Abstract

A boosted regression tree model was developed to predict pH conditions in three dimensions throughout the glacial aquifer system of the contiguous United States using pH measurements in samples from 18,386 wells and predictor variables that represent aspects of the hydrogeologic setting. Model results indicate that the carbonate content of soils and aquifer materials strongly controls pH and, when coupled with long flowpaths, results in the most alkaline conditions. Conversely, in areas where glacial sediments are thin and carbonate-poor, pH conditions remain acidic. At depths typical of drinking-water supplies, predicted pH > 7.5—which is associated with arsenic mobilization—occurs more frequently than predicted pH < 6—which is associated with water corrosivity and the mobilization of other trace elements. A novel aspect of this model was the inclusion of numerically based estimates of groundwater flow characteristics (age and flowpath length) as predictor variables. The sensitivity of pH predictions to these variables was consistent with hydrologic understanding of groundwater flow systems and the geochemical evolution of groundwater quality. The model was not developed to provide precise estimates of pH at any given location. Rather, it can be used to more generally identify areas where contaminants may be mobilized into groundwater and where corrosivity issues may be of concern to prioritize areas for future groundwater monitoring.

Introduction

The pH of water influences geochemical processes such as cation exchange, sorption, complexation, and mineral dissolution or precipitation (Stumm and Morgan 1981; Appelo and Postma 2005). The influence of these pH-dependent processes is important in aquifer systems used for drinking-water supply because they can affect the mobility and fate of geogenic constituents such as manganese (Mn), arsenic (As), other trace elements and radionuclides (Kent et al. 2000; Stollenwerk 2003; Brown et al. 2019; Szabo et al. 2020), and anthropogenic contaminants such as pesticides (Site 2001). The pH of groundwater also affects corrosivity, which, if left untreated, can dissolve lead (Pb) and other trace metals from pipes and other components in water distribution systems (Belitz et al. 2016). Previous investigators have used geostatistical interpolation methods to estimate groundwater pH at the catchment scale (Nas and Berktay 2010; Gunarathna et al. 2016). This paper demonstrates the use of machine learning (ML) to predict pH conditions at the sub-continental scale, in the stratigraphically complex glacial aquifer system (GLAC), which is an important source of drinking water in the United States. The size and complexity of the GLAC prohibits the direct simulation of groundwater flow and reactive solute transport with a continuous domain model (Parkhurst et al. 2010; Starn and Belitz 2018). Empirically based methods such as ML are amenable to making predictions from large, complex datasets through the process of statistical learning (Hastie et al. 2017).
(response variables) based on a set of input terms (predictor variables) (Hastie et al. 2017). Examples of supervised ML include artificial neural networks (ANN), Bayesian networks (BN), support vector machines, and ensemble-tree methods such as random forest and gradient boosting (Hastie et al. 2017). Here we use boosted regression trees (BRT)—a form of gradient boosting—to predict pH conditions throughout the GLAC. Previous investigators have demonstrated that ensemble-tree methods provide more accurate predictions of groundwater quality as compared to ANN or BN (Nolan et al. 2015), and other statistical methods such as linear or logistic regression (Nolan et al. 2014; Wheeler et al. 2015; Ayotte et al. 2016; Ransom et al. 2017; Yan et al. 2019). Tree-based models use a series of IF-THEN statements (decision rules) to split the data multiple times until specified criteria are met (e.g., number of trees, number of splits, minimum number of observations per node). A prediction is then generated as the average value of observations in each terminal node (Kuhn and Johnson 2016). BRT iteratively fits many simple trees and then combines them for prediction (Elith et al. 2008). The trees are built on the residual errors of the existing tree assemblage and, thus, increasingly focus on the most difficult observations to predict. BRT models do not require data transformations, can fit complex nonlinear relations, and automatically incorporate interaction effects among predictors (Elith et al. 2008). BRT models are, therefore, well suited for predicting water-quality conditions in complex, heterogeneous settings such as the GLAC (Erickson et al. 2018) and other aquifer settings (Ransom et al. 2017; Knoll et al. 2019).

BRT predictions are based on patterns identified in a training dataset with no assumptions of the processes controlling the response variable. This can lead to spurious or inaccurate predictions, particularly if predictions are made outside the range of data used to train the model (Read et al. 2019). Furthermore, overreliance on predictor variables that represent surface or near-surface conditions may adversely affect the ability of ML to generate accurate predictions of conditions in the subsurface (Ayotte et al. 2016). To counter these issues, output from process-based (numerical) and geostatistical models can be incorporated into a BRT model to integrate information on controlling processes or conditions and, thus, to help guide predictions (Fienen et al. 2015; Read et al. 2019). For example, numerical models of groundwater flow can provide predictor variables that constrain three-dimensional aspects of the flow system such as water fluxes, flowpath lengths, and groundwater age (Walter and Starn 2013; Ransom et al. 2017; Starn and Belitz 2018) and interpolated empirical observations can describe and constrain subsurface conditions that vary spatially and with depth such as aquifer texture characteristics (Nolan et al. 2014; Ayotte et al. 2016; Bayless et al. 2017; Haj et al. 2018).

The GLAC consists of a complex collection of heterogeneous, discontinuous aquifers in quaternary sediments north of the line of maximum glacial advance in the northern contiguous United States (Yager et al. 2018). The GLAC, which covers an area of \(1.87 \times 10^6 \text{ km}^2\) including parts of 25 States, has been subdivided into 17 hydrogeologic terranes, each of which contains quaternary sediment that has a common depositional history and can be characterized by similar age, texture, and thickness (Figure 1A; Table S1) (Yager et al. 2018). The GLAC ranks first in the Nation as a source of groundwater for both public and domestic supplies, with a combined pumped of about 2.6 billion gallons per day for these purposes (Maupin and Barber 2005; Maupin and Arnold 2010). About 30 million people are reliant on the GLAC for their water supplies (Erickson et al. 2019).

The pH of groundwater in the GLAC is of interest because the pH-dependent constituents Mn and As have been found to occur at concentrations greater than their human-health benchmarks in samples from public- and domestic-supply well depths more frequently than other constituents (Warner and Ayotte 2015; Stackelberg 2017; Erickson et al. 2019). The occurrence of Mn and As in drinking water is a human-health concern because Mn has been identified as a neurotoxin and potential carcinogen (Spangler and Reid 2010; Harischandra et al. 2019) and As is a known bladder and skin carcinogen (Baris et al. 2016; Mayer and Goldman 2016) and can cause a variety of other adverse health effects (NRC 2013). Water corrosivity is primarily of concern for the estimated 6.3 million people who rely on water from the GLAC obtained from private, domestic wells (Erickson et al. 2019) because water from private wells is unregulated in most states and often untreated. Public water supplies are regulated and often treated to control corrosion and other undesirable qualities (USEPA 2016).

The objective of this study was to develop a BRT model to predict pH conditions in three dimensions in the GLAC and to map those predictions at depths consistent with domestic- and public-supply use. Secondary objectives were to evaluate the accuracy of ML for these purposes and whether hydrogeologic insight can be gained from such models. ML has not often been used to predict groundwater pH. Notable exceptions include the use of BRT to predict and map pH in California’s Central Valley aquifer system (Rosecrans et al. 2017), the North Atlantic Coastal Plain (DeSimone et al. 2020), and the Mississippi Embayment (Kingsbury et al. 2020). Estimates of pH conditions throughout the GLAC do not currently exist. Predictions of pH in the GLAC can be used (1) to identify areas where geogenic contaminants may be mobilized into groundwater and where groundwater corrosivity issues may be of greatest concern; and (2) as a predictor variable in subsequent models describing the occurrence and distribution of pH-dependent constituents of concern in the GLAC. An important contribution of this paper is to demonstrate the utility of ML to predict pH conditions throughout the GLAC whose size and complexity precludes simulation with single domain, process-based modeling approaches.

As part of our predictor data, we incorporate and evaluate input terms derived from previously existing models. Characteristics of the subsurface such as (1)
the transmissivity of the shallow groundwater system, (2) the depth to the water table, (3) groundwater discharge as base flow, and (4) long-term average water content of the unsaturated zone were obtained from a continental-scale groundwater flow and transport model (Zell and Sanford 2020). Estimates of groundwater age and flowpath lengths in the GLAC were derived from a metamodelling approach (Fienen et al. 2015) that was used to bridge the gap between numerical simulation and statistical learning. Specifically, a ML model was trained on four groundwater flow metrics (fraction of young water, mean age of the young fraction of water, median age of the old fraction of water, and flowpath length) which were generated from 115 small-domain numerical models (Starn and Belitz 2018; Starn et al. 2020). Together, these 115 numerical models represent the range of hydrogeologic conditions across the GLAC. The ML model was then used to estimate these four groundwater metrics throughout the entire GLAC system (Starn et al. 2020). Information on subsurface characteristics such as the thickness and texture of glacial sediments was derived from the interpolation of subsurface observations from about 14 million water-well drillers’ records (Bayless et al. 2017; Haj et al. 2018). Finally, geochemical modeling is used to gain further
insight into conditions that favor either low or high pH conditions.

Methods

Groundwater Quality Data

Groundwater quality data were obtained from three sources: (1) the U.S. Geological Survey (USGS) National Water Information System Database (USGS 2019); (2) the USEPA Safe Drinking Water Information System (USEPA 2013); and (3) numerous agencies and organizations at the state, regional, and local level. Details on the compilation of these data are available in Erickson et al. (2019) and Wilson et al. (2019). pH measurements from 18,836 wells sampled between 1988 and 2018 were used for model development (Figure 1B). Most samples were collected between 1995 (25th percentile) and 2012 (75th percentile). Only measurements with 3.5 ≤ pH ≤ 10.5 were considered to avoid samples not representative of ambient groundwater-quality conditions (Erickson et al. 2019). For sites with multiple samples, only the most recent sample was retained. Seventy percent of the sites were monitoring (33%), public-supply (20%), or domestic-supply (17%) wells. The well type for 27% of the sites was unknown and the remaining 3% of sites were other miscellaneous water uses.

Data from 9655 wells with necessary water-quality data and a charge imbalance ≤10% were used to calculate mineral saturation index (SI) values for calcite and kaolinite using the computer program PHREEQC (Parkhurst and Appelo 2013) with the WATEQ4F thermodynamic database (Ball and Nordstrom 1991), where SI = log(ion activity product/solubility product). The weathering of carbonate minerals such as calcite will cause an increase in the calcite SI and a concomitant increase in pH. Similarly, the weathering of aluminosilicates such as potassium feldspar, which consumes acidity and also raises pH, increases the SI of its weathering product, kaolinite. Thus, the SIs for calcite and kaolinite can illustrate the evolution of pH along flowpaths in the GLAC. For sites where groundwater temperature was not reported, the median value from the host hydrogeologic terrane was assigned. All water-quality data used herein for predictive and geochemical modeling are available in Brown et al. (2020).

Predictor Variables

To map predictions of pH throughout the GLAC, predictor variables were considered if: (1) they related to processes and (or) conditions that govern pH and (2) estimates of their values were available across the landscape or throughout the aquifer system. Latitude and longitude were omitted as predictor variables because they are strictly locational without hydrologic context. Conversely, predictor variables extracted from numerically based (Zell and Sanford 2020; Starn et al. 2020) and empirical models (Bayless et al. 2017), such as groundwater age or sediment texture, were included as predictors because they characterize subsurface conditions and, thus, provide insight into the evolution of pH within the GLAC. In some instances, estimates of predictors were unavailable sporadically across the GLAC due to insufficient information. Examples of select predictor variable rasters are illustrated in Figure S1.

One hundred and forty-two variables that were compiled from existing geospatial data, process-based models, and empirical models were considered for potential inclusion in the model. This set of predictor variables includes information on geology, hydrology, climate, and other factors that might affect pH. If a predictor variable was determined to be highly correlated (Spearman’s rho ≥ 0.95) with another, one or the other was removed from consideration to avoid redundancy and the diminishing of variable importance values (Kuhn and Johnson 2016). Remaining predictor variables were evaluated using the “recipes” package of Kuhn and Wickham (2019) to determine if (1) they had a small number of unique values relative to the number of observations, or (2) the ratio of the frequency of the two most common values was large. All such “near zero variance” predictors were removed from further consideration. This process reduced the number of predictor variables used for model development to 97. Categorical variables with factor levels that occurred at <2% were pooled into an “other” category using the recipes package. Descriptions of all 97 variables and their sources are available in Table S2.

Several predictor variables require additional description. The variable “Lithology” is a generalized representation of bedrock geology underlying the GLAC and was compiled from multiple sources. The primary source of bedrock geology information was obtained from Haj et al. (2018) who presented an amalgamation of several “integrated geologic map databases for the United States” which were generalized into five bedrock geology types (Table S2). Where bedrock geology was categorized as “unmapped” or “water” by Haj et al. (2018), bedrock geology information was filled in from the state geologic map compilation of the conterminous United States (Horton et al. 2017) and generalized to the same five bedrock geology types.

Screen interval length has been shown to be an effective predictor of groundwater quality (Ransom et al. 2017; Erickson et al. 2018). Thirty-five percent of the sites with available pH measurements were missing screen length information. Missing screen length values were imputed using the “missForest” package (Stekhoven and Bühlmann 2011) by training a random forest model on related factors, including water use, well depth, land surface elevation, mean annual (1981–2010) precipitation, thickness of quaternary sediments, percentage of coarse material in quaternary sediments, percentage of well-drained soils, percent sand content of soils, and hydrogeologic terrane. The population of imputed screen lengths is comparable to reported screen lengths for each category of water use (Figure S2).

Only sites with well depth information were used in this study. Well depth measurements from land
BRT Model Development

A BRT model for predicting pH was developed from available pH measurements and data for 97 predictor variables using the Classification and Regression Training “caret” (Kuhn et al. 2019) and Generalized Boosted Regression Models “gbm” packages (Greenwell et al. 2019) in R. The data were randomly split into training (80%) and testing (20%) datasets. Tenfold cross validation (CV) on the training dataset was used in caret for model training and parameter tuning. CV systematically adjusts tuning parameters to control for potential model overfit and to maximize prediction accuracy. Four tuning parameters were adjusted: the number of trees, the number of splits (interaction depth), learning rate (which determines how much each tree contributes toward minimizing the loss function), and the minimum number of observations per node. From 1260 possible combinations of tuning parameters, CV identified the combination that results in a model with minimal root mean square error (RMSE). To allow for an independent assessment of prediction error, the optimal set of tuning parameters from 10-fold CV were used to fit a model to the training data in gbm—the resulting model is referred to as the “gbm” model.

Breiman et al. (1984) suggested that the optimal tuning parameters from CV may over fit to new (testing) data. To further minimize the potential for model overfit, all models with less complex tuning parameters and a RMSE within one standard error of that from the optimal CV tuned model were identified. Tuning parameters were considered less complex if interaction depth, learning rate, or the number of trees decreased and (or) the minimum number of observations per node increased. From an ensemble of 278 models with less complex tuning parameters, the model with minimal percent bias (described below) was selected (Figure S3) and its tuning parameters were used to refit a model to the training data in gbm. The resulting model is referred to herein as the “oneSE” model.

The recursive feature elimination (rfe) function in caret was used to explore reducing the number of predictor variables in a final model formulation to decrease model complexity. The rfe function implements backward selection of predictors based on variable importance rankings (described below). Beginning with the least important predictors, variables are sequentially removed and the resulting model RMSE recorded (Figure S4). A reduced number of predictor variables were then used, along with the tuning parameters from the gbm model, to refit a model to the training data. This model is referred to as the “rfe” model.

The gbm, oneSE and rfe models were used to predict pH values for both the training and testing datasets from which the coefficient of determination (R^2), RMSE, percent bias (%bias), and the Nash-Sutcliffe efficiency (NSE) index (Nash and Sutcliffe 1970) were calculated using the following equations:

\[
R^2_{(y, \hat{y})} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}
\]

\[
RMSE_{(y, \hat{y})} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}}
\]

\[
\%bias_{(y, \hat{y})} = 100 \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)}{\sum_{i=1}^{n} y_i}
\]

\[
NSE_{(y, \hat{y})} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}
\]

where \(i\) is an observation, \(y\) is the observed value, \(\hat{y}\) is the predicted value, and \(\bar{y}\) is the mean observed value. \(R^2\) is the square of the Pearson correlation between predicted and measured pH values and indicates the proportion of the variance in pH described by the predictor variables. RMSE is the standard deviation of the model prediction error. Smaller RMSE values indicate better model performance. \%bias measures the average tendency of predicted values to be larger or smaller than their measured values. The optimal \%bias is 0, with low-magnitude values indicating accurate model predictions. The NSE index describes the proportion of variance in observations that is predicted by the model (Nash and Sutcliffe 1970). NSE can take on values from \(-\infty\) to 1 with values closer to 0 indicating a more accurate model.

When developing ensemble tree models with a large number of predictor variables, it is useful to identify the influence of each predictor on the response because some predictors may be largely irrelevant (Hastie et al. 2017). Variable importance and partial dependence plots (PDPs) were generated to determine the relative importance and direction of influence of predictor variables, respectively. The “gbm” package calculates the relative importance of each variable in reducing the loss function—which for continuous response terms is based on variable selection counts during splitting and the reduction of squared error (over all trees) attributable to each predictor (Friedman 2001; Yan et al. 2019).

PDPs provide a visualization of the relationships discovered in complex ML models and thus aid in model interpretation (Elith et al. 2008; Greenwell 2017). PDPs show the average direction and nature of influence for a predictor variable in the model, while holding all other predictor variables constant (Friedman 2001; Greenwell 2017). Continuous predictor variables are
Model Training and Testing Results

| BRT Model | # of Predictor Variables | NT | ID | LR | MO | Training | Testing |
|-----------|--------------------------|-----|----|----|----|----------|---------|
|           |                          |     |    |    |    | $R^2$   | RMSE    | $R^2$ | RMSE | %bias | NSE |
| gbm       | 97                       | 4000| 18 | 0.012 | 10 | 0.86  | 0.27 | 0.60  | 0.44 | −0.05 | 0.60 |
| oneSE     | 97                       | 2500| 18 | 0.006 | 10 | 0.73  | 0.37 | 0.57  | 0.52 | −0.30 | 0.44 |
| rfe       | 50                       | 4000| 18 | 0.012 | 10 | 0.85  | 0.28 | 0.55  | 0.5  | −0.01 | 0.48 |

# number; %bias, percent bias; BRT, boosted regression tree; ID, interaction depth; LR, learning rate; MO, minimum number of observations per node; NSE, Nash-Sutcliffe efficiency index; NT, number of trees; $R^2$, coefficient of determination; RMSE, root mean square error.

#### Prediction and Uncertainty

The process for selecting a final model is described in Results section. Once selected, the final model was used to make predictions of pH at depths typical of domestic- and public-supply wells across the GLAC by applying the model to continuous raster grid representations of the predictor variables with 1 km$^2$ resolution. The raster grids of predicted pH were produced with the package “raster” (Hijmans 2019) and the “gbm” package predict function (Greenwell et al. 2019). Raster grids representing depths typical of domestic and public supply wells across the GLAC were generated using the methods described in Appendix S1. These rasters were attributed with values of groundwater age metrics and flow-path lengths (Starn and Belitz 2018; Starn et al. 2020) which vary with depth across the GLAC.

Uncertainty in predicted pH values was assessed using the bootstrapping approach described in Ransom et al. (2017). In short, model parameters were held constant and 199 models were trained on bootstrapped replicates (created with replacement) of the training data. Resampling with replacement results in changes to the variables selected for tree splitting, the split levels at internal tree nodes, and the predictions at terminal nodes for each of the 199 bootstrap samples. Each of the 199 models created with the bootstrapped samples was used to make predictions at the testing sites and at each of the 1 km$^2$ raster cells, resulting in a range of pH predictions. A randomly selected value from the training data residuals was subtracted from each bootstrap prediction to include the additional source of error associated with model fit. At each testing site or raster grid cell, the lower and upper prediction intervals (PIs) were calculated from the quantile of the distribution of model error (Schwarz et al. 2006) and PI widths were calculated by subtracting the lower PI from the upper.

#### Results

Model parameters and performance statistics are presented in Table 1 for the gbm, oneSE and rfe models. The gbm and rfe models performed similarly when applied to the training data as indicated by $R^2$ and RMSE goodness-of-fit statistics whereas the oneSE model had...
lower $R^2$ and higher RMSE. When generalized to new (testing) data, the gbm model described 60% of the variance in measured pH values ($R^2 = 0.60$). Predicted pH values from the gbm model at the testing sites plot more closely along the 1:1 line with measured pH values as compared to those from the oneSE and rfe models (Figure 2). Both the oneSE and rfe models described less of the variance in measured pH values at the testing sites, as indicated by lower NSE values (Table 1), and predictions at these sites were more oblique to the 1:1 line with measured pH values (Figure 2). The appearance of low bias for the rfe model is due to offsetting errors in prediction—overprediction at the low end and underprediction at the high end of pH (Figure 2).

Removal of predictor variables from 97 to about 44 had little effect on model RMSE (Figure S4). To be conservative, the gbm model was re-fit using predictor variables 1 through 50 to produce the rfe model. The result that predictions of pH from the rfe model at testing sites were biased and oblique to the 1:1 line with measured pH values (Figure 2C) indicates that in some instances the removal of predictor variables with relatively low importance scores can adversely affect model performance when predicting to new data. This finding contrasts with other studies in which the number of features in a model could be simplified without losing predictive accuracy, as long as the most important features remained in the model (Fienen et al. 2018; Starn and Belitz 2018).

Because the oneSE and rfe models produced predicted pH values at testing sites that were biased and more oblique to the 1:1 line with measured pH values, they were discarded from further consideration and the gbm model was selected for the purpose of predicting pH in groundwater throughout the GLAC. The following discussion examines the 12 most influential predictor variables in the gbm model as indicated by their variable importance values (Figure 3). ML methods fundamentally rely on identifying patterns in complicated datasets without regard to causative effects. However, these 12 variables can broadly be categorized as representing (1) recharge characteristics such as precipitation chemistry, (2) hydrogeologic terranes and hydrologic position on the land surface, and (3) residence time and hydrologic position within the aquifer system, all of which may affect the pH of groundwater. The PDPs for the two most important variables in the model are shown in Figure 4. PDPs for all 12 of the most influential predictors are shown in Figure S5. Descriptions and importance ranks for all predictor variables are listed in Table S2.

Recharge Characteristics

The chemical composition of precipitation can affect groundwater quality (Neal and Kirchner 2000; Löfgren et al. 2011; Boumans et al. 2014) and the annual mean concentration of Na+ (Na+_ppt) and H+ (pH_ppt) in
precipitation for the period 1985–1995 were the first and third most important predictors of pH, respectively (Figure 3). Their PDPs indicate increasingly acidic groundwater with increasing Na\(^+\) and H\(^+\) in precipitation (Figures 4 and S5). The greatest concentration of Na\(^+\) in precipitation occurs in eastern and Western coastal areas of the conterminous United States, where the amount of Na\(^+\) in precipitation is influenced by oceanic effects (Reid et al. 1981; Neal and Kirchner 2000). In coastal areas with poorly buffered, sandy soils such as occur in the northeastern United States, Na\(^+\) can contribute to acidification of groundwater as it takes part in cation exchange reactions in soils and aquifer materials (Wiklander 1975; Norton et al. 2014).

The coastal effect is reflected in the rapid increase in predicted pH with increasing elevation (LSelev; Figure S5). Mean annual potential evapotranspiration (PET; Figures 3 and S5) varies north to south across the GLAC. Predictions of pH decrease at PET values approximately 25 in/year, which may help distinguish areas to the north with higher pH values from areas to the south; however, PET can also contribute to groundwater acidification by concentrating ions such as Na\(^+\) (Humphries et al. 2011). Predicted pH increases with decreasing concentration of H\(^+\) in precipitation (“pH_ppt”; Figure S5). Lower values of H\(^+\) in precipitation occur primarily in continental portions of the GLAC (Figure S1) indicating that the concentration of H\(^+\) in precipitation serves to refine predictions of groundwater pH in noncoastal areas.

Hydrogeologic Terranes and Hydrologic Position on the Land Surface

Hydrogeologic terrane was the second most important predictor of pH (Figure 3). As previously described, terranes represent regions with quaternary sediments that have a common age and mode of deposition and are of similar texture and thickness (Yager et al. 2018) (Table S1). The mineral composition of sediments in these terranes is governed by the bedrock type over which glacial ice lobes traveled (Rodgers 1989; Grasby et al. 2010) or from which residuum formed in areas of terrane 1G (Yager et al. 2018). Together these factors control the stratigraphic complexity and composition of sediments in each terrane, which in turn affects the movement and quality of groundwater.

Figure 4B illustrates that the terranes can broadly be grouped as having predicted pH either greater than or less than 7.3. Terranes with predicted pH <7.3 contain quaternary sediments that are generally less stratigraphically complex (thinner) and lower in carbonate content, and, thus, contain groundwater that is less geochemically evolved as compared to terranes with predicted pH >7.3 (Figure S6). The influence of sediment thickness and composition on predicted pH is illustrated by terranes 1A and 4A.

Terrane 1A is characterized by the thinnest sediments on average in the GLAC whereas terrane 4A is characterized by thick sediments with the longest flowpaths on average (Figure S7; Table S1). Sediments in terrane 4A are rich in quartz and poor in calcareous material (Coston et al. 1995; Brown et al. 2000; Bau et al. 2004) (Figure S7). Sediments in terrane 1A are derived from a variety of bedrock types with varying carbonate content. In the eastern part of terrane 1A, sediments are largely derived from igneous and metamorphic bedrock (Rodgers 1989) (Figure S8) that are low in calcareous material. In the Western part of terrane 1A, sediments are derived from sedimentary bedrock with a higher carbonate content, albeit classified as noncarbonate sedimentary bedrock by Haj et al. (2018) (Figure S8). Where sediments in terrane 1A are rich in carbonate, groundwater becomes saturated with respect to calcite increasing its buffering capacity—despite having water-sediment contact times limited by short flowpaths (Figure 5). In terrane 4A, and where sediments in terranes 1A are poor in calcareous material, groundwater remains undersaturated with respect to calcite and, thus, poorly buffered—regardless of flowpath length (Figure 5). The positive correlation between kaolinite and calcite SI where groundwater is poorly buffered reflects the gradual release of bicarbonate as aluminosilicate minerals weather—a process that is slower than the weathering of carbonate minerals and that generally does not result in saturation with respect to calcite (Figure 5). Where groundwater flowpaths are long, and the carbonate content of aquifer materials is high (e.g., terranes 1F and 2E), the continued dissolution of carbonate minerals through processes such as cation exchange result in the highest pH conditions, on average, across the GLAC (Figure S7). Thus, while lithologic complexity partly determines water-sediment contact times, it alone is insufficient for controlling pH. Rather, the carbonate content of aquifer materials more strongly controls pH and when coupled with long water-sediment contact times along extended flowpaths results in the most elevated pH.

Figure 5. Groundwater in terranes 1A and 4A remains undersaturated with respect to calcite regardless of flowpath length in areas where sediments are low in calcareous material. In areas of terrane 1A where sediments are carbonate rich, groundwater is saturated with respect to calcite despite having relatively short flowpaths.
Hydrologic position on the landscape has been shown to be an important predictor of groundwater chemistry (Landon et al. 2011; Ayotte et al. 2016). DSD₈ represents the distance from stream to divide and is an indicator of hydrologic position on the landscape within an eighth order stream watershed (Belitz et al. 2019). In the GLAC, DSD₈ serves to partition some hydrogeologic terranes to refine predictions of pH. DSD₈ has relatively constant influence on predicted pH at values ≤200,000 km and ≥300,000 km, with a systematic decrease in between (Figure S5). Values of DSD₈ between 200,000 and 300,000 km occur in terranes 1A, 4A, and 1F. These terranes also have a wide range and (or) a spatially uneven distribution of pH values (Figures 1B and S7). For example, sites in Western parts of terrane 1A tend to have higher pH values as compared to sites in eastern parts of terrane 1A (Figure 1B). Values of DSD₈ between 200,000 and 300,000 km effectively separate areas of lower and higher pH measurements in terranes 1A, 4A, and 1F thereby refining predictions of pH within these terranes (Figures S1 and S9).

Residence Time and Hydrologic Position Within the Aquifer System

Predicted values of pH generally increase with increasing relative well depth, screen length, flowpath length, groundwater age (GWage, young fraction) and soil clay content, and decreasing recharge amounts (Figure S5). Collectively these variables describe residence time and hydrologic position within the aquifer system. Higher pH with increasing well depth, flowpath length, and groundwater age indicates, as expected, that pH generally increases with groundwater age along flowpaths and that these conditions occur with increasing depth in the aquifer system. Low recharge and poorly conductive soils as indicated by higher clay content in the C horizon inhibit subsurface flow providing greater water-sediment interaction time and, thus, more geochemically evolved water quality with higher pH. The effect of longer screen lengths likely reflects the fact that wells finished in the shallow part of the aquifer system where groundwater is generally younger, and pH values lower, often have shorter screen lengths as compared to wells located deeper in the aquifer where groundwater is older and pH values higher.

Predictions and Implications

The gbm model was used to predict pH conditions across the GLAC at depths that are typical for domestic- and public-supply wells (Figure 6). In the GLAC, the range of depths for domestic- and public-supply wells often overlap (Erickson et al. 2019) (Figure S10). This reflects the fact that all types of public-supply wells were considered in this analysis including municipal systems that serve large populations year-round as well as nontransient and transient systems that serve smaller populations, for shorter periods of time (e.g., schools, gas stations, or campgrounds with their own water systems). These later categories of public-supply wells are likely to have well construction characteristics comparable to domestic wells. Where quaternary sediments are thick (e.g., terrane 4A) public-supply wells are usually located deeper in the aquifer system as compared to domestic wells (Figure S10).

Predicted values of pH vary spatially across the GLAC, but the differences between domestic- and public-supply wells are typically small (Figure 6) reflecting the overlap in depths of domestic- and public-supply wells common across much of the GLAC. Where aquifer sediments are thin (e.g., terrane 1A), the difference between pH predictions at depths typical of domestic and public supplies is minimal (Figure 7). The difference in average predicted pH between these depth intervals for terrane 1A is 0.03 standard pH units. However, in the presence of abundant carbonate minerals, groundwater rapidly becomes saturated with respect to calcite at near neutral pH (∼7.0) (Figure 5) and if the dissolution of carbonate minerals continues in response to processes such as cation exchange, pH can increase to concentrations >7.0 (Figure 7). Where aquifer sediments are thick and carbonate-poor (e.g., terrane 4A), pH is predicted to be greater in public supply than in domestic wells reflecting the gradual increase in pH associated with the weathering of aluminosilicate minerals at the greater depths of public-supply wells (Figure 7). The difference in average predicted pH between these domestic- and public-supply well depth intervals for terrane 4A is 0.13 standard pH units.

In previous studies of drinking-water quality in the GLAC, Mn was the constituent found to occur most frequently at concentrations greater than its health-based benchmark, occurring in 25.6% of the assessed area that serves an estimated 4.1 million people (Stackelberg 2017; Erickson et al. 2019). High concentrations of Mn were found to occur more commonly under reducing conditions across the GLAC, except in terrane 4A where elevated Mn was driven by pH <7.0. A recent study of elevated Mn concentrations in groundwater across the United States found that in shallow or poorly buffered aquifers with pH <6, the oxidation of ammonium in agricultural areas can lower pH and enhance Mn mobility (McMahon et al. 2019).

In this study, Mn concentrations in poorly buffered terranes (1A, 1G, and 4A) are significantly higher (Wilcoxon rank sum test, p = 0.004) in samples with pH <6 as compared to those with pH ≥6. In these areas NO₃ concentrations are inversely related to pH (Spearman’s rho = −0.45; p < 0.001). These findings corroborate the findings of (McMahon et al. 2019) and indicate that poorly buffered areas of the GLAC with pH <6 may be susceptible to elevated Mn concentrations regardless of redox condition. At depths typical of domestic and public supplies, predicted values of pH <6 occur primarily in the southeastern portion of terrane 1A where poorly buffered sediments support low pH conditions (Figure 6).
Figure 6. Predicted values of pH at depths typical of (A) domestic wells and (B) public-supply wells. [Corrections added after online publication, 11 April 2021: The first number in the Figure legend was corrected to $<6.5$ from $<−6.5$.]

Figure 7. Predicted values of pH are similar between depths typical of domestic and public supplies where aquifer sediments are thin (terrane 1A) but vary by depth where aquifer sediments are thick (terrane 4A).

Arsenic was the second most frequently occurring constituent at concentrations greater than its health-based benchmark in previous studies of drinking-water quality in the GLAC (Stackelberg 2017, Erickson et al. 2019). Arsenic occurred in 10.4% of the assessed area that serves an estimated 2.9 million people (Erickson et al. 2019). Sorption is the predominant mechanism affecting the mobility of As in many groundwater systems, and pH strongly influences the adsorption/desorption of As to aquifer solids (Manning and Goldberg 1997; Welch et al. 2000; Stollenwerk 2003). At pH $>7.5$, As is more likely to be present at high concentrations. For the depth zone used for domestic supply, the gbm model predicts that 20% of the GLAC has pH $>7.5$. For the depth zone used for public supply, the percentage is 23%. Predicted values of pH $>7.5$ predominate in terranes 1F and 2E (Figure 6) where Erickson et al. (2019) reported the highest median As concentrations, corroborating the importance of pH in mobilizing As into groundwater. The pH of water is one of several factors that affect corrosivity and the presence of trace metals in drinking supplies (Singley 1981). In particular, source water with pH $<6.5$ is more likely to corrode Pb, if present, from distribution systems, plumbing systems, and (or) fixtures. Areas of the GLAC with pH $<6.5$ occur primarily in terrane 1A (Figure 6).

Sensitivity of the pH Model to Groundwater Flow Variables

The gbm model includes four predictor variables related to groundwater flow: fraction of young water;
mean age of the young fraction of water; median age of the old fraction of water; and flowpath length (Table S2). The effect of the four variables on predictions of pH was determined by developing a new model using the same parameters as the original model but which excluded these variables (altered model). In turn, the predictions from the altered model were subtracted from the original predictions (full model). By excluding these four variables, the altered model essentially treated all areas of the GLAC as having the same groundwater age and flowpath lengths. By developing a new model, one can determine what information these four variables provide that the remaining 93 variables cannot explain. In recharge areas where groundwater is young and flowpaths short, the altered model is expected to overestimate the evolution of pH conditions and predict higher values of pH than the full model (negative residuals) and in discharge areas where groundwater is older and flowpaths longer the altered model is expected to underestimate the evolution of pH conditions and predict lower values of pH (positive residuals).

Terranes 1A and 4A serve as useful end members, illustrating the effect of these four groundwater flow variables. The Tug Hill Plateau, which is in terrane 1A, is a bedrock upland area encompassing 2500 km² in north-central New York and which shares a boundary along the southwest extent of the Adirondack Mountains (Figure 8A). The plateau is mantled with a thin layer of till and glacial-drift deposits which are generally only 6–15 m thick (Miller et al. 2007), consistent with the thin nature of glacial deposits in terrane 1A (Table S1). Most recharge to the Tug Hill glacial-drift aquifer is from direct precipitation or from losing streams that flow westward across the plateau (Miller et al. 2007). Groundwater ages in the Tug Hill glacial-drift aquifer have been estimated by metamodeling to generally be <20 years (Starn et al. 2020).

Cape Cod, which is in terrane 4A, encompasses 1100 km² in southeastern Massachusetts (Figure 8B). Glacially derived sediments on Cape Cod range in thickness from approximately 30 to >300 m (LeBlanc et al. 1986), consistent with the thick nature of glacial deposits in this terrane. Freshwater lenses in these sediments constitutes the Cape Cod aquifer system and are generally 30 to about 90 m thick. Groundwater ages in this aquifer system have been estimated by numerical modeling methods to range from <1 year in areas adjacent to discharge locations to >100 years near groundwater divides (Walter et al. 2004; Starn et al. 2020).

In both areas, as expected, residuals of pH predictions are mostly negative where groundwater is young and flowpaths short. These conditions occur across much of the Tug Hill Plateau whereas on Cape Cod they are concentrated in areas where the water table has greatest elevation (Masterson 2004) (Figure 8). On Cape Cod, groundwater flows radially from the tops of these mounds and increases in age as it flows toward the coastline and other points of discharge (LeBlanc et al. 1986; Masterson 2004). Conversely, residuals of pH predictions tend to be positive in lower-lying, downgradient areas which correspond to areas of discharge and older groundwater. These areas occur along the northern boundary of the Tug Hill Plateau and in the low-lying areas between the plateau and the southern extent of the Adirondack Mountains (Figure 8A). On Cape Cod, these conditions occur in areas of lower water table elevations which represent downgradient portions of the groundwater flow system (Figure 8B). Thus, the inclusion of these four groundwater flow variables—in a model with 97 variables overall—has a small, but systematic effect on predictions of pH which is consistent with our understanding of groundwater flow systems and the evolution of pH conditions along flowpaths.

Uncertainty and Limitations

Predictions generated from ML regression methods will inherently have reduced variance as compared to observed values (Zhang and Lu 2012). This can be seen in Figure 2 where ML predictions are not as high or as low as the highest and lowest observed values of pH. Uncertainty in a predicted concentration is expressed in terms of the PI defined by the 2.5th and 97.5th percentiles (Figure 9A). As expected, most—but not all—of measured pH values fall within the 95th PI. The size of the PI at each site is represented as the ratio of the upper PI to the predicted pH value (Figure 9B). Outliers represent sites where one or more predictor variables were relatively extreme in value when compared with values from the rest of the sites. These sites, therefore, exhibited greater uncertainty in their predicted pH values than did the others.

PIs were mapped at the depths typical of domestic and public supply (Figure S11). PIs tend to be smaller in areas where GLAC sediments contain abundant calcareous material reflecting the strong control calcareous material has on pH. Where aquifer materials are carbonate-poor, predictions of pH are more uncertain. The difference in PIs between depths of domestic and public supplies was small due to the overlap of these depths across much of the GLAC.

The amount of variability in measured pH conditions explained by the gbm model when applied to new (testing) data was substantial. However, the fact that 40% of the measured variability in pH remains unexplained (Table 1) indicates that factors that are important to controlling pH are not completely represented by the predictor variables for which data are available across the entire GLAC. In addition, the sub-continental scale at which this model was developed could result in erroneous predictions if the model is applied at more local scales because it may not adequately reflect local variations in important predictor variables that affect pH. Local factors that might affect pH conditions require consideration before results from this model can be interpreted for any specific location. Also, the application of the model to sites with fundamental differences in hydrogeologic conditions or to sites whose values of predictor variables differ from those used for
model training will result in greater uncertainty in model predictions. These limitations, and the abovementioned reduced variance and uncertainty in predictions of pH, indicate that the model presented herein should not be used to characterize absolute pH conditions at specific locations. However, the model can be used to identify areas where pH conditions could affect the mobility of specific contaminants and may therefore merit further investigation.

**Conclusions**

The glacial aquifer system provides more water—about 2.6 billion gallons per day—for domestic and public supplies than any other aquifer in the United States. About 30 million people are reliant on the glacial aquifer for their water supplies. It is important to understand the pH of this resource because pH-dependent geochemical processes can mobilize both geogenic and anthropogenic contaminants.

BRT model was developed to provide predictions of pH in three dimensions throughout the aquifer system. The BRT model incorporates predictor variables extracted from numerically based and geostatistical models which characterize subsurface conditions such as groundwater age, flowpath lengths, and sediment textures, to constrain pH predictions and to provide insight into the evolution of pH conditions within the aquifer system. The model explains 60% of the variability observed in pH measurements when generalized to new (testing) data.
Results from this study indicate that, as expected, pH conditions vary spatially across the GLAC. The stratigraphic complexity and mineral composition of aquifer sediments across the GLAC affects the movement and quality of groundwater. Although lithologic complexity partly controls water-sediment contact times, it alone is insufficient for controlling pH. Rather, the carbonate content of aquifer materials more strongly controls pH, and when coupled with long water-sediment contact times along extended flowpaths, results in the most elevated pH conditions. Conversely, thin, carbonate-poor sediments support more acidic pH conditions. The pH of groundwater will generally evolve and increase along flowpaths; however, pH conditions were predicted to be similar between depths typical of domestic and public supplies in many areas due to the overlap in depths of these two well types across much of the GLAC.

Many trace elements, including Mn, can be mobilized into groundwater under acidic conditions. Acidic groundwater also contributes to issues with corrosivity. The most acidic groundwater was predicted to occur primarily in the northeastern United States. Other trace elements, such as As, can be mobilized into groundwater under alkaline conditions. Predictions of pH > 7.5 occur across 20% and 23% of the GLAC at depths typical of domestic or public supplies, respectively. These areas occur primarily in the upper Midwest and correspond to areas known to have elevated As concentrations.

Predictions of pH were shown to be sensitive to predictor variables that represent flow characteristics of the GLAC such as groundwater age and flowpath lengths. The inclusion of these variables resulted in small but systematic changes in pH predictions consistent with our understanding of the importance of hydrologic position within an aquifer system and geochemical evolution of groundwater quality along flowpaths. Where aquifer sediments are thin, groundwater is young and flowpaths are short, predictions of pH with these flow variables were consistently lower than predictions without these flow variables. Conversely, where aquifer sediments are thick, groundwater is generally older and flowpaths longer, predictions with these flow variables were consistently higher than predictions without these flow variables. These findings illustrate the ability of ML to identify patterns between pH, groundwater age and the length of flowpaths and to use those patterns to refine predictions of pH across the GLAC.

These results demonstrate the utility of empirically based ML methods such as BRT for predicting water-quality conditions in an aquifer system too large and stratigraphically complex to support single-domain simulation of groundwater flow and reactive solute transport. Just as this model benefitted from the inclusion of predictor variables derived from previously existing numerically based and geostatistical models that describe aspects of the subsurface, pH predictions generated from this model may be useful for inclusion as predictors of water-quality constituents such as Mn or As in subsequent ML models. The BRT model can also be used to identify areas where pH conditions may be such as to mobilize geogenic contaminants into groundwater or support corrosivity. Such predictions can be used for the design of future monitoring programs or for setting priorities among different areas for further research.

Acknowledgments

The authors wish to thank Leslie DeSimone and Michael Fienen (both with the U.S. Geological Survey) for colleague reviews, and Walter Niccoli and Ty Ferre for journal reviews. Their comments substantially improved the content and presentation of this article. Any use of trade, firm, or product names is for descriptive purposes.
only and does not imply endorsement by the U.S. government. Open access funding enabled and organized by Projekt DEAL.

Authors’ Note
The authors do not have any conflicts of interest or financial disclosures to report. Funding for this work was provided by the U.S. Geological Survey’s National Water Quality Assessment Project, a component of the National Water Quality Program.

Supporting Information
Additional supporting information may be found online in the Supporting Information section at the end of the article. Supporting Information is generally not peer reviewed.

Appendix S1. Methods for determining well characteristics in the glacial aquifer system

Figure S1. Example rasters showing (A) Na+ in precipitation, (B) H+ in precipitation, and (C) the distance from stream to divide for 8th order stream networks (DSD8). A description of the variables and their sources is provided in Table S2.

Figure S2. Distribution of measured and imputed screen lengths by well types (values above boxplots represent the number of screen length measurements; 1 foot = 0.3048 m).

Figure S3. Complexity of model parameters and root mean square error for the ensemble of one standard error models. The model with the least percent bias is highlighted in red.

Figure S4. Recursive feature elimination plot showing the model RMSE as each predictor variable is iteratively removed from the model.

Figure S5. Partial dependence plots for the 12 most important predictor variables in the gbm model. Descriptions of the variables are available in Table S2. Red hash marks represent deciles.

Figure S6. The thickness of Quaternary sediments, the carbonate content of the soil C-horizon and the molar ratio of Na:Cl are all significantly ($p < 0.05$) lower in hydrogeologic terranes where the partial dependence of pH is $< 7.3$ as compared to those terranes where the partial dependence of pH is $> 7.3$.

Figure S7. Distributions of (A) pH, (B) sediment thickness, and (C) carbonate content of the soil C-horizon by hydrogeologic terrane. Distributions of (D) flow path length, (E) mean age of the young fraction of water and (F) the Na:Cl molar ratio by hydrogeologic terrane.

Figure S8. Generalized bedrock lithology modified from Haj et al. (2018).

Figure S9. Values of the distance from stream to divide (DSD) for 8th order streams between 200,000 and 300,000 km serve to distinguish areas of higher and lower pH in terranes 1A, 4A, and 1F.

Figure S10. Distribution of well depths by hydrogeologic terrane and well type (1 foot = 0.3048 m).

Figure S11. 95th prediction intervals for depths typical of (A) domestic and (B) public-supply wells.

Table S1. Generalized characteristics of quaternary sediments in the map units within hydrogeologic terranes in the glaciated area of the conterminous United States (Figure 1A), simplified and summarized from Yager et al. (2018, pp. 30–66). Reprinted from Erickson et al. (2019), with permission.

Table S2. Description and variable importance rank for predictor variables

References
Appelo, C.A.J., and D. Postma. 2005. Geochemistry, Groundwater and Pollution, 2nd ed. Boca Raton, Florida: CRC Press.

Ayotte, J.D., B.T. Nolan, and J.M. Gronberg. 2016. Predicting arsenic in drinking water wells of the Central Valley, California. Environmental Science & Technology 50, no. 14: 7555–7563. https://doi.org/10.1021/acs.est.6b01914

Ball, J.W., and D.K. Nordstrom. 1991. User’s manual for WATEQ4F, with revised thermodynamic database and test cases for calculating speciation of major, trace, and redox elements in natural waters. U.S. Geological Survey Open-File Report 91-183. https://doi.org/10.3133/ofr911183

Baris, D., R. Waddell, L.E. Beane Freeman, M. Schwenn, J.S. Colt, J.D. Ayotte, M.H. Ward, J.R. Nuckols, A. Schned, B. Jackson, C. Clerkin, N. Rothman, L.E. Moore, A. Taylor, G. Robinson, G.M. Hosain, K.R. Armenti, R. McCoy, C. Samanic, R.N. Hoover, J.J.F. Fraumeni, A. Johnson, M.R. Karagas, and D.T. Silverman. 2016. Elevated bladder cancer in northern New England: The role of drinking water and arsenic. Journal of the National Cancer Institute 108, no. 9: 1–9. https://doi.org/10.1093/jnci/djw099

Bau, M., B. Alexander, J.T. Chesley, P. Dulski, and S.L. Brantley. 2004. Mineral dissolution in the Cape Cod aquifer, Massachusetts, USA: I. Reaction stoichiometry and impact of accessory feldspar and glauconite on stromium isotopes, solute concentrations, and REY distribution. Geochimica et Cosmochimica Acta 68, no. 5: 1199–1216. https://doi.org/10.1016/j.gca.2003.08.015

Bayless, E.R., L.D. Arihood, H.W. Reeves, B.J.S. Sperl, S.L. Qi, V.E. Stipe, and A.R. Bunch. 2017. Maps and grids of hydrogeologic information created from standardized water-well drillers’ records of the glaciated United States. U.S. Geological Survey Scientific Investigations Report 2015-5105. https://doi.org/10.3133/sir20155105

Belitz, K., R.B. Moore, T.L. Arnold, J.B. Sharpe, and J.J. Starn. 2019. Multidonor hydrologic position in the conterminous United States: A set of metrics in support of groundwater mapping at the regional and national scales. Water Resources Research 55, no. 12: 11,188–11,207. https://doi.org/10.1029/2019WR025908

Belitz, K., B.C. Jurgens, and T.D. Johnson. 2016. Potential corrosivity of untreated groundwater in the United States. U.S. Geological Survey Scientific Investigations Report 2016-5092. https://doi.org/10.3133/sir20165092

Boumans, L.J.M., E.J.W. Wattel-Koekkoek, and E. van der Swaluw. 2014. Changes in rainwater and groundwater quality as a result of atmospheric emission reductions—Acidification and eutrophication, 1989–2010. National Institute for Public Health and the Environment, RIVM Report 680720007/2014:51

Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. 1984. Classification and regression trees. Belmont, California: Wadsworth International Group.
Brown, C.J., K. Belitz, M.L. Erickson, S.M. Elliott, L.J. Kauffman, K.M. Ransom, J.E. Reddy, and P.E. Stackelberg. 2020. Data for machine learning predictions of pH in the glacial aquifer system, Northern USA. U.S. Geological Survey Data Release. https://doi.org/10.5066/P9RF0R6E

Brown, C.J., J.R. Barlow, C.A. Cravotta III., and B.D. Lindsey. 2019. Factors affecting the occurrence of lead and manganese in untreated drinking water from Atlantic and Gulf Coastal Plain aquifers, eastern United States—Dissolved oxygen and pH framework for evaluating risk of elevated concentrations. Applied Geochemistry 101: 88–102. https://doi.org/10.1016/j.apgeochem.2018.10.017

Brown, C.J., J. Rakovan, and M.A.A. Schoonen. 2000. Heavy minerals and sedimentary organic matter in Pleistocene and Cretaceous sediments on Long Island, New York, with emphasis on pyrite and marcasite in the Magothy aquifer. U.S. Geological Survey Water-Resources Investigations Report 99-4216. https://doi.org/10.3133/wri994216

Costen, J.A., C.C. Fuller, and J.A. Davis. 1995. Pb2+ and Zn2+-adsorption by a natural aluminum- and iron-bearing surface coating on an aquifer sand. Geochimica et Cosmochimica Acta 59, no. 17: 3535–3547. https://doi.org/10.1016/0016-7037(95)00231-N

Elith, J., J. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. Journal of Animal Ecology 77: 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Erickson, M.L., R.M. Yager, L.J. Kauffman, and J.T. Wilson. 2019. Drinking water quality in the glacial aquifer system, northern USA. Science of The Total Environment 694: 133735. https://doi.org/10.1016/j.scitotenv.2019.133735

Erickson, M.L., S.M. Elliott, C.A. Christenson, and A.L. Krall. 2018. Predicting geogenic arsenic in drinking water wells in glacial aquifers, north-central USA: Accounting for depth-dependent factors. Water Resources Research 54, no. 12: 10172–10187. https://doi.org/10.1002/2018WR023106

Fienen, M.N., B.T. Nolan, L.J. Kauffman, and D.T. Feinstein. 2018. Metamodeling for groundwater age forecasting in the Lake Michigan Basin. Water Resources Research 54, no. 7: 4750–4766. https://doi.org/10.1002/2017WR023287

Fienen, M.N., B.T. Nolan, D.T. Feinstein, and J.J. Starn. 2015. Metamodels to bridge the gap between modeling and decision support. Groundwater 53, no. 4: 511–512. https://doi.org/10.1111/gwat.12339

Friedman, J.H. 2001. Greedy function approximation: A gradient boosting machine. Annals of Statistics 29, no. 5: 1189–1232. https://doi.org/10.1214/aoas1013203451

Grasby, S., J. Osborn, Z. Chen, and P. Wozniak. 2010. Influence of till provenance on regional groundwater geochemistry. Chemical Geology 273: 225–237. https://doi.org/10.1016/j.chemgeo.2010.02.024

Greenwell, B., B. Boehmke, J. Cunningham, and GBM Developers. 2019. gbm: Generalized boosted regression models. R package version 2.1.5. https://CRAN.R-project.org/package=gbm (accessed April 10, 2020)

Greenwell, B.M. 2017. pdp: An R package for constructing partial dependence plots. The R Journal 9, no. 1: 421–436. https://doi.org/10.32614/RJ-2017-016

Gunarathna, M.H.J.P., M.K.N. Kumari, and K.G.S. Nirmanee. 2016. Evaluation of interpolation methods for mapping pH of groundwater. International Journal of Latest Technology in Engineering, Management & Applied Science V, no. III: 1–5.

Haj, A.E., D.R. Soller, J.E. Reddy, L.J. Kauffman, R.M. Yager, and C.A. Buchwald. 2018. Hydrogeologic framework for characterization and occurrence of confined and unconfined aquifers in quaternary sediments in the glaciated conterminous United States—A digital map compilation and database. U.S. Geological Survey Data Series 1090. https://doi.org/10.3133/ds1090

Harischandra, D.S., S. Ghaisas, G. Zenisky, H. Jin, A. Kanthasamy, V. Anantharam, and A.G. Kanthasamy. 2019. Manganese-induced neurotoxicity: New insights into the triad of protein misfolding, mitochondrial impairment, and neuroinflammation. Frontiers in Neuroscience 13, no. 654. https://doi.org/10.3389/fnins.2019.00654

Hastie, T., R. Tibshirani, and J. Friedman. 2017. The Elements of Statistical Learning—Data Mining, Inference, and Prediction, 2nd ed. New York: Springer. https://doi.org/10.1007/978-0-387-84896-2

Hijmans, R.J. 2019. raster: Geographic data analysis and modeling. R package version 2.9-23. https://cran.r-project.org/web/packages/raster/index.html (accessed May 20, 2020)

Horton, J.D., C.A. San Juan, and D.B. Stoeser. 2017. The State Geologic Map Compilation (SGMC) geodatabase of the conterminous United States. U.S. Geological Survey Data Series 1052. https://doi.org/10.3133/ds1052

Humphries, M.S., A. Kindness, W.N. Ellery, and J.C. Hughes. 2011. Water chemistry and effect of evapotranspiration on chemical sedimentation on the Mkuse River floodplain, South Africa. Journal of Arid Environments 75: 555–565. https://doi.org/10.1016/j.jaridenv.2011.01.013

Kent, D.B., R.H. Abrams, J.A. Davis, J.A. Coston, and D.R. LeBlanc. 2000. Modeling the influence of variable pH on the transport of zinc in a contaminated aquifer using semiempirical surface complexation models. Water Resources Research 36, no. 12: 3411–3425. https://doi.org/10.1029/2000wr900244

Kingsbury, J.A., K.J. Knierim, and C.J. Haugh. 2020. Prediction grids of pH for the Mississippi River Valley Alluvial and Claiborne aquifers. U.S. Geological Survey data release. https://doi.org/10.5066/P9CX7LN

Knoll, L., L. Breuer, and M. Bach. 2019. Large scale prediction of groundwater nitrate concentrations from spatial data using machine learning. Science of The Total Environment 668: 1317–1327. https://doi.org/10.1016/j.scitotenv.2019.03.045

Kuhn, M., and H. Wickham. 2019. recipes: Preprocessing tools to create design matrices. R package version 0.1.6. https://CRAN.R-project.org/package=recipes (accessed May 8, 2020)

Kuhn, M., J. Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, the R Core Team, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y. Tang, C. Candan, and T. Hunt. 2019. caret: Classification and regression training. R package version 6.0-84. https://CRAN.R-project.org/package=caret (accessed January 15, 2020)

Kuhn, M., and K. Johnson. 2016. Applied Predictive Modeling. New York: Springer. https://doi.org/10.1007/978-1-4614-6849-3

Landon, M.K., C.T. Green, K. Belitz, M.J. Singleton, and B.K. Esser. 2011. Relations of hydrogeologic factors, groundwater reduction-oxidation conditions, and temporal and spatial distributions of nitrate, central-eastside San Joaquin Valley, California, USA. Hydrogeology Journal 19, no. 6: 1203–1224. https://doi.org/10.1007/s10040-011-0750-1

LeBlanc, D.R., J.H. Gusswa, M.H. Frimpter, and C.J. Londoquist. 1986. Ground-water resources of Cape Cod, Massachusetts. U.S. Geological Survey Hydrologic Investigations Atlas HA-692 (4 Sheets).

Löfgren, S., M. Aastrup, L. Brinmark, H. Hultberg, L. Lewin-Pihlbad, L. Lundin, G.P. Karlsson, and B. Thunholm. 2011. Recovery of soil water, groundwater, and streamwater from acidification at the Swedish integrated monitoring
Norton, S.A., J. Kopáček, and I.J. Fernandez. 2014. Acid
Nolan, B.T., M.N. Fienen, and D.L. Lorenz. 2015. A statistical
Neal, C., and J.W. Kirchner. 2000. Sodium and chloride
Nolan, B.T., J.M. Gronberg, C.C. Faunt, S.M. Eberts, and
Manning, B.A., and S. Goldberg. 1997. Adsorption and stability
NGWA.org P.E. Stackelberg et al. Groundwater 59, no. 3: 352–368 367
Nas, B., and A. Berktay. 2010. Groundwater quality mapping in
Molnar, C. 2020. Interpretable machine learning—A guide for
Mayer, J.E., and R.H. Goldman. 2016. Arsenic and skin
cancer in the USA: The current evidence regarding arsenic-contaminated drinking water. International Journal of Dermatology 55: e858–e891. https://doi.org/10.1111/ijd.13318
McMahon, P.B., K. Belitz, J.E. Reddy, and T.D. Johnson. 2019. Elevated manganese concentrations in United States groundwater, role of land surface–soil–aquifer connections. Environmental Science & Technology 53, no. 1: 29–38. https://doi.org/10.1021/acs.est.8b04055
Miller, T.S., E.F. Bugliosi, K.K. Hettcher-Aguila, and D.A. Eckhardt. 2007. Hydrogeology of two areas of the Tug Hill glacial-drift aquifer, Oswego County, New York. U.S. Geological Survey Scientific Investigations Report 2007–5169. https://doi.org/10.3133/si20075169
Molnar, C. 2020. Interpretable machine learning—A guide for making black box models explainable. https://christophm.github.io/interpretable-ml-book/ (accessed June 1, 2020).
Nas, B., and A. Berkty. 2010. Groundwater quality mapping in
urban groundwater using GIS. Environmental Monitoring and Assessment 160, no. 1: 215–227. https://doi.org/10.1007/s10661-008-0689-4
Nash, J.E., and J.V. Sutcliffe. 1970. River flow forecasting through conceptual models part I—A discussion of principles. Journal of Hydrology 10, no. 3: 282–290. https://doi.org/10.1016/0022-1694(70)90255-6
Neal, C., and J.W. Kirchner. 2000. Sodium and chloride levels in rainwater, mist, streamwater and groundwater at the Plynlimon catchments, mid-Wales: Inferences on hydrological and chemical controls. Hydrology and Earth System Sciences 4: 295–310. https://doi.org/10.5194/hess-4-295-2000
Nolan, B.T., M.N. Fienen, and D.L. Lorenz. 2015. A statistical learning framework for groundwater nitrate models of the Central Valley, California, USA. Journal of Hydrology 531: 902–911. https://doi.org/10.1016/j.jhydrol.2015.10.025
Nolan, B.T., J.M. Gronberg, C.C. Faunt, S.M. Eberts, and K. Belitz. 2014. Modeling nitrate at domestic and public-supply well depths in the Central Valley, California. Environmental Science & Technology 48, no. 10: 5643–5651. https://doi.org/10.1021/es405452q
Norton, S.A., J. Kopáček, and I.J. Fernandez. 2014. Acid rain—Acidification and recovery. In Treatise on Geochemistry—Second ed., ed. Heinrich D. Holland, and Karl K. Turekian, 379–414. Oxford, UK: Elsevier. https://doi.org/10.1016/B978-0-08-095975-7.00910-4
NRC. 2013. National Research Council: Critical Aspects of EPA’s IRIS Assessment of Inorganic Arsenic: Interim Report. Washington, DC: The National Academies Press.
Parkhurst, D.L., and C.A.J. Appelo. 2013. Description of input
and examples for PHREEQC version 3: A computer program for speciation, batch-reaction, one-dimensional
transport, and inverse geochemical calculations. U.S. Geological Survey Techniques and Methods 6-A43. https://doi.org/10.3133/tm6A43
Parkhurst, D.L., K.L. Kipp, and S.R. Charlton. 2010. PHAST
version 2-A program for simulating groundwater flow, solute transport, and multicomponent geochemical reactions. U.S. Geological Survey Techniques and Methods 6-A35. https://doi.org/10.3133/tm6A35
Ransom, K.M., B.T. Nolan, J. Traum, C.C. Faunt, A. Bell, J.M. Gronberg, D. Wheeler, C. Rosecrans, B. Jurgens, G.E. Schwarz, K. Belitz, S.M. Eberts, G. Kourakos, and T. Harter. 2017. A hybrid machine learning model to predict and visualize nitrate concentration throughout the Central Valley aquifer, California, USA. Science of The Total Environment 601–602: 1160–1172. https://doi.org/10.1016/j.scitotenv.2017.05.192
Read, J.S., X. Jia, J. Willard, A.P. Applping, J.A. Zwart, S.K. Oliver, A. Karpatne, G.J.A. Hansen, P.C. Hansen, W. Watkins, M. Steinchab, and V. Kumar. 2019. Process-guided deep learning predictions of lake water temperature. Water Resources Research 55: 9173–9190. https://doi.org/10.1029/2019WR024922
Reid, J.M., D.A. MacLeod, and M.S. Cresser. 1981. Factors affecting the chemistry of precipitation and river water in an upland catchment. Journal of Hydrology 50: 129–145. https://doi.org/10.1016/0022-1694(81)90064-0
Rodgers, R.J. 1989. Geochemical comparison of groundwater in areas of New England, New York and Pennsylvania. Groundwater 27, no. 5: 690–712. https://doi.org/10.1111/j.1745-6584.1989.tb00483.x
Rosecrans, C.Z., B.T. Nolan, and J.M. Gronberg. 2017. Predicted pH at the domestic and public supply drinking water depths, Central Valley, California. U.S. Geological Survey Scientific Investigations Map 3377. https://doi.org/10.3133/sim3377
Schwarz, G.E., A.B. Hoos, R.B. Alexander, and R.A. Smith. 2006. The SPARROW surface water-quality model: Theory, application and user documentation. U.S. Geological Survey Techniques and Methods, Book 6, Chapter 3, https://doi.org/10.3133/tm6B3
Singh, J.E. 1981. The search for a corrosion index. Journal American Water Works Association 73: 579–582. https://www.jstor.org/stable/41270609
Site, A. 2001. Factors affecting sorption of organic compounds in natural sorbent/water systems and sorption coefficients for selected pollutants. A review. Journal of Physical and Chemical Reference Data 30: 187–439. https://doi.org/10.1063/1.1347984
Spangler, J.G., and J.C. Reid. 2010. Environmental manganese and cancer mortality rates by county in North Carolina: An ecological study. Biological Trace Element Research 133, no. 2: 128–135. https://doi.org/10.1007/s12011-009-8415-9
Stackberg, P.E. 2017. Groundwater Quality in the Glacial Aquifer System, United States. U.S. Geological Survey Fact Sheet. https://doi.org/10.3133/fs2017305
Starn, J.J., L.J. Kauffman, C.S. Carlson, J.E. Reddy, and M.N. Fienen. 2020. Data for three-dimensional distribution of groundwater residence time metrics in the glaciated United States using metamodels trained on general numerical simulation models: U.S. Geological Survey data release https://doi.org/10.5066/P9BNWWCU
Starn, J., and K. Belitz. 2018. Regionalization of groundwater residence time using metamodeling. Water Resources Research 54, no. 9: 6357–6373. https://doi.org/10.1029/2017WR021531
Stekhoven, D.J., and P. Bühlmann. 2011. MissForest—Non-parametric missing value imputation for mixed-type data. Bioinformatics 28, no. 1: 112–118. https://doi.org/10.1093/bioinformatics/btr597
NGWA.org
P.E. Stackelberg et al. Groundwater 59, no. 3: 352–368
367
Stollenwerk, K.G. 2003. Geochemical processes controlling transport of arsenic in groundwater: A review of adsorption. In *Arsenic in Ground Water—Geochemistry and Occurrence*, eds. A.H. Welch, and K.G. Stollenwerk. Boston: Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47956-7_3

Stumm, W., and J.J. Morgan. 1981. *Aquatic Chemistry—An Introduction Emphasizing Chemical Equilibria in Natural Waters*, 2nd ed. New York: John Wiley & Sons.

Szabo, Z., P.E. Stackelberg, and C.A. Cravotta III. 2020. Occurrence and geochemistry of lead-210 and polonium-210 radionuclides in public-drinking-water supplies from principal aquifers of the United States. *Environmental Science & Technology* 27: 729–752. https://dx.doi.org/10.1021/acs.est.0c00192

USEPA. 2016. Basic information about lead in drinking water. https://www.epa.gov/your-drinking-water/basic-information-about-lead-drinking-water#getinto (accessed August 8, 2020).

USEPA. 2013. Safe drinking water information system (SDWIS). https://www.epa.gov/enviro/sdwis-search (accessed July 12, 2020).

Walter, D.A., and J.J. Starn. 2013. The use of process models to inform and improve statistical models of nitrate occurrence, Great Miami River Basin, southwestern Ohio. U.S. Geological Survey Scientific Investigations Report 2012-5001. doi: http://pubs.usgs.gov/sir/2012/5001

Walter, D.A., J.P. Masterson, and K.M. Hess. 2004. Groundwater recharge areas and travel times to pumped wells, ponds, streams, and coastal water bodies, Cape Cod, Massachusetts. U.S. Geological Survey Scientific Investigations Map J-2857. https://doi.org/10.3133/sim2857

Warner, K.L., and J.D. Ayotte. 2015. The quality of our Nation’s waters: Water quality in the glacial aquifer system, northern United States, 1993–2009. U.S. Geological Survey Circular 1352. https://doi.org/10.3133/cir1352

Welch, A.H., D.B. Westjohn, D.R. Helsel, and R.B. Wanty. 2000. Arsenic in ground water of the United States: Occurrence and geochemistry. *Groundwater* 38, no. 4: 589–604. https://doi.org/10.1111/j.1745-6584.2000.tb00251.x

Wheeler, D.C., B.T. Nolan, A.R. Flory, C.T. DellaValle, and M.H. Ward. 2015. Modeling groundwater nitrate concentrations in private wells in Iowa. *Science of The Total Environment* 536: 481–488. https://doi.org/10.1016/j.scitotenv.2015.07.080

Wiklander, L. 1975. The role of neutral salts in the ion exchange between acid precipitation and soil. *Geoderma* 14, no. 2: 93–105. https://doi.org/10.1016/0016-7061(75)90068-3

Wilson, J.T., L. J. Kauffman, M. L. Erickson, and J.B. Sharpe. 2019. Data used to evaluate drinking water quality in the glacial aquifer system, Northern USA. U.S. Geological Survey Data Release. https://doi.org/10.5066/P9POL486

Yager, R.M., L.J. Kauffman, D.R. Soller, A.E. Haj, P.M. Heisig, C.A. Buchwald, S.M. Westenbrook, and J.E. Reddy. 2018. Characterization and occurrence of confined and unconfined aquifers in glacial cover sediments in the glaciated conterminous United States. U.S. Geological Survey Scientific Investigations Report 2018-5091. https://doi.org/10.3133/sir20185091

Yan, J., S. Jia, A. Lv, and W. Zhu. 2019. Water resources assessment of China’s transboundary river basins using a machine learning approach. *Water Resources Research* 55, no. 1: 632–655. https://doi.org/10.1029/2018WR023044

Zell, W.O., and W.E. Sanford. 2020. Calibrated simulation of the long-term average surficial groundwater system and derived spatial distributions of its characteristics for the contiguous United States. *Water Resources Research* 56, no. 8: 1–16. https://doi.org/10.1029/2019WR026724

Zhang, G., and Y. Lu. 2012. Bias-corrected random forests in regression. *Journal of Applied Statistics* 39, no. 1: 151–160. https://doi.org/10.1080/02664763.2011.578621