced training sample problem, in *Progress in Pattern Recognition Research*, edited by A. Sanfeliu and J. Ruiz-Shulcloper, pp. 424–431, Springer, Berlin.

Baumgardner, R. E., T. F. Lavery, C. M. Rogers, and S. S. Isil (2002), Estimates of the atmospheric deposition of sulfur and nitrogen species: Clean air status and trends network, 1990–2000, *Environ. Sci. Technol.*, *36*(12), 2614–2629.

Benčoková, A., J. Hruška, and P. Krám (2011), Modeling anticipated climate change impact on biogeochemical cycles of an acidified headwater catchment, *Appl. Geochem.*, *26*, suppl., S6–S8.

Bivand, R., L. Anselin, O. Berke, A. Bernat, M. Carvalho, Y. Chun, C. Dormann, S. Dray, R. Halbersma, and N. Lewin-Koh (2011), spdep: Spatial dependence: Weighting schemes, statistics and models, R Package Version 0.5–31. [Available at http://cran.r-project.org/web/packages/spdep/index.html.].

Breiman, L. (1984), *Classification and Regression Trees*, Chapman & Hall/CRC, New York, N. Y.

Breiman, L. (2001a), Random forests, *Mach. Learn.*, *45*(1), 5–32.

Breiman, L. (2001b), Statistical modeling: The two cultures (with comments and a rejoinder by the author), *Stat. Sci.*, *16*(3), 199–231.

Byun, D., and K. L. Schere (2006), Review of the governing equations, computational algorithms, and other components of the models: 3. Community Multiscale Air Quality (CMAQ) modeling system, *Appl. Mech. Rev.*, *59*(2), 51–77.

Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer (2002), SMOTE: Synthetic minority over-sampling technique, *J. Artif. Intell. Res.*, *16*(1), 321–357.

Christophersen, N., and C. Neal (1990), Linking hydrological, geochemical, and soil chemical processes on the catchment scale: An interplay between modeling and field work, *Water Resour. Res.*, *26*(12), 3077–3086.

Christophersen, N., C. Neal, and J. Mulder (1990), Reversal of stream acidification at the Birkenes catchment, southern Norway: Predictions based on potential ANC changes, *J. Hydrol.*, *116*(1–4), 77–84.

Clark, J. M., A. Heinemeyer, P. Martin, and S. H. Bottrell (2012), Processes controlling DOC in pore water during simulated drought cycles in six different UK peats, *Biogeochemistry*, *109*(1–3), 253–270.

Cosby, B., G. Hornberger, J. Galloway, and R. Wright (1985), Modeling the effects of acid deposition: Assessment of a lumped parameter model of soil water and streamwater chemistry, *Water Resour. Res.*, *21*(1), 51–63.

Cosby, B., J. Webb, J. Galloway, and F. Deviney (2006), Acidic deposition impacts on natural resources in Shenandoah National Park, Tech. Rep. NPS/NER/NRTR-2006/066, Philadelphia, Pa.

Cronan, C. S., W. A. Reiners, R. C. Reynolds, and G. E. Lang (1978), Forest floor leaching: Contributions from mineral, organic, and carbonic acids in New Hampshire subalpine forests, *Science*, *200*(4339), 309–311.

De'ath, G. (2007), Boosted trees for ecological modeling and prediction, *Ecology*, *88*(1), 243–251.

De'ath, G., and K. E. Fabricius (2000), Classification and regression trees: A powerful yet simple technique for ecological data analysis, *Ecology*, *81*(11), 3178–3192.

Driscoll, C. T., B. J. Wyskowski, C. C. Cosentini, and M. E. Smith (1987), Processes regulating temporal and longitudinal variations in the chemistry of a low-order woodland stream in the Adirondack Region of New York, *Biogeochemistry*, *3*(1/3), 225–241.

Driscoll, C. T., K. M. Driscoll, K. M. Roy, and M. J. Mitchell (2003), Chemical response of lakes in the Adirondack region of New York to declines in acidic deposition, *Environ. Sci. Technol.*, *37*(10), 2036–2042.

Duan, L., J. Hao, S. Xie, and K. Du (2000), Critical loads of acidity for surface waters in China, *Sci. Total Environ.*, *246*(1), 1–10.

Dunford, R. W., D. N. M. Donoghue, and T. P. Burt (2012), Forest land cover continues to exacerbate freshwater acidification despite decline in sulphate emissions, *Environ. Pollut.*, *167*, 58–69.

Elith, J., and J. R. Leathwick (2009), Species distribution models: Ecological explanation and prediction across space and time, *Ann. Rev. Ecol. Evol. Syst.*, *40*, 677–697.

Elith, J., J. Leathwick, and T. Hastie (2008), A working guide to boosted regression trees, *J. Anim. Ecol.*, *77*(4), 802–813.

Evans, C. D. (2005), Modelling the effects of climate change on an acidic upland stream, *Biogeochemistry*, *74*(1), 21–46.

Franklin, J., and J. A. Miller (2009), *Mapping Species Distributions: Spatial Inference and Prediction*, Cambridge Univ. Press, New York, NY.

Gahegan, M. (2003), Is inductive machine learning just another wild goose (or might it lay the golden egg)?, *Int. J. Geogr. Inf. Sci.*, *17*(1), 69–92.

Galloway, J. (2001), Acidification of the world: Natural and anthropogenic, *Water Air Soil Pollut.*, *130*(1), 17–24.

Galloway, J. N., Z. Dianwu, X. Jiling, and G. E. Likens (1987), Acid rain: China, United States, and a remote area, *Science*, *236*(4808), 1559–1562.

Gee, A., and J. Stoner (1989), A review of the causes and effects of acidification of surface waters in Wales and potential mitigation techniques, *Arch. Environ. Contam. Toxicol.*, *18*(1), 121–130.

Grimm, J. W., and J. A. Lynch (2004), Enhanced wet deposition estimates using modeled precipitation inputs, *Environ. Monit. Assess.*, *90*(1), 243–268.

Guerold, F., J.-P. Boudot, G. Jacquemin, D. Vein, D. Merlet, and J. Rouiller (2000), Macroinvertebrate community loss as a result of headwater stream acidification in the Vosges Mountains (N-E France), *Biodivers. Conserv.*, *9*(6), 767–783.

Hargrove, W., and F. Hoffman (2004), A flux atlas for representativeness and statistical extrapolation of the Ameriflux network, Oak Ridge Natl. Lab. Tech. Memo. ORNL-TM-2004/112, pp. 1–152.

Harriman, R., and B. Morrison (1982), Ecology of streams draining forested and non-forested catchments in an area of central Scotland subject to acid precipitation, *Hydrobiologia*, *88*(3), 251–263.

Hastie, T., R. Tibshirani, J. Friedman, and J. Franklin (2005), The elements of statistical learning: Data mining, inference and prediction, *Math. Intell.*, *27*(2), 83–85.

Hayhoe, K., C. Wake, B. Anderson, X. Z. Liang, E. Maurer, J. Zhu, J. Bradbury, A. DeGaetano, A. M. Stoner, and D. Wuebbles (2008), Regional climate change projections for the Northeast USA, *Mitig. Adapt. Strat. Global Change*, *13*(5), 425–436.

He, H., and E. A. Garcia (2009), Learning from imbalanced data, *IEEE Trans. Knowl. Data Eng.*, *21*(9), 1263–1284.

Hendershot, W. H., P. Warfvinge, F. Courchesne, and H. U. Sverdrup (1991), The mobile anion concept—Time for a reappraisal?, *J. Environ. Qual.*, *20*(3), 505–509.

Henriksen, A., and M. Posch (2001), Steady-state models for calculating critical loads of acidity for surface waters, *Water Air Soil Pollut.: Focus*, *1*(1), 375–398.

Henriksen, A., M. Posch, H. Hultberg, and L. Lien (1995), Critical loads of acidity for surface waters: Can the ANC limit be considered variable?, *Water Air Soil Pollut.*, *85*(4), 2419–2424.

Herlihy, A. T., P. R. Kaufmann, M. R. Church, P. J. Wigington Jr., J. R. Webb, and M. J. Sale (1993), The effects of acidic deposition on streams in the Appalachian Mountain and Piedmont Region of the Mid-Atlantic United States, *Water Resour. Res.*, *29*(8), 2687–2703.

Herlihy, A. T., J. L. Stoddard, and C. B. Johnson (1998), The relationship between stream chemistry and watershed land cover data in the mid-Atlantic region, US, *Water Air Soil Pollut.*, *105*(1), 377–386.

Homer, C., J. Dewitz, J. Fry, M. Coan, N. Hossain, C. Larson, N. Herold, A. McKerrow, J. N. VanDriel, and J. Wickham (2007), Completion of the 2001 National Land Cover Database for the conterminous United States, *Photogramm. Eng. Remote Sens.*, *73*(4), 337–341.

Hruška, J., F. Moldan, and P. Krám (2002), Recovery from acidification in central Europe—Observed and predicted changes of soil and streamwater chemistry in the Lysina catchment, Czech Republic, *Environ. Pollut.*, *120*(2), 261–274.

Jenkins, A., B. J. Cosby, R. C. Ferrier, T. A. B. Walker, and J. D. Miller (1990), Modelling stream acidification in afforested catchments: An assessment of the relative effects of acid deposition and afforestation, *J. Hydrol.*, *120*(1–4), 163–181.

Jenness, J. S. (2004), Calculating landscape surface area from digital elevation models, *Wildlife Soc. B.*, *32*(3), 829–839.

Jenson, S., and J. Domingue (1988), Extracting topographic structure from digital elevation data for geographic information system analysis, *Photogramm. Eng. Remote Sens.*, *54*(11), 1593–1600.

Johnson, L. B., and G. E. Host (2010), Recent developments in landscape approaches for the study of aquatic ecosystems, *J. North Am. Benthol. Soc.*, *29*(1), 41–66.

Karl, T. R., J. M. Melillo, and T. C. Peterson (2009), *Global Climate Change Impacts in the United States*, Cambridge Univ. Press, New York, NY.

Krug, E. C., and C. R. Frink (1983), Acid rain on acid Soil: A new perspective, *Science*, *221*(4610), 520–525.

Lawrence, G. B., M. B. David, G. M. Lovett, P. S. Murdoch, D. A. Burns, J. L. Stoddard, B. P. Baldigo, J. H. Porter, and A. W. Thompson (1999), Soil calcium status and the response of stream chemistry to changing acidic deposition rates, *Ecol. Appl.*, *9*(3), 1059–1072.

Levin, S. (1992), The problem of pattern and scale in ecology: The Robert H. MacArthur award lecture, *Ecology*, *73*(6), 1943–1967.

Liaw, A., and M. Wiener (2002), Classification and regression by randomForest, *R News*, *2*(3), 18–22.

Lien, L., G. Raddum, and A. Fjellheim (1992), Critical loads of acidity to freshwater fish and invertebrates, *Naturens Talegrenser 23*.

Lynch, D. D., and N. B. Dise (1985), Sensitivity of stream basins in Shenandoah National Park to acid deposition, Rep. 85–4115, U.S Dep. of the Inter., U.S. Geol. Surv., Richmond, Va.

Maclure, M., and W. C. Willet (1987), Misinterpretation and misuse of the Kappa statistic, *Am. J. Epidemiol.*, *126*(2), 161–169.

McDonnell, T., B. Cosby, T. Sullivan, S. McNulty, and E. Cohen (2010), Comparison among model estimates of critical loads of acidic deposition using different sources and scales of input data, *Environ. Pollut.*, *158*, 2934–2939.

McDonnell, T. C., B. J. Cosby, and T. J. Sullivan (2012), Regionalization of soil base cation weathering for evaluating stream water acidification in the Appalachian Mountains, USA, *Environ. Pollut.*, *162*, 338–344.

Menz, F. C., and H. M. Seip (2004), Acid rain in Europe and the United States: An update, *Environ. Sci. Policy*, *7*(4), 253–265.

Miles, J. (1986), What are the effects of trees on soils?, in *Trees and Wildlife in the Scottish Uplands*, edited by D. Jenkins, pp. 55–62, NERC/ITE, Swindon, UK.

Moore, I. D., P. E. Gessler, G. A. Nielsen, and G. A. Peterson (1993), Soil attribute prediction using terrain analysis, *Soil Sci. Soc. Am. J.*, *57*(2), 443–452.

National Atlas of the United States (2006), Federal Lands of the United States: National Atlas of the United States, Reston, Va.

Neal, C., B. Reynolds, and A. J. Robson (1999), Acid neutralisation capacity measurements within natural waters: Towards a standardised approach, *Sci. Total Environ.*, *243–244*, 233–241.

Nihlgård, B. (1970), Precipitation, its chemical composition and effect on soil water in a beech and a spruce forest in south Sweden, *Oikos*, *21*, 208–217.

Nilsson, J., and P. Grennfelt (1988), *Critical Loads for Sulfur and Nitrogen*, 31 pp., Nordic Counc. of Minist., Copenhagen, Denmark.

NRCS Soil Survey Staff (2010a), Soil Survey Geographic (SSURGO) database for southern Appalachian Region. Available online at http://soildatamart.nrcs.usda.gov. Accessed [October/10/2010].

NRCS Soil Survey Staff (2010b), U.S. General Soil Map State Soil Geographic (STATSGO) database for the southern Appalachian Region. Available online at http://soildatamart.nrcs.usda.gov. Accessed [October/10/2010].

Olden, J. D., J. J. Lawler, and N. Poff (2008), Machine learning methods without tears: A primer for ecologists, *Q. Rev. Biol.*, *83*(2), 171–194.

Omernik, J. M. (1987), Ecoregions of the conterminous United States, *Ann. Assoc. Am. Geogr.*, *77*(1), 118–125.

Prasad, A. M., L. R. Iverson, and A. Liaw (2006), Newer classification and regression tree techniques: Bagging and random forests for ecological prediction, *Ecosystems*, *9*(2), 181–199.

Puckett, L. J., and O. P. Bricker (1992), Factors controlling the major ion chemistry of streams in the blue ridge and valley and ridge physiographic provinces of Virginia and Maryland, *Hydrol. Processes*, *6*(1), 79–97.

R Development Core Team (2011), *R: A Language and Environment for Statistical Computing*, Vienna, Austria.

Rago, P. J., and J. G. Wiener (1986), Does pH affect fish species richness when lake area is considered?, *Trans. Am. Fish. Soc.*, *115*(3), 438–447.

Reuss, J. O., B. J. Cosby, and R. F. Wright (1987), Chemical processes governing soil and water acidification, *Nature*, *329*(6134), 27–32.

Reynolds, K. M., P. F. Hessburg, T. J. Sullivan, N. A. Povak, T. C. McDonnell, B. J. Cosby, and W. Jackson (2012), Spatial decision support for assessing impacts of atmospheric sulfur deposition on aquatic ecosystems in the southern Appalachian Region, paper presented at 45th annual Hawaii International Conference on System Sciences, IEEE, Wailea, Maui, Hawaii, 4–7 January.

Ridgeway, G. (2006), Generalized boosted regression models, Documentation on the R Package 'gbm', Version 1·5, 7.

Schindler, D. W. (1988), Effects of acid rain on freshwater ecosystems, *Science*, *239*(4836), 149–157.

Schöpp, W., M. Posch, S. Mylona, and M. Johansson (2003), Long-term development of acid deposition (1880–2030) in sensitive freshwater regions in Europe, *Hydrol. Earth Syst. Sci.*, *7*(4), 436–446.

Solomon, S. (2007), *Climate Change 2007: The Physical Science Basis: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge Univ. Press, New York, NY.

Steel, E. A., R. M. Hughes, A. H. Fullerton, S. Schmutz, J. A. Young, M. Fukushima, S. Muhar, M. Poppe, B. E. Feist, and C. Trautwein (2010), Are we meeting the challenges of landscape-scale riverine research? A review, *Living Rev. Landscape Res.*, *4*(1), 1–60.

Sullivan, T., J. Webb, K. Snyder, A. Herlihy, and B. 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*(1), 57–71.

Sullivan, T. J. (2000), *Aquatic Effects of Acidic Deposition*, CRC Press, Boca Raton, Fla.

Sullivan, T. J., D. F. Charles, J. A. Bernert, B. McMartin, K. B. Vaché, and J. Zehr (1999), Relationship between landscape characteristics, history, and lakewater acidification in the Adirondack Mountains, New York, *Water Air Soil Pollut.*, *112*(3), 407–427.

Sullivan, T. J., B. J. Cosby, A. T. Herlihy, J. R. Webb, A. J. Bulger, K. U. Snyder, P. F. Brewer, E. H. Gilbert, and D. L. Moore (2004), Regional model projections of future effects of sulfur and nitrogen deposition on streams in the southern Appalachian Mountains, *Water Resour. Res.*, *40*, W02101, doi:10.1029/2003WR001998.

Sullivan, T. J., B. J. Cosby, A. T. Herlihy, C. T. Driscoll, I. J. Fernandez, T. C. McDonnell, C. W. Boylen, S. A. Nierzwicki-Bauer, and K. U. Snyder (2007), *Assessment of the Extent to Which Intensively Studied Lakes Are Representative of the Adirondack Mountain Region*, N. Y. State Energy Res. and Dev. Auth.

Sullivan, T. J., B. J. Cosby, J. R. Webb, R. L. Dennis, A. J. Bulger, and F. A. Deviney (2008), Streamwater acid-base chemistry and critical loads of atmospheric sulfur deposition in Shenandoah National Park, Virginia, *Environ. Monit. Assess.*, *137*(1), 85–99.

Sullivan, T. J., B. J. Cosby, W. A. Jackson, K. U. Snyder, and A. T. Herlihy (2011), Acidification and prognosis for future recovery of acid-sensitive streams in the southern Blue Ridge Province, *Water Air Soil Pollut.*, *219*(1), 11–26.

Turner, M. (1989), Landscape ecology: The effect of pattern on process, *Annu. Rev. Ecol. Syst.*, *20*, 171–197.

United States Environmental Protection Agency (2009), *Risk and Exposure Assessment for Review of the Secondary National Ambient Air Quality Standards for Oxides of Nitrogen and Oxides of Sulfur: Final*, U.S. EPA, Cent. for Environ. Assess., Off. of Res. and Dev., Research Triangle Park, N. C.

United States Environmental Protection Agency and United States Geological Survey (2005), National Hydrography Dataset Plus—NHDPlus Version 1.0. Available from http://www.horizon-systems.com/NHDPlus/NHDPlusV1_home.php. Accessed on September 9, 2010.

United States Geological Survey (2005a), Preliminary integrated geologic map databases for the United States: Kentucky, Ohio, Tennessee, and West Virginia, Open-File Rep. 2005/1324.

United States Geological Survey (2005b), Preliminary integrated geologic map databases for the United States: Delaware, Maryland, New York, Pennsylvania, and Virginia, Open-File Rep. 2005/1325.

Whitehead, P. G., B. Reynolds, M. Hornung, C. Neal, J. Cosby, and P. Paricos (1988), Modelling long term stream acidification trends in upland Wales at Plynlimon, *Hydrol. Processes*, *2*(4), 357–368.

Wood, E. F., M. Sivapalan, and K. Beven (1990), Similarity and scale in catchment storm response, *Rev. Geophys.*, *28*(1), 1–18.

Wright, R. F., et al. (2006), Modelling the effect of climate change on recovery of acidified freshwaters: Relative sensitivity of individual processes in the MAGIC model, *Sci. Total Environ.*, *365*(1–3), 154–166.