AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2013WR014203

Key Points:
- Regression models used to extrapolate catchment-level BCw to the region
- Random forest models reduced error rates by half compared to linear models
- BCw rates influenced by lithology, soil, sulfur deposition, forest, and climate

Supporting Information:
- Readme
- Supplementary texts 1 and 2
- Table S1
- Figures S1 and S2

Correspondence to:
N. A. Povak,
npovak@fs.fed.us

Citation:
Povak, N. A., P. F. Hessburg, T. C. McDonnell, K. M. Reynolds, T. J. Sullivan, R. B. Salter, and B. J. Cosby (2014), Machine learning and linear regression models to predict catchment-level base cation weathering rates across the southern Appalachian Mountain region, USA, *Water Resour. Res.*, *50*, 2798–2814, doi:10.1002/2013WR014203.

Received 30 MAY 2013
Accepted 15 FEB 2014
Accepted article online 25 FEB 2014
Published online 1 APR 2014

Machine learning and linear regression models to predict catchment-level base cation weathering rates across the southern Appalachian Mountain region, USA

Nicholas A. Povak[1], Paul F. Hessburg[1], Todd C. McDonnell[2], Keith M. Reynolds[3], Timothy J. Sullivan[2], R. Brion Salter[1], and Bernard J. Cosby[4]

[1]US Forest Service, Pacific Northwest Research Station, Wenatchee Forestry Sciences Laboratory, Wenatchee, Washington, USA, [2]E&S Environmental Chemistry Inc., Corvallis, Oregon, USA, [3]US Forest Service, Pacific Northwest Research Station, Corvallis Forestry Sciences Laboratory, Corvallis, Oregon, USA, [4]Department of Environmental Sciences, University of Virginia, Charlottesville, Virginia, USA

Abstract Accurate estimates of soil mineral weathering are required for regional critical load (CL) modeling to identify ecosystems at risk of the deleterious effects from acidification. Within a correlative modeling framework, we used modeled catchment-level base cation weathering (BC_w) as the response variable to identify key environmental correlates and predict a continuous map of BC_w within the southern Appalachian Mountain region. More than 50 initial candidate predictor variables were submitted to a variety of conventional and machine learning regression models. Predictors included aspects of the underlying geology, soils, geomorphology, climate, topographic context, and acidic deposition rates. Low BC_w rates were predicted in catchments with low precipitation, siliceous lithology, low soil clay, nitrogen and organic matter contents, and relatively high levels of canopy cover in mixed deciduous and coniferous forest types. Machine learning approaches, particularly random forest modeling, significantly improved model prediction of catchment-level BC_w rates over traditional linear regression, with higher model accuracy and lower error rates. Our results confirmed findings from other studies, but also identified several influential climatic predictor variables, interactions, and nonlinearities among the predictors. Results reported here will be used to support regional sulfur critical loads modeling to identify areas impacted by industrially derived atmospheric S inputs. These methods are readily adapted to other regions where accurate CL estimates are required over broad spatial extents to inform policy and management decisions.

1. Introduction

Atmospheric sulfur (S) inputs to streams in the southeastern United States (US) have reduced native fish and invertebrate populations, and altered aquatic community structure in favor of acid tolerant species [*United States Environmental Protection Agency* (*US EPA*), 2009]. Deposited S is largely derived from industrial coal and oil combustion [*National Acid Precipitation Assessment Program (NAPAP)*, 1991] and affects soil and surface water chemistry through base cation leaching, increased sulfate concentrations, reduced pH and acid neutralizing capacity (ANC), and mobilization of potentially toxic inorganic aluminum (Ali) [*Driscoll et al*., 2001; *US EPA*, 2009].

Sulfur deposition levels (wet + dry) have declined by 44% across the eastern US beginning in the early 1990s [*Greaver et al*., 2012; *United States Environmental Protection Agency* (*US EPA*), 2008]. However, current levels of S deposition and lagged effects of past deposition continue to threaten aquatic ecosystems across the region [*Greaver et al*., 2012; *US EPA*, 2008, 2009].

The U.S. Environmental Protection Agency (EPA) and federal land management agencies are charged with assessing the ecological effects of S emissions and implementing restorative management to areas most susceptible to chronic deposition. The standard approach used to identify acid-sensitive ecosystems is to develop critical load (CL) estimates [*United Nations Economic Commission for Europe* (*UNECE*), 2004]. The CL is defined as the level of sustained atmospheric deposition of strong acids (here, S) above which harmful effects are expected to occur [*Nilsson and Grennfelt*, 1988]. Several studies have used aquatic critical load calculations to identify potential exceedances in aquatic habitats [*Henriksen et al*., 2002, 1995; *McDonnell et al*., 2012; *Reynolds et al*., 2012; *Sullivan et al*., 2005]. The most common approach to estimate regional

steady state CL for surface waters is with the Steady State Water Chemistry model (SSWC; equation (1)) [*Henriksen and Posch*, 2001], which is calculated as:

$$CL(S) = BC_{dep} + BC_{w} - Bc_{u} - Cl_{dep} - ANC_{limit} \quad (1)$$

where BC_{dep} and Cl_{dep} represent the total atmospherically deposited major base cations Ca, K, Mg, Na, and chloride, respectively; BC_w is the base cation weathering rate, Bc_u represents the uptake and removal of nutrient base cations by tree harvesting, and ANC_{limit} is the critical ANC leaching rate.

The research presented here focuses on estimating the catchment-level BC_w rate, which is considered to be the most critical [*Hodson and Langan*, 1999] and often the most poorly estimated term in the CL calculation equation (1) [*Li and McNulty*, 2007; *McDonnell et al*., 2010; *Skeffington*, 2006]. BC_w is generally calculated as an intermediate output in models that estimate CLs. Models currently used to estimate BC_w rates fall into two categories: (1) process-based, and (2) empirical steady state [*Hall et al*., 2001; *Hodson and Langan*, 1999].

Process-based models such as PROFILE [*Sverdrup and Warfvinge*, 1993; *Warfvinge and Sverdrup*, 1992] and MAGIC (Model of Goundwater Acidification in Catchments) [*Cosby et al*., 1985] are able to generate accurate plot or catchment-level BC_w predictions. However, the unavailability of needed data precludes their applicability to larger spatial extents.

Conversely, empirical steady state models such as SSWC are able to predict CL (and consequently BC_w) at regional extents using empirical relationships between broad-scale soil texture and geological data. However, these models suffer high error rates related to variability in regional biogeoclimatic conditions, low spatial accuracy or inconsistent scaling of soils and geological data, and high dissimilarity between regions where models are developed and applied [*Li and McNulty*, 2007; *McDonnell et al*., 2010; *Skeffington*, 2006]. Furthermore, error rates for these models are not often reported, bringing into question the validity of their application [*Li and McNulty*, 2007].

The goals of this research were to improve estimates of BC_w across the southern Appalachian Mountain region [*Sullivan et al*., 2008] using a hybrid modeling approach that combined process-based and empirical steady state modeling, and that built upon a previous modeling effort in Virginia and West Virginia [*McDonnell et al*., 2012]. Statistical and machine learning (ML) regression models were used to correlate remotely sensed biogeochemical and climatic variables (the independent variables) to catchment-level BC_w estimates (the response variable) made using a dynamic process-based model. Variable reduction procedures were used to identify the main environmental correlates of BC_w while balancing model performance with model parsimony. These relationships were then used to extrapolate estimates of catchment-level BC_w across the study region. We evaluated the environmental correlates and the degree of correlation to confirm known and hypothesized relationships between catchment-level BC_w levels and the biophysical setting.

Machine learning algorithms are relatively new to ecological research [*Hastie et al*., 2005; *Olden et al*., 2008]. These algorithms identify patterns in a data set while treating the data-generating mechanism as unknown [*Breiman*, 2001b]. This, in turn, leads to the development of predictive models that make few assumptions regarding the underlying structure of the data. This is particularly beneficial where patterns in a response emerge from interactions among covariates that can exist at several scales of observation, and often generate nonlinear response functions. ML incorporates a variety of modeling approaches that correlate predictor variables to a response variable for extracting information on the relationships between the predictors and response, and for the purposes of creating accurate predictions outside the training data. Examples of ML algorithms include support vector machines, neural networks, decision trees, random forest, and boosted regression trees, among others [*Berk*, 2008; *Hastie et al*., 2005].

Building on the work of *McDonnell et al*. [2012], the current assessment includes: (1) an expanded study area with additional catchment-level BC_w estimates derived from the MAGIC model, (2) inclusion of new climatic, productivity, edaphic, topographic, and sulfur deposition predictor variables, (3) incorporation of ML algorithms, and (4) spatial assessment of model uncertainty. Model estimates of BC_w were incorporated along with modeled ANC [*Povak et al*., 2013] into a decision support application [*Reynolds et al*., 2012] that can be used to generate CL and identify acid-sensitive streams across the same region.

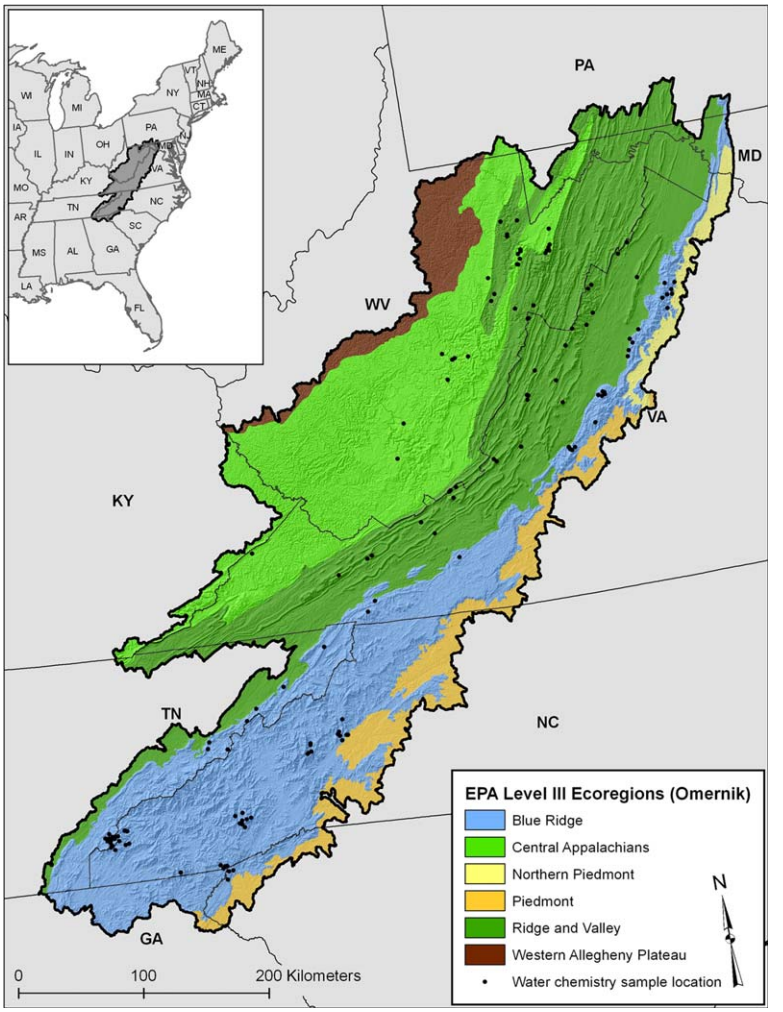

Figure 1. Distribution of BC_w sample points within *Omernik* [1987] ecoregions for the southern Appalachian Mountains region.

2. Methods

2.1. Study Area

The study was conducted across a broad geographic area (14.3×10^6 ha) encompassing much of the southern Appalachian Mountain region of the eastern United States, from northern Georgia to southern Pennsylvania, and from eastern Kentucky and Tennessee to central Virginia and western North Carolina (Figure 1). The study area includes several *Omernik* [1987] Level III ecoregions; the three main ecoregions are the Blue Ridge, Ridge and Valley, and Central Appalachian regions. Elevations in the study region range from 300 to 2000 m. Forests are primarily composed of oak (*Quercus* spp.), hickory (*Carya* spp.), and pine (*Pinus* spp.). Mixed and pure spruce-fir and northern hardwood forests occupy the highest elevation sites with cool, moist climates. Forests are interspersed with agricultural and urban lands, which generally occupy lowland valleys.

2.2. MAGIC Model

MAGIC is a process model used to estimate effects of acidic deposition on soils and surface waters in catchments [*Cosby et al*., 1985]. BC_w estimates calculated by MAGIC represent the catchment average net supply of base cations (BC; see Table 1 for a list of common acronyms) derived from soil mineral weathering draining to surface waters. Within the model, physical and chemical processes associated with the acid-base status of soil and stream water chemistry are averaged across a catchment to predict its potential response to acidic inputs. MAGIC predictions are calibrated using observed stream, soil, and atmospheric deposition

Table 1. A Selection of Acronyms and Their Definitions

ANC	Acid neutralizing capacity	A measure of the acid-base status of a solution, similar to pH [*Hemond*, 1990]
BC	Base cations	e.g., Ca, K, Mg, Na, and chloride
BC_w	Base cation weathering rate	Base cation supply through mineral weathering [*Ouimet and Duchesne*, 2005]
BR	Blue Ridge	Omernik [*Omernik*, 1987] ecoregion within the study area
BRT	Boosted regression trees	Machine learning algorithm [*Friedman*, 2002]
CA	Central Appalachian	Omernik [*Omernik*, 1987] ecoregion within the study area
CL	Critical Loads	The level of pollution below which negative ecological effects are not anticipated to occur [*Nilsson and Grennfelt*, 1988]
GWR	Geographically weighted regression	Linear regression analysis where local model coefficients vary spatially [*Fotheringham et al*., 2002]
GS	Growing season	Used to describe time period for some predictor variables used in modeling
LM	Linear model	Ordinary least squares regression
MAGIC	Model of groundwater acidification in catchments	Process-based model used to calculate catchment-level soil and stream water chemistry [*Cosby et al*., 1985]
MARS	Multiplicative adaptive regression splines	Machine learning algorithm [*Friedman*, 1991]
ML	Machine learning	Branch of statistics that uses algorithms to model complex data [*Hastie et al*., 2005]
OOB	Out-of-bag samples	The bootstrap samples withheld from the training of individual regression trees within random forest
NGS	Nongrowing season	Used to describe time period for some predictor variables used in modeling
PROFILE		Steady state soil chemistry computer model used to calculate the Critical Load of acidity for soil and surface water systems [*Warfvinge and Sverdrup*, 1992]
RF	Random Forest	Machine learning algorithm (Breiman, 2001)
RMSE	Root mean squared error	Measure of model error for regression models
RV	Ridge and Valley	Omernik [*Omernik*, 1987] ecoregion within the study area
SSWC	Steady State Water Chemistry model	Model used to calculate regional critical loads estimates [*Henriksen and Posch*, 2001]

data. After calibration, MAGIC can be used to simulate historical and future soil, soil solution, and surface water chemistry at monthly and annual time steps.

In the current study, MAGIC was calibrated to 140 catchments, using methods described by *McDonnell et al*. [2012]. For each catchment, 10 MAGIC simulations were performed to calibrate the model across the ranges of soil and stream chemistry, and atmospheric deposition inputs. Effects of uncertainty in MAGIC model assumptions and in the available input data were assessed using results from all successful calibrations for a given site. The average difference between the maximum and minimum BC_w estimates within a catchment was 21.4 meq m^{-2} yr^{-1} across all catchments, which is a measure of overall MAGIC model uncertainty.

The study sites were compiled from a variety of previous studies. Most of these studies targeted areas of known acidic stream conditions and likely lower than average BC_w rates (see section 3.1). Nonetheless, this data set represented the most comprehensive analysis of catchment-level BC_w rates for the study region to date. To assess the representativeness of the environmental settings of the modeled catchments compared to those present across the study domain, we compared the estimated kernel density function for each predictor variable from the 140 sites to those constructed from 4000 randomly selected catchments (Table 2). The proportion of overlap among the density functions indicated the degree of correspondence among the distributions.

2.3. Predictor Variables

An initial set of 55 environmental predictor variables was chosen to represent broad- to fine-scale climatic, lithologic, geomorphic, topographic, edaphic, vegetative, land ownership, and atmospheric S deposition variables that were potentially influential in explaining geographical patterns of BC_w. Variables are briefly described below, but also see supporting information Text S1 and Table S1 for a more complete description.

Climate variables (1 km resolution) included estimates of the growing season (GS) and nongrowing season (NGS) precipitation, temperature, and insolation regime, and measures of overall productivity [*Hargrove and Hoffman*, 2004]. Vegetation data (30 m) included percentage composition of conifer, deciduous, and all forest types [*Homer et al*., 2007]. Soil variables (1 km) included percentage clay, soil pH, and soil depth [*NRCS Soil Survey Staff*, 2010a, 2010b], and amount of organic matter, mean Kjeldahl N content, and mean soil plant-available water [*Hargrove and Hoffman*, 2004].

A lithologic classification provided by *Sullivan et al*. [2007] for the southern Appalachian region was used to capture the percentage composition of geologic parent materials across the study area. Classes included siliceous, argillic, felsic, mafic, and carbonate substrates. A 30 m digital elevation model (DEM) [*Gesch et al*.,

Table 2. Proportional Overlap in the Kernel Density Functions for a Selection of Environmental Covariates Used in Modeling BC_w Rates ($n = 140$) Compared to Those Developed Using 4000 Randomly Select Catchments[a]

Variable Shortname	Description	Proportion Overlap
AB90GROW	Mean number of days above 32.2°C during the local growing season (days×0.02)	0.72
DDAYHMAX	Mean penultimate maximum degrees × days heating > 18°C (days×°C)	0.63
TDAYMEAN	Mean daytime temperature (°C)	0.62
PRECIPNG	Mean precipitation sum during the local nongrowing season (mm×0.01)	0.68
PDAYMAX	Mean penultimate maximum days with precipitation (>0.3 cm) while >10°C (days)	0.90
DIFF95GR	Mean 95th percentile of maximum diurnal surface temperature difference during the local growing season (°C×0.02)	0.61
GPPGROW	Mean gross primary production (GPP) integrated over the local growing season ((kg C/m^2/8 days)×days×0.0001)	0.80
GPPNG	Mean gross primary production (GPP) integrated over the local nongrowing season ((kg C/m^2/8 days)×days×0.0001)	0.70
CON42	Percentage contributing area in conifer cover (% cover)	0.65
DECID41	Percentage contributing area in conifer (1) + mixed (0.5) cover (weighted by number in parentheses; % cover)	0.80
FAC	Flow accumulation (number of pixels)	0.79
LITH_CAR	Percentage contributing area in carbonaceous lithology (% cover)	0.81
LITH_SIL	Percentage contributing area in siliceous lithology (% cover)	0.64
SOIL_DEP	Mean soil depth (cm)	0.89
SOIL_CLAY	Mean soil clay (%)	0.65
SOIL_PH	Mean soil pH	0.70
OMNEW	Mean organic matter content to 50 cm depth (kg×ha^{-1})	0.83
NITRONEW	Mean soil Kjeldahl nitrogen to 50 cm depth (kg×ha^{-1})	0.66
S_WET	Mean wet sulfur deposition (meq m^2×yr^{-1})	0.87
S_DRY	Mean dry sulfur deposition (meq m^2×yr^{-1})	0.85

[a]High overlap indicates good concordance between the two distributions. See supporting information Text S1 and Table S1 for a complete description of the predictor variables used in the analysis.

2002, *Gesch*, 2007] was used to derive attributes related to topographic setting such as the surface area roughness, topographic wetness index, and flow accumulation (catchment size) metrics. Levels of wet and dry S deposition were also included [*Byun and Schere*, 2006; *Grimm and Lynch*, 2004].

The raster layer of each predictor variable was resampled to 30 m resolution and the values were averaged across the contributing upslope area for each grid cell (see Table S1 for native raster resolutions). Upslope averaging incorporated the catchment influence of the predictor variables on BC_w as calibrated to the stream sampling location, or pour point, at the base of each catchment [*McDonnell et al*., 2012]. The equation for upslope averaging is:

$$\overline{P}_i = \frac{\left(P_i + \sum_{j=1}^{N} P_j\right)}{(N+1)} \tag{2}$$

in which $\overline{P}_i$ is the upslope averaged value for the candidate cell (P_i), $\sum_{j=1}^{N} P_j$ is the summation of values for all cells upslope from P_i, and N is the total number of upslope cells. Upslope averaging enabled us to incorporate the average of each predictor variable to each pour point.

2.4. Statistical Modeling

We applied several statistical modeling approaches, ranging from conventional linear regression to more complex ML algorithms, to objectify the process of identifying the models with the lowest error rates. Statistical models tested included: linear regression (LM) [stats package in R v 2.12.2, *R Development Core Team*, 2011], boosted regression trees (BRT) [gbm package in R v2.12.2, *Ridgeway*, 2010], random forest (RF) [randomForest package in R v2.12.2, *Breiman*, 2001a; *Liaw and Wiener*, 2002], multivariate adaptive regression splines (MARS) [mda package in R v2.12.2, *Hastie and Tibshirani*, 2011], and geographically weighted regression (GWR) [*Bivand et al*., 2010]. Each statistical model has been used in previous landscape modeling assessments [e.g., *Prasad et al*., 2006]. Each model is briefly described below, but see also supporting information Text S2 for an in-depth discussion of the models used here.

The RF [*Breiman*, 2001a] and BRT [*Friedman*, 2002] models are ensemble versions of traditional regression tree analyses, in which hundreds to thousands of individual regression trees are assembled, and final predictions are averaged across the individual trees. In RF, individual regression trees are built using a random subset of the data points and of the predictor variables; resulting trees are not pruned. Model error rate is

assessed using out-of-bag (OOB) estimates, which are model predictions to the subset of data that was randomly withheld at each model iteration.

BRT uses the entire set of predictors and a portion of the data to build each individual tree, and the algorithm is sequential, with higher weights given to poorly classified observations and final predictions are weighted means of each tree [*De'ath*, 2007; *Elith et al*., 2008]. MARS is a nonparametric extension of traditional linear regression models [*Friedman*, 1991], which consist of a series of piecewise linear functions that allow for response nonlinearities. GWR is another extension of linear regression developed by *Fotheringham et al*. [2002], which allows the parameters in a global regression model to be estimated locally at every point by giving higher weights to geographically proximal data points.

2.5. Model Building and Validation

Multicollinearity in the initial set of predictor variables was reduced by retaining only those with Pearson's correlation scores <0.7, leading to a modeling set of 33 predictors. Among correlated variables, those with the highest Pearson's correlation with BC_w were retained. Higher and lower correlation cutoff values were tested, but none improved the final models. Many of the variables removed following the Pearson's analysis were temperature and precipitation or productivity variables that were highly correlated with others retained in the model.

RF models were then run using the reduced predictor set for each iteration of the modeling with successively fewer predictor variables. After each model iteration, the highest scoring variables from the prior run were reentered into each of the statistical models tested. RF models were used in variable selection to (1) avoid potential bias associated with traditional stepwise reduction procedures [*Whittingham et al*., 2006], and (2) allow for direct comparison of model performance (e.g., linear, nonlinear) with the same model predictors.

Separate statistical models were built using the reduced set of 33 predictors. Model validation statistics, including root mean squared error (RMSE) and R-squared, were calculated using both (1) predictions made to the training data (hereafter, training error), and (2) 10-fold cross validation (hereafter, $CV_{10\text{-fold}}$), which is a method used to reveal model instability and overfitting when the sample size is not large enough to withhold a separate testing set [*Cutler et al*., 2007; *Elith et al*., 2008]. These statistics were calculated for each model using sequentially fewer predictors (25, 20, 15, 10, 5, and 3) to help balance model parsimony with prediction accuracy.

We also built ecoregion-specific models using the same methods, for the three main Omernik ecoregions within the study area. Ecoregion models were developed to identify differences in BC_w covariates, the sufficiency of sampling by ecoregion, and model error rates that were related to the physical geography. Variable reduction procedures proceeded as outlined above. All ecoregion-specific models included the same number of predictors because (1) most ecoregions converged on the same number of predictor variables, and (2) this allowed for model comparisons across ecoregions. Excluded from the ecoregional analysis were the Northern Piedmont, Piedmont, and Western Allegheny Plateau ecoregions, which each included <5 BC_w MAGIC sampling locations, and together comprised only 12.5% of the study area. Results of ecoregional modeling were compared to those of *McDonnell et al*. [2012].

3. Results and Discussion

3.1. BC_w Summary

MAGIC-modeled BC_w rates ranged from 3.1 to 257.1 meq m^{-2} yr^{-1}, with a median of 66.4 meq m^{-2} yr^{-1}. The estimated kernel density function for the data was somewhat right-skewed, with approximately 20% of catchment-level weathering rates >100 meq m^{-2} yr^{-1} (Figure 2). Median elevation for study catchments was 753 m (interquartile range (IQR): 585, 937 m), and MAGIC-modeled catchments were located at slightly higher elevations, in slightly cooler and moister climates, with higher percent forest cover, higher percent siliceous lithology and lower soil clay levels, in comparison with the rest of the study area. Table 2 describes the proportional overlap in kernel density functions for the key environmental covariates. We found reasonably good concordance between the sampled locations and those present across the study region (mean overlap = 0.74). A high degree of overlap indicated high concordance between the distributions of modeling sites ($n = 140$) and the broader landscape ($n = 4000$ randomly selected watersheds).

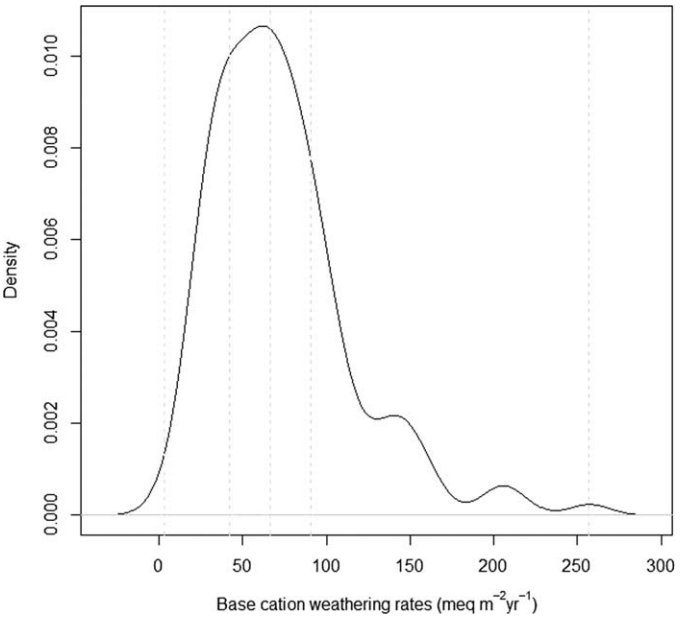

Figure 2. Kernel density plot for the distribution of base cation weathering rates calculated for the 140 catchments within the southern Appalachian study region. Gray vertical lines indicate minimum, first quartile, median, second quartile, and maximum BC_w values.

3.2. Model Calibration and Validation

Machine learning algorithms (RF, BRT, and MARS) consistently outperformed linear regression and GWR models based on model error rates (Figure 3). Most models, with the exception of RF, showed a steep increase in model error as the number of predictor variables decreased (Figure 3). Here we note that all models tested chose a similar set of predictor variables, and therefore, the comparatively low error rate for the RF model was due to the performance of the algorithm itself, and not due to the fact that the RF model was used to identify the important subset of predictor variables. The RF model was least influenced by the number of predictor variables included in the model (Figure 3), and showed the best overall performance among those tested (Figure 4). Training model RMSE averaged 16.5 $\pm$ 1.2 meq m^{-2} yr^{-1} across all models (Figure 4). BRT and MARS models both showed inconsistent behavior; performance was high when many predictor variables were included, but model error rates increased with fewer predictors (Figure 3).

LM and GWR models recorded the poorest overall mean training error rates when averaged across all models tested (28.8 $\pm$ 3.9 and 25.7 $\pm$ 4.7 meq m^{-2} yr^{-1} for LM and GWR, respectively; Figure 4). The bandwidth (i.e., search distance to collect neighboring data points used to train the local GWR regression models) selected by GWR model cross validation was $\sim$250 km, a distance that incorporated a high degree of environmental heterogeneity, and likely precluded any increase in model performance over the LM model.

Cross-validation estimates of RMSE for all models ranged from $\sim$30 to 35 meq m^{-2} yr^{-1}, and RF models reported the lowest RMSE rates of the models tested. $CV_{10\text{-fold}}$ error rates of the LM and GWR models were

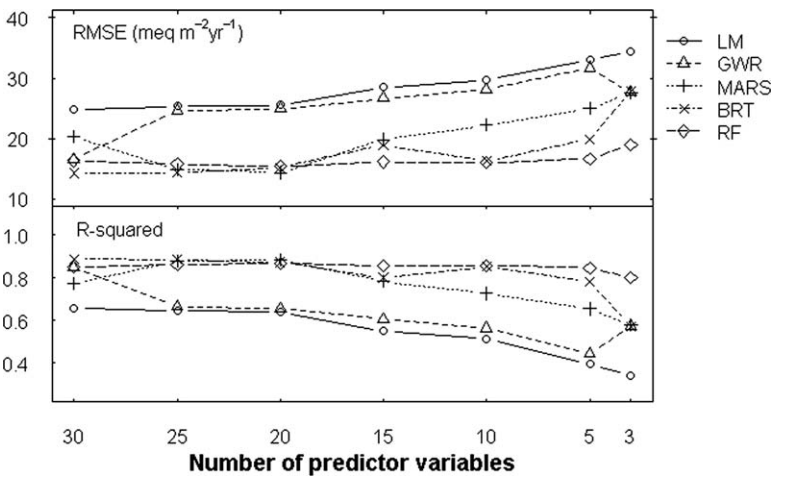

Figure 3. Training model error rates for five statistical models tested (LM = linear regression, GWR = geographically weighted regression, MARS = multivariate adaptive regression splines, BRT = boosted regression trees, RF = random forest). RMSE is root mean squared error.

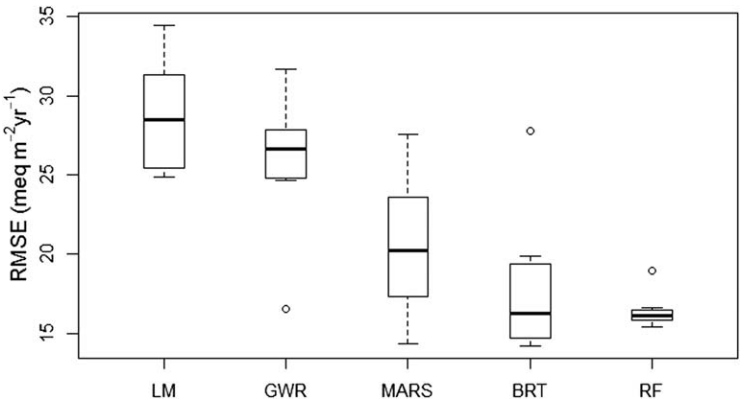

Figure 4. Boxplot summary of training model error rates averaged across all combinations of number of variables entered into each statistical model. Each boxplot constructed using training RMSE from models built with successively fewer predictor variables for each regression model; see Figure 3. Abbreviations are the same as Figure 3.

<8% higher than those based on estimates from the training data set. $CV_{10\text{-fold}}$ error rates for the ML algorithms increased from 16 to 22 meq m^{-2} yr^{-1} to 30 to 35 meq m^{-2} yr^{-1} (Table 3), indicating that some model instability remained. $CV_{10\text{-fold}}$ also indicated that there was more evidence of model overfitting among the ML algorithms compared to LM and GWR models, despite their superior overall performance.

Instability observed across all models was likely due to two common factors: (1) under-sampled BC_w across the geographic range of the predictors, and (2) inconsistencies in spatial accuracy of some remotely sensed predictor variables. The under-sampling of BC_w across the geographic range of the predictors was a by-product of our ad hoc sampling design, which incorporated a large number of MAGIC-modeled catchment-level data sets into a single global data set (see section 2.2). Catchments showing high BC_w rates were poorly represented in this global data set, leading to higher error rates in MAGIC model predictions of high BC_w.

Inconsistency in spatial accuracy of some predictor variables was related to errors in the remotely sensed data itself; a problem common to all spatial modeling applications [*Elith and Leathwick*, 2009; *Guisan et al*., 2007]. For example, raster data in our study were resampled to 30 m resolution to establish a common map resolution, but native resolutions ranged from 30 m to 1 km. Accuracy of some raster layers, such as lithology, varied from state to state due to differences in mapping protocol. Such inconsistencies will surely influence the generalizability of model predictions to areas outside the training set. Upslope averaging of the predictor variables reduced but could not eliminate the influence of data resolution on model behavior, as each predictor was summarized to individual pour points across upslope catchments that were much larger than the spatial resolution of the predictor variables.

3.3. Ecoregional Models

Individual ecoregion models did not improve overall model goodness-of-fit in comparison to a single global model. RMSE training rates were lowest for Blue Ridge (BR) and Ridge and Valley (RV) ecoregions (~15 meq m^{-2} yr^{-1}) and highest for the Central Appalachians (CA; 24 meq m^{-2} yr^{-1}), based on RF model predictions (Table 4). The poor performance of the CA model is likely due to the low sample size compared to its environmental variability. Moreover, the CA model included several predictor variables not associated with the other ecoregional models, or the full model (Tables 4 and 5). For example, >70% of the predictor variables in the BR and RV models matched those included in the full model (see below). But, for the CA model, only one variable (wet S deposition) was included in the full model. Other variables, unique in the CA model alone, described aspects of the vegetation (% cover), temperature, and precipitation regimes (Table 4). At face value, these results might suggest that environmental drivers of BC_w in the CA ecoregion may differ from those of other ecoregions. However, sampling within this ecoregion was relatively low (21 sites), suggesting that more sampling is needed to confirm the main environmental drivers of BC_w rates in this ecoregion.

3.4. Model Selection

Based on the model validation statistics, we selected a single global RF model with 10 predictor variables, built with 1000, three predictor variable regression trees (Table 3). This model was chosen due to its high

Table 3. Model Validation Statistics Calculated for Linear Regression (LM), Geographically Weighted Regression (GWR), Multivariate Adaptive Regression Splines (MARS), Boosted Regression Trees (BRT) and Random Forest (RF) Models[a]

Model	R-Squared	RMSE (meq m^{-2} yr^{-1})	$CV_{10\text{-fold}}$ RMSE
LM	51.2	29.6	30.6 ± 9.1
GWR	56.2	28.1	30.2 ± 9.7
MARS	73.8	21.7	34.9 ± 12.3
BRT	81.8	18.1	31.8 ± 11.2
RF	85.6	16.1	29.5±10.2

[a]RMSE is root mean squared error. The final column is the mean (±1 SD) RMSE resulting from 10-fold cross validation ($CV_{10\text{-fold}}$). All models included 10 predictor variables.

prediction accuracy and relatively few predictor variables. The model had a training error rate of 16.1 meq m^{-2} yr^{-1}, and an R^2 of 0.86. The OOB error rate was 31.4 meq m^{-2} yr^{-1} (R^2_{OOB} =0.452; data not shown), similar to the $CV_{10\text{-fold}}$ results (Table 3). The final RF model performed best when the observed BC_w was <125 meq m^{-2} yr^{-1} (Figure 5). Model errors were inflated for BC_w estimates above this value, presumably due to the scarcity of data points in relatively high BC_w environments. We note that it is more important to accurately identify areas of low BC_w because the inability to maintain acidic deposition levels below the natural resupply rate of base cations is of greatest ecological and regulatory consequence. Hence, higher error rates were more acceptable for relatively high BC_w rates.

3.5. Biophysical Predictors of BC_w

The selected predictor variables represented key aspects of the local climatic regime, atmospheric S inputs, underlying lithology, and the soil characteristics of the upslope catchment. Low BC_w rates were predicted for catchments that received relatively low precipitation, in areas underlain by siliceous lithology, with low soil clay, low Kjeldahl nitrogen, low organic matter content, and relatively high levels of canopy cover of mixed coniferous and deciduous forest (Figure 6). Response functions from the RF models were generally nonlinear over the range of the predictors, but distinct directional patterns were obvious for most variables (Figure 6).

Low BC_w was consistently associated with the presence of siliceous bedrock, which is characteristically low in base cations [*Duan et al*., 2002; *Posch et al*., 2003]. Siliceous lithology was identified as the most influential predictor variable in all global models. This result is consistent with independent geological sensitivity mapping by *Sullivan et al*. [2007]. The RF modeled response curve for the percent siliceous lithology was nonlinear in the left tail. These relations suggested that BC_w declined precipitously in catchments with ~10% siliceous lithology, and additional increases in the proportion of siliceous bedrock did not lead to significantly greater declines in BC_w rates (Figure 6).

Table 4. Training Model Root Mean Squared Error (RMSE) and Variable Importance Measures for Ecoregional Models Used in Random Forest Models to Predict BC_w Rates Across the Southern Appalachian Study Region[a]

Variable	Description	BR	CA	RV	ALL
S_WET	Wet sulfur deposition	7 (+)	7 (+)	2 (+)	2 (+)
LITH_SIL	Percent siliceous lithology	4 (−)		1 (−)	1 (−)
PRECIPNG	NGS precipitation	3 (+)		4 (+)	5 (+)
PDAYMAX	Maximum days with precipitation (when >10°C)	1 (+)			4 (+)
OMNEW	Amount of soil organic matter	6 (+)		5 (+)	3 (+)
SOIL_CLAY	Percent soil clay				7 (+)
SOIL_PH	Soil pH	2 (+)			
NITRONEW	Soil Kjeldahl nitrogen			3 (+)	6 (+)
NEWCTI	Compound topographic index		6 (+)		
LITH_FEL	Percent felsic lithology	5 (−)			
LSTGROW	Number of GS days >5.6°C		2 (+)		
PSUMMAX	Maximum annual precipitation amount (when >10°C)		3 (+)		
AB90GROW	Number of days >32.2°C		4 (+)		
GPPNG	NGS gross primary productivity			6 (−)	
RIGROW	Growing season respiration index			7 (−)	
CON42	Percent conifer forest cover		1 (−)		
FOREST	Percent forest cover		5 (−)		
Sample size		79	21	38	140
RMSE		15.7	23.8	15.2	16.8

[a]Numbers indicate the ranking of each variable in the model by its variable importance score within the random forest model. The (+/−) signs indicate the approximate positive/negative effect of the predictor on BC_w. Ecoregions are: Blue Ridge Mountains (BR), Central Appalachian (CA), and Ridge and Valley (RV). See supporting information Table S1 for descriptions of each predictor variable. Models were built using the top seven predictor variables, including the ALL ecoregion model, which is shown for comparison. A ranking of one indicates the variable with the highest variable importance score, and seven the lowest.

Table 5. Root Mean Squared Error (RMSE) Rates (meq m^{-2} yr^{-1}) for Random Forest Models Built With Seven Predictor Variables for Each Individual Ecoregion and All Ecoregions Combined[a]

	Training Data			
Testing Data	Blue Ridge	Central Appalachian	Ridge and Valley	All
Blue Ridge	15.7	46.5	31.5	14.9
Central Appalachian	54.1	23.8	51.5	26.4
Ridge and Valley	44.1	46.9	15.2	13.9
All	33.2	44.4	32.0	16.8

[a]Each of the resulting four models was then tested in turn on each ecoregion and all ecoregions combined. RMSE scores on the diagonal represent estimates of model error based on estimates made on the training data for that ecoregion (Table 4).

Organic matter, Kjeldahl nitrogen, and soil clay content were positively correlated with BC_w rate (Figure 6). Predicted BC_w rose sharply and linearly in catchments where soil organic matter ranged from ~500 to 750 kg ha^{-1}, however, there was a negligible increase associated with higher organic matter levels. A significant interaction was found between organic matter and precipitation (Figure 7), with the lowest BC_w rates occurring in areas with low to moderate levels of precipitation and low soil organic matter. Organic matter may accelerate weathering rates, thereby leading to the positive correlation between OM levels and BC_w [*Drever and Stillings*, 1997], but this relationship is complex and influenced by pH and parent material composition [*Drever*, 1994; *Kump et al*., 2000]. For Kjeldahl N levels, BC_w increased slowly and linearly in catchments with ~2500–5000 kg ha^{-1}, but it rose sharply beyond this point. Similarly, the percentage of soil clay had little influence on predicted BC_w in catchments with <15% soil clay, but BC_w rose sharply in catchments with soil clay ranging from 15% to 23%.

Besides soils and lithology, climate variables reflecting temperature and precipitation regimes are known to influence weathering rates [*Johnson et al*., 1994; *Kump et al*., 2000; *Peltier*, 1950], and several climate variables were retained in the models. A positive correlation was identified between BC_w, precipitation [total consecutive days with precipitation (PDAYMAX) and nongrowing season accumulation (PRECIPNG)], and temperature (mean 95th percentile of maximum diurnal surface temperature difference during the local nongrowing season (DIFF95NG); Figure 6). Large increases in BC_w rates were predicted for catchments with >90 days with precipitation. BC_w also increased linearly with the amount of precipitation falling in the nongrowing season. Precipitation likely influences rates of BC transport within the soil solution and alters chemical concentrations in soil water over time [*Drever*, 1994]. Large linear increases in BC_w were also predicted for catchments with >8°C maximum diurnal surface temperature difference during the local nongrowing season. Inclusion of the variable DIFF95NG in the model (Figure 6) likely indicated that large diurnal differences in surface temperatures may be related to enhanced physical weathering by freeze-thaw cycles [*Peltier*, 1950]. Temperature also influences chemical reaction and respiration rates in soils and overall weathering rates [*Lloyd and Taylor*, 1994; *Winkler et al*., 1996; *Wright et al*., 2006].

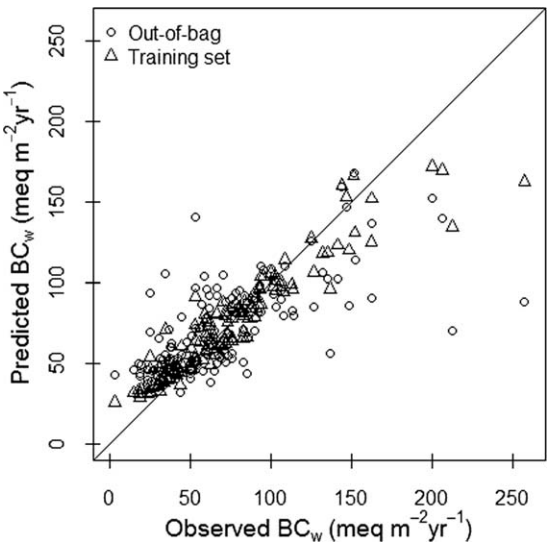

Figure 5. Final random forest model predictions of BC_w plotted against observed (MAGIC calibrated) BC_w. Black circles indicate out-of-bag predictions (similar to cross validation, but see text) and black triangles indicate predictions made on the original training set of data.

RF model interactions between precipitation (PDAYMAX), siliceous lithology (LITH_SIL), and organic matter (OMNEW) indicated that high levels of precipitation were consistently related to high BC_w rates, regardless of soil or lithology (Figure 7). This suggests that BC inputs and/or enhanced weathering associated with higher precipitation may counteract low BC_w rates in inherently low-BC soil minerals. Similarly, interactions were identified between wet S deposition, precipitation, and siliceous lithology, which indicated that high wet S deposition can either exacerbate the influence of

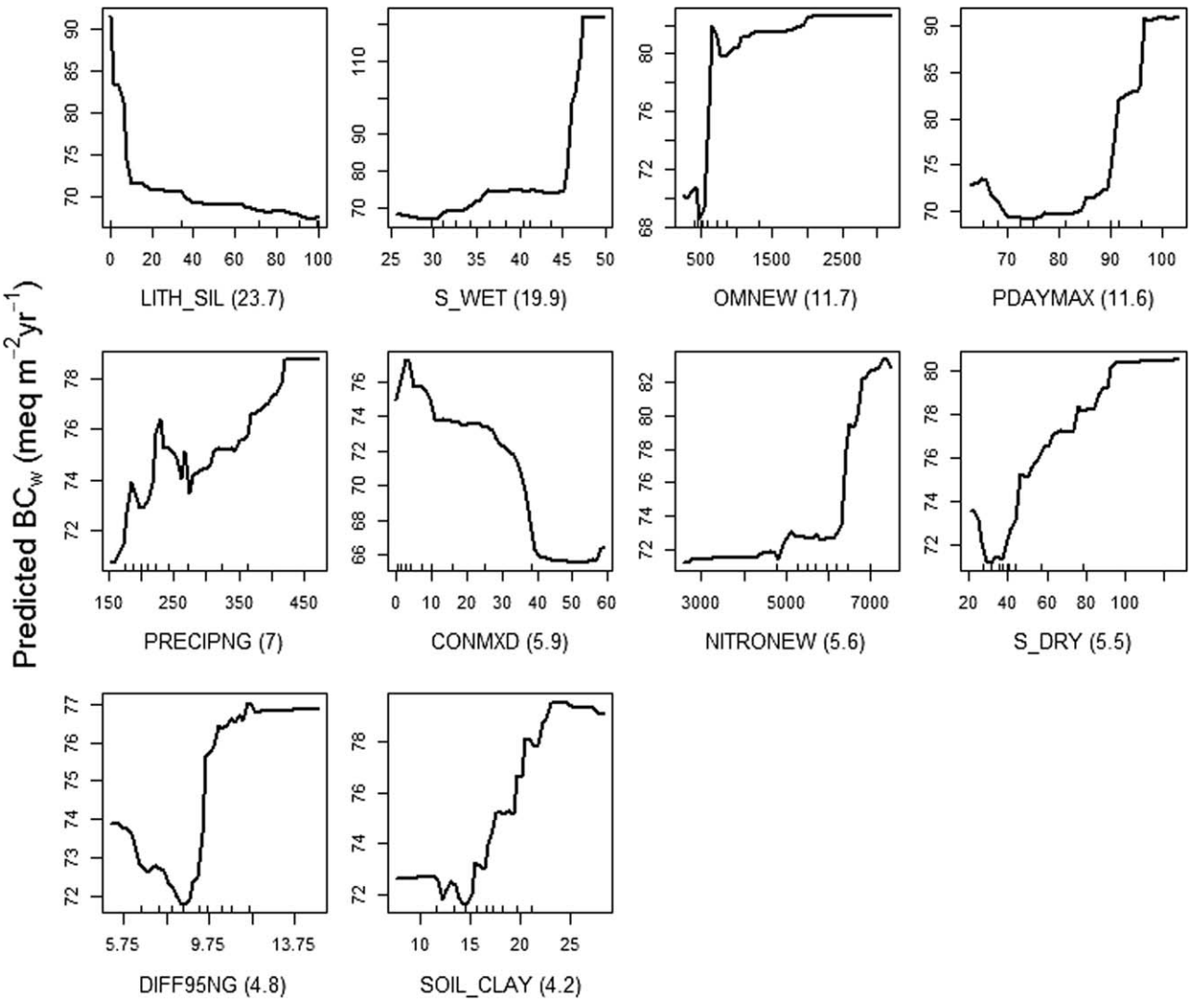

Figure 6. Response curves showing relations between predicted BC_w and individual predictor variables included in the final RF model. Black tick marks (rug plot) on the *x* axis indicate decile classes for the predictor variables. The *y* axis indicates the relative effect of the predictor on BC_w. In general, higher *y* axis values indicate higher predicted BC_w. Numbers in parentheses indicate the variable importance measure (i.e., mean decrease in squared error) scaled to sum to 100% over all variables included in the model. Variables are: LITH_SIL, percent of catchment in siliceous lithologies (% cover); S_WET, amount of wet sulfur deposition (meq m^{-2} yr^{-1}); OMNEW, mean soil organic matter (kg ha^{-1}); PDAYMAX, mean penultimate maximum days with precipitation (days); PRECIPNG, amount of precipitation in the nongrowing season (mm 0.01); CONMXD, percent of catchment in mixed-conifer (%); S_DRY, amount of dry sulfur deposition (meq m^{-2} yr^{-1}); DIFF95NG, Mean 95th percentile of maximum diurnal surface temperature difference during the local nongrowing season (°C); NITRONEW, mean soil Kjeldahl nitrogen to 50 cm depth (kg ha^{-1}); SOIL_CLAY, mean percent clay of catchment soils (%). All variables are reported in their native units (see Table S1).

precipitation or otherwise increase BC_w in BC-poor catchments (Figure 7). Taken together our results indicate that climate and the level of S deposition may moderate BC_w rates by influencing soil process rates and influxes of atmospherically deposited strong acids.

The importance of identifying climatic correlates of BC_w with our models has implications for future monitoring and continued modeling of stream water acid-base status [*Benčoková et al.*, 2011; *Evans*, 2005; *Wright et al.*, 2006]. Climate projections for the southern Appalachian Mountain region suggest both warmer and drier (south) and warmer and wetter (north) futures (year 2100 projections) [*Hayhoe et al.*, 2008; *Karl et al.*, 2009; *Solomon et al.*, 2007]. BC_w modeling results reported here indicate that areas with relatively low precipitation and warm temperatures display relatively low BC_w rates. Drier conditions would tend to decrease wet S deposition and leaching losses of BC from soils. Rising temperatures in general should lead to higher weathering rates [*Wright et al.*, 2006]. However, our model indicates that temperature alone (TDAYMEAN; data not shown) has a negative influence on BC_w rates, and large differences in diurnal surface

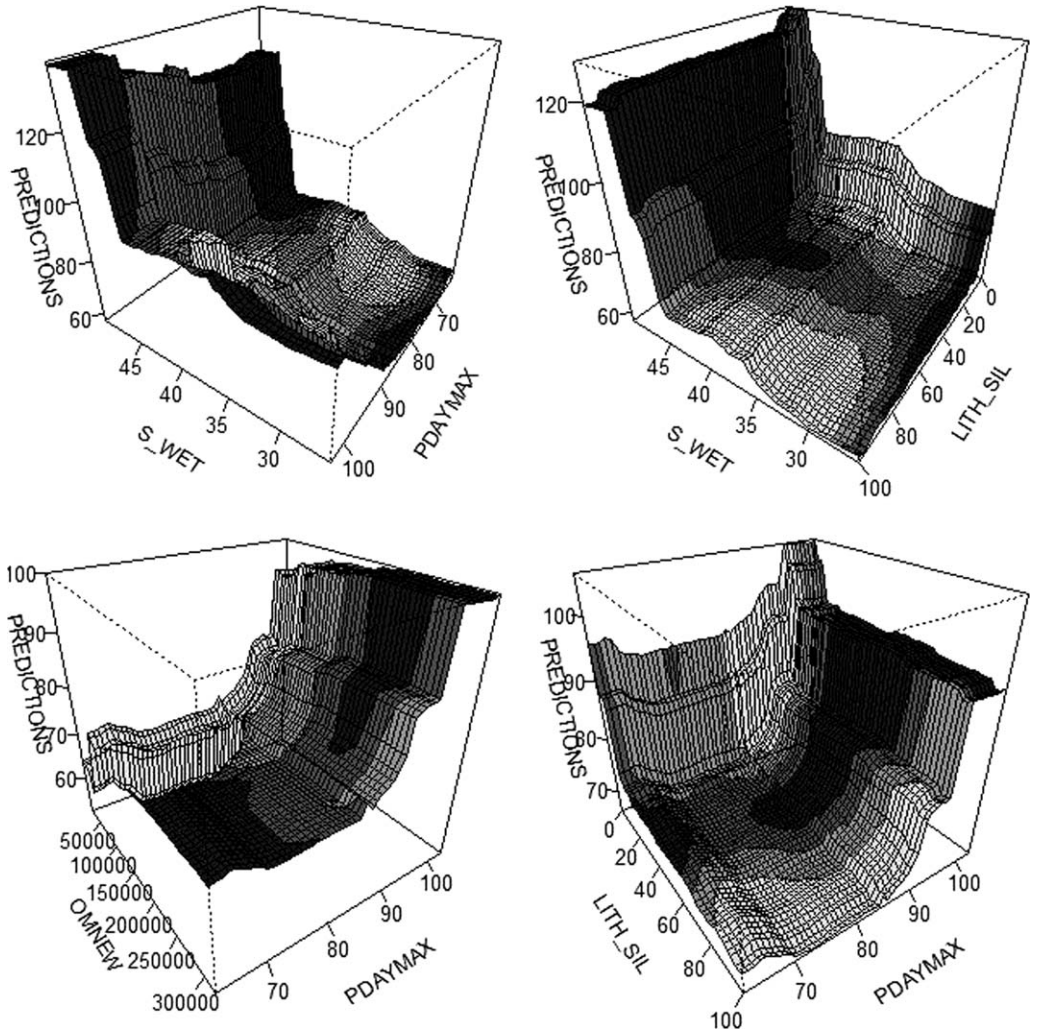

Figure 7. Surface plots displaying interactions between two continuous predictor variables used in modeling BC_w rates (meq m^{-2} yr^{-1}). Top interactions were calculated for final RF models using methods similar to those of *Elith et al*. [2008] for gradient boosted regression trees. Areas with light shading indicate regions in data space where sample sizes were relatively high. Darker areas indicate regions in data space associated with few data, where model estimates should be interpreted with caution. Variables are: S_WET (meq m^{-2} yr^{-1}), amount of wet sulfur deposition; LITH_SIL, percent of catchment in siliceous lithologies (%); PDAYMAX, mean penultimate maximum days with precipitation (days); and OMNEW, mean soil organic matter (kg ha^{-1} 0.01). All variables are reported in their native units (see Table S1).

temperatures lead to higher BC_w rates. Likely further sampling is needed across heterogeneous environmental conditions to better understand causal connections between climatic driving variables and corresponding BC_w rates.

Ecosystem responses to long-term climatic change are largely unknown due to the complexity of interactions among the covariates. Changes in temperature, precipitation, and insolation regimes, and accompanying changes in site productivity, vegetative communities, plant cover, carbon and nutrient uptake by plants, and influence of disturbances may all affect BC_w within the study domain. Thus, it will be important to continue to monitor and model BC_w and its relations with covariates. Likewise, new combinations or changes in the importance of predictor variables will likely necessitate continued stream water chemistry

Table 6. Comparisons Among Random Forest (RF) and Linear Regression (LM) Model Predictions Across the Study Region[a]

BC_w Classes	RF	LM[b]
<20	0.0	2.2
20–50	4.0	4.8
50–100	67.6	47.7
100–150	26.8	42.1
>150	1.6	3.2
Sum	100.0	100.0

[a]Values represent the percentage of 30 m grid within predefined BC_w classes.
[b]See supporting information Figure S2 for predictions from this mode.

sampling that evenly samples the shifting variability of conditions represented by the predictor variables.

Vegetation influences BC_w through physical processes that fracture bedrock, and chemical processes that control the deposition of organic matter, uptake of water and nutrients, and release of carbon dioxide into soil water [*Drever*, 1994; *Hultberg*, 1985; *Miles*, 1986; *Wright et al*., 2006]. Coniferous forests, in general, develop humus layers high in organic acids, which can increase BC leaching, lower soil pH, and increase chemical weathering [*Augusto et al*., 2002; *Hornung*, 1985; *Johnson et al*., 1994], which in turn may influence BC_w rates. Accordingly, our models identified a negative and nonlinear relationship between the amount of mixed deciduous-coniferous forest cover and the predicted BC_w rate (Figure 6). The relation indicated that BC_w rates fell sharply for catchments with >30% coverage of mixed forest. A similar relationship was found for pure conifer cover types.

Sulfur deposition was positively correlated with BC_w, and wet S deposition was included as a main predictor variable in the global model, and in all ecoregion models. However, the influence of S deposition on BC_w remains unclear [*Johnson et al*., 1994]. An interaction between wet S deposition and PDAYMAX suggested that S deposition can confound a positive correlation between precipitation and BC_w, particularly at high levels of S deposition.

Sulfur deposition levels are strongly influenced by orography [*Byun and Schere*, 2006]; hence, this variable may "code" for several topographic and climatic variables, thereby contributing to a high variable importance score in the models. To better understand these relations, we ran additional RF models where S deposition was excluded from the predictor set. RF model error rates increased <10% by eliminating S deposition, and several additional predictor variables were substituted for S deposition. These included average catchment soil pH (positive correlation with BC_w), nongrowing season solar insolation (negative correlation), and the fraction of photosynthetically active radiation (FPAR) absorbed by vegetation during the nongrowing season (negative correlation). Low predicted BC_w rates were associated with high solar insolation, low catchment soil pH, and high nongrowing season FPAR. The inclusion of FPAR likely differentiated coniferous from deciduous forests, and high from low elevations. The negative relationship between solar insolation and BC_w rates suggested that north facing and less exposed catchments may benefit from cool and moist conditions, which may moderate otherwise warm and dry climatic influences. Together, these variables represent climatic processes that occur at finer scales than S deposition.

Regardless of the amount of S deposition and its lagged effects on streams, this information may be useful to managers when considering mitigation measures. To illustrate, we plotted RF and LM predicted BC_w maps, where S deposition was included and excluded from the models (Figures S1 and S2, respectively). From these maps, it is apparent that there are fairly large differences in fine-scaled patterns of predicted BC_w driven by the inclusion or exclusion of deposited S. These results suggest that BC_w rates may vary considerably across small spatial extents due to edaphic, topographic, and orographic influences.

3.6. Summary of Model Predictions

Models predicted that BC_w rates ranged between 25.2 and 174.2 meq m^{-2} yr^{-1}, and that 72% of the southern Appalachian Mountain region displayed predicted rates <100 meq m^{-2} yr^{-1} (Table 6). RF model predictions underestimated BC_w rate at the upper end of the distribution (>100 meq m^{-2} yr^{-1}), as evidenced by higher MAGIC model simulated values at many locations (Figure 5). Areas with high BC_w were generally concentrated in the northwest portion of the study area, where high levels of precipitation during the nongrowing season likely contribute to relatively high weathering rates (Figure 8). Strongly patterned siliceous lithologies were apparent throughout the RV ecoregion, where some of the lowest BC_w rates were also found on NE/SW trending, parallel, sandstone ridges (Figure 8).

Model uncertainty was also highest in the northwestern portion of our study area. Locations showing the highest uncertainty were associated with high predicted BC_w rates, low MAGIC sampling density, and location within the CA ecoregion (Figure 8). This ecoregion model also displayed the highest error rate and included a unique set of predictors, in comparison with other ecoregion models. This observation was consistent with the results of *McDonnell et al*. [2012]; however, model RMSE rates reported here were 50% lower. As expected, model uncertainty was lowest for areas near the locations of MAGIC-modeled catchments, and in areas dominated by siliceous lithologies (e.g., NE/SW trending sandstone ridges in the Ridge and Valley ecoregion of Virginia).

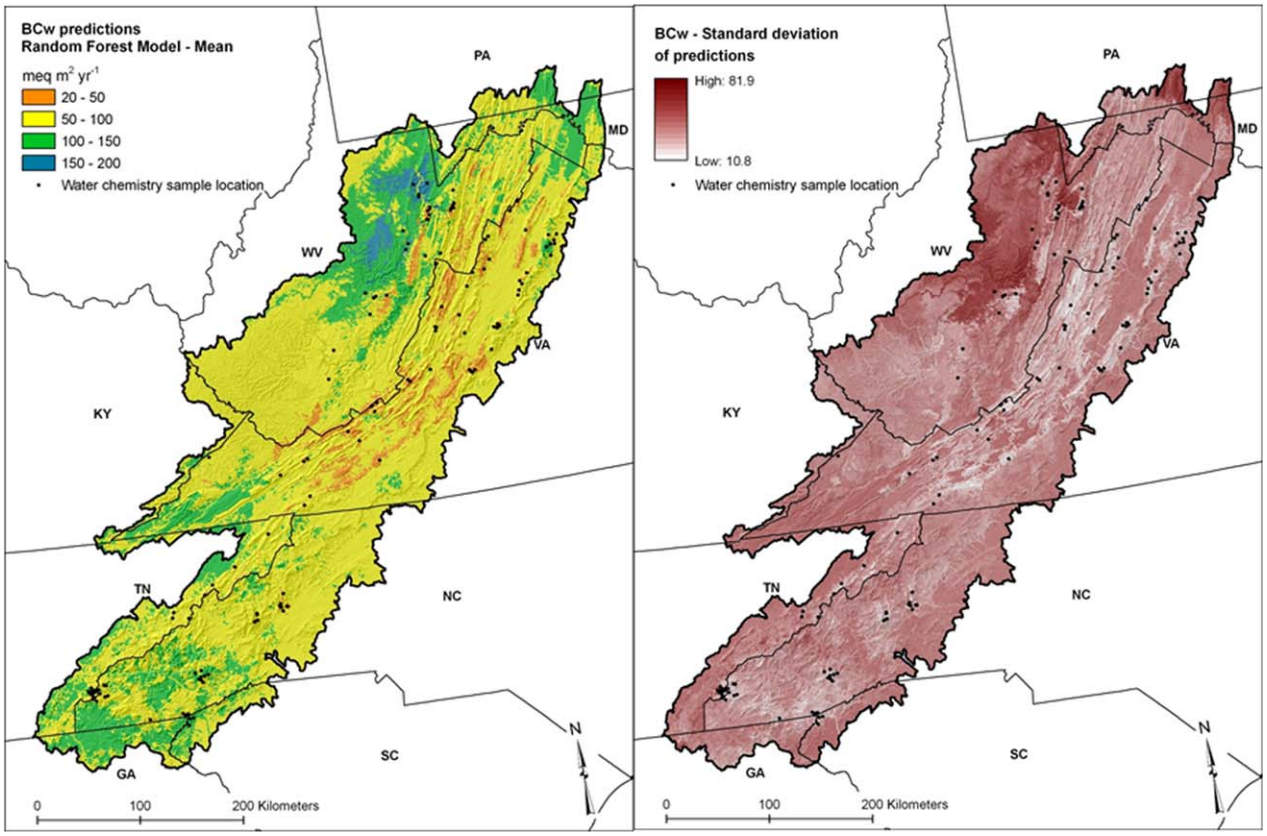

Figure 8. (left) Continuous surface of predicted BC_w and (right) standard deviation of predictions from the final model. Standard deviations were calculated from the predictions made from the 1000 individual regression trees within the RF algorithm.

Our assessment builds on previous work completed by *McDonnell et al*. [2012], which also used MAGIC model estimates to quantify site-specific BC_w rates. *McDonnell et al*. [2012] used three separate ecoregional models to predict BC_w to their study domain (~70% of our study domain). The authors concluded that less than one-third of their study region displayed BC_w rates <100 meq m^{-2} yr^{-1}, compared to our finding of more than three-quarters of the study region (Table 6). Furthermore, the authors reported nearly one-quarter of the landscape with BC_w rates >200 meq m^{-2} yr^{-1}, compared to our finding of $<2\%$. There are several likely explanations for the observed differences. The *McDonnell et al*. [2012] study used (1) ecoregion-specific models to predict to the study region, (2) linear models with stepwise variable selection, (3) 33% fewer data points (92 versus 140) in model training, and (4) a more restricted geographic extent.

As a direct comparison, we developed a LM with the 10 variables chosen by the RF model (Figure 6). Predictions from the LM were more evenly distributed among BC_w classes (Table 6), but were still concentrated in the 50–100 and 100–150 meq m^{-2} yr^{-1} classes. This suggests that our model predictions were fairly robust to the choice of statistical model used.

4. Conclusions

Correlative landscape modeling was used to identify relationships between environmental predictors and individual catchment-level BC_w rates across the southern Appalachian Mountain region. Results from this analysis suggest that several broad- to fine-scale cofactors are related to chemical weathering rates, including precipitation and temperature regimes, lithology, soil properties and vegetation, and interactions among these variables. ML algorithms outperformed traditional linear regression analyses. In particular, RF

was least influenced by model parameterization and number of predictor variables included in the model, and it helped uncover several nonlinear relationships among predictors and BC_w rates, as well as important interactions among predictors. Furthermore, when model parsimony and predictive ability were considered, the RF model outperformed all other models.

Predictions from the RF model indicate that the lowest BC_w rates generally occur in dry areas underlain by siliceous lithologies that experience high levels of S deposition, and in small catchments that contain low levels of soil pH, N, and organic matter. Model predictions indicated that more than three-quarters of the study region displayed BC_w rates <100 meq m^{-2} yr^{-1} and suggested that weathering rates across the region were inherently low. To improve model error rates and accuracy, MAGIC model estimates should be made at additional stream water sampling locations throughout the study area, and particularly in currently under-sampled geographic areas (areas showing high SD in Figure 8). Further model enhancements based on sample augmentation may increase our understanding of broad and fine-scale biophysical drivers that are not directly modeled in the process-based calculations of BC_w, and enable more refined predictions of weathering rates across the landscape.

Acknowledgments
The authors would like to thank the two anonymous reviewers who contributed valuable input into the revisions of this manuscript. We would also like to acknowledge the U.S. EPA for providing the funding for this project.

References

Augusto, L., J. Ranger, D. Binkley, and A. Rothe (2002), Impact of several common tree species of European temperate forests on soil fertility, *Ann. For. Sci.*, *59*(3), 233–253.
Benčoková, A., J. Hruška, and P. Krám (2011), Modeling anticipated climate change impact on biogeochemical cycles of an acidified headwater catchment, *Appl. Geochem.*, *26*, supplement, S6–S8.
Berk, R. A. (2008), *Statistical Learning From a Regression Perspective*, Springer, New York.
Bivand, R., D. Yu, T. Nakaya, and M. Garcia-Lopez (2010), spgwr: Geographically weighted regression, R package version 0.6-19. [Available at: http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=nrcs142p2_053629.]
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.
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.
Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler (2007), Random forests for classification in ecology, *Ecology*, *88*(11), 2783–2792.
De'ath, G. (2007), Boosted trees for ecological modeling and prediction, *Ecology*, *88*(1), 243–251.
Drever, J. I. (1994), The effect of land plants on weathering rates of silicate minerals, *Geochim. Cosmochim. Acta*, *58*(10), 2325–2332.
Drever, J. I., and L. L. Stillings (1997), The role of organic acids in mineral weathering, *Colloids Surf. A*, *120*(1–3), 167–181.
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*(3), 180–198.
Duan, L., J. Hao, S. Xie, Z. Zhou, and X. Ye (2002), Determining weathering rates of soils in China, *Geoderma*, *110*(3–4), 205–225.
Elith, J., and J. R. Leathwick (2009), Species distribution models: Ecological explanation and prediction across space and time, *Annu. 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.
Fotheringham, A. S., C. Brunsdon, and M. Charlton (2002), *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, John Wiley, New York.
Friedman, J. H. (1991), Multivariate adaptive regression splines, *Ann. Stat.*, *19*(1), 1–67.
Friedman, J. H. (2002), Stochastic gradient boosting, *Comput. Stat. Data Anal.*, *38*(4), 367–378.
Gesch, D., M. Oimoen, S. Greenlee, C. Nelson, M. Steuck, and D. Tyler (2002), The national elevation dataset, *Photogramm. Eng. Remote Sens.*, *68*(1), 5–11.
Gesch, D. B. (2007), The national elevation dataset, in *Digital Elevation Model Technologies and Applications: The DEM Users Manual*, 2nd ed., edited by D. Maune, pp. 99–118, Am. Soc. for Photogramm. and Remote Sens., Bethesda, Md.
Greaver, T. L., T. J. Sullivan, J. D. Herrick, M. C. Barber, J. S. Baron, B. J. Cosby, M. E. Deerhake, R. L. Dennis, J. J. B. Dubois, and C. L. Goodale (2012), Ecological effects of nitrogen and sulfur air pollution in the US: What do we know?, *Frontiers Ecol. Environ.*, *10*(7), 365–372.
Grimm, J. W., and J. A. Lynch (2004), Enhanced wet deposition estimates using modeled precipitation inputs, *Environ. Monit. Assess.*, *90*(1), 243–268.
Guisan, A., C. H. Graham, J. Elith, and F. Huettmann (2007), Sensitivity of predictive species distribution models to change in grain size, *Diversity Distrib.*, *13*(3), 332–340.
Hall, J., B. Reynolds, S. Langan, M. Hornung, F. Kennedy, and J. Aherne (2001), Investigating the uncertainties in the Simple Mass Balance Equation for acidity critical loads for terrestrial ecosystems in the United Kingdom, *Water Air Soil Pollut.*, *1*(1), 43–56.
Hargrove, W., and F. Hoffman (2004), A flux atlas for representativeness and statistical extrapolation of the Ameriflux network, *ONRL Tech. Memo. ORNL-TM-2004*/112, pp. 1–152, Oak Ridge Natl. Lab., Oak Ridge, Tenn.
Hastie, T., and R. Tibshirani (2011), mda: Mixture and flexible discriminant analysis, R package version 0.4-2. [Available at: http://CRAN.R-project.org/package=mda.].
Hastie, T., R. Tibshirani, J. Friedman, and J. Franklin (2005), The elements of statistical learning: Data mining, inference and prediction, *Math. Intel.*, *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, *Mitigation Adaption Strategies Global Change*, *13*(5), 425–436.

Hemond, H. F. (1990), Acid neutralizing capacity, alkalinity, and acid-base status of natural waters containing organic acids, *Environ. Sci. Technol.*, *24*(10), 1486–1489.

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

Henriksen, A., P. Dillon, and J. Aherne (2002), Critical loads of acidity for surface waters in south-central Ontario, Canada: Regional application of the Steady-State Water Chemistry (SSWC) model, *Can J. Fish. Aquat. Sci.*, *59*(8), 1287–1295.

Hodson, M., and S. Langan (1999), Considerations of uncertainty in setting critical loads of acidity of soils: The role of weathering rate determination, *Environ. Pollut.*, *106*(1), 73–81.

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.

Hornung, M. (1985), Acidification of soils by trees and forests, *Soil Use Manage.*, *1*(1), 24–27.

Hultberg, H. (1985), Budgets of base cations, chloride, nitrogen and sulphur in the acid Lake Gårdsjön catchment, SW Sweden, *Ecol. Bull.*, *37*, 133–157.

Johnson, C. E., M. I. Litaor, M. F. Billett, and O. P. Bricker (1994), Chemical weathering in small catchments: Climatic and anthropogenic influences, in *Biogeochemistry of Small Catchments: A Tool for Environmental Research*, edited by B. Moldan and J. Cerny, chap. 14, pp. 323–341, John Wiley, Hoboken, N. J.

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

Kump, L. R., S. L. Brantley, and M. A. Arthur (2000), Chemical weathering, atmospheric CO_2, and climate, *Annu. Rev. Earth Planet. Sci.*, *28*(1), 611–667.

Li, H., and S. G. McNulty (2007), Uncertainty analysis on simple mass balance model to calculate critical loads for soil acidity, *Environ. Pollut.*, *149*(3), 315–326.

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

Lloyd, J., and J. A. Taylor (1994), On the temperature dependence of soil respiration, *Funct. Ecol.*, *8*(3), 315–323.

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.

Miles, J. (1986), What are the effects of trees on soils?, in *Trees and Wildlife in the Scottish Uplands*, edited by D. Jenkins, ITE Symposium No. 17, pp. 55–62. Huntingdon, U. K.

National Acid Precipitation Assessment Program (NAPAP) (1991), *1990 Integrated Assessment Report*, Off. of the Dir., Gov. Print. Off., Washington, D. C.

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

NRCS Soil Survey Staff (2010a), U.S. General Soil Map State Soil Geographic (STATSGO2) database. [Available at: http://soildatamart.nrcs.usda.gov, last accessed 6 Oct 2011.].

NRCS Soil Survey Staff (2010b), Soil Survey Geographic (SSURGO) database for southern Appalachian Region. [Available at: http://soildatamart.nrcs.usda.gov, last accessed 6 Oct 2011.].

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.

Ouimet, R., and L. Duchesne (2005), Base cation mineral weathering and total release rates from soils in three calibrated forest watersheds on the Canadian Boreal Shield, *Can. J. Soil Sci.*, *85*(2), 245–260.

Peltier, L. C. (1950), The geographic cycle in periglacial regions as it is related to climatic geomorphology, *Ann. Assoc. Am. Geogr.*, *40*(3), 214–236.

Posch, M., J. Hettelingh, and J. Slootweg (2003), *Manual for Dynamic Modelling of Soil Response to Atmospheric Deposition Report*, 71 pp., Coord. Cent. for Eff., Bilthoven, Netherlands.

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*, 3531–3546, doi:10.1002/wrcr.20308.

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.

R Development Core Team (2011), *R: A Language and Environment for Statistical Computing*, R Found. for Stat. Comput., Vienna, ISBN: 3-900051-07-0. [Available at: http://www.R-project.org/.].

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 the 45th annual Hawaii International Conference on System Sciences, IEEE, Wailea, Maui, Hawaii, 4–7 Jan.

Ridgeway, G. (2010), gbm: Generalized Boosted Regression Models, R package version 2.0-8. [Available at: http://CRAN.R-project.org/package=gbm.].

Skeffington, R. (2006), Quantifying uncertainty in critical loads: Aliterature review, *Water Air Soil Pollut.*, *169*(1), 3–24.

Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller (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, Cambridge, U. K.

Sullivan, T., B. Cosby, K. Tonnessen, and D. Clow (2005), Surface water acidification responses and critical loads of sulfur and nitrogen deposition in Loch Vale watershed, Colorado, *Water Resour. Res*, *41*, W01021, doi:10.1029/2004WR003414.

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., 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.

Sverdrup, H., and P. Warfvinge (1993), Calculating field weathering rates using a mechanistic geochemical model PROFILE, *Appl. Geochem.*, *8*(3), 273–283.

United Nations Economic Commission for Europe (UNECE) (2004), *Manual on Methodologies and Criteria for Modelling and Mapping Critical Loads and Levels and Air Pollution Effects, Risks and Trends, ICP Modelling and Mapping, UNECE Convention on Long-range Transboundary Air Pollution*, p. 254, Umweltbundesamt, Dessau, Germany.

United States Environmental Protection Agency (US EPA) (2008), *US EPA's Report on the Environment (Final Report)*, EPA/600/R-07/045F (NTIS PB2008-112484), U. S. Environmental Protection Agency, Washington, D. C.

United States Environmental Protection Agency (US EPA) (2009), *Risk and exposure assessment for review of the secondary National Ambient Air Quality Standards for Oxides of Nitrogen and Oxides of Sulfur*, Cent. for Environ. Assess., Off. of Res. and Dev., Res. Triangle Park, N. C.

Warfvinge, P., and H. Sverdrup (1992), Calculating critical loads of acid deposition with PROFILE—A steady-state soil chemistry model, *Water Air Soil Pollut.*, *63*(1), 119–143.

Whittingham, M. J., P. A. Stephens, R. B. Bradbury, and R. P. Freckleton (2006), Why do we still use stepwise modelling in ecology and behaviour?, *J. Anim. Ecol.*, *75*(5), 1182–1189.

Winkler, J. P., R. S. Cherry, and W. H. Schlesinger (1996), The Q10 relationship of microbial respiration in a temperate forest soil, *Soil Biol. Biochem.*, *28*(8), 1067–1072.

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.