

Mapped Predictions of Manganese and Arsenic in an Alluvial Aquifer Using Boosted Regression Trees

by Katherine J. Knierim¹, James A. Kingsbury², Kenneth Belitz³, Paul E. Stackelberg⁴, Burke J. Minsley⁵, and J.R. Rigby⁶

Abstract

Manganese (Mn) concentrations and the probability of arsenic (As) exceeding the drinking-water standard of 10 $\mu\text{g/L}$ were predicted in the Mississippi River Valley alluvial aquifer (MRVA) using boosted regression trees (BRT). BRT, a type of ensemble-tree machine-learning model, were created using predictor variables that affect Mn and As distribution in groundwater. These variables included iron (Fe) concentrations and specific conductance predicted from previously developed BRT models, groundwater flux and age estimates from MODFLOW, and hydrologic characteristics. The models also included results from the first airborne geophysical survey conducted in the United States to target an entire aquifer system. Predictions of high Mn and As occurred where Fe was high. Predicted high Mn concentrations were correlated with fraction of young groundwater (less than 65 years) computed from MODFLOW results. High probabilities of As exceedance were predicted where groundwater was relatively old and airborne electromagnetic resistivity was high, typically proximal to streams. Two-variable partial-dependence plots and sensitivity analysis were used to provide insight into the factors controlling Mn and As distribution in groundwater. The maps of predicted Mn concentrations and As exceedance probabilities can be used to identify areas where these constituents may be high, and that could be targeted for further study. This paper shows that incorporation of a selected set of process-informed data, such as MODFLOW results and airborne geophysics, into a machine-learning model improves model interpretability. Incorporation of process-rich information into machine-learning models will likely be useful for addressing a wide range of problems of interest to groundwater hydrologists.

Introduction

High concentrations of manganese (Mn) and arsenic (As) in drinking water pose a risk to human health. The World Health Organization (2011) recommends a limit of 400 $\mu\text{g/L}$ for Mn and 10 $\mu\text{g/L}$ for As. In the United States,

300 $\mu\text{g/L}$ is a nonenforceable health-based screening level (HBSL) for Mn (Toccalino et al. 2018), whereas 10 $\mu\text{g/L}$ is an enforceable maximum contaminant level (MCL) for As (U.S. Environmental Protection Agency [U.S. EPA] 2020a). Hundreds of millions of people globally are exposed to high As concentrations in groundwater, mostly in Asia and South America (Amini et al. 2008; Podgorski and Berg 2020). High Mn concentrations in drinking water have been noted in Bangladesh (Ghosh et al. 2020), Finland (Kousa et al. 2020), and the United States (McMahon et al. 2019). High Mn concentrations also cause nonhealth related problems; for example, concentrations more than 100 $\mu\text{g/L}$ contribute to fouling of irrigation equipment (Haman 2017). In the United States, the U.S. EPA has a secondary MCL (SMCL) of 50 $\mu\text{g/L}$ for Mn in drinking water because of staining of fixtures (U.S. EPA 2020b).

The greatest prevalence of high Mn groundwater in the United States is found in the Mississippi River Valley alluvial aquifer (MRVA) (McMahon et al. 2019), which also includes areas of high As (Kingsbury et al. 2015) (Figure 1). Groundwater from the MRVA is used for drinking-water supply, but availability is limited by high Mn and As (Kingsbury et al. 2015). The MRVA is

¹Corresponding author: U.S. Geological Survey, PO Box 25303, Little Rock, AR 72221; kknierim@usgs.gov

²U.S. Geological Survey, Nashville, TN

³U.S. Geological Survey, Carlisle, MA

⁴U.S. Geological Survey, Troy, NY

⁵U.S. Geological Survey, Denver, CO

⁶U.S. Geological Survey, Oxford, MS

Article impact statement: Machine-learning models trained on output from geophysical surveys, MODFLOW, and water-quality predictions can be used to map trace metals.

Received September 2021, accepted December 2021.

Published 2022. This article is a U.S. Government work and is in the public domain in the USA. *Groundwater* published by Wiley Periodicals LLC on behalf of National Ground Water Association.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

doi: 10.1111/gwat.13164

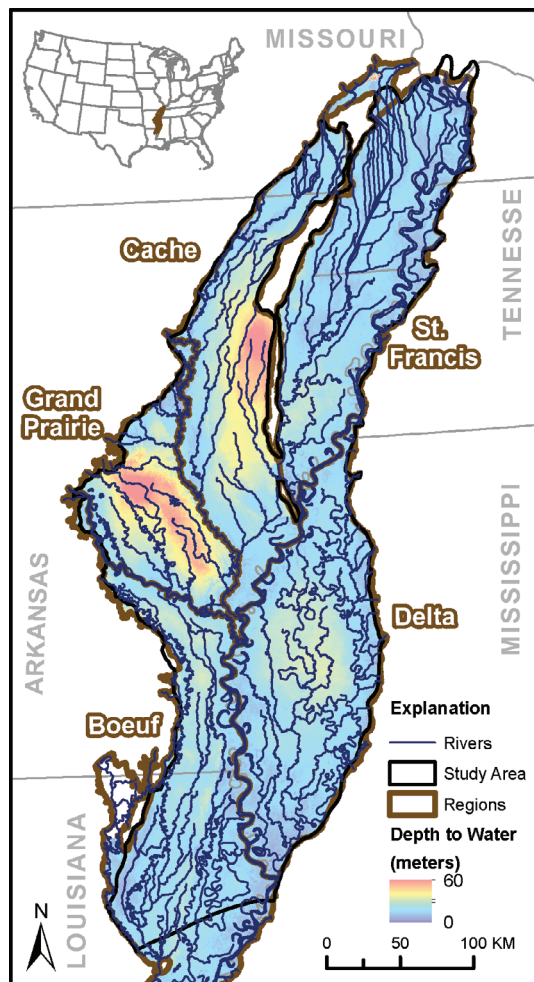


Figure 1. Location of the Mississippi River Valley alluvial aquifer and regions associated with the Mississippi Alluvial Plain. Depth to water from 2018 calculated from McGuire et al. (2020).

the primary source of irrigation supply that supports a \$12 billion agricultural economy (Alhassan et al. 2019). The presence of Mn can foul irrigation equipment (Haman 2017). The presence of As can be a concern for irrigating of food crops, particularly rice (Zavala and Duxbury 2008), which is grown in the MRVA. A better understanding of Mn and As distribution in this highly stressed (Clark and Hart 2009) and important groundwater resource is needed.

Previous investigations of As in the MRVA have primarily focused at the field or sub-aquifer scale or been included as part of continental and global studies. Studies in Louisiana and Arkansas found that As was mobilized via dissimilatory iron (Fe) reduction from Fe and Mn oxide coatings on sediment (Sharif et al. 2008; Borrok et al. 2018). At the continental scale, predictions of As focused on domestic drinking-water wells and found low As in the MRVA (Ayotte et al. 2017; Lombard et al. 2021). At the global scale, As has been studied due to its significant health effects; low probabilities of high As were predicted in the MRVA (Amini et al. 2008; Podgorski and Berg 2020). The low predicted probabilities

in continental and global maps do not exclude the possibilities of more localized areas of high As in the MRVA.

There are fewer studies of Mn relevant to the MRVA as compared to As. A drinking water-quality assessment of the MRVA found high Mn associated with anoxic conditions (Kingsbury et al. 2015). McMahon et al. (2019) noted that pumping could affect Mn distribution because of changes in groundwater flow directions and recharge sources; this could have important implications in the heavily pumped MRVA (Clark and Hart 2009). No continental or global predictions of Mn are available, though there is growing concern over high Mn in drinking water (Ramachandran et al. 2021).

The objectives of this work were to (1) map the distribution of Mn and As across the entire MRVA (both spatially and with depth) using limited data and (2) better understand the hydrogeologic drivers of high Mn and As concentrations in the aquifer. Currently, maps of Mn are not available and there are no aquifer-specific maps of As in the MRVA. A better understanding of the hydrogeologic drivers of Mn and As in the MRVA is important to help identify where the resource can be developed and how concentrations might change in the future due to stresses imposed on the system. We used Boosted Regression Trees (BRT) to accomplish these objectives.

BRT is a type of ensemble-tree machine-learning method that combines many simple tree-based models to improve predictive performance (Elith et al. 2008). Previous research has shown that ensemble-tree methods, such as BRT, are able to accurately predict groundwater quality across regional aquifers (Nolan et al. 2015; Stackelberg et al. 2020), the continental United States (Ransom et al. 2021), and the globe (Podgorski and Berg 2020). The algorithms can identify complex, nonlinear relations among predictor variables (also known as “features”) to make spatially continuous predictions of groundwater quality (Nolan et al. 2015). Hence, BRT was used to map Mn and As in the MRVA. Testing the suitability of different types of machine-learning models for predicting Mn and As was not a goal of this work.

Machine-learning models are sometimes thought of as uninterpretable “black boxes,” but they need not be (Molnar 2019; Stackelberg et al. 2020). Tools, such as variable influence and partial-dependence plots, are available to quantitatively evaluate how the BRT models use predictor variables to make predictions. In addition, the interpretability of a machine-learning model can be improved by careful selection, or curation, of predictor variables. Hence variable influence, partial dependence plots, and model sensitivity testing were used to obtain a better understanding of the hydrogeologic drivers of Mn and As in the MRVA.

Predictor variables were selected using prior knowledge of the geochemical and hydrogeologic conditions affecting Mn and As. Geochemical conditions affecting Mn and As include redox and pH (Welch et al. 2000; Smedley and Kinniburgh 2002; McMahon et al. 2019). Hydrogeologic characteristics include groundwater age,

depth to groundwater, recharge source, and sediment texture (Stute et al. 2007; Aziz et al. 2008; Sharif et al. 2008; Yang et al. 2014; McMahon et al. 2019; Nghiem et al. 2019). The BRT models were developed with direct indicators and surrogates for these characteristics. The incorporation of selected process-oriented variables permits subsequent interpretation of the models from a hydrogeologic perspective.

Among the surrogate variables used in the BRT models were information from a MODFLOW groundwater-flow model and an aquifer-wide airborne geophysical survey. Groundwater variables included groundwater fluxes and age estimates obtained from a MODFLOW model of the region (Haugh et al. 2020a, 2020b, 2020c). The geophysical survey was the first in the United States to collect airborne electromagnetic (AEM) data across an entire regional aquifer system (Minsley et al. 2021). The survey produced three-dimensional models of subsurface electrical resistivity that provide new information about aquifer structure and texture. The geophysical survey also used an airborne gamma-ray spectrometer to measure shallow soil radiochemistry, which helps constrain surficial geology (James and Minsley 2021; Minsley et al. 2021). The BRT models of Mn and As represent the first use of aquifer-wide geophysical data to predict groundwater quality in the United States at a regional aquifer scale.

Methods

The BRT models were developed in five steps: (1) acquire Mn and As groundwater data (response variables); (2) assign predictor variables to response variables; (3) tune BRT models, make predictions at wells, and assess performance; (4) predict Mn and As across the aquifer; and (5) inspect predictor variable importance, influence, interaction, and sensitivity. The workflow for steps three through five is available as R scripts (R Core Team 2019) in Knierim et al. (2021a).

Response Variables

Groundwater quality data were acquired from the USGS National Water Information System (U.S. Geological Survey 2019), U.S. EPA's Safe Drinking Water Information System (U.S. EPA 2013), and state agencies that maintain groundwater monitoring networks. Mn and As data from wells sampled between 1960 and 2019 were used for model development. When multiple samples were collected at a well, data from the most recent sample were used.

Approximately 2% of the 1022 Mn samples were censored at 100 $\mu\text{g/L}$ or less, as multiple censoring limits occur when analytical methods and detection limits change over time. There is concentration information in censored values, although the exact concentration is not known. Censored Mn values were imputed using the NADA package in R (Lee 2020), which is based on methods for analyzing nondetects in environmental samples from Helsel and Hirsch (2002). Mn data were natural log transformed for prediction with a regression

BRT model. The response variable may benefit from a transformation because BRT performance metrics are ultimately evaluated around a central tendency, such as RMSE.

Approximately 13% of the 764 As samples were censored below 10 $\mu\text{g/L}$ or less. The relatively higher censoring rate for As required a classification BRT model, where the probability of As concentration exceeding a threshold of 10 $\mu\text{g/L}$ was predicted. Arsenic values ≤ 10 $\mu\text{g/L}$ were classified as a nonevent and more than 10 $\mu\text{g/L}$ were classified as an event. There were 10 As samples with a reported concentration of 0 $\mu\text{g/L}$, which is assumed to be reported without a censoring limit. Censored and 0 $\mu\text{g/L}$ values were not imputed because the values are classified as nonevents for the BRT classification model.

Mn and As datasets were split into a training dataset (80%) for model tuning and holdout dataset (20%) to assess model performance. Training and holdout data were split randomly in R using the *caret* package; a seed was set so that the random split can be reproduced to allow model verification (Kuhn et al. 2019; R Core Team 2019). Input data for the BRT model are available in Knierim et al. (2021a).

Predictor Variables

The BRT modeling focused on producing accurate predictions of Mn and As in the MRVA using a targeted set of predictor variables to increase model interpretability. Nineteen predictor variables were selected, which represent indicators or surrogates of conditions that have been found to control Mn and As in groundwater (Table S1, Supporting Information). Predictor variables must be available as spatially continuous rasters across the study area (Figure 1) so that the BRT models can be used to predict groundwater quality in the MRVA. For the purposes of discussion, the BRT predictor variables were grouped into four categories: geochemical, hydrologic characteristics, groundwater model (input and output), and geophysical. All variables were used in the final BRT models and recursive feature elimination was not performed.

It is assumed that the magnitude of any error associated with an input variable is similar across the model domain. The effect of such error on Mn and As predictions was not quantified.

Geochemical conditions were extracted from previously developed BRT models of pH (Kingsbury et al. 2020a, 2020b), specific conductance (SC) (Knierim et al. 2020a, 2020b), and Fe (Knierim et al. 2020c, 2021b) (Table S1). The SC, Fe, and pH BRT models were trained using similar predictor variables as the Mn and As BRT models, with the exception of the geophysical datasets. BRT methods are able to handle co-linearity and interaction among predictors (Elith et al. 2008). Mn and As observations were not used to create any predictor variables used in the SC, Fe, and pH BRT models, so feature leakage—where information about the response variable is unintentionally encoded in one or more

variables (Kaufman et al. 2012)—is not expected. In the MRVA, SC provides a surrogate of rock-water interaction and, thus, position along flowpath (Knierim et al. 2020b). Iron serves as a surrogate for redox conditions (Chapelle et al. 1995) and was hypothesized to be more diagnostic of redox than dissolved oxygen, which is generally low throughout the MRVA (Knierim et al. 2021b). Groundwater pH can exert important control on As mobilization (Smedley and Kinniburgh 2002).

Nine groundwater model variables (Table S1) were derived from a regional MODFLOW groundwater-flow model (Haugh et al. 2020a, 2020b, 2020c). Input and output variables were extracted from the flow model using the *flopy* package in Python (Bakker et al. 2016, 2021) and assigned to wells by intersecting the bottom of the well with the corresponding groundwater-flow model layer grid cell. Cell-by-cell flow was used to describe the magnitude, direction, and vertical to horizontal ratio of groundwater flow at each model cell (Table S1). Variables for groundwater residence time—a surrogate of groundwater age—were derived from particle tracks (Haugh et al. 2020c) using MODPATH6 (Pollock 2012) following methods from Starn and Belitz (2018).

Three explanatory variables from the geophysical survey were included in the BRT models: two from the AEM and one from the gamma-ray instruments (Table S1). AEM data collected across the MRVA were inverted to derive three-dimensional electrical resistivity models along flightpaths and then kriged to the 1-km² resolution National Hydrologic Grid (Clark et al. 2018) to provide an aquifer-wide characterization of geologic structure (Minsley et al. 2021). Bulk electrical resistivity is sensitive to subsurface properties related to changes in sediment texture, lithology, and salinity (Palacky 1987). AEM-derived average bulk resistivity of the MRVA was included in the BRT models (Table S1). The degree of surficial connectivity or confinement, a variable also included in the BRT models, was interpreted from shallow resistivity values to map the thickness and extent of the surficial clay unit that overlies the MRVA (James and Minsley 2021); this unit exerts important control on areal recharge (Renken 1998). A gamma-ray spectrometer was used to measure the abundance of natural radioelements (including potassium, uranium, and thorium) in the upper approximately 30 cm. The ratio of thorium to potassium was also included in the BRT models (Table S1).

Hydrologic characteristics included average recharge from a continental United States model (Reitz et al. 2017), multiorder hydrologic position (MOHP) variables for third order streams (MOHP-3) (Belitz et al. 2019; Moore et al. 2019), and well depth (Table S1). Well depth was either included with well metadata (70% for Mn and 49% for As) or extracted from an interpolated surface of MRVA withdrawal zones (Knierim et al. 2019).

Values for predictor variables were assigned to the well location for Mn and As samples using a point, buffer, or grid extraction from the predictor raster. MOHP was assigned to response data using a point extraction at the well location because of the high

resolution of the source rasters (Belitz et al. 2019; Moore et al. 2019). Recharge, geophysical variables, and BRT-derived groundwater quality were assigned to response data using a 500-m buffer around the well and the *zonepy* package in Python (Clark et al. 2019; Python Software Foundation 2019). The buffer approach represents a surrogate for contributing recharge area to the well (Johnson and Belitz 2009). As described, groundwater flow-model derived variables were assigned to wells using an intersection with the MODFLOW model grid. The extracted values were stored in flat files for input to the BRT models and are available in Knierim et al. (2021a). Descriptions and source references for all predictor variables are available in Table S1.

Boosted Regression Tree Modeling

BRT is an ensemble-tree machine-learning method that utilizes decision trees (Elith et al. 2008; Kuhn and Johnson 2016). In a decision tree, a dataset is sequentially divided through a series of splits into smaller sets to reduce variance. At each split, the predictor variable that most effectively reduces variance in the response variable is used to divide the data. In practice, a single, complex decision tree can be trained to a high level of accuracy but performs poorly when making new predictions. Ensemble learning methods, including BRT, combine a large number of simple decision trees to improve predictive accuracy. These simple trees are sometimes called weak learners. The BRT algorithm differs from some other ensemble methods in that it sequentially grows trees from the residuals of the previous predictions. The BRT approach increasingly focuses on points that are the most difficult to predict (Elith et al. 2008).

BRT model complexity and performance are controlled by five hyperparameters: number of trees, minimum observations per tree node, interaction depth (how many times the trees split), learning rate (how much of the model residual is used in boosting), and bagging fraction (random fraction of training data used) (Elith et al. 2008; Kuhn and Johnson 2016). Bagging fraction was held at 0.5. A tuning grid with ranges of values for the other four hyperparameters was used to find the combination of hyperparameters with the lowest root-mean-square error (RMSE) for natural log Mn or highest area under the receiving operator curve (AUROC) for the As-exceedance probability.

Model performance of hyperparameter combinations was calculated via 10-fold cross-validation (CV) tuning using the *caret* and *gbm* packages in R (Kuhn et al. 2019; R Core Team 2019; Ridgeway 2019). CV tuning randomly subsets the training dataset a specified number of times, such as 10 subsets for 10-fold CV tuning (Kuhn and Johnson 2016). During 10-fold CV tuning, a random 10% of the training data is used as a “testing” dataset to evaluate model performance, which is repeated 10 times. Testing performance from CV tuning will generally be similar to holdout performance. The CV-tuning grids for Mn and As models were different because of the different dataset sizes and type of BRT model.

The hyperparameter tuning grid for Mn evaluated the predictive performance of 2160 possible models: interaction depth ranged from 4 to 14 by 2, learning rate of 0.001, 0.005, and 0.01, minimum observations per node of 6 to 10 by 2, and number of trees from 50 to 2000 by 50. The “best” model was identified by the hyperparameter combination with the lowest RMSE. Models with an RMSE within one standard error (1-SE) of the “best” model’s RMSE were inspected to find simpler models that may have similar or better predictive performance for the holdout data (Kuhn et al. 2019). Models were considered less complex for hyperparameter combinations with lower interaction depth, learning rate, and number of trees and higher minimum observations per node. Percent proportion bias, defined as the summed difference between predicted and observed values divided by observed values (Knierim et al. 2021b), and range of predicted concentrations were also considered when choosing the final BRT model. Final selection of hyperparameters is ultimately at the discretion of the modeler. Many of the 1-SE models could be selected as reasonable models because the variation in RMSE was generally small (less than 0.15 natural log $\mu\text{g/L}$).

BRT predictions typically overpredict low observations and underpredict high observations (Zhang and Lu 2012). Retransforming predicted natural log Mn to original concentration units can introduce additional bias (Belitz and Stackelberg 2021). An empirical distribution matching (EDM) procedure from Belitz and Stackelberg (2021) was used to correct bias and retransform predicted Mn values. The EDM method uses the empirical distribution of the training dataset observations to transform Mn predictions so the resulting cumulative distribution function matches that of the observations (Belitz and Stackelberg 2021). In a series of case studies, the EDM method was found to outperform other transformation methods, including for retransforming predicted natural log Fe concentration in the study area (Belitz and Stackelberg 2021). Following EDM bias correction, Mn predictions were retransformed from natural log space to concentration units.

The hyperparameter tuning grid for As evaluated the predictive performance of 1890 possible models: interaction depth ranged from 4 to 8 by 2, learning rate of 0.001, 0.005, and 0.01, minimum observations per node of 4 to 12 by 2, and number of trees from 10 to 700 by 10. Initial BRT model runs indicated that more than 1000 trees caused the As models to be overfit (e.g., perfect prediction of training data) so the tuning grid was selected to reduce overfitting. The “best” model was identified by the hyperparameter combination with the highest AUROC. Models with AUROC values within 1-SE of the “best” model’s AUROC were inspected to find simpler models that may have similar or better predictive performance for the holdout data (Kuhn et al. 2019).

In addition to AUROC, the sensitivity (correct event prediction) and specificity (correct nonevent prediction) of holdout data were used to select the final BRT model. To assess sensitivity and specificity, predicted probabilities

are classified into events and nonevents using a probability cutoff (typically 0.5). Alternate probability cutoffs were calculated to increase model sensitivity, which is a recommended method for biased classes (Kuhn and Johnson 2016) and has been used for global As prediction maps (Podgorski and Berg 2020). The final probability cutoff was calculated so that the predicted event rate of the training dataset was the same as the observed prevalence of events (13%). The final model also balanced false positive (observed nonevents predicted as events) and false negative (observed events predicted as nonevents) predictions, while maintaining model sensitivity.

Prediction Mapping

Rasters of predictor variables were resampled to the 1-km² resolution National Hydrologic Grid (Clark et al. 2018) for mapping Mn and As distribution. Resampling was required to ensure all predictor variable datasets were in the same resolution and extent, so predictions can be made across the aquifer. Resampling was completed in Python using the *gdal* package methods of bilinear interpolation (for most variables) or nearest neighbor (where interpolation computed values outside of the range of published values; GDAL/OGR contributors 2019). Most variables have similar or lower resolution as the 1-km² prediction resolution.

The *gbm* package in R (Ridgeway 2019) was used to predict Mn concentration and As-exceedance probability at raster cells across the MRVA. Predictions were made using the final hyperparameters selected for each BRT model and a flat file of predictor variable values at each raster cell (Knierim et al. 2021a). The depth represented by each raster cell (and hence depth of BRT model prediction) was the mid-point depth of MRVA saturated thickness, calculated using the 2018 potentiometric groundwater altitude (McGuire et al. 2020) and the bottom of the MRVA (Torac and Painter 2019a, 2019b). Wells used in model training may have different depths than the prediction depth of the coincident raster cell where the well is located, but BRT methods can handle this sort of complexity. Prediction depths ranged from 5 to 65 m below land surface. Mn predictions at raster cells were also bias corrected using the EDM method and retransformed to concentration units. Predicted As probabilities were categorized into nonevents (\leq probability cutoff value) and events ($>$ probability cutoff value).

The prediction mapping may introduce error in addition to error associated with predictor variables and the BRT models. The propagation of these errors and resulting uncertainty in Mn and As predictions was not quantified. However, the modeling effort sought to produce predictions with similar proportions of exceedance (Mn $>300 \mu\text{g/L}$ and As $>10 \mu\text{g/L}$) as the observations.

Importance, Interaction, Influence, and Thematic Sensitivity of Predictor Variables

Quantifying how predictor variables affect prediction of response variables facilitates interpretation of BRT

Table 1

Boosted Regression Tree Hyperparameters and Holdout Performance for Manganese (Mn) Concentration and Arsenic (As) Exceedance Probability (id, interaction depth; mo, minimum observations per node; lr, learning rate; nt, number of trees; RMSE, root-mean-square-error; AUROC, area under receiving operator curve)

Response Variable	Number of Samples		Model	Hyperparameters				RMSE or AUROC	Sensitivity ¹	Specificity ¹
	Training	Holdout		id	mo	lr	nt			
Mn	819	203	Best	12	6	0.01	700	1.60	NA	NA
			Simplest	4	10	0.005	700	1.65	NA	NA
			Final	8	10	0.01	1500	1.62	NA	NA
As	611	153	Best	8	4	0.01	700	0.79	0.26	0.96
			Simplest	4	4	0.01	660	0.78	0.26	0.90
			Final	6	8	0.01	250	0.78	0.37	0.96

¹Using a cutoff probability value based on 13% prevalence event rate in training dataset.

models and a better understanding of the drivers of Mn and As distribution. The relative importance of a variable for predicting Mn or As was calculated using the summary method in *gbm* (Ridgeway 2019). Variables with greater importance are used more frequently to split trees because of subsequent improvement to the loss function (Friedman 2001; Elith et al. 2008). Variable interaction quantifies the fraction of variance explained by interaction between two variables, which is not explained via the two variables independently (Friedman and Popescu 2008). The H-statistic of variable interaction was calculated pairwise for all variables using the *interact* method in *gbm*, and ranges from 0 (no interaction) to larger values for stronger interaction (Ridgeway 2019).

Partial-dependence plots show the direction and magnitude of predictor variable influence for predicting the response variable if all other predictor variables are held at their measured or calculated values (Friedman 2001). Two-variable partial dependence plots show how two variables interact to predict the response variable. In a two-variable partial dependence plot, each predictor variable occurs on the *x*- and *y*-axes, and the magnitude of the predicted response variable is shown as a color ramp in *x*-*y* space. Tick marks on the *x*-axis of partial-dependence plots show the deciles of predictor variable data. Regions of the plot with less tick marks are not well defined because of the limited data. One and two-variable partial-dependence plots were created with the *pdp* package in R (Greenwell 2017).

The effects that predictor variables have on Mn and As predictions was additionally explored thru development of alternative models following methods in Stackelberg et al. (2020). BRT models were retrained without either (1) nine variables representing groundwater characteristics from the groundwater-flow model, or (2) three variables derived from the geophysical survey (Table S1). Model residuals were calculated as the difference in predictions between the original and alternative (retrained) models. Retraining the models without groups of predictor variables and comparing model residuals reveals what effect the excluded variables have on Mn and As

predictions, thus quantifying the sensitivity of the BRT model to the excluded variables (Stackelberg et al. 2020). Note that this sensitivity analysis is different than the sensitivity performance metric used to evaluate categorical BRT models (Table 1), and hereafter will be referred to as “thematic sensitivity” for clarity. Accuracy of the alternative models was not quantified because the models were retrained without tuning, or finding new combinations of hyperparameters. One important variable from each variable category was also inspected in detail with model residuals to illustrate how variable influence and thematic sensitivity can be used to interpret BRT models.

Results

Mn Predictions and Model Performance

Observed Mn concentration ranged from 0.21 to 25,000 $\mu\text{g/L}$ with a median of 431 $\mu\text{g/L}$. High Mn values were underpredicted and low values were overpredicted (Figure 2A), as is typical for tree-base models such as BRT (Zhang and Lu 2012). Following the EDM method for bias correction (Belitz and Stackelberg 2021), predicted Mn at groundwater wells ranged from 0.21 to 25,000 $\mu\text{g/L}$ with a median of 410 $\mu\text{g/L}$ (Figure 2B). The EDM transformation corrected bias in the low and high predicted Mn concentrations using the empirical distribution of Mn observations in the training dataset. For the holdout data, BRT predictions transformed with the EDM method more closely matched the empirical distribution of observed Mn concentrations compared to uncorrected values (Figure S1).

The final BRT model was simpler than the best CV-tuned model, based on a smaller interaction depth and higher number of minimum observations per node (Table 1). The simplest Mn model identified by the 1-SE method was not chosen as the final model because of the higher RMSE on holdout data (Table 1). Depending on learning rate, models were adequately tuned (i.e., the RMSE did not further decrease) after approximately 1000 trees. The BRT model was able to predict the variability

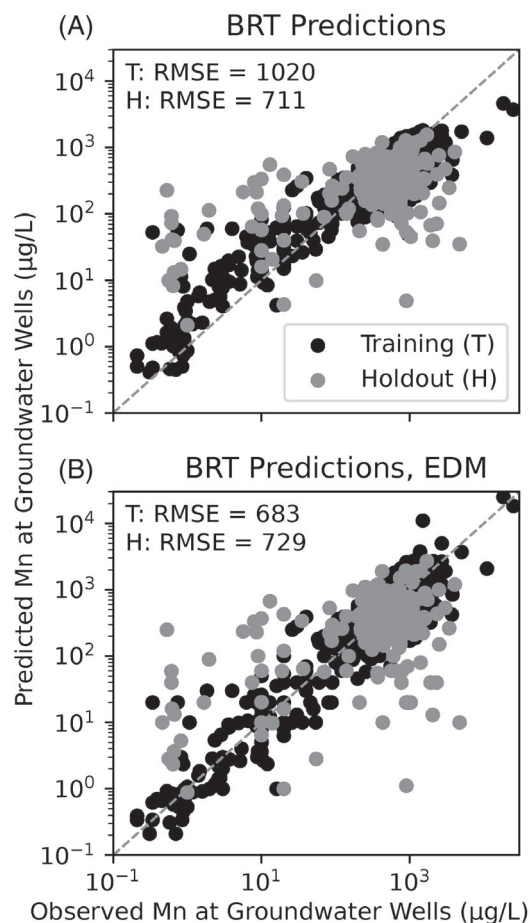


Figure 2. Predicted vs. observed Mn concentration transformed from natural log space for (A) the final Boosted Regression Trees (BRT) model and (B) the final BRT model with Empirical Distribution Matching (EDM) bias correction.

in Mn concentrations (even prior to the EDM transformation), despite 66% of the samples being more than 300 $\mu\text{g/L}$ (Figure 2). The RMSE of EDM-transformed holdout data based on the empirical cumulative distribution function (Belitz and Stackelberg 2021) was 343 $\mu\text{g/L}$ (Figure S1).

Predicted Mn in raster cells across the MRVA ranged from 0 to 21,060 $\mu\text{g/L}$ and had a similar cumulative distribution function as predictions at wells (Figure S1). Predicted Mn concentration was generally less than 300 $\mu\text{g/L}$ in the northern part of MRVA in southeastern Missouri, in the Grand Prairie region of central Arkansas, and in a small area of northeastern Louisiana (Figure 3B). Approximately 52% of the footprint of the aquifer was predicted to have Mn more than 300 $\mu\text{g/L}$, compared to 66% of observations. Less than 1% was greater than 3000 $\mu\text{g/L}$ (10 times the HBSL for drinking water) and limited to a clustering of raster cells in southeastern Missouri, northeastern Arkansas, and along the Mississippi River (Figure 3B). For wells with depths similar to the prediction depth of the coincident raster cell where the well is located (Figure 3A), predictions at points and coincident

raster cells showed a similar distribution compared to observed Mn (Figure S2).

As Predictions and Model Performance

Observed As concentrations ranged from less than 0.1 to 103.2 $\mu\text{g/L}$. Predicted probability of As concentration exceeding a threshold of 10 $\mu\text{g/L}$ in groundwater wells ranged from 0.03 to 0.83 (Figure 4) with a median of 0.1. The low predicted probabilities reflect the relatively high nonevent rate in the As observations. A probability cutoff value of 0.22 ensured a 13% prevalence of predicted events (i.e., As > 10 $\mu\text{g/L}$) in the training dataset, which is equivalent to the prevalence in observations.

The final As BRT model was simpler than the best CV-tuned model based on smaller interaction depth, higher minimum observations per node, and less trees (Table 1). The simplest 1-SE As model was not chosen as the final model because an alternative model was found with a higher sensitivity (0.37) for the holdout data (Table 1). Using a probability cutoff of 0.22, the final model correctly predicted 7 events and 128 nonevents in the holdout dataset, with 6 false positives and 12 false negatives (Figure 4).

Arsenic events were predicted across 4% of the MRVA (Figure 5B). Approximately 10% of wells with depths similar to the depth of the coincident raster cell were classified as events (Figures 5A and S3). Of these, 6 wells were false positives (Figure S3), and all 6 wells were within 1 km (or one raster cell) of a predicted nonevent cell. Fifteen other wells were false negatives (Figure S3), but were within a median distance of 2 km (or two raster cells) of a predicted event cell. The BRT models did not accurately predict As events at every well, but general spatial patterns of high As were captured in the prediction maps. Arsenic events are noted along the Mississippi River corridor and in the Grand Prairie and Cache regions of Arkansas (Figure 5B).

Variable Importance for Predicting Mn and As

Important predictor variables for predicting Mn and As were predicted Fe, predicted SC, fraction of young groundwater (less than 65 years) within each model cell (estimated from the MODFLOW model), AEM-derived bulk resistivity, and well depth (Figure 6). For both Mn and As, aquifer hydraulic conductivity (K) from the MODFLOW model and degree of surficial confinement from AEM surveys were the least influential predictors (Figure 6). For variables with similar relative influence, different 1-SE models showed variable ordering of importance. However, the most and least influential variables were generally similar among models.

Predicted Fe was an important variable for predicting both Mn and As (Figure 6), so was inspected with other important variables in two-variable partial-dependence plots. The two-variable partial dependence plots show how Fe (y-axis) interacts with a second variable (i.e., SC, resistivity, or fraction of young groundwater on the x-axis) to predict either natural log Mn (Figure 7) or

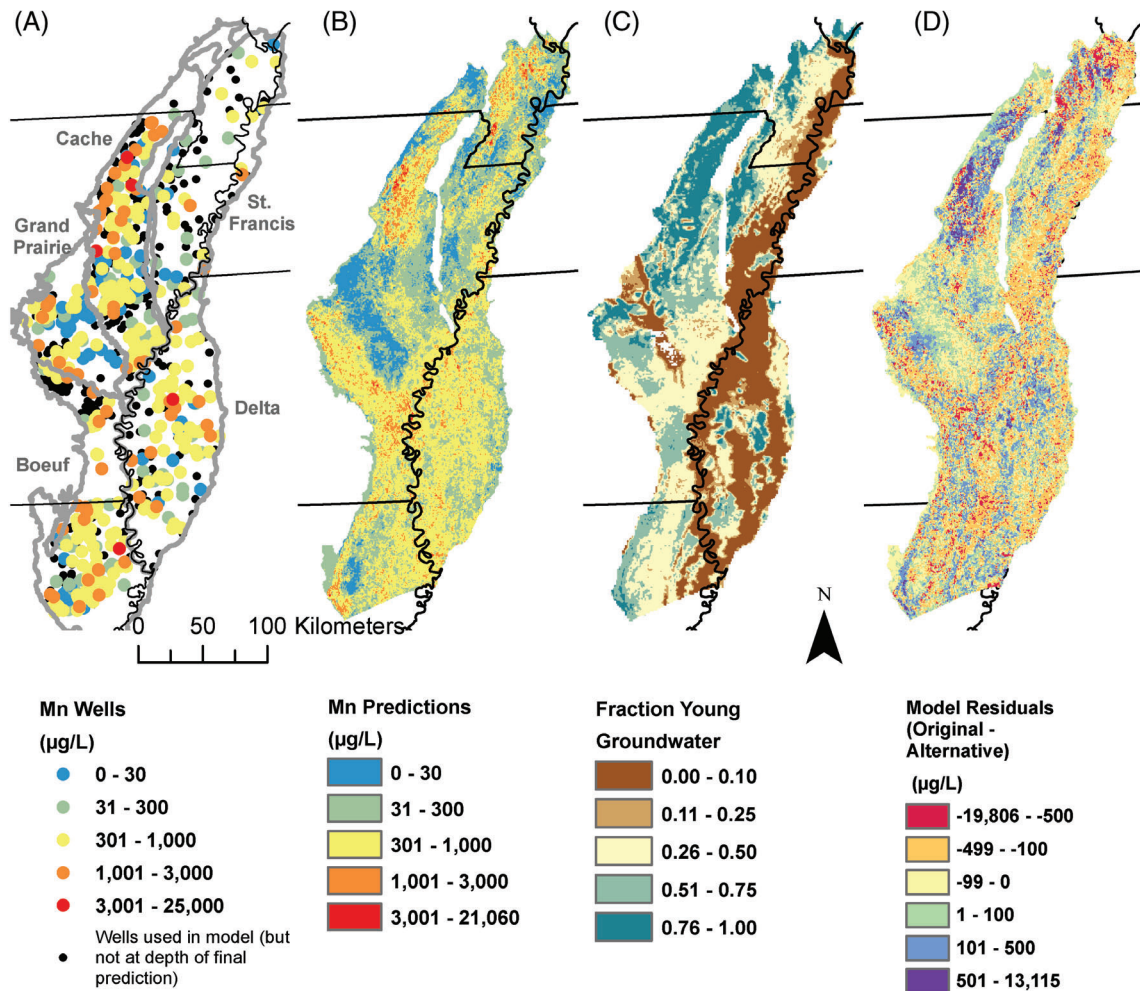


Figure 3. (A) Mn observations in groundwater wells, (B) predicted Mn concentrations from the BRT regression model, (C) fraction of young groundwater in each raster cell (from Haugh et al. 2020c), and (D) the difference in model prediction between the BRT model and an alternative model where all groundwater-flow model variables were withheld during model training.

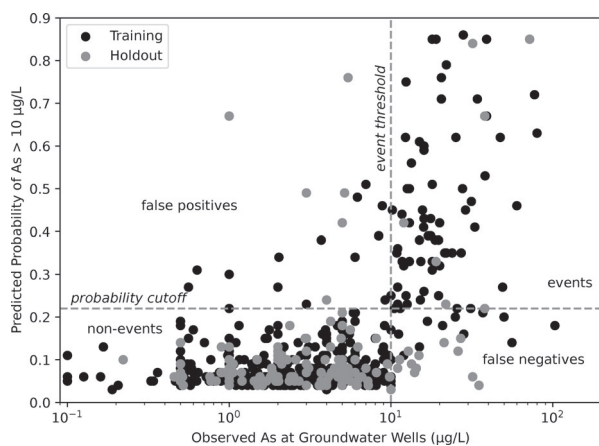


Figure 4. Predicted probability of As concentration exceeding a 10 µg/L threshold vs. observed As concentrations.

As-exceedance probability (Figure 8). One-variable partial dependence plots are included as line plots in Figures 7 and 8.

Discussion

Partial Dependence Plots Illustrate How Geochemical Conditions Influence Mn and As Predictions

Geochemical conditions incorporated into the BRT models—SC and Fe—had a clear effect on predicted Mn. Approximately 90% of SC values were between 400 and 1000 µS/cm; Mn generally increased with SC (Figure 7D). Specific conductance serves as a proxy for position along groundwater flowpath in the MRVA (Knierim et al. 2020b). Geochemical conditions also generally become more reduced along flowpaths. The increase in Mn with SC represents increasingly reduced conditions. Fe also is a surrogate for redox (Chapelle et al. 1995), and Mn increased with Fe, especially from 0 to 5000 µg/L (Figure 7A). The correlation is causative: Mn is released from Fe and Mn oxide coatings due to bacteria reducing Fe and Mn (Sharif et al. 2008). The two-dimensional partial dependence plots show that high Mn generally occurs with increasing SC and where Fe more than 5000 µg/L (Figure 7B).

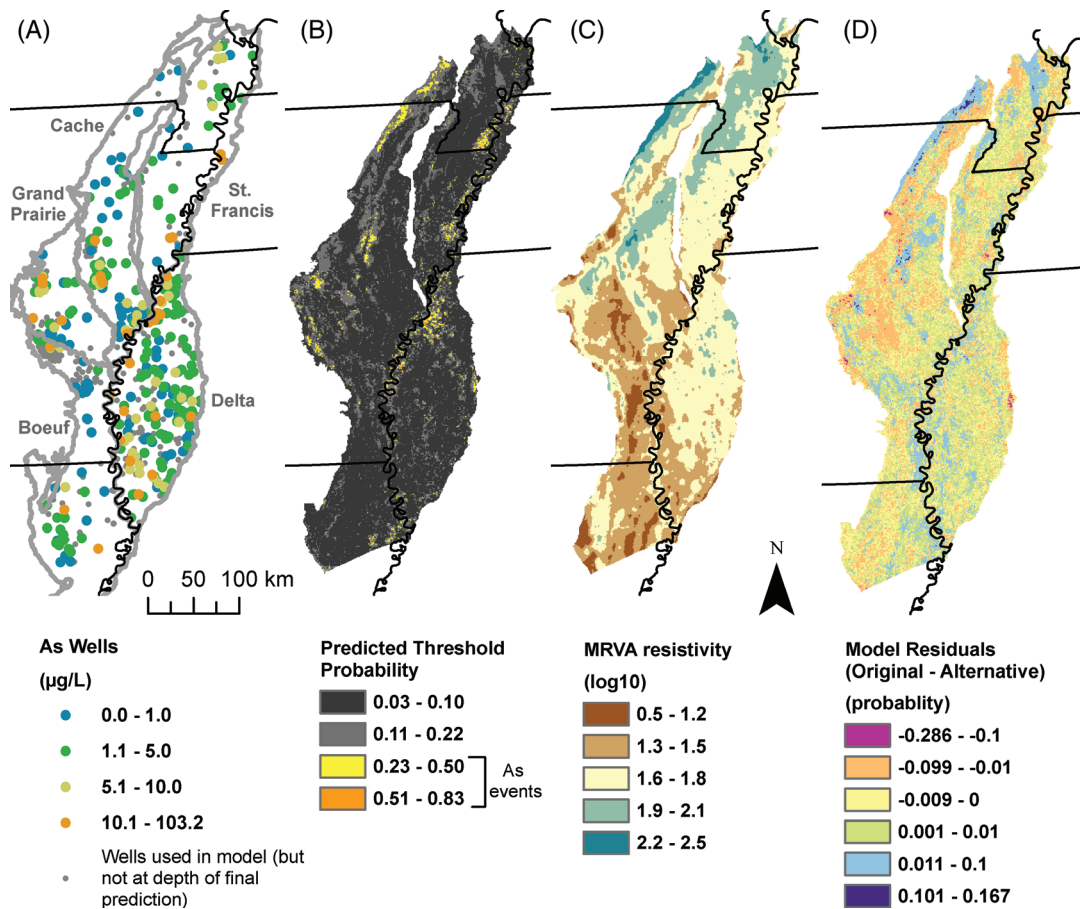


Figure 5. (A) As observations in groundwater wells, (B) predicted As probabilities from the BRT classification model, (C) Log_{10} resistivity values for the MRVA (from Minsley et al. 2021), and (D) the difference in model prediction between the BRT model and an alternative model where all AEM variables were withheld during model training.

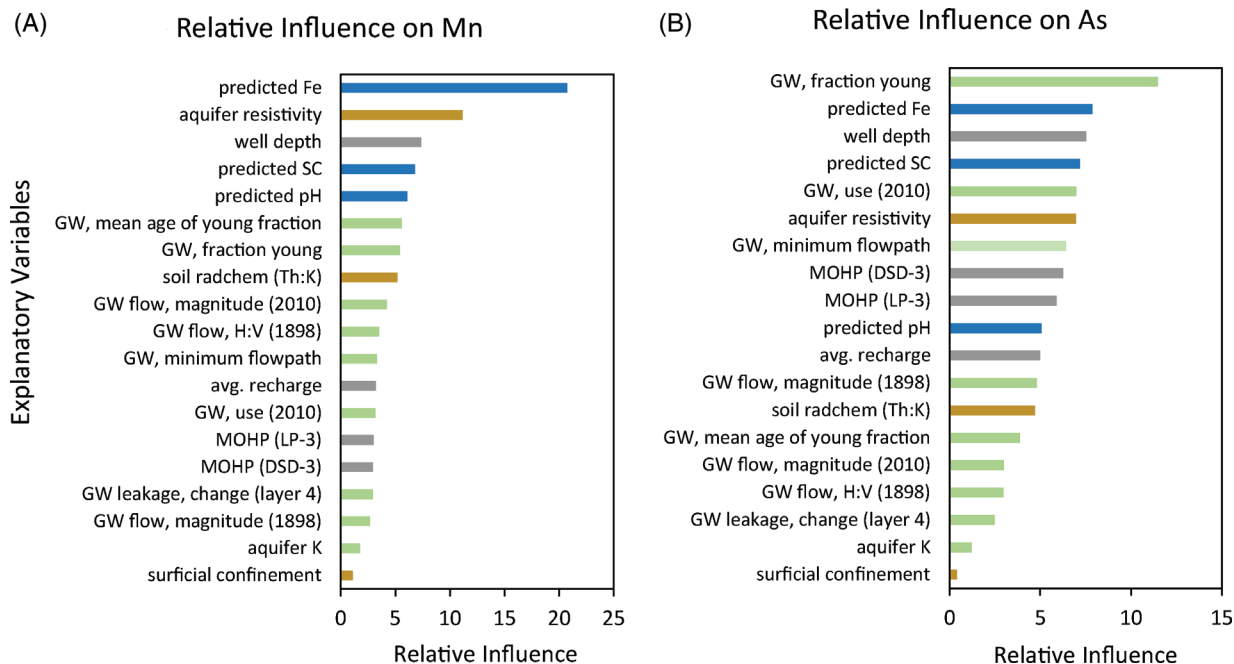


Figure 6. Relative influence of predictor variables for predicting (A) Mn and (B) As. See Table S1 in Supplemental Information for variable dataset sources. Blue, green, gold, and gray variables describe geochemical, groundwater-flow model, geophysical, and hydrologic variables, respectively.

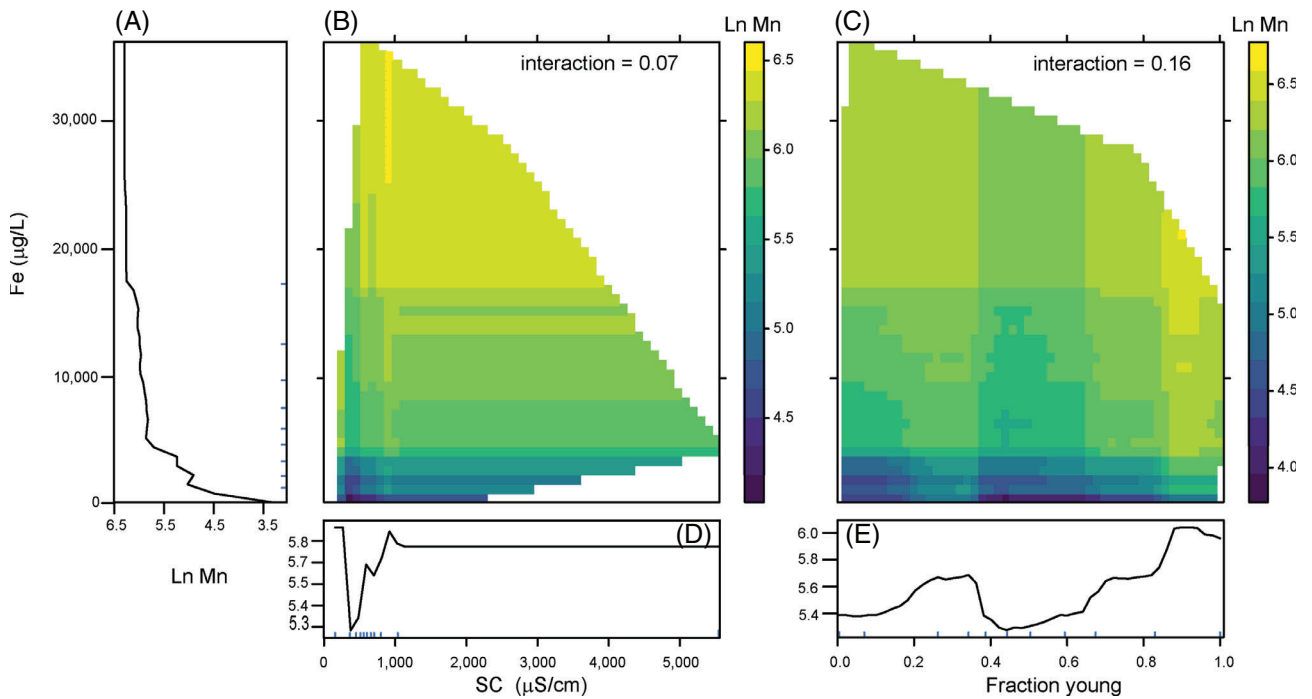


Figure 7. Partial dependence plots showing the variation in predicted natural log of Mn with (A) Fe, (B) Fe and SC, (C) Fe and fraction of young groundwater, (D) SC, and (E) fraction of young groundwater.

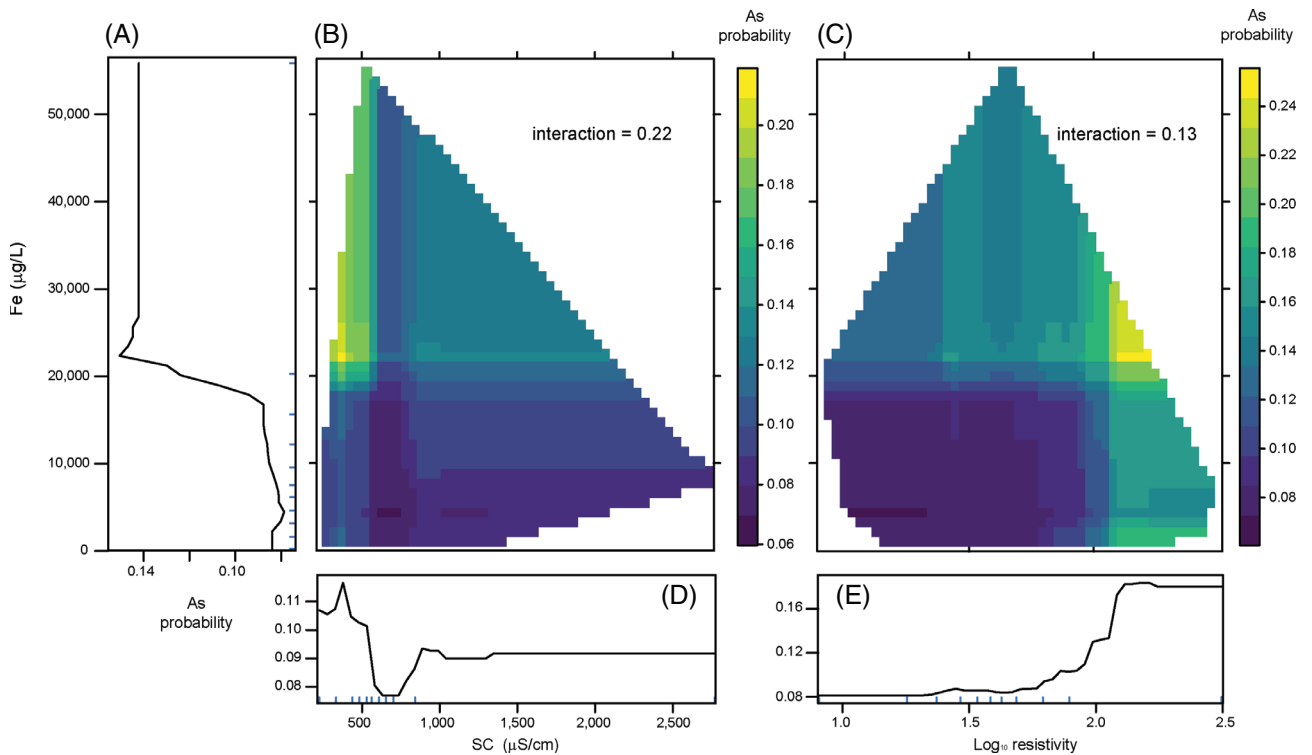


Figure 8. Partial dependence plots showing the variation in predicted probability of As greater than 10 µg/L with (A) Fe, (B) Fe and SC, (C) Fe and average aquifer resistivity, (D) SC, and (E) average aquifer resistivity.

The predicted probability of As exceeding 10 $\mu\text{g/L}$ was also influenced by SC and Fe. High As was predicted where mapped SC was approximately 400 $\mu\text{S/cm}$ and Fe more than 20,000 $\mu\text{g/L}$ (Figure 8B). Arsenic concentrations increase as MRVA groundwater becomes more reducing, until total dissolved solids reach approximately 400 mg/L, and then As decreases thereafter due to precipitation of minerals that sequester As (Kresse and Fazio 2003). This general pattern is observed in the partial-dependence plot, with high As predicted approximately 400 $\mu\text{S/cm}$ (Figure 8D). Additionally, there is an increase in arsenic as Fe increases from 15,000 to 22,000 $\mu\text{g/L}$ (Figure 8A). Arsenic, like Mn, can be released from Fe and Mn oxide coatings under reducing conditions (Sharif et al. 2008). Thus, the partial-dependence plots show that SC and Fe affect model predictions consistent with our understanding of geochemistry.

Partial Dependence and Thematic Sensitivity Analysis Illustrate How Groundwater Model Variables Influence Mn Predictions

Groundwater model characteristics included nine variables (Table S1), and variables estimating groundwater age were influential predictors of Mn (Figure 6). Fraction of young groundwater varied systematically across the MRVA; from less than 0.1 along the Mississippi River, along major streams within the Delta region, and within the Grand Prairie region; to more than 0.75 throughout the western parts of the MRVA (Figure 3C). Groundwater flow generally originates at the edges of the MRVA and flows toward cones of depression at pumping centers and toward major rivers (Haugh et al. 2020a); these areas subsequently have a lower fraction of young water.

Fraction of young groundwater influenced Mn predictions, but not in a manner that would be expected. Manganese is generally expected to increase with estimated groundwater age; however, Mn increased with fraction of young groundwater (Figure 7E). The inverse relationship occurs because other surrogates of groundwater age, particularly mapped Fe, were more important for predicting Mn (Figure 6). Iron was predicted using a BRT model that included groundwater age predictor variables (Knierim et al. 2020c, 2021b) and Mn increased with Fe (Figure 7A). The partial-dependence plot for Fe shows that natural log Mn varied from 3.2 to 6.3 with changes in Fe (Figure 7A), whereas natural log Mn varied from 5.3 to 6.0 with fraction of young groundwater (Figure 7E); fraction young water modifies the influence of Fe (Figure 7C). These results show that one cannot always count on a machine-learning model to reproduce the expected relationship for a given variable because other explanatory variables can capture similar underlying processes. In this particular case, Fe integrates groundwater age as well as other factors, such as availability of Mn in the aquifer matrix and electron acceptors in the water. Consequently, Fe is more influential than fraction of young water for predicting Mn.

Given that the relationship between Mn and fraction of young groundwater was inverse to expectation, the thematic sensitivity of Mn predictions to the groundwater model predictor variables was evaluated. An alternative model was trained without nine groundwater model characteristics, and predictions from the alternative model were compared to those from the original model. Withholding groundwater characteristics from the BRT model is equivalent to assuming that groundwater flow and age are the same throughout the aquifer. Manganese residuals (Figure 3D) were generally positive—underpredicting Mn—where fraction of young groundwater was high (Figure S4). Therefore, the groundwater model variables had a systematic influence on Mn predictions, despite the unexpected relationship between groundwater age and Mn. The systematic pattern of the residuals illustrates that the information contained in the groundwater model variables was not captured by other variables in the BRT model.

Partial Dependence and Thematic Sensitivity Analysis Illustrate How AEM-Derived Resistivity Influence As Predictions

Geophysical survey derived predictor variables included in the BRT model were average bulk resistivity, soil radiochemistry, and an interpretation of the degree of surficial confinement (Table S1). Resistivity is a surrogate for texture and was the sixth most influential variable for predicting As (Figure 6). Near-surface radiochemistry is a surrogate for soil mineralogy and was less influential than resistivity (Figure 6). The degree of surficial confinement was interpreted from resistivity and ranked least influential (Figure 6). Consequently, the following discussion focuses on resistivity and its influence on predicting As.

The predicted probability of As exceeding 10 $\mu\text{g/L}$ increased with resistivity (Figure 8E). Two localized areas of elevated As probability were predicted in the Cache region of Arkansas (Figure 5B): one with observed high values (southern Cache) and one without (northern Cache along northwestern margin of MRVA, Figure 5A). The BRT model integrated these observations and assigned a high probability to areas where \log_{10} -resistivity >1.9 (Figure 5C). Very high resistivity values occur where sediments of the Mississippi embayment or Ozark Plateaus aquifer systems subcrop the MRVA. Predictions of As events in areas of very high resistivity—especially in the northern Cache area along the northwestern margins of the study area where the MRVA is thin—may be spurious.

High As probabilities were also predicted where \log_{10} -resistivity values are between 1.6 and 1.8 (Figure 5B). Such resistivity occurs near streams in the Boeuf region of central Arkansas and along sections of the Mississippi River (Figure 5C). \log_{10} -resistivity >1.4 in the study area is associated with coarse-grained sediments such as sands and gravels (Minsley et al. 2021). Therefore, the high resistivities near streams are interpreted as coarse-grained materials associated with natural levee and point bar deposits. Previous research has found

increased As near rivers, driven by variation in sediment texture and organic carbon (Yang et al. 2014; Nghiem et al. 2019).

High As probabilities were predicted near streams, associated with high resistivity values. Other variables also captured position relative to streams, including MOHP-3 and groundwater age estimates. High As probability was predicted where values of MOHP-3 lateral position were low, or close to streams. Approximately 80% of correctly predicted As events were located where fraction of young groundwater less than 0.1, which occurs predominantly along the Mississippi River (Figure 3C). Arsenic-exceedance probability was predicted to vary from 0.08 to 0.14 with changing fraction young. Comparatively, As probability was predicted to vary from 0.08 to 0.18 with changes in resistivity (Figure 8E).

Because resistivity was predicted to affect As over a larger probability range compared to other variables, the thematic sensitivity of the BRT model to geophysical predictor variables was evaluated. Predictions were compared to an alternative model trained without three variables derived from the geophysical survey. Arsenic residuals (Figure 5D) were generally positive—As was underpredicted—where resistivity was high (Figure S5), which is associated with more resistive or coarse-grained sediments. Withholding geophysical variables from BRT model training is equivalent to assuming there is no variability resistivity—or sediment texture—throughout the MRVA. Residuals are small likely because other variables capture differences in texture across the system; however, the systematic variation in residuals highlights that the AEM-derived resistivity adds unique textural information to the model.

Conclusions

BRT models were used to predict Mn concentration and the probability of As exceeding a 10 $\mu\text{g/L}$ threshold across the MRVA. The models were trained with predictor variables representing geochemical, groundwater model (input and output), geophysical, and hydrologic characteristics. Iron and SC obtained from previously developed BRT models were important predictors of Mn and As; Fe and SC are surrogates for redox and/or flowpath length. Fraction of young groundwater, derived from a MODFLOW groundwater-flow model, was the most important variable for predicting As. Fraction of young groundwater was also important for predicting Mn because it modified the influence of Fe. Variables with unique signatures near streams—such as AEM-derived resistivity, fraction of young groundwater, and MOHP-3—were important predictors of As.

The thematic sensitivity of predicted Mn and As was assessed by developing alternative models trained without groundwater-flow model and geophysical variables, respectively. In turn, the results of the alternative models were subtracted from the original model. Manganese was underpredicted by the alternative model where fraction of young groundwater was high. Arsenic was underpredicted

by the alternative model where resistivity was high. The systematic variation in alternative model residuals shows that other variables were unable to fully represent the information contained in the excluded variables. Thus, thematic sensitivity analysis confirmed the importance of including variables that describe groundwater flow and aquifer texture.

The maps of predicted Mn concentration and As-exceedance probabilities are useful for identifying regional patterns, but are not intended for site-specific predictions. Model results help inform where potentially high Mn or As warrant further investigation. For example, rice is grown in the MRVA and As bioaccumulates in rice (Zavala and Duxbury 2008). Additionally, high concentrations of Mn can cause fouling of irrigation equipment (Haman 2017) and require treatment for drinking water.

BRT models were developed using 19 predictor variables, selected for their inferred influence on Mn and As in groundwater. The use of this curated set of process-informed variables increased model interpretability. Although BRT models are empirical, this paper provides insight into factors controlling Mn and As distribution in groundwater. The BRT approach described here, and in other studies, is applicable for understanding groundwater quality and could be applied to understanding other aspects of groundwater systems. The BRT method requires a sufficient number of observations and associated predictor variables. To develop maps of predictions in groundwater aquifers, predictor variables need to be available on a wall-to-wall and top-to-bottom basis.

Acknowledgments

The authors developed these models as part of a larger team at the U.S. Geological Survey National Water Quality Program and appreciate the thoughtful insight and guidance from colleagues. Review comments from Connor Haugh, Peter McMahon, Peter Knappett, anonymous reviewers, and associate editors strengthened this manuscript. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Authors' Note

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. The authors do not have any conflicts of interest or financial disclosures to report.

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.

Table S1. Predictor variables used in manganese and arsenic boosted regression tree models.

Figure S1. Flowchart representing the main steps of the modeling and prediction workflow. Numbers refer to steps in the Methods section. See Methods section for more details.

Figure S2. Empirical distributions of manganese observations, holdout predictions, holdout predictions bias corrected with an Empirical Distribution Matching (EDM) method, and predictions at all raster cells.

Figure S3. Observed vs. predicted Mn concentration transformed from natural log space for predictions at wells and predictions at coincident raster cells where wells are located. The dataset was filtered for wells with depths similar to the depth of prediction at the coincident raster cell where the well is located.

Figure S4. Observed As concentrations vs. predicted probability of As concentration exceeding a 10 $\mu\text{g/L}$ threshold for predictions at wells and predictions at coincident raster cells where wells are located. The dataset was filtered for wells with depths similar to the depth of prediction at the coincident raster cell where the well is located.

Figure S5. Predicted Mn at raster cells from the original BRT model vs. the alternative model where all groundwater-flow model variables were withheld, compared to the fraction of young groundwater (less than 65 years) at coincident raster cells.

Figure S6. The predicted probability of As exceeding a concentration of 10 $\mu\text{g/L}$ at raster cells from the original BRT model vs. the alternative model where all geophysical survey derived variables were withheld, compared to MRVA resistivity at coincident raster cells.

References

- Alhassan, M., C. B. Lawrence, S. Richardson, and E. J. Pindilli. 2019. The Mississippi Alluvial Plain aquifers—An engine for economic activity. U.S. Geological Survey Fact Sheet 2019–3003. <https://doi.org/10.3133/fs20193003>.
- Amini, M., K.C. Abbaspour, M. Berg, L. Winkel, S.J. Hug, E. Hoehn, H. Yang, and C.A. Johnson. 2008. Statistical modeling of global geogenic arsenic contamination in groundwater. *Environmental Science & Technology* 42, no. 10: 3669–3675. <https://doi.org/10.1021/es702859e>
- Ayotte, J.D., L. Medalie, S.L. Qi, L.C. Backer, and B.T. Nolan. 2017. Estimating the high-arsenic domestic-well population in the conterminous United States. *Environmental Science & Technology* 51, no. 21: 12443–12454. <https://doi.org/10.1021/acs.est.7b02881>
- Aziz, Z., A. van Geen, M. Stute, R. Versteeg, A. Horne-man, Y. Zheng, S. Goodbred, M. Steckler, B. Weinman, I. Gavrieli, M.A. Hoque, M. Shamsudduha, and K.M. Ahmed. 2008. Impact of local recharge on arsenic concentrations in shallow aquifers inferred from the electro-magnetic conductivity of soils in Araihasar, Bangladesh. *Water Resources Research* 44, no. 7: 1–15. <https://doi.org/10.1029/2007WR006000>
- Bakker, M., V. Post, C. D. Langevin, J. D. Hughes, J. T. White, J. J. Starn, and M. N. Fiennen. 2021. FloPy v3.3.4—Release candidate. U.S. Geological Survey Software Release. <https://doi.org/10.5066/F7BK19FH>.
- Bakker, M., V. Post, C.D. Langevin, J.D. Hughes, J.T. White, J.J. Starn, and M.N. Fiennen. 2016. Scripting MODFLOW model development using python and FloPy. *Groundwater* 54, no. 5: 733–739. <https://doi.org/10.1111/gwat.12413>
- Belitz, K., and P.E. Stackelberg. 2021. Evaluation of six methods for correcting bias in estimates from ensemble tree machine learning regression models. *Environmental Modelling & Software*, 139, no. 2: 105006. <https://doi.org/10.1016/j.envsoft.2021.105006>
- Belitz, K., R.B. Moore, T.L. Arnold, J.B. Sharpe, and J.J. Starn. 2019. Multi-order hydrologic position in the conterminous United States: A set of metrics in support of groundwater mapping at regional and national scales. *Water Resources Research* 55: 11188–11207. <https://doi.org/10.1029/2019WR025908>
- Borrok, D.M., R.M. Lenz, J.E. Jennings, M.L. Gentry, J. Steensma, and D.S. Vinson. 2018. The origins of high concentrations of iron, sodium, bicarbonate, and arsenic in the Lower Mississippi River Alluvial Aquifer. *Applied Geochemistry* 98, no. 11: 383–392. <https://doi.org/10.1016/j.apgeochem.2018.10.014>
- Chapelle, F.H., P.B. McMahon, N.M. Dubrovsky, R.F. Fujii, E.T. Oaksford, and D.A. Vroblesky. 1995. Deducing the distribution of terminal electron-accepting processes in hydrologically diverse groundwater systems. *Water Resources Research* 31, no. 2: 359–371. <https://doi.org/10.1029/94WR02525>
- Clark, B. R., K. J. Knierim, and L. Duncan. 2019. Zoney: Zonal analysis tools for raster and vector datasets (version 0.0). <https://github.com/brclark-usgs/zoney>. (accessed January 5 2021).
- Clark, B. R., P. M. Barlow, S. M. Peterson, J. D. Hughes, H. W. Reeves, and R. Viger. 2018. National-scale grid to support regional groundwater availability studies and a national hydrogeologic framework. U.S. Geological Survey Data Release. <https://doi.org/10.5066/f7p84b24>.
- Clark, B. R., and R. M. Hart. 2009. The Mississippi Embayment Regional Aquifer Study (MERAS): Documentation of a groundwater-flow model constructed to assess water availability in the Mississippi Embayment. U.S. Geological Survey Scientific Investigations Report 2009–5172. <https://doi.org/10.3133/sir20095172>.
- Elith, J., J.R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77, no. 4: 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>
- Friedman, J.H., and B.E. Popescu. 2008. Predictive learning via rule ensembles. *The Annals of Applied Statistics* 2, no. 3: 916–954. <https://doi.org/10.1214/07-AOAS148>
- Friedman, J.H. 2001. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics* 29, no. 5: 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- GDAL/OGR contributors. 2019. GDAL/OGR geospatial data abstraction software library. Open Source Geospatial Foundation. (version 2.4). <https://gdal.org>. (accessed September 18, 2020).
- Ghosh, G.C., M.J.H. Khan, T.K. Chakraborty, S. Zaman, A.H.M.E. Kabir, and H. Tanaka. 2020. Human health risk assessment of elevated and variable iron and manganese intake with arsenic-safe groundwater in Jashore, Bangladesh. *Scientific Reports* 10, no. 1: 5206. <https://doi.org/10.1038/s41598-020-62187-5>
- Greenwell, B.M. 2017. Pdp: An R package for constructing partial dependence plots. *The R Journal* 9, no. 1: 421–436.
- Haman, D., 2017, *Causes and Prevention of Emitter Plugging in Microirrigation Systems*, University of Florida Institute of Food and Agricultural Sciences Extension BUL258, Department of Agricultural and Biological Engineering, Gainesville, FL: University of Florida, Institute of Food and Agricultural Sciences, <https://edis.ifas.ufl.edu/publication/AE032>. (accessed June 1, 2021).

- Haugh, C. J., C. D. Killian, and J. R. B. Barlow. 2020a. MODFLOW-2005 model used to evaluate water-management scenarios for the Mississippi Delta. U.S. Geological Survey Data Release. <https://doi.org/10.5066/P9906VM5>.
- Haugh, C. J., C. D. Killian, and J. R. B. Barlow. 2020b. Simulation of water-management scenarios for the Mississippi Delta. U.S. Geological Survey Scientific Investigations Report 2019–5116. <https://doi.org/10.3133/sir20195116>.
- Haugh, C. J., K. J. Knierim, and J. A. Kingsbury. 2020c. Simulated groundwater residence times in two principal aquifers of the Mississippi embayment physiographic region. U.S. Geological Survey Data release. <https://doi.org/10.5066/P9X513WO>.
- Helsel, D.R. and R. M. Hirsch, 2002. Statistical methods in water resources techniques of water resources investigations, Book 4, chapter A3. U.S. Geological Survey. 522 pages. <https://doi.org/10.3133/twri04A3>.
- James, S. R., and B. J. Minsley. 2021. Combined results and derivative products of hydrogeologic structure and properties from airborne electromagnetic surveys in the Mississippi Alluvial Plain. U.S. Geological Survey Data Release. <https://doi.org/10.5066/P9382RCI>.
- Johnson, T.D., and K. Belitz. 2009. Assigning land use to supply wells for the statistical characterization of regional groundwater quality: Correlating urban land use and VOC occurrence. *Journal of Hydrology* 370, no. 1–4: 100–108. <https://doi.org/10.1016/j.jhydrol.2009.02.056>
- Kaufman, S., S. Rosset, C. Perlich, and O. Stitelman. 2012. Leakage in data mining: Formulation, detection, and avoidance. *ACM Transactions on Knowledge Discovery from Data* 6, no. 4: 1–21. <https://doi.org/10.1145/2382577.2382579>
- Kingsbury, J. A., K. J. Knierim, and C. J. Haugh. 2020a. Predicted pH of groundwater in the Mississippi River Valley alluvial and Claiborne aquifers, south-central United States. U.S. Geological Survey Scientific Investigations Map 3465. <https://doi.org/10.3133/sim3465>.
- Kingsbury, J. A., K. J. Knierim, and C. J. Haugh. 2020b. Prediction grids of pH for the Mississippi River Valley Alluvial and Claiborne Aquifers. U.S. Geological Survey Data release. <https://doi.org/10.5066/P9CXX7LN>.
- Kingsbury, J. A., J. R. Barlow, B. G. Katz, H. L. Welch, R. W. Tollett, and L. S. Fahlquist. 2015. The quality of our Nation's waters: Water quality in the Mississippi embayment-Texas coastal uplands aquifer system and Mississippi River Valley alluvial aquifer, south-central United States, 1994–2008. U.S. Geological Survey Circular 1356. <https://doi.org/10.3133/cir1356>.
- Knierim, K. J., J. A. Kingsbury, and C. J. Haugh. 2020a. Machine-learning model predictions and groundwater-quality rasters of specific conductance, total dissolved solids, and chloride in aquifers of the Mississippi embayment. U.S. Geological Survey Data Release. <https://doi.org/10.5066/P9WBFR1T>.
- Knierim, K.J., J.A. Kingsbury, C.J. Haugh, and K.M. Ransom. 2020b. Using boosted regression tree models to predict salinity in Mississippi embayment aquifers, Central United States. *Journal of the American Water Resources Association* 56, no. 6: 1010–1029. <https://doi.org/10.1111/1752-1688.12879>
- Knierim, K. J., J. A. Kingsbury, and C. J. Haugh. 2020c. Machine-learning model predictions and rasters of dissolved oxygen probability, iron concentration, and redox conditions in groundwater in the Mississippi River Valley alluvial and Claiborne aquifers. U.S. Geological Survey Data release. <https://doi.org/10.5066/P9N108JM>.
- Knierim, K. J., J. A. Kingsbury and K.M. Ransom. 2021a. Machine-learning model predictions and rasters of arsenic and manganese in groundwater in the Mississippi River Valley alluvial aquifer. U.S. Geological Survey Data Release. <https://doi.org/10.5066/P9PRLNA3>.
- Knierim, K. J., J. A. Kingsbury, and C. J. Haugh. 2021b. Machine-learning predictions of redox conditions in groundwater in the Mississippi River Valley alluvial and Claiborne aquifers, south-central United States. U.S. Geological Survey Scientific Investigations Map 3468. <https://doi.org/10.3133/sim3468>.
- Knierim, K., J. A. Kingsbury, and C. J. Haugh. 2019. Groundwater withdrawal zones for drinking water from the Mississippi River Valley alluvial aquifer and Mississippi embayment aquifers. U.S. Geological Survey Data release. <https://doi.org/10.5066/P9FZV4ED>.
- Kousa, A., H. Komulainen, T. Hatakka, B. Backman, and S. Hartikainen. 2020. Variation in groundwater manganese in Finland. *Environmental Geochemistry and Health* 43, no. 7: 1193–1211. <https://doi.org/10.1007/s10653-020-00643-x>
- Kresse, T., and J. Fazio. 2003. Occurrence of arsenic in ground waters of Arkansas and implications for source and release mechanisms. Arkansas Department of Environmental Quality, Water Quality Report WQ03-03-01. <https://www.adeq.state.ar.us/water/planning/pdfs/publications/WQ03-03-1.pdf>. (accessed May 30, 2018).
- 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. Scrucra, Y. Tang, C. Candan, and T. Hunt. 2019. caret: Classification and Regression Training (version 6.0-85), <https://cran.r-project.org/web/packages/caret/index.html>. (accessed September 20, 2020).
- Kuhn, M., and K. Johnson. 2016. *Applied Predictive Modeling*. New York, New York: Springer. <https://doi.org/10.1007/978-1-4614-6849-3>
- Lee, L. 2020. NADA package (version 1.6). <https://CRAN.R-project.org/package=NADA>. (accessed August 20, 2020).
- Lombard, M.A., M.S. Bryan, D.K. Jones, C. Bulka, P.M. Bradley, L.C. Backer, M.J. Focazio, D.T. Silverman, P. Toccalino, M. Argos, M.O. Gribble, and J.D. Ayotte. 2021. Machine learning models of arsenic in private wells throughout the conterminous United States as a tool for exposure assessment in human health studies. *Environmental Science & Technology* 55, no. 8: 5012–5023. <https://doi.org/10.1021/acs.est.0c05239>
- McGuire, V. L., R. C. Seanor, W. H. Asquith, A. M. Nottmeier, S. C. David, R. W. Tollett, W. H. Kress, and K. R. Strauch. 2020. Datasets used to map the potentiometric surface, Mississippi River Valley alluvial aquifer, spring 2018. U.S. Geological Survey Data Release. <https://doi.org/10.5066/P992HD1R>.
- 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>
- Minsley, B.J., J.R. Rigby, S.R. James, B.L. Burton, K.J. Knierim, M.D.M. Pace, P.A. Bedrosian, and W.H. Kress. 2021. Airborne geophysical surveys of the lower Mississippi Valley demonstrate system-scale mapping of subsurface architecture. *Communications Earth & Environment* 2, no. 1: 131. <https://doi.org/10.1038/s43247-021-00200-z>
- Molnar, C. 2019. Interpretable machine learning. <https://christophm.github.io/interpretable-ml-book/>. (accessed April 15, 2019).
- Moore, R., K. Belitz, T. L. Arnold, J. B. Sharpe, and J. J. Starn. 2019. National Multi Order Hydrologic Position (MOHP) predictor data for groundwater and groundwater-quality modeling. U.S. Geological Survey Data release. <https://doi.org/10.5066/P9HLU4YY>.
- 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, no. 3: 902–911. <https://doi.org/10.1016/j.jhydrol.2015.10.025>
- Nghiem, A.A., M.O. Stahl, B.J. Mailloux, T.T. Mai, P.T. Trang, P.H. Viet, C.F. Harvey, A. Geen, and B.C. Bostick. 2019. Quantifying riverine recharge impacts on redox conditions and arsenic release in groundwater aquifers along the Red River, Vietnam. *Water Resources Research* 55, no. 8: 6712–6728. <https://doi.org/10.1029/2019WR024816>
- Podgorski, J., and M. Berg. 2020. Global threat of arsenic in groundwater. *Science* 368, no. 6493: 845–850. <https://doi.org/10.1126/science.aba1510>
- Pollock, D. W. 2012. User guide for MODPATH version 6—A particle-tracking model for MODFLOW. U.S. Geological Survey Techniques and Methods 6–A41. <https://doi.org/10.3133/tm6A41>
- Palacky, G.J. 1987. Resistivity characteristics of geologic targets. In *Electromagnetic Methods in Applied Geophysics: Volume 1, Theory*, ed. M.N. Nabighian, 53–129. Tulsa: Society of Exploration Geophysicists.
- Python Software Foundation. 2019. Python Language Reference, v 3.7. Python.Org, 2019, <https://www.python.org>. (accessed June 1 2019).
- R Core Team. 2019. *R: A Language and Environment for Statistical Computing*, Vol. v. 4.0. Vienna, Austria: R Foundation for Statistical Computing. <https://www.r-project.org/>. (accessed June 20 2019)
- Ramachandran, M., K.A. Schwabe, and S.C. Ying. 2021. Shallow groundwater manganese merits deeper consideration. *Environmental Science & Technology* 55, no. 6: 3465–3466. <https://doi.org/10.1021/acs.est.0c08065>
- Ransom, K.M., B.T. Nolan, P.E. Stackelberg, K. Belitz, and M.S. Fram. 2021. Machine learning predictions of nitrate in groundwater used for drinking supply in the conterminous United States. *Science of the Total Environment* 807, no. 3: 151065. <https://doi.org/10.1016/j.scitotenv.2021.151065>
- Reitz, M., W.E. Sanford, G.B. Senay, and J. Cazenias. 2017. Annual estimates of recharge, quick-flow runoff, and evapotranspiration for the contiguous U.S. using empirical regression equations. *JAWRA Journal of the American Water Resources Association* 53, no. 4: 961–983. <https://doi.org/10.1111/1752-1688.12546>
- Renken, R. A. 1998. Ground Water Atlas of the United States: Segment 5, Arkansas, Louisiana, Mississippi. U.S. Geological Survey Hydrologic Atlas 730-F. 10.3133/ha730F.
- Ridgeway, G. 2019. gbm: Generalized boosted regression models (version 2.1.5). <https://cran.r-project.org/web/packages/gbm/gbm.pdf>. (accessed September 15, 2020).
- Sharif, M.U., R.K. Davis, K.F. Steele, B. Kim, P.D. Hays, T.M. Kresse, and J.A. Fazio. 2008. Distribution and variability of redox zones controlling spatial variability of arsenic in the Mississippi River Valley alluvial aquifer, southeastern Arkansas. *Journal of Contaminant Hydrology* 99, no. 1: 49–67. <https://doi.org/10.1016/j.jconhyd.2008.03.001>
- Smedley, P.L., and D.G. Kinniburgh. 2002. A review of the source, behaviour and distribution of arsenic in natural waters. *Applied Geochemistry* 17, no. 5: 517–568. [https://doi.org/10.1016/S0883-2927\(02\)00018-5](https://doi.org/10.1016/S0883-2927(02)00018-5)
- Stackelberg, P.E., K. Belitz, C.J. Brown, M.L. Erickson, S.M. Elliott, L.J. Kauffman, K.M. Ransom, and J.E. Reddy. 2020. Machine learning predictions of pH in the glacial aquifer system, Northern USA. *Groundwater* 59, no. 3: 352–368. <https://doi.org/10.1111/gwat.13063>
- Starn, J.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>
- Stute, M., Y. Zheng, P. Schlosser, A. Horneman, R.K. Dhar, S. Datta, M.A. Hoque, A.A. Seddique, M. Shamsudduha, K.M. Ahmed, and A. van Green. 2007. Hydrological control of as concentrations in Bangladesh groundwater. *Water Resources Research* 43, no. 9: 1–11. <https://doi.org/10.1029/2005WR004499>
- Toccalino, P. L., J. E. Norman, and K. M. Schoephoester. 2018. Health-based screening levels for evaluating water-quality data. U.S. Geological Survey, <https://doi.org/10.5066/F71C1TWP>.
- Torak, L. J., and J. A. Painter. 2019a. Digital surfaces of the bottom altitude and thickness of the Mississippi River Valley alluvial aquifer and site data within the Mississippi Alluvial Plain project region. U.S. Geological Survey Data Release. <https://doi.org/10.5066/P9D9XR5F>.
- Torak, L. J., and J. A. Painter. 2019b. Geostatistical estimation of the bottom altitude and thickness of the Mississippi River Valley alluvial aquifer. U.S. Geological Survey Scientific Investigations Map 3426. <https://doi.org/10.3133/sim3426>.
- U.S. Environmental Protection Agency. 2013. Safe Drinking Water Information System (SDWIS). Data and Tools, 2013. <https://www.epa.gov/enviro/sdwis-search>.
- U.S. Environmental Protection Agency. 2020a. National Primary Drinking Water Regulations. Overviews and Factsheets. <https://www.epa.gov/ground-water-and-drinking-water/national-primary-drinking-water-regulations>. (accessed December 12, 2020).
- U.S. Environmental Protection Agency. 2020b. Secondary Drinking Water Standards: Guidance for nuisance chemicals. 2020. <https://www.epa.gov/sdwa/secondary-drinking-water-standards-guidance- nuisance-chemicals#table>. (accessed December 12, 2020).
- U.S. Geological Survey. 2019. USGS water data for the Nation: U.S. Geological Survey National Water Information System database, 2019. <https://doi.org/10.5066/F7P55KJN>.
- 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>
- World Health Organization. 2011. *Guidelines for Drinking-Water Quality*, 4th ed. WHO Press: Geneva, Switzerland. https://apps.who.int/iris/bitstream/handle/10665/44584/9789241548151_eng.pdf?sequence=1 (accessed January 2, 2020)
- Yang, N., L.H.E. Winkel, and K.H. Johannesson. 2014. Predicting geogenic arsenic contamination in shallow groundwater of South Louisiana, United States. *Environmental Science & Technology* 48, no. 10: 5660–5666. <https://doi.org/10.1021/es405670g>
- Zavala, Y.J., and J.M. Duxbury. 2008. Arsenic in rice: I. Estimating normal levels of total arsenic in rice grain. *Environmental Science & Technology* 42, no. 10: 3856–3860. <https://doi.org/10.1021/es702747y>
- 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>