Evaluating Sources of Arsenic in Groundwater in Virginia using a Logistic Regression Model

Tiffany Jebson VanDerwerker

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
In
Geosciences

Madeline Schreiber

Donald Rimstidt

Mark Widdowson

April 29, 2016

Blacksburg, Virginia

Keywords: Arsenic, Groundwater, Aquifer Vulnerability, Logistic Regression

Evaluating Sources of Arsenic in Groundwater in Virginia using a Logistic Regression Model

Tiffany Jebson VanDerwerker

Abstract (Academic)

For this study, I have constructed a logistic regression model, using existing datasets of environmental parameters to predict the probability of As concentrations above 5 parts per billion (ppb) in Virginia groundwater and to evaluate if geologic or other characteristics are linked to elevated As concentrations. Measured As concentrations in groundwater from the Virginia Tech Biological Systems Engineering (BSE) Household Water Quality dataset were used as the dependent variable to train (calibrate) the model. Geologic units, lithology, soil series and texture, land use, and physiographic province were used as regressors in the model. Initial models included all regressors, but during model refinement, attention was focused solely on geologic units. Two geologic units, Triassic-aged sedimentary rocks and Devonian-aged shales/sandstones, were identified as significant in the model; the presence of these units at a spatial location results in a higher probability for As occurrences in groundwater. Measured As concentrations in groundwater from an independent dataset collected by the Virginia Department of Health were used to test (validate) the model. Due to the structure of the As datasets, which included As concentrations mostly (95-99%) ≤ 5 ppb, and thus few (1-5%) data in the range > 5 ppb, the regression model cannot be used reliably to predict As concentrations in other parts of the state. However, our results are useful for identifying areas of Virginia, defined by underlying geology, that are more likely to have elevated As concentrations in groundwater. Results of this work suggest that homeowners with wells installed in these geologic units have their wells tested for As and regulators closely monitor public supply wells in these areas for As.

Abstract (Public)

Arsenic (As) is a naturally-occurring toxin and carcinogen, associated with cardiovascular, pulmonary, immunological, and neurological disease as well as skin, lung and bladder cancer. One of the dominant routes of exposure to As is through drinking water. In many regions worldwide, including notable examples in Southeast Asia, tens of millions of people are drinking groundwater with elevated As concentrations. In Virginia, more than 1/3 of counties depend on groundwater for their water supply. Homeowners are responsible for having their wells tested but they may not be aware of what water quality parameters should be analyzed. Because As is naturally occurring, exists in many types of aquifers, is odorless and colorless, and generally does not cause immediate illness, it is challenging to evaluate if As is a concern in groundwater supplies.

For this study, I have constructed a logistic regression model, using existing datasets of environmental parameters to predict the probability of As concentrations above 5 parts per billion (ppb) in Virginia groundwater and to evaluate if geologic or other environmental characteristics are linked to elevated As concentrations. Measured As concentrations in groundwater from the Virginia Tech Biological Systems Engineering (BSE) Household Water Quality dataset were used as the dependent variable to train (calibrate) the model. Geologic units, lithology, soil series and texture, land use, and physiographic province were used as regressors (parameters) in the model. Initial models included all regressors, but during model refinement, attention was focused solely on geologic units. Two geologic units, Triassic-aged sedimentary rocks and Devonian-aged shales/sandstones, were identified as significant in the model; the presence of these units at a spatial location results in a higher probability for As occurrences in groundwater. Measured As concentrations in groundwater from an independent dataset collected by the Virginia Department of Health were used to test (validate) the model. Due to the structure of the As datasets, which included As concentrations mostly (95-99%) $\leq$ 5 ppb, and thus few (1-5%) data in the range > 5 ppb, the regression model cannot be used reliably to predict As concentrations in other parts of the state. However, our results are useful for identifying areas of Virginia, defined by underlying geology, that are more likely to have elevated As concentrations in groundwater. Results of this work suggest that homeowners with wells installed in these geologic units have their wells tested for As and regulators closely monitor public supply wells in these areas for As.

Acknowledgments

A huge thank you to my adviser, Madeline Schreiber, who has guided me through these past two years. You have been the key to my success during my graduate studies. Thank you to Lin Zhang who was my statistical guru and co-collaborator. Thank you to my committee members, Don Rimstidt and Mark Widdowson, both of whom were fabulous teachers and great committee members. Thank you to Scott Bruce, Zack Oremland, and Randy McFarland, for their assistance in identifying aquifers and their geologic and geochemical interpretations. Thank you to Erin Ling, Brian Benham, and Asa Spiller of the VT BSE Department for access to their dataset and for asking the really difficult questions. Thank you to Maya Whitaker and John Aulbach for their access to the VDH Dataset.

Thank you to the Virginia Tech Department of Geosciences and the Geological Society of America for their funding of this project.

Thank you to my husband for listening to my troubles and for letting me pursue my Master's Degree, as my professional rap career didn't pan out. Thank you to coffee for making me productive. Thank you to my parents who have always supported me in all my endeavors.

Table of Contents

List of Abbreviations

As – Arsenic

BSE – Biological Systems Engineering

DEQ – Virginia Department of Environmental Quality

EPA – Environmental Protection Agency

GIS – Geographic Information System

LASSO - Least Absolute Shrinkage and Selection Operator

MCL – Maximum Contaminant Level

MSMA -Monosodium methanearsonate

NLCD2011 – National Land Cover Database 2011

ppb – Parts per Billion

ppm – Parts per Million

STORET – EPA Storage and Retrieval database

SSURGO - Soil Survey Geographic Database

USGS – United States Geological Survey

VDH – Virginia Department of Health

VT IRB – Virginia Tech Institutional Review Board

Chapter 1: Introduction

Elevated arsenic (As) concentrations in groundwater occur worldwide, with notable examples in Southeast Asia, including regions in Bangladesh, India, China, Cambodia, and Vietnam (Bhattacharya et al., 2007). In the U.S., As concentrations in groundwater are also elevated in some areas, with affected regions including New England, the upper Midwest, the Plains, and the West, among others (Welch et al., 2000). Arsenic is of particular concern as it is odorless, colorless, and tasteless; ingestion of As through drinking water is one of the most common exposure pathways (Singh et al., 2007). Worldwide, it is estimated that up to 140 million people drink water with unsafe levels of As (Michael, 2013).

Health Impacts of Arsenic Exposure

Arsenic is a known toxin and carcinogen (Lan et al., 2011; NRC 2001; Smith et al., 2000; Fajčíková et al., 2014). Adverse health effects have been associated with ingestion of As concentrations above 10 parts per billion (ppb), which is the current Environmental Protection Agency Maximum Contaminant Level (EPA MCL) and drinking water standard for the World Health Organization. However, more recent studies indicate that adverse health effects may also be caused by ingesting As concentrations lower than 10 ppb (Shiber, 2005; Gong and O'Bryant, 2012). For example, the lifetime risk of bladder and lung cancer from aqueous As exposure of 10 ppb is about one in 333 individuals (NRC, 2001). This is nearly 30 times the standard of one in 10,000 individuals set for other carcinogens. At 5 ppb, the risk is 1 in 667 (Shiber, 2005). Because of the carcinogenic nature of As, New Jersey has initiated a statewide MCL for As at 5 ppb, the lowest in the country (Barringer et al. 2010). New Jersey now requires over 600 public community water systems and 900 non-transient, non-community water systems in New Jersey to monitor periodically for As and to comply with the revised MCL (New Jersey Department of Health, 2004).

In addition to cancer, exposure to As is also linked with heart disease (Gong and O'Bryant, 2012; Farzan et al., 2015). Low levels of As may also be associated with complications during pregnancy and the quality of life in children exposed to As while still in utero. In pregnant women, gestational diabetes has been associated with low levels (less than 25 ppb) of As ingestion (Ettinger et al., 2009). Fei et al. (2013) showed that for every 1 ppb increase of As ingested, birth weight decreases by approximately 1.30 grams. Wasserman et al. (2014) observed that ingestion of As concentrations as low as 5 ppb is associated with a lowered intelligence quotient in children.

The association between low-level As ingestion and diabetes has been debated. A study performed in Denmark indicates that exposure to low concentrations of As (less than 10 ppb) may increase the risk for diabetes where the risk could possibly increase by 2–3% for every ppb increase in As in drinking water (Brauner et al., 2014). However, several other studies observe no correlation with diabetes and ingestion of low concentration of As (Ruiz-Navarro et al. 1998; Penga et al., 2015).

Naturally Occurring Sources for Arsenic

Arsenic is a major and/or minor component of over 200 naturally occurring minerals including sulfides, oxides, and silicates (Smedley and Kinniburgh, 2002). Average As concentrations in sediments range from 5 to 10 ppm (Smedley and Kinniburgh, 2002). More information about these minerals, with attention to the As-bearing minerals found in Virginia, is discussed below. Table 1 summarizes the potential As sources in Virginia, including both natural and anthropogenic sources, and their associated release mechanisms.

Table 1. Summary of potential arsenic sources and release mechanisms in Virginia.

	Source	Release Mechanism	Example
Naturally-Occurring	Sulfide Deposits	Oxidation of sulfides	Paikaray, 2012; Polizzotto et al., 2005
	Iron Oxides	Reductive dissolution of iron oxides to release As	Nickson et al. 1998; Berg et al. 2008
	Glauconite	Reductive dissolution of iron oxides to release As	Barringer et al., 2010; Drummond and Bolton, 2010
Anthropogenic	Arsenical Pesticides and Herbicides	Mostly insoluble; release can be facilitated by influx of phosphate	Peryea and Creger, 1994; Whitmore et al., 2008
	Animal Feed Additives	Leaching from poultry litter	Li and Chen, 2005; Rutherford et al., 2003
	Wood Preservation	Leaching from treated wood	Mercer and Frostick, 2014; Dubey et al., 2009
	Mining Activities	Oxidation of sulfides	Williams, 2001; Schrenk et al., 1998
	Superfund Sites	Site Activities, Reductive dissolution of iron oxides	deLemos et al., 2006; Kent and Fox, 2004
	Landfills, Petroleum Spills, and other Organic Material	Site Activities, Reductive dissolution of iron oxides	Cozzarelli et al., 2015; Powell et al., 2015

Sulfide deposits

Sulfides, including As-bearing pyrite, arsenopyrite (FeAsS), realgar (AsS), and orpiment (As_2S_3), are present in some sedimentary, igneous and metamorphic rocks. Oxidative dissolution of As-bearing sulfides and release of As to groundwater has been associated with increased exposure to As in several areas worldwide where As concentrations in sulfide minerals range from 5 to 126,000 parts per million (ppm) (Smedley and Kinniburgh, 2002). For example, in Taiwan elevated As in groundwater is linked to increased occurrences of skin cancer (Liu et al., 2006), where the As source has been associated with sulfide-rich black shales (Nordstrom, 2002). The mining of sulfide deposits has also resulted in releasing As to natural waters, discussed below. Oxidation of As-bearing sulfides has also been linked with elevated As concentrations in

Wisconsin (Schreiber et al., 2000). In black shales of Kentucky, As-bearing pyrites release As upon oxidation (Tuttle et al., 2009). Within the Newark Basin (New Jersey) elevated As in groundwater has been observed with black shales containing pyrite (Serfes, 2005). The As content in black shales is correlated with sulfur and organic material (Serfes, 2005).

Oxides

Iron oxides are widespread and are commonly associated with elevated As due to the strong adsorption of As to iron oxides (Welch et al., 2000). The release of As from iron oxides due to microbially-mediated iron(III) reduction coupled with oxidation of organic matter has been identified as the dominant mechanism causing elevated As concentrations in groundwater in parts of Southern Asia, the Midwest of the U.S., and other locations where reducing groundwater conditions exist (Michael, 2013; Korte, 1991; Nickson et al., 1998; Berg et al., 2008). Previous studies have shown As concentrations in iron oxides can be up to 76,000 ppm (Smedley and Kinniburgh, 2002).

Another As-bearing oxide is scorodite ($FeAsO_4 \cdot H_2O$), which is a secondary weathering product of arsenopyrite oxidation (Dove and Rimstidt 1985; Harvey et al., 2006). Scorodite can undergo both congruent and incongruent dissolution. In the case of incongruent dissolution, scorodite breaks down into iron oxides and aqueous arsenate (AsO_4) (Dove and Rimstidt, 1985).

Glauconite

Glauconite is an iron potassium phyllosilicate that, due to its adsorptive properties, has been used as a sorbent for As removal in water (Thompson and Hower, 1975; Ismael and Kharbish, 2013). Although not extensively studied, glauconite has been suggested as a source of As to groundwater in the Coastal Plain of Maryland and New Jersey (Barringer et al., 2010; Mumford et al., 2012). Elevated iron concentrations in groundwater, in combination with the presence of iron-reducing bacteria, suggests that iron reduction in minerals may be a mobilization mechanism for As; microbes are likely the driving force behind the mobilization of As from glauconitic subsurface sediments into groundwater (Mumford et al., 2012).

Anthropogenic Sources for Arsenic in Virginia

There are many anthropogenic sources for As in groundwater including agricultural application of arsenical herbicides and pesticides, animal feed additives, wood preservation, mining activities, and waste sites, which will be discussed in this thesis. Many other anthropogenic sources for As exist that will not be discussed in this thesis, for example - smelter operations, combustion of fossil-fuels, mildew resistant paints, bullets, lead batteries, and some types of glass (Smedley and Kinniburgh, 2002; Garelick et al., 2008).

Agricultural Application of Arsenical Herbicides and Pesticides

Arsenical herbicides have been applied to golf courses and other agricultural areas since the early 1900's (Whitmore et al., 2008). Several studies have shown that As in monosodium methanearsonate (MSMA)-treated soils at golf courses may be mobilized into natural waters (Ma et al. 2000; Cai et al. 2002) due to herbicide application and frequent irrigation (Whitmore et al., 2008). However, other studies show little to no correlation between cultivated land with a history of MSMA application and As concentrations in groundwater (Reedy et al., 2007; Bednar et al., 2002; Welch et al. 2000).

The use of arsenical pesticides, such as lead arsenate ($PbHAsO_4$), on apple orchards occurred from the late 1800's to the 1960's to manage the codling moth, a destructive insect pest (Schooley et al. 2008). Peryea and Creger (1994) found that the movement of As from lead arsenate is greater in soils with little clay and organic material, frequent irrigation, and high rates of application of lead arsenate. However, several studies have not observed elevated As in groundwater downgradient from former orchards, indicating low mobility of lead arsenate in groundwater (Robinson et al., 2007; Reedy et al., 2007).

Animal Feed Additives

The organic As-containing feed additive, roxarsone, is fed to poultry to control intestinal parasites (Rutherford et al., 2003). Arsenic is not retained in tissue in significant concentrations, resulting in elevated As concentrations in poultry urine, which is then mixed with wood chips and feathers to create poultry litter. Poultry litter is commonly applied to agricultural fields as a fertilizer. Studies have suggested that As can leach from litter into surface and groundwater (Rutherford et al. 2003; Silbergeld and Nachman, 2008); however other studies (Oyewumi and Schreiber, 2012) observed extensive uptake of As onto clay-rich soils, suggesting that the release of As from litter to water supplies is influenced by subsurface geology and hydrogeology.

Wood Preservation

The southeastern region of the U.S. is the largest producer and user of chromated copper arsenate (CCA)-treated wood due to the humid climate and presence of parasitic insects (Solo-Gabriele and Townsend, 1999). In 2004, CCA was no longer approved for use in treating wood for residential use. However, millions of homes in the U.S. likely have decks treated with CCA (Mercer and Frostick, 2014). In addition, CCA-treated wood has been used for telephone poles, play structures, and vineyard posts (Robinson, et al. 2006). Existing CCA-treated wood can leach CCA into surrounding soils (Robinson et al., 2006; Mercer and Frostick, 2014), which can then potentially leach into groundwater (Dubey et al., 2009; Powell et al., 2015).

Mining Activities

Approximately 10% of As added to soils are derived from mining activities (Nriagu and Pacyna, 1988). Arsenic is approximately 10 to 50 times higher in areas of sulfide mineralization and mining activities (Mandal and Suzuki, 2002). As discussed in the section on geogenic sources, As can be released from sulfide minerals by oxidation, a process that is exacerbated in mining environments (Williams, 2001). Oxidation of sulfides also generates acidity, which can maintain high aqueous concentrations of As and other trace elements (Schrenk et al., 1998).

Waste sites

Nearly one-third of Superfund sites in the U.S. list As as a constituent of possible concern (EPA, 2003). Groundwater downgradient of these toxic waste sites may be impacted by As from site-derived activities, but can also be impacted as result of secondary As release from a change in groundwater chemistry (Welch et al., 2000). For example, organic matter released from Superfund sites can facilitate reducing conditions in groundwater, which can release As as a result of reductive dissolution of iron oxides (Kent and Fox, 2004). This release mechanism is likely occurring at many sites nationwide, as many Superfund sites have no known anthropogenic sources for As but 80% of Superfund sites have organic contaminants that could facilitate reducing conditions in groundwater (deLemos et al., 2006).

Other waste sites, including landfills, sewage release sites, and petroleum spill sites, contain organic material that can promote reducing conditions and release naturally-occurring As in aquifers, as mentioned above. For example, petroleum spills have also been shown to potentially release naturally occurring As to groundwater due to creation of reducing conditions (Cozzarelli et al, 2015; Burgess and Pinto, 2005). While landfills may have anthropogenic-derived As buried in the refuse, many studies indicate that the source for As in groundwater is likely derived from naturally occurring sources that are released under reducing conditions (Welch et al., 2000; Powell et al., 2015; Dubey et al., 2009).

Predictive Models

As discussed above, there are many natural and anthropogenic sources of As; as a result, it is often difficult to determine which source (or sources) is/are responsible for causing elevated As in groundwater. Thus, sampling and testing of wells for As has been the primary approach for evaluating the existence of an As problem in groundwater. Because sampling and testing is expensive, and many homeowners do not regularly sample and test their wells, if at all, other approaches for evaluating As risk are needed. Predictive modeling is one approach that has been used to identify areas susceptible to As contamination and factors that associated with elevated As in groundwater (Amini et al., 2008; Ayotte et al., 2006). For predictive models, there are three general approaches that have been used to predict As occurrences: GIS-based, logistic regression, and miscellaneous decision tree and index models.

A GIS-based approach is useful for modeling complicated interactions between independent variables (regressors). For a GIS model, the data do not need to be normally distributed, but adequate spatial distribution is needed. For example, Amini et al. (2008), and Hossain and Piantanakulchai (2013), used GIS approaches to predict As occurrences in groundwater worldwide and in Bangladesh. Amini et al. (2008) observed that their GIS model explained over 67% of As variation in a variety of groundwater conditions and the model correlated well with known contaminated regions worldwide. Hossain and Piantanakulchai (2013) observed that 87.9% of data were found to be correctly predicted by their model of As concentrations in groundwater in Bangladesh.

Logistic regression models can predict binary outcomes for a threshold (As concentration in groundwater. These models quantify the importance of each regressor (environmental parameters input into model) with As concentrations. For example, Ayotte et al. (2006) constructed a logistic regression model to evaluate the influence of geologic as well as other environmental factors on As concentrations in groundwater in New England. The model was also used to predict spatial distribution of As in areas where there were no data. Results showed that specific geologic units, geochemical indicators, and groundwater residence time were significant regressors. Using their model, Ayotte et al. (2006) accurately predicted 79.8% of existing groundwater As measurements. In another example, Yang et al. (2014) constructed a logistic regression model to predict shallow groundwater As contamination in Louisiana. Results showed that type of sediment deposit, soil texture and clay content, and distance to rivers were associated with a higher probability of elevated As concentrations in groundwater. Their model correctly predicted 79% of existing shallow groundwater As measurements; results agreed with the known spatial distribution of As contamination in Louisiana.

The last approach to evaluate regions of As susceptibility is to use a decision tree or index; this approach is often used when few spatial locations exist but locations have been sampled frequently over time. When using decision trees or an index, mean concentrations may be used in place of true concentrations and models may produce less sensitive results, which reflects the model’s ability to correctly detect As concentrations. For example, Singh and Vedwan (2015) developed a “composite vulnerability index” to examine the population who are at the highest risk of impacts from As exposure using data from various other studies in India. Results showed that by using this index, which included both social and environmental parameters, As vulnerability to groundwater in different areas of India can be better understood for planning and decision-making purposes.

Study Objective

Although extensive research on As occurrences have been conducted elsewhere in the U.S., there are few studies on As in groundwater in Virginia, likely because national surveys such as the one conducted by Welch et al. (2000) have not identified Virginia as a “hot spot” for As. However, a recent study by Chapman et al. (2013) identified elevated As in sedimentary rocks in the Mesozoic Basins of the Eastern U.S. and in metamorphosed clastic sedimentary units in the Piedmont and Blue Ridge aquifers, all of which are present in Virginia.

Evaluating As in groundwater in Virginia is important due to the heavy reliance on groundwater for drinking water supply. Of the 95 counties in Virginia, 38 depend on groundwater for their primary potable water supply from both domestic and public wells. Homeowners are responsible for having their own wells tested, but they may not be aware of what water quality parameters should be analyzed. Because As is naturally occurring, exists in many types of aquifers, is odorless and colorless, and generally does not cause immediate illness, it is difficult to evaluate if As is a concern in groundwater supplies.

The objective of this study was to evaluate the occurrence of As in groundwater in Virginia, to examine the spatial distribution of As concentrations in groundwater with a GIS, and to develop and apply a logistic regression model to evaluate if elevated As occurrences can be predicted in Virginia using geologic and other environmental parameters.

Chapter 2: Methods

Datasets

Existing groundwater chemistry datasets from the Virginia Department of Health (VDH) and the Biological Systems Engineering Household Water Quality Program (BSE) were used for this project. Each of these datasets are stored and maintained at Virginia Tech; the VDH dataset is maintained by Madeline Schreiber (Geos); the BSE dataset is maintained by Erin Ling (BSE). Prior to use in the statistical model, datasets were checked for duplicate samples. If more than one sample was collected at an individual location, the maximum As concentration for each location was retained and the remainder of the samples were discarded. This was done in an effort to construct the most conservative model.

Virginia Department of Health (VDH)

The VDH dataset contains As concentrations from unfiltered samples collected from public water supply wells in Virginia from 1973 to 2013. The original dataset did not include latitude and longitude for the well locations, but did include a meta-data description. For this study, locations were assigned using Google Earth and the meta-data description. If locations could not be not clearly identified, the data were discarded. The reporting limit for As for most of the samples was 5 ppb. Approximately 13 samples (duplicates removed) had reporting limits higher than 5 ppb, and in order to limit censored data, those samples were removed from the dataset. See Table 2 for more information regarding the dataset. The exact methods for sample collection and analysis could not be determined for the VDH dataset, but samples were likely collected using EPA drinking water sample collection methods (Appendix A).

Biological Systems Engineering Household Water Quality Program (BSE)

The BSE dataset was obtained from the Biological Systems Engineering Department Household Water Quality Program (BSE). This program was created to improve the water quality and health of Virginians. Since 1989, drinking water clinics have been conducted in 45 counties. Through the clinics, homeowners can have their drinking water tested for various contaminants; the results are returned confidentially. Water samples included in this dataset were collected from wells, springs, and cisterns and were not filtered prior to analysis. Homeowners complete a survey with information regarding their water source i.e., water treatment systems, sources for drinking water, and other pertinent information. It should be noted that not all homeowners are aware of details regarding their water source; thus, supplemental data collected in the survey were evaluated with discretion. Arsenic was analyzed in samples collected from 2008 to 2015. The reporting limit for As samples in this dataset is 1 ppb, which corresponds to the lowest concentration standard used to calibrate the laboratory instrument. Two samples were collected at each location: a first draw sample and a flushed sample. The first draw sample was collected after stagnation in the plumbing (typically overnight). The flushed sample was collected after water was flushed through pipes for at least 5 minutes. Data from flushed samples were used in this study as they likely represent a more accurate depiction of groundwater chemistry with less influence from household plumbing. See Table 2 for more information regarding the dataset.

Samples collected and analyzed in this dataset contain confidential information and are thus protected under a Virginia Tech Institutional Review Board protocol (VTIRB #12-267). To maintain confidentiality for participants, sample location (latitudes and longitudes) were geocoded by staff members of Household Water Quality Program using addresses of participants and the following website: http://www.findlatitudeandlongitude.com/batch-geocode/#.VovimPkrJaQ. This geocoding method generalized sample locations to protect homeowner's information.

Table 2. Summary of datasets used for this study.

	VDH Data Set	**BSE Data Set**
Number of Samples in original dataset	10,261	6,739
Number of Duplicates Removed	9,043	2,325
Final Number of Samples used in Model	1,218	4,414
Arsenic Reporting Limit (RL)	5 ppb	1 ppb
Sample Collection Dates	1973-2013	2008-2015
Sample Type	Unfiltered	Unfiltered
Number of Samples Below RL	1,157	4,144
Number of Samples Above RL	61	270

GIS Data

Location of each sample in the datasets was mapped spatially in ArcGIS version 10.2 in separate project files (ESRI, 2013). Geologic unit, lithology, physiographic province, soil series and particle size, and land use layers were added to each project file to represent environmental attributes. A spatial join was performed to associate each As sample with its corresponding environmental attributes. Upon completion of the spatial join, the data were exported to an excel file in .csv format. See Appendix B for further discussion on GIS procedures.

Geologic Unit GIS Layer – USGS, digitized from King and Beikman 1974

The geologic unit layer was obtained from the U.S. Geological Survey (USGS) website (https://mrdata.usgs.gov/geology/state/state.php?state=VA) as a shapefile. The shapefile includes bedrock geologic unit name, spatial locations, and a short description, for example - Cambrian limestone.

These geologic units were classified first by age, then stratigraphy. Stratigraphic units are only used where they illustrate a special geologic feature and where the age of the units is uncertain. In general, most of the geologic systems that form outcrops can be separated into several comprehensive time-stratigraphic units. In the Eastern U.S., hybrid nomenclature is used to describe units that form outcrop bands too narrow to be separated, or that the two units form a homogeneous body of rocks. Figure 1 shows the geologic units based on age.

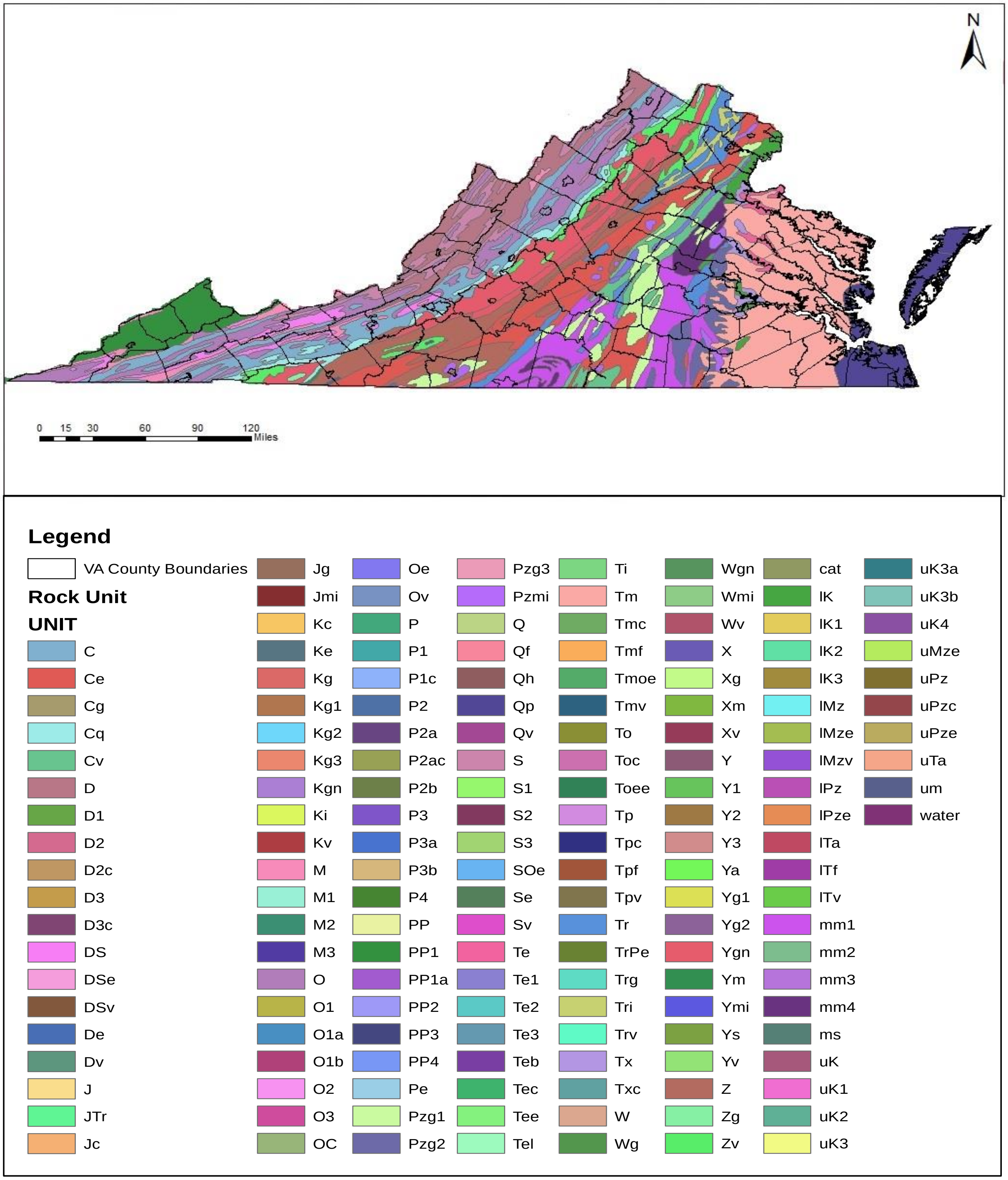


Figure 1. Spatial extent of geologic units in Virginia based on age. Data obtained from the USGS.

Lithology GIS Layer – USGS (Dicken et al. 2008)

The lithology layer was obtained from the USGS website as a shapefile. The shapefile includes information about spatial locations for generalized lithology, for example sandstone or shale. Three fields, rocktype1, rocktype2, and rocktype3, are present in the lithology attribute table to create a generalized dominant lithology map. Rocktype1 is the most abundant lithology in the unit; rocktype2 is the second-most abundant lithology; and remaining lithologies are included in rocktype3. Figure 2 shows the lithologies in Virginia.

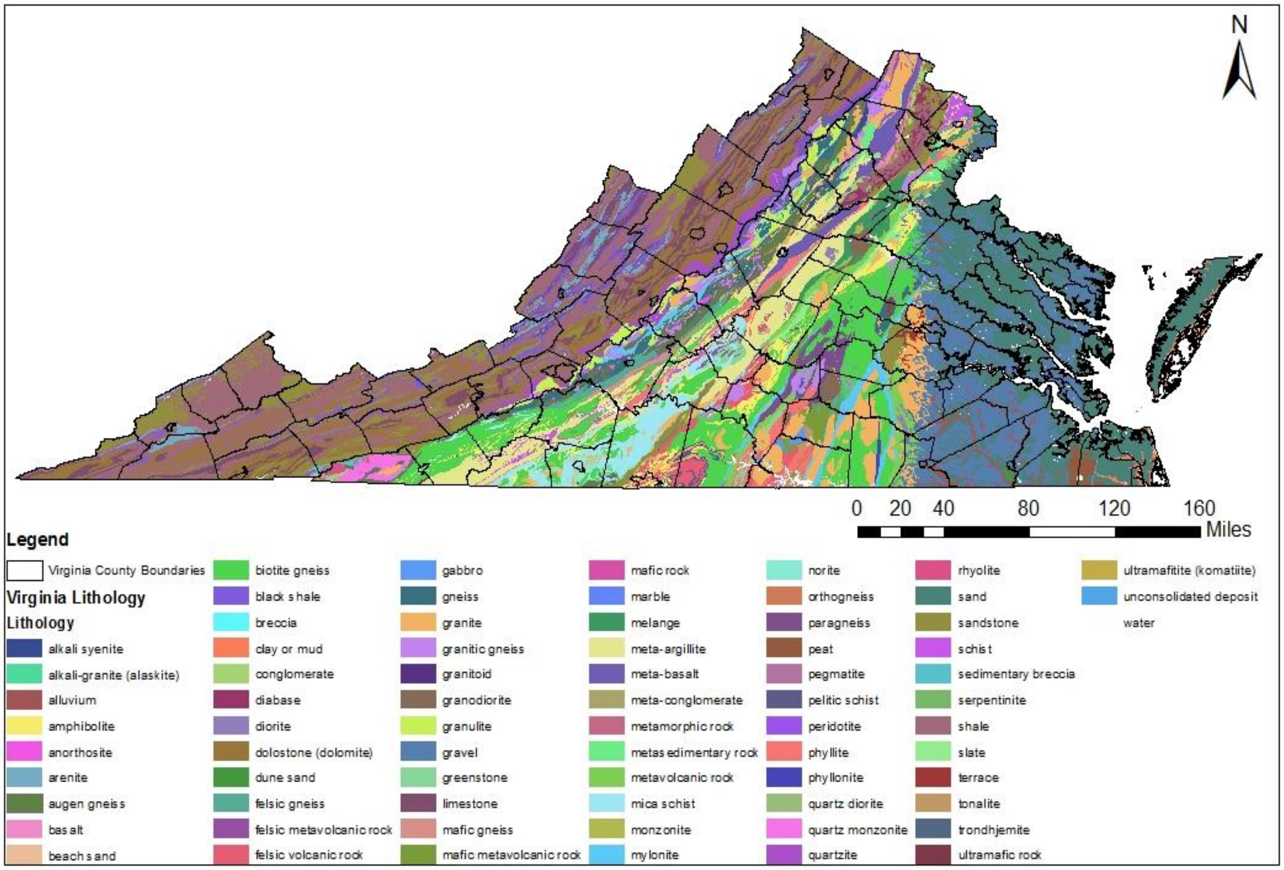


Figure 2. Spatial extent of lithologies in Virginia. Data obtained from the USGS.

Soil Series and Particle Size GIS Layer – Soil Survey Geographic Database

The soil layer was obtained from the Soil Survey Geographic Database (SSURGO), (http://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm) as a shapefile. The shapefile includes information about soil series, particle size, classification, and other soil attributes, as well as spatial locations. Each map unit describes soils and other components that have unique properties, interpretations, and productivity. The attribute table contains one to three major components and potentially minor components, although map units are named for major components. SSURGO datasets consist of map data, tabular data, and information about how the maps and tables were created. Figure 3 shows the spatial extent of soil series in Virginia.

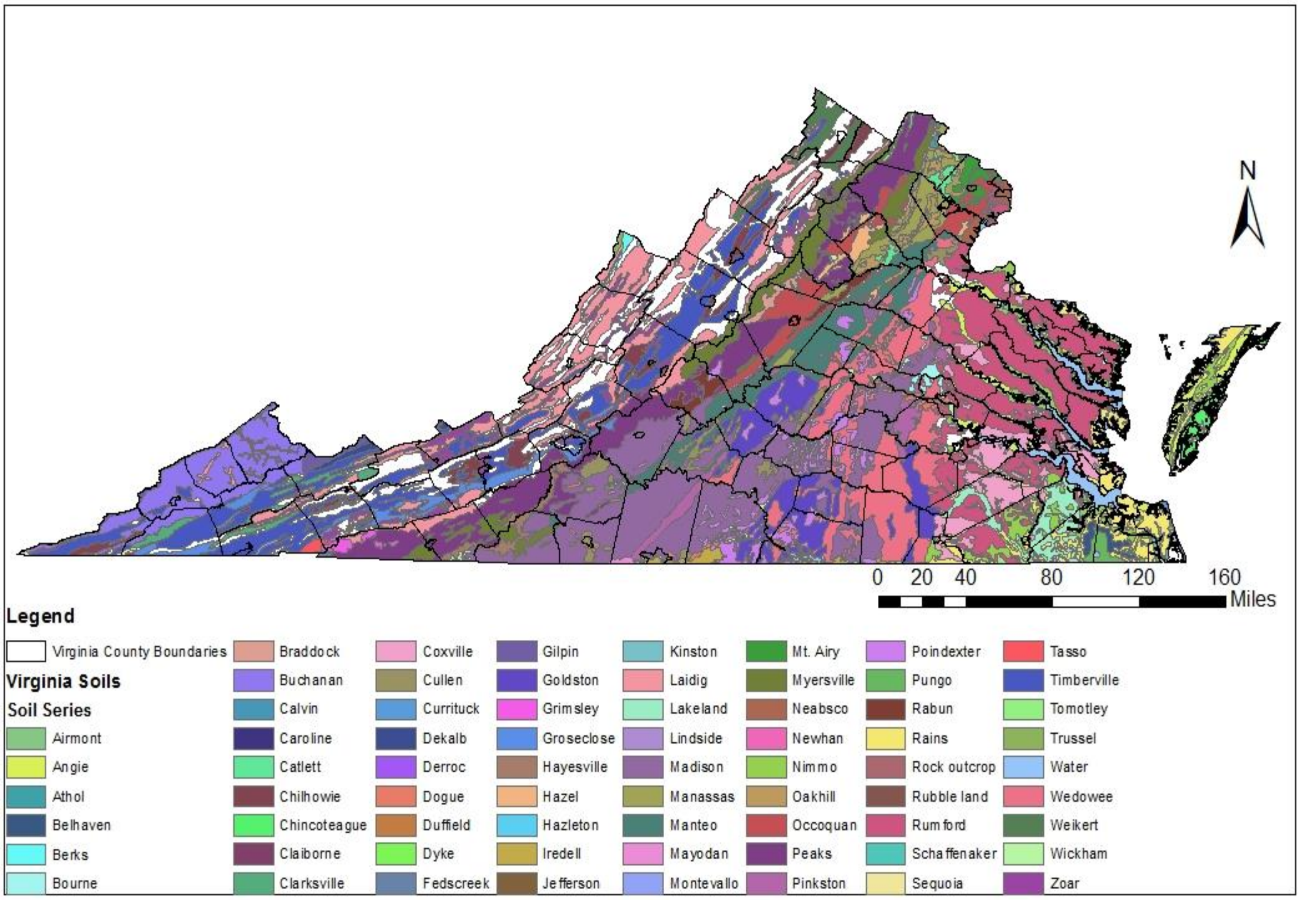


Figure 3. Spatial extent of soil series in Virginia. Data taken from SSURGO and the Web Soil Survey.

Physiographic Province GIS Layer – USGS (Fenneman and Johnson, 1946)

The physiographic province layer was obtained from the USGS website as a shapefile. The shapefile encompasses the conterminous U.S. and was derived from Fenneman's 1:7,000,000-scale map, "Physical Divisions of the United States". This map was based on eight major divisions, 25 provinces, and 86 sections representing distinctive areas, having similar topography, lithologies, structure, and geologic history. Figure 4 shows the physiographic provinces in Virginia.

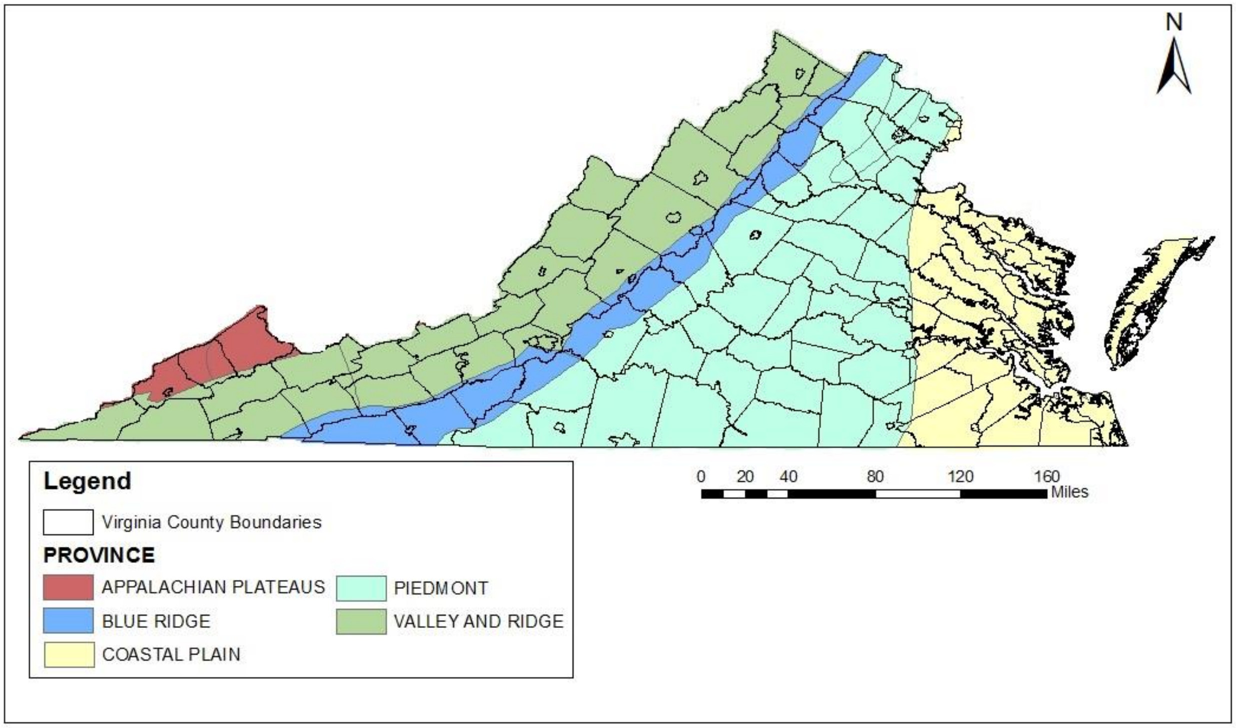


Figure 4. Physiographic Provinces of Virginia. Data obtained from the USGS.

Land Use GIS Layer – National Land Cover Database 2011 (NLCD2011)

The land use layer was obtained from the United States Department of Agriculture (USDA) Geospatial Data Gateway (https://gdg.sc.egov.usda.gov/) as a raster file. The raster file included classification of land use data for the state of Virginia. Examples of land use classification include hay pasture, wetlands, and developed low to high intensity. The classification system used by NLCD2011 is modified from the Anderson Land Cover Classification System. Figure 5 shows the land use cover in Virginia.

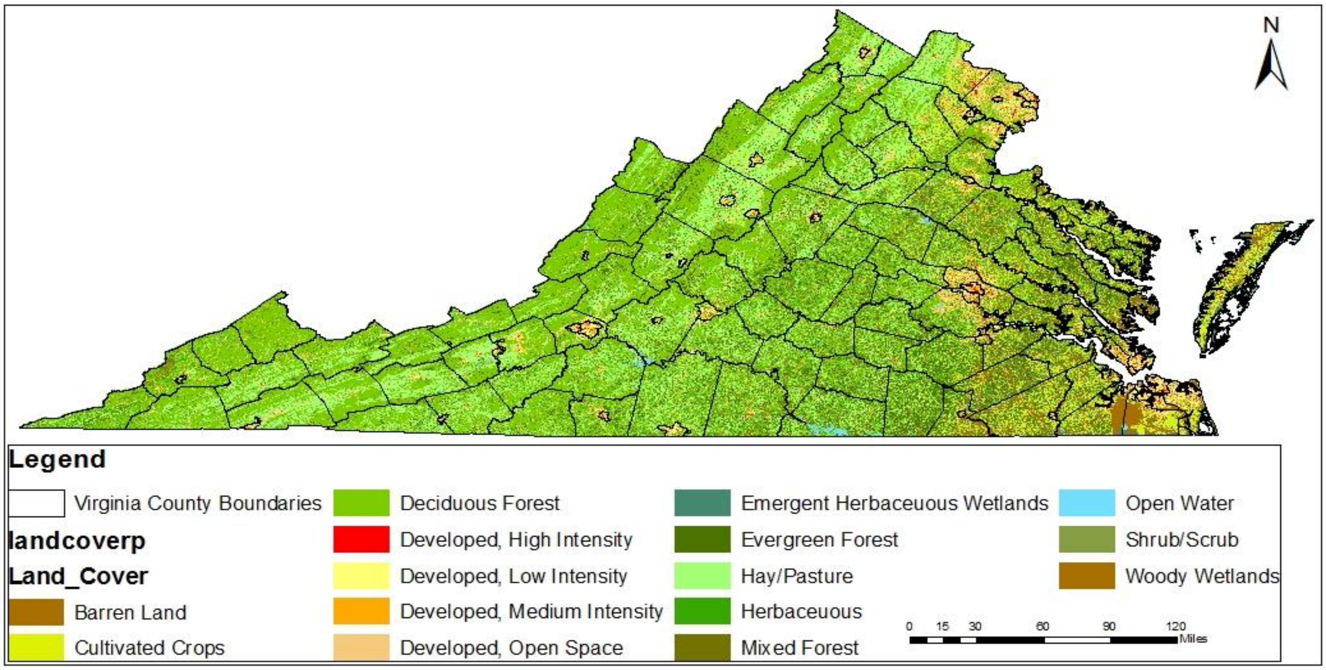


Figure 5. Spatial extent of land use in Virginia. Data taken from the USDA Agriculture Geospatial Data Gateway.

Model Creation and Variable Selection

Because of the type and quality of datasets available for this study, it was decided that logistic regression would be the most accurate and well-suited method for predictive modeling. Another advantage of logistic regression is that the results identify which regressors have a higher probability to be associated with elevated As concentrations in groundwater, which is useful not only for scientific purposes, but also for managing groundwater resources.

As with any model, logistic regression models have inherent assumptions, including: 1) the probabilities are a logistic function of the regressors; 2) all important regressors are included; 3) non-significant regressors are not included; 4) the regressors are measured without error; 5) the observations are independent and; 6) the regressors are not linear combinations of each other.

Structure of Logistic Regression Model

For this project, the logistic regression model was created in statistical software R for the probability that As concentrations exceeds a given threshold (R. Core Team, 2012).

$$P(y = 1[x]) = \frac{e^{\beta_0+\beta_1x_1+\beta_2x_2+\ \dots+\beta_kx_k}}{1 + e^{\beta_0+\beta_1x_1+\beta_2x_2+\ \dots+\beta_kx_k}} \tag{1}$$

Where P is the probability that an event occurred, y is binary variable; when y = 1, an As sample has a concentration greater than the threshold (5 ppb); when y = 0, an As sample that has a concentration less than or equal to the threshold. x_1, x_2, x_3… are the responses (As concentrations), and β1, β2, β3… are the regressors (regressors, discussed below). See Appendix C for the R code used to construct the model.

Regressors chosen for the model included geologic units, lithology, soil series and particle size, physiographic province, and land use. These regressors are categorical, which means they have a fixed number of possible values, which are then assigned to a particular group or “category”.

These categorical variables were then input into R to create binary variables or "indicator variables" for each level. The levels indicate how many choices per regressor exist (i.e. how many geologic units exist in Virginia) and are summarized in Table 3.

Table 3. List of regressors input into the logistic regression model.

Regressor	**Maximum Number of Levels in VA**
Geologic Unit	160
Lithology	72
Physiographic Province	5
Soil Series	69
Soil Particle Size	10
Land Use	15

The model was trained (calibrated) using the BSE dataset. Approximately 99% of the 4,414 observations used in the training model were ≤ 5 ppb (threshold for model). To test the model, the VDH dataset was used, in which approximately 95% of the 1,218 observations were ≤ 5 ppb. Using two independent datasets to train and test the model helps to better evaluate model suitability. Other studies that only have one dataset separate the data into two sections (e.g., 85% of data and 15% of data) in order to train and test the model, respectively (see Ayotte et al., 2006 for example).

For the BSE dataset, the reporting limit is 1 ppb, and for the VDH dataset the reporting limit is 5 ppb, so to simplify the model, we used the 5 ppb as our threshold. Other thresholds that were higher than the reporting limit (e.g., 10 ppb) were tested during model development, but produced less sensitive models with fewer significant regressors identified. Models with higher thresholds also have a higher uncertainty as the probability of an As concentration greater than the threshold is smaller.

Preliminary model results using all regressors shown in Table 3 showed multicollinearity between many of the regressors. Multicollinearity is a statistical phenomenon where two or more regressors are highly correlated and can be linearly predicted from other regressors. To reduce multicollinearity, soil series and particle size were omitted; the reasoning for their removal from the model is that soil units originate from bedrock geology, which is already reflected in the model and wells are open to bedrock and not soil. Land use was also omitted during the model development stage. The reasoning for this omission is that land use categories in the GIS file were not detailed enough to reflect land uses that would be impacted by As, for example, orchards and golf courses, to which arsenical pesticides may have been applied, are not differentiated in the land use dataset. Physiographic provinces were also omitted as they were too large to be an accurate predictor of As occurrences. In the final model, only geologic units were used as regressors.

Because multicollinearity still existed between different geologic units, the Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression was used to select variables and ridge logistic regression was used to run the final model and obtain significant regressors (significant regressors were identified using a p-value less than 0.10). When multicollinearity occurs, the least squares estimates are unbiased, but the variances are large and may be far from the true value. Both ridge and LASSO regression models add a degree of bias to these regression estimates, which reduce the standard errors, thus resulting in more reliable estimates. LASSO regression places a penalty (λ) on β (the regressor), pushing β towards zero and in some cases equates β to zero, thus excluding regressors from the model. Ridge regression places a similar penalty (λ) on β, pushing β towards one (Tibsharani, 1996).

K-fold cross validation was used to select the best λ for the ridge regression model, where the data were randomly partitioned into 10 equal subsamples (10 = k in k-fold cross validation).

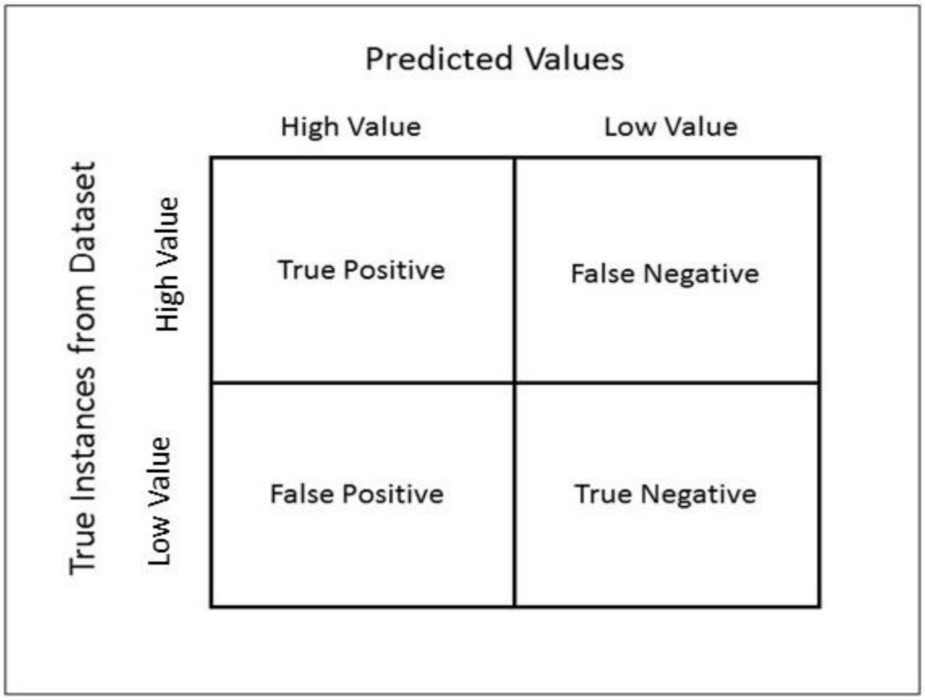
Nine of the subsamples were used to train the model and the remaining subsamples were used to test the model. This process was repeated 10 times using each of the subsamples to test the model and select the best value for λ. The best λ indicates the smallest deviance for all of the calibrated models. The LASSO regression model uses a Semi-Automatic Method to select λ, which varies slightly from the k-fold cross validation method in the ridge regression model (Cule and Iorio, 2012).

To evaluate model fit, classification functions were computed upon completion of the regression model (see Table 4). The classification functions are a way to examine type I and type II errors in the model. A type I error is the incorrect rejection of a true null hypothesis (false positive). A type I error may predict relationships between regressors and the response that do not exist. A type II error is the failure to reject a false null hypothesis (false negative). A type II error may predict relationships that do not include important regressors. Probabilities of predicted As occurrences from the model were determined, where probabilities greater than 0.5 were determined to be As concentrations above the threshold (5 ppb) and probabilities less than 0.5 were determined to be As concentrations less than the threshold (5 ppb). Using the probabilities calculated, classification functions were tabulated (Table 4). Figure 6 shows a contingency table of type I and type II errors. In a perfect model, only true positives and true negatives would be predicted.

The number of true negatives and positives (observed in the dataset) was compared to the model predicted negatives and positives to determine the sensitivity of the model, which reflects the model's ability to correctly detect As concentrations greater than 5 ppb and the specificity of the model, which reflects the model's ability to detect As concentrations less than or equal to 5 ppb.

Table 4. Classification function descriptions. TP = True Positive; P = Positive Instances; TN = True Negative; N= Negative Instances; FP = False Positive and; FN = False Negative.

Classification Function	**Formula**
True Positive Rate (Sensitivity)	TP / P
True Negative Rate (Specificity)	TN/N
Positive Predictive Value (Precision)	TP/(TP+FP)
Negative Predictive Value	TN/(TN+FN)
False Positive Rate	FP/N
False Negative Rate	FN/(TP+FN)
False Discovery Rate	FP/(TP+FP)
Accuracy	(TP+TN)/(TP+FN+FP+TN)



	High Value	Low Value
High Value	True Positive	False Negative
Low Value	False Positive	True Negative

Figure 6. Contingency table for classification functions.

Chapter 3: Results

Table 5 summarizes the distribution of As data in the VDH and BSE datasets. As shown in Table 5, the majority of samples (95% in the VDH dataset; 99% in the BSE dataset) had As concentrations ≤ 5 ppb. Only a small percentage (2.8% in the VDH and 0.6% in the BSE) of samples had concentrations between 5 and 10 ppb; similar percentages (2.2% in VDH; 0.4% in BSE) of samples had As concentrations > 10 ppb. Overall, this is a positive result for groundwater quality in Virginia.

Table 5. Concentration ranges for each dataset and their respective percentages.

Concentration Range	VDH	BSE
≤ 5ppb	95%	99%
5.01-10.00 ppb	2.8%	0.6%
> 10 ppb	2.2%	0.4%

Figure 7 shows the spatial distribution of As concentrations in groundwater in Virginia using the VDH and BSE datasets. The highest As concentrations (greater than 10 ppb) were observed in three geographic areas: 1) the central/northern region of Virginia; 2) the northern Shenandoah Valley and; 3) the Coastal Plain.

Table 6 shows the results of the variable selection using the BSE dataset and LASSO regression. The geologic units are identified as well as the coefficient (β term in logistic regression equation), where significant regressors were identified with a coefficient greater than one. P-values cannot be computed using the LASSO regression model. Coefficients in Table 6 with a dash indicate geologic units that have one or no As samples with concentrations greater than 5 ppb, and therefore, a coefficient was not calculated for the geologic unit.

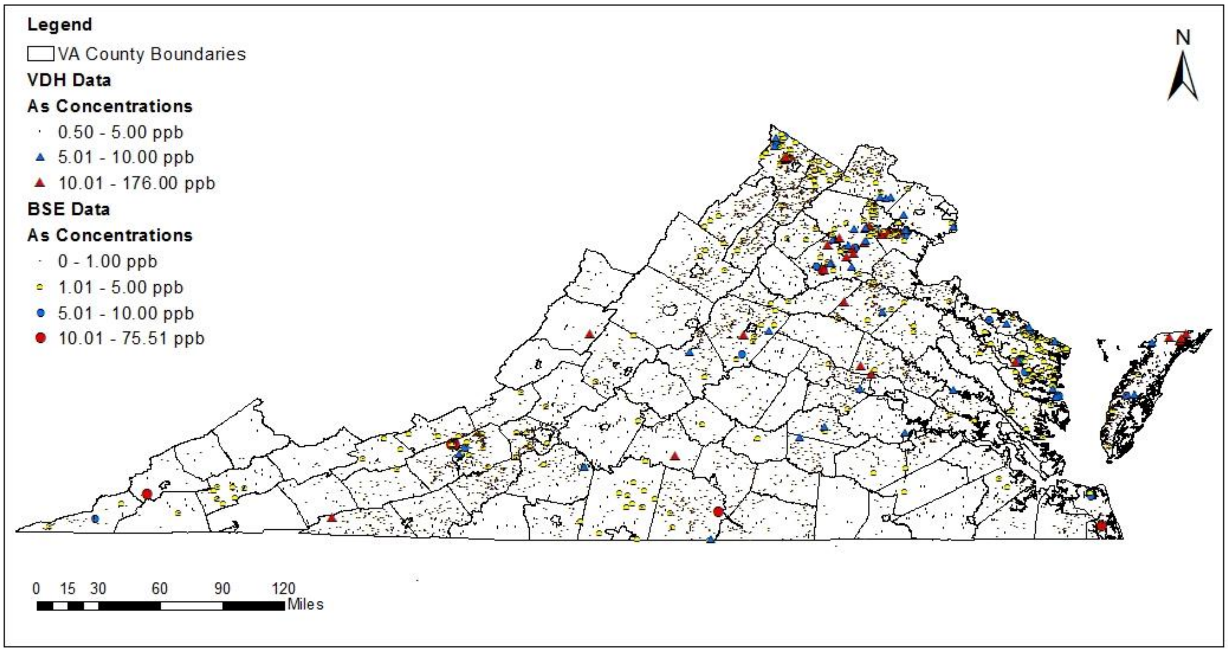


Figure 7. Arsenic concentrations in groundwater in Virginia.

Table 6. Results of variable selection using the BSE dataset and LASSO regression.

Geologic Unit	Coefficient
(Intercept)	-6.275
C – Cambrian-aged shales and limestones	-
Ce – Cambrian metamorphic and volcanic rocks	1.596
Cq – lower Cambrian clastic rocks	<0.0001
Cv – Cambrian-aged Volcanic rocks	-0.430
D – Devonian-aged shales/sandstones	2.671
DS – Devonian and Silurian-aged shales and limestones	3.390
lK – lower Cretaceous-aged metamorphic rocks	-
M - Mississippian-aged dolostone and sandstone	3.080
Mm1 – felsic paragneiss and schist	-0.697
Mm4 – granite gneiss	-0.334
O – Ordovician-aged shales and dolostones	-
Oe – Ordovician metamorphic rocks	-

PP1 – Atokan and Morrowan Series (shale and sandstone)	-
Pzg1 – lower Paleozoic granitic and metamorphic rocks	1.573
Pzg2 – middle Paleozoic granitic and metamorphic rocks	-
Pzmi – Paleozoic mafic intrusives	-
Qp – Pleistocene-aged sands	1.812
S – Silurian-aged shales and limestones	2.328
Te – Eocene-aged sands and gravels	-
Tm – Tertiary gravels and sands	2.171
Tr - Triassic-aged sedimentary rocks	3.106
Tri – Triassic mafic intrusives	2.586
Tx – Paleocene-aged sands and gravels	-
Um – ultramafic rocks	-
Ya - Anorthosite	-
Ygn – Proterozoic-aged volcanic and metamorphic rocks	-1.010
Ym – Paragneiss and Schist	-
Z – sedimentary and metamorphic rocks	-0.833
Zg – granitic and metamorphic rocks	-
Zv – volcanic rocks	-0.437

Table 7 summarizes the significant regressors from the model trained using the BSE dataset and ridge regression. Regressors with a p-value less than 0.10 were identified as significant in the model. Regressors identified in Table 7 were input into the model and tested using the VDH dataset. Results from the testing dataset and final model are summarized in Table 8. Regressors from the VDH test with p-values less than 0.10 were identified as significant.

Table 7. Results from the model using the BSE dataset and ridge regression.

Geologic Unit	Coefficient	Exp. Coefficient	Standard Error	t-statistic	p-value
Z – sedimentary and metamorphic rocks	-0.0741	0.9285	0.1369	-1.717	0.0858
Ygn – Proterozoic-aged volcanic and metamorphic rocks	-0.0885	0.9152	0.1359	-2.051	0.0402
Tm – Tertiary gravels and sands	0.1163	1.123	0.1367	2.685	0.0072
D -Devonian-aged shales/ sandstones	0.1081	1.114	0.1418	2.489	0.0128
M - Mississippian-aged dolostone and sandstone	0.1196	1.127	0.1357	2.743	0.0060
Tr - Triassic-aged sedimentary rocks	0.2095	1.233	0.1416	4.805	<0.0001
DS – Devonian and Silurian-aged shales and limestones	0.1454	1.156	0.1357	3.325	0.0008

Table 8. Significant geologic units (regressors) and associated values from the final model using both the BSE and VDH datasets

Geologic Unit	Coefficient	Exp. Coefficient	Standard Error	t-statistic	p-value
D - Devonian-aged shales/sandstones	2.408	11.19	1.220	1.974	0.0483
Tr - Triassic-aged sedimentary rocks	4.585	98.08	1.214	3.775	0.0002
Ygn – Proterozoic-aged volcanic and metamorphic rocks	-1.866	0.7868	0.1359	-1.787	0.0779

In the final regression model developed for this study (Table 8), two geologic units were identified as having a higher probability of elevated groundwater As occurrences: Unit D (Devonian-aged shales/sandstones) and Unit Tr (Triassic-aged sedimentary rocks). Geologic Unit Ygn was identified as having a lower probability of elevated As in groundwater. Figure 8 shows the spatial extent of the geologic units that have a higher probability of observing elevated As in groundwater. Geologic Unit Tr makes up the sedimentary portion of the Mesozoic rift basin complex that extends from North Carolina to Connecticut. Geologic Unit D runs along the Valley and Ridge province and is composed of Devonian-aged shales and sandstones.

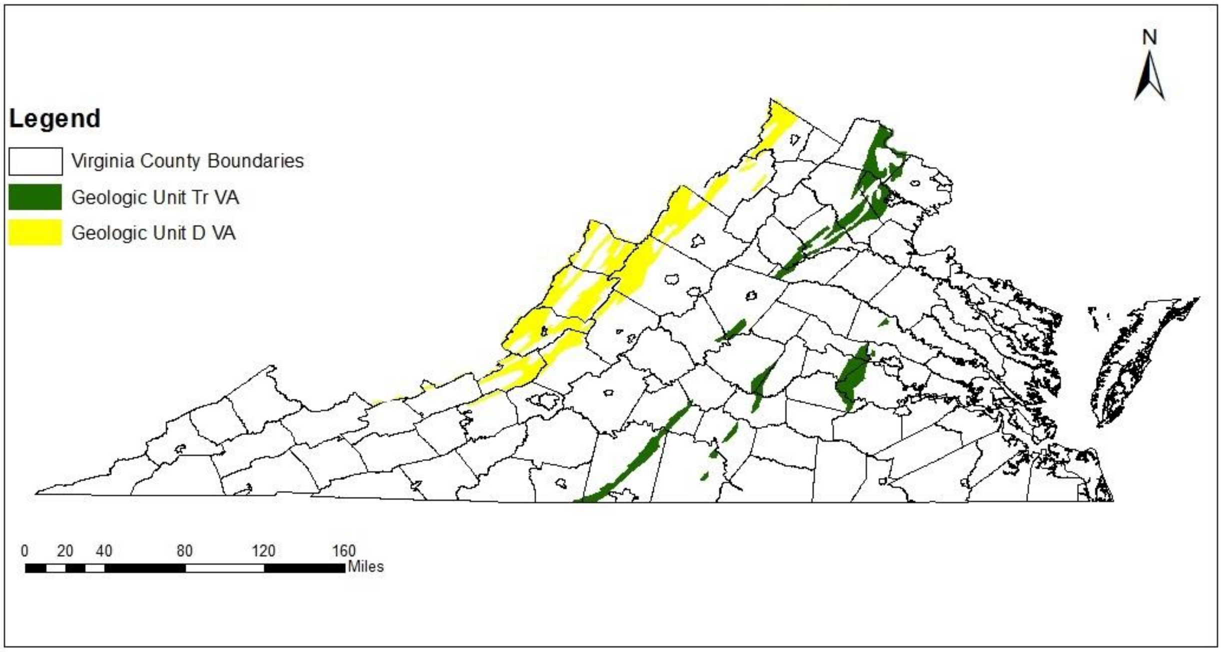


Figure 8. Spatial extent of significant geologic units from the final logistic regression model. See Table 8.

The final equation for the logistic regression model with only significant regressors is as follows:

$$Logit\ (y) = -2.9807 + 0.5763x_1 + 0.5719x_2 - 0.2055x_3$$

where $Logit = \frac{P(y=1)}{1-P(y=1)}$ and the x-values can be interpreted as the impact from each of the regressors on As concentrations (x_1 is geologic unit Tr; x_2 is geologic unit D; x_3 is geologic unit Ygn).

Even though elevated As concentrations were clustered within the Coastal Plain (Figure 7), geologic units in the Coastal Plain were not significant in the final model (Table 8). This is likely due to the large spatial extent of the geologic units (Tertiary gravels, sands, and clays) present within the Coastal Plain, the multi-layered aquifer system, and the fact that other factors, such as geochemical conditions, likely influence As release to groundwater.

Results from the evaluation of model fit analysis are shown in Table 9. Our model was 95% accurate, which is a direct reflection of the distribution of the VDH (testing/validation) dataset, as 95% of As concentrations were below the threshold. The VDH dataset was input into the model and predicted how many elevated As concentrations (>5 ppb) were present in Virginia groundwater. The sensitivity of the model is 0% because no true positives were predicted in the model, thus making the true negative rate 100%. These results show that this logistic regression model, using these two datasets, is not able to predict As concentrations in groundwater in Virginia. The primary reason for this lack of sensitivity is related to a) the distribution of As concentrations in the dataset and b) that most As concentrations are below the threshold value of 5 ppb (see Table 5). Despite the inability of the model to accurately predict As concentrations in groundwater, the model is still useful for evaluating the geologic sources of As, which was the primary goal of this study.

Table 9. Classification functions for model. TP = True Positive; P = Positive Instances; TN = True Negative; N= Negative Instances; FP = False Positive and; FN = False Negative.

Classification Function	Formula	Value
True Positive Rate (Sensitivity)	TP / P	0%
True Negative Rate (Specificity)	TN/N	100%
Positive Predictive Value (Precision)	TP/(TP+FP)	0%
Negative Predictive Value	TN/(TN+FN)	94.99%
False Positive Rate	FP/N	0%
False Negative Rate	FN/(TP+FN)	100%
False Discovery Rate	FP/(TP+FP)	0%
Accuracy	(TP+TN)/(TP+FN+FP+TN)	94.99%

Chapter 4: Discussion

Significant Geologic Units

Triassic-aged Sedimentary Rocks

Results from the logistic regression model show that Triassic-aged sedimentary rocks are 98.08 times (Exp. Coefficient from Table 8) more likely to contain elevated As concentrations in groundwater than other geologic units in Virginia. These Triassic-aged sedimentary rocks are part of the Mesozoic rift basin complex, which extends from North Carolina to Connecticut. This basin complex is composed of Triassic-aged sediments and intrusive Jurassic diabases (Kent et al., 1995). Figure 9 shows the locations of smaller rift basins within the Mesozoic rift basin complex.

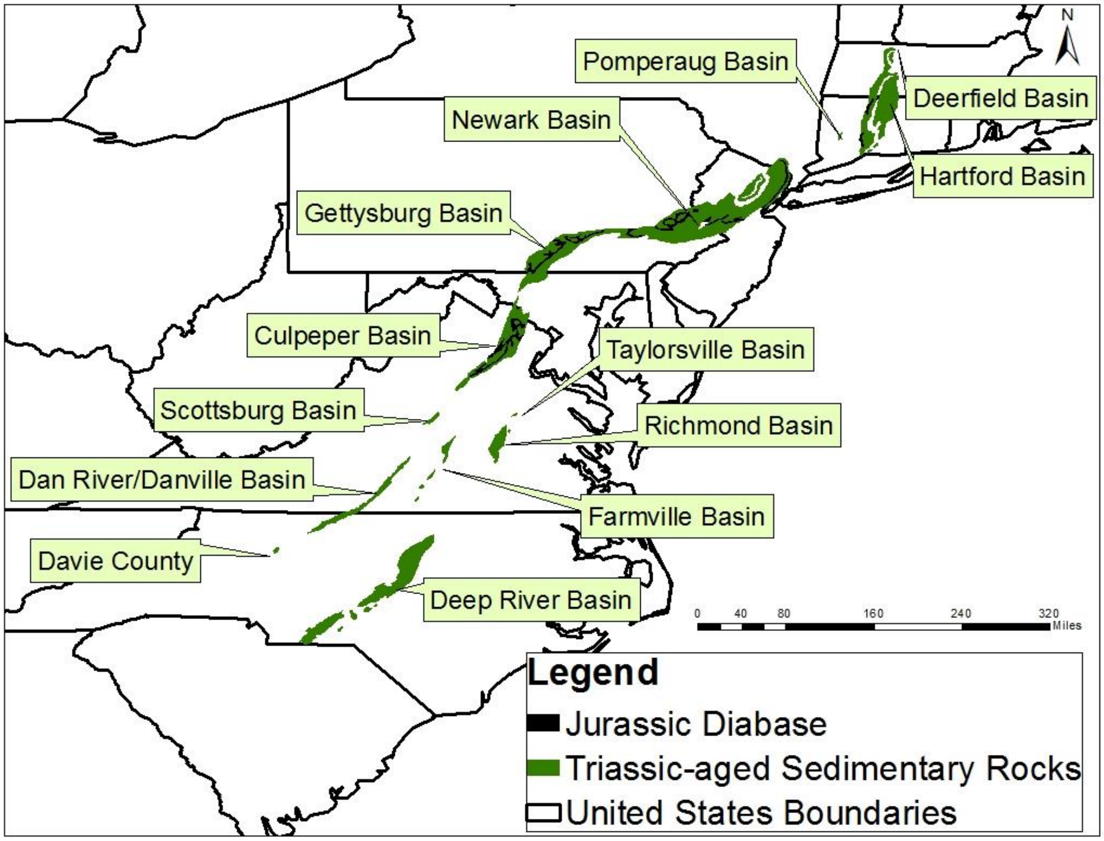


Figure 9. Mesozoic rift basin complex extent along the east coast of the United States. GIS data obtained from the USGS.

One of the largest and most well-studied of the Mesozoic rift basins is the Newark Basin, located in New Jersey and Pennsylvania. The Newark Basin is composed of thick, lacustrine and fluvial sedimentary rocks (Kent et al., 1995). The Newark, Gettysburg, and Culpeper Basins, within the Mesozoic rift basin complex, are thought to have been connected during sedimentation and therefore, sediments within these three basins are likely very similar (Olsen et al., 1996; Lee, 1977). Both the Newark and Gettysburg Basins display elevated As concentrations in groundwater associated with clastic lacustrine rocks and metamorphosed sedimentary rocks (O'Day et al., 2005; Senior and Sloto, 2006; Chapman et al., 2013). Specifically, the Newark Basin contains the Lockatong (black and grey shale) and Passaic (red mudstone/siltstone) geologic formations, which are thought the be the source rocks for the elevated As in groundwater; these two units generally correlate with the Balls Bluff Siltstone and Bull Run formations, respectively, found in the Culpeper Basin (Blake and Peters, 2014; Lee 1977). Figure 10 shows the Culpeper Basin with As concentrations in groundwater from the VDH and BSE datasets.

Identifying the source of As in the Mesozoic rift basin complex is challenging as the literature suggest that there are two possible hosts for As: sulfides (pyrite) and iron oxides (hematite). Serfes et al. (2010) examined sources, mobilization, and transport of As in the Passaic (red mudstone/siltstone) and Lockatong (black and grey shale) formations in the Newark Basin. Results from electron microanalyses on the Lockatong Formation showed that pyrite was observed to contain approximately 4 weight percent of As. In the Passaic Formation, pyrite was not observed to be the As source, but rather As was associated with clay and iron oxides.

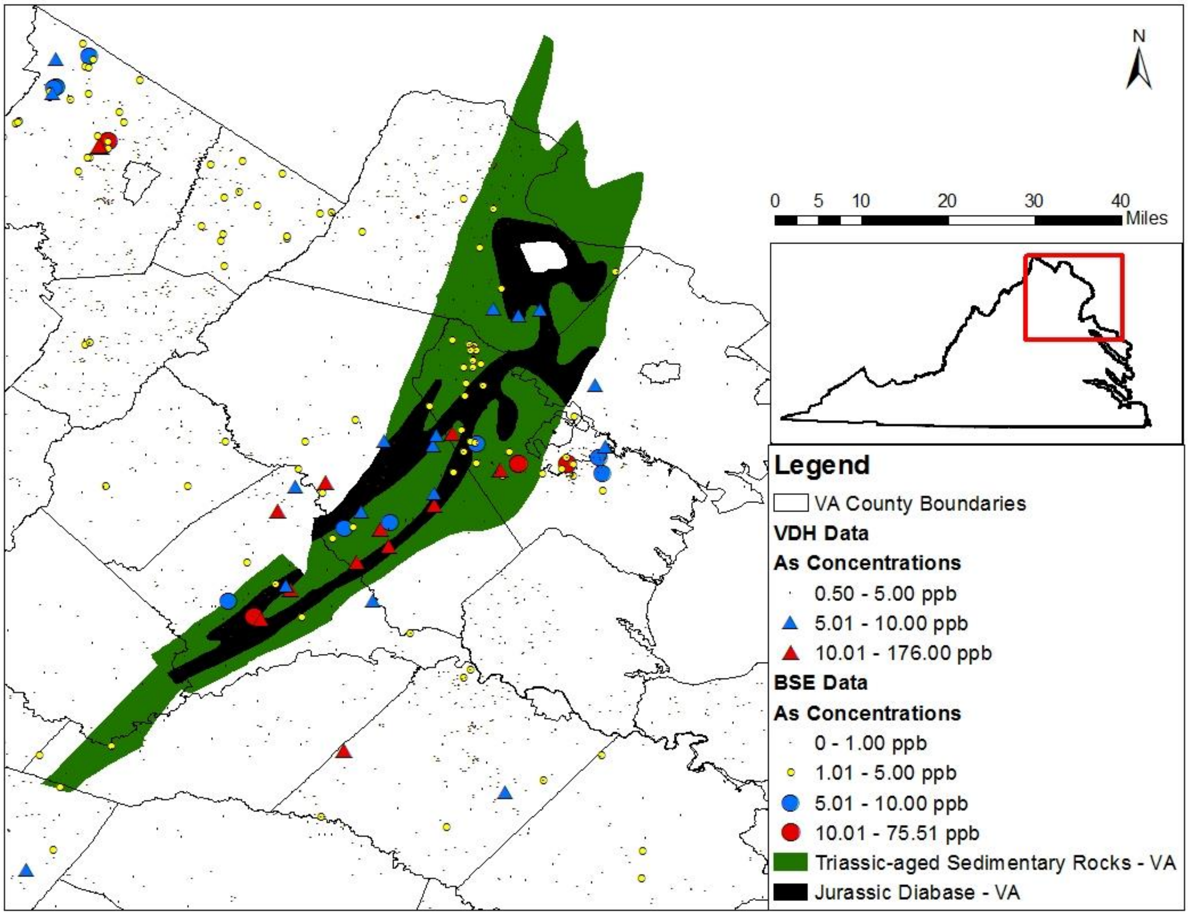


Figure 10. The Culpeper Basin with overlays of Triassic-aged sedimentary rocks and the Jurassic diabase, with As concentrations in groundwater from the VDH and BSE datasets.

Although As is present in both sulfides (in the black/grey shale units) and oxides (in the red mudstone units) in the Gettysburg and Newark basins, the dominant release mechanism for As to groundwater in the has been suggested to be pH-dependent desorption from iron oxides in the mudstone units (Blake and Peters, 2014; Peters and Burkert, 2008; Serfes et al., 2010). The original source of As in the mudstones is most likely from As-rich pyrite in the shales (Peters and Burkert, 2008). Blake and Peters (2014) report that both the Gettysburg and Newark basins have similar depositional histories; however, the As occurrences in groundwater varied greatly between the two basins. The authors suggest that the difference in As concentrations are likely influenced by the paleo-depositional environment (Blake and Peters, 2014; Peters and Burkert, 2008).

Devonian Shales/Sandstones

The model results show that that Devonian-aged shales and sandstones are 11.19 times (Exp. Coefficient from Table 8) more likely to have elevated As concentrations in groundwater than other geologic units. A cluster of elevated As samples from both the BSE and VDH datasets occurs in northwestern Virginia in Frederick County (Figure 7). While this region is home to many historic orchards to which lead arsenate was likely applied, it is unlikely that As concentrations in groundwater are derived from anthropogenic sources because lead arsenate has low solubility and is not highly mobile in groundwater (Robinson et al., 2007).

One of the geologic units in the Devonian shales and sandstones in northwestern Virginia (Frederick County) is the Hampshire Formation. (Figure 11). The Hampshire Formation is composed of brown and green sandy shales, with thin bedded sandstones and red beds (Cleaves et al., 1968). In a study performed by the Maryland Geological Survey (Bolton et al., 2014)

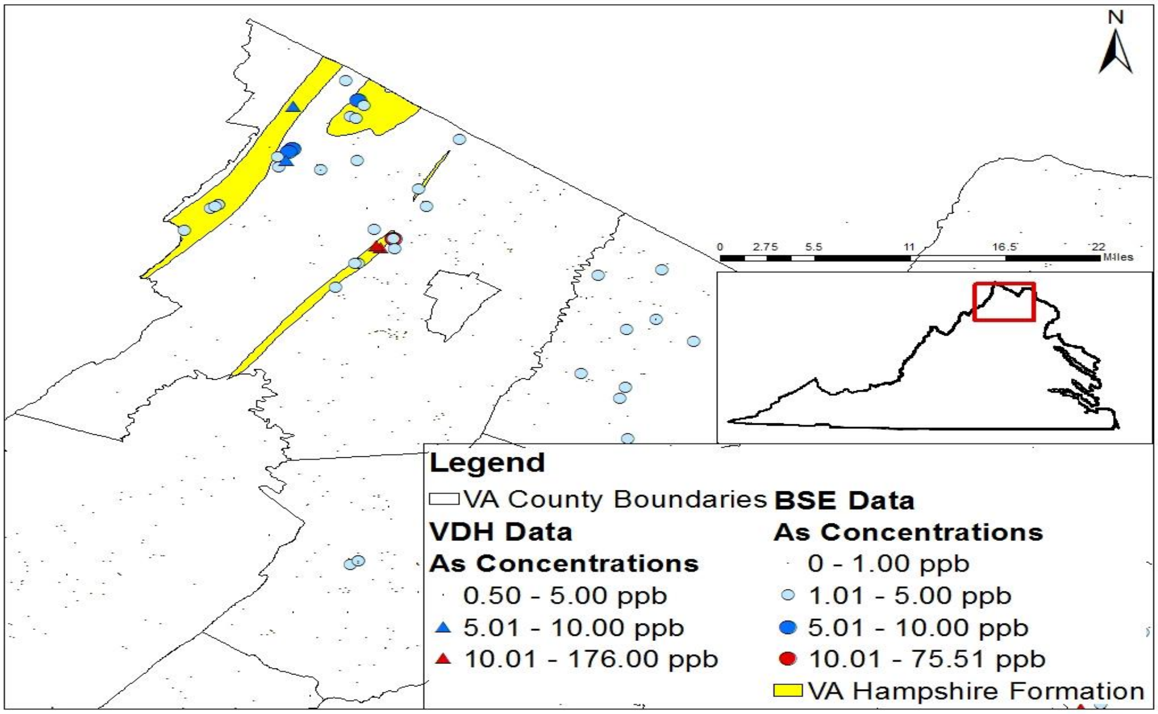


Figure 11. Spatial extent of the Hampshire Formation in northwestern Virginia with As concentrations in groundwater from the VDH and BSE datasets.

approximately 20% of groundwater samples collected from the Hampshire Formation had As concentrations greater than the EPA MCL of 10 ppb. Bolton et al. (2014) also showed that Devonian-aged rocks had a higher percentage of As concentrations (15%) above 10 ppb than Mississippian and Pennsylvanian-aged rocks (2-3%). Black shales, similar to the Devonian-aged shales, have been identified as sources for As release into groundwater in other areas (Paikaray, 2012). These black shales have elevated As concentrations due to their enhanced pyrite content, and may release As when in contact in organic-rich, reducing groundwater conditions (Smedley and Kinniburgh, 2002). In the New Albany Shale of Kentucky, As may be present in more than one mineral, however, pyrite is abundant in these Devonian shales and is the likely the source for As (Tuttle et al., 2009).

Devonian-aged shales and sandstones extend further down the Valley and Ridge Province. However, because our datasets have sparse data in these regions, the association between other Devonian-aged formations and elevated As cannot be evaluated in this study.

Other Areas of Interest

Coastal Plain

In the Coastal Plain, elevated As concentrations in groundwater occur mostly in the Northern Neck and on the Eastern Shore (Figure 12). Although the final logistic regression model did not identify any of the geologic units (various Tertiary gravel, sand, and clay units) in the Coastal Plain aquifers to be significant with respect to As in groundwater, we investigated wells along the Northern Neck and identified their respective aquifers as elevated As concentrations appeared to cluster in this area based on visual inspection of the data (see Figure 7).

This investigation was completed with the assistance of the Virginia DEQ and the USGS, who provided information on the screened intervals of individual wells. Once the elevation of the screened interval was identified, the top and bottom elevations of the aquifers were evaluated in ArcGIS to identify the aquifer to which each well was open. See Appendix D for figures displaying the aquifer elevations.

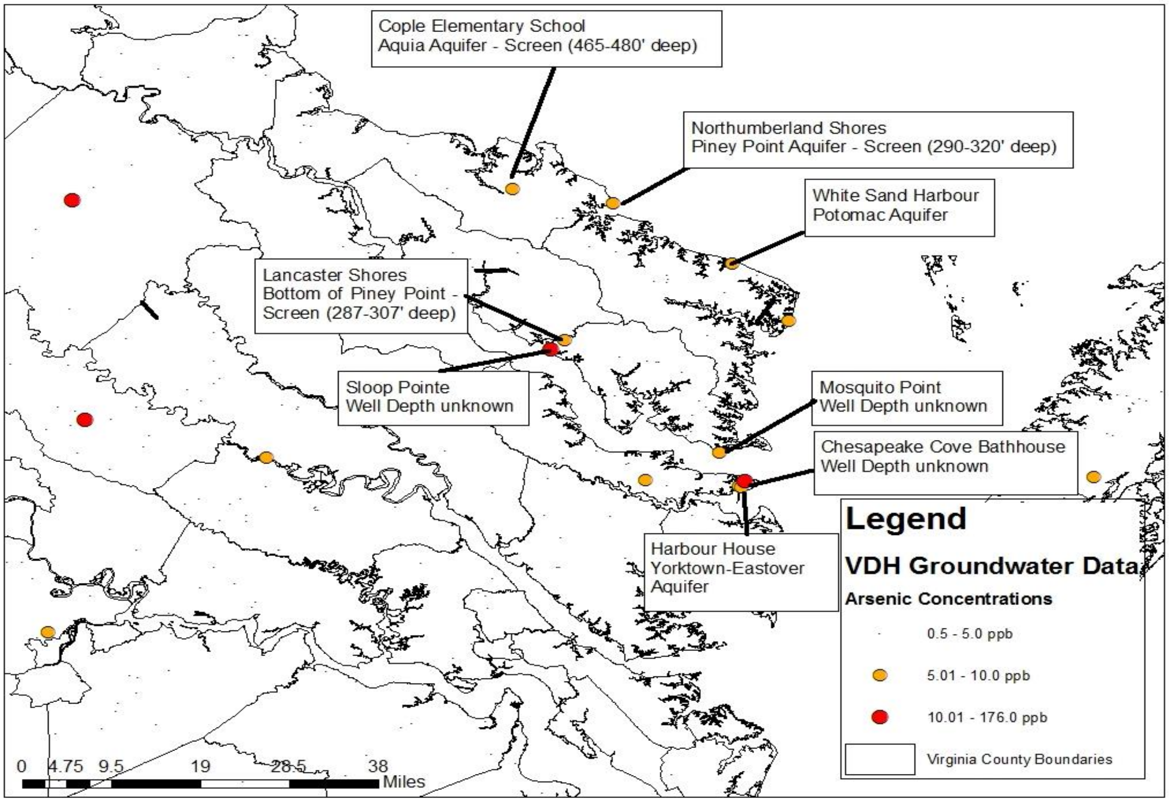


Figure 12. Locations and aquifers of samples in the Northern Neck of the Coastal Plain.

Previous studies by the Maryland Geological Survey identified the Aquia and Piney Point aquifers in the Coastal Plain as sources for elevated As in groundwater (Drummond and Bolton, 2010; Pearcy et al., 2011). Although the As source has not been definitively identified in those studies, there is some evidence to suggest that glauconite in Coastal Plain aquifers may be a source for As, with release of As through reductive dissolution of iron oxides (Barringer et al., 2010; Drummond and Bolton, 2010; Pearcy et al., 2011; Dooley et al., 2001). In addition to glauconite, As may be associated with at least two secondary phases, including iron hydroxides and siderite (Mumford et al., 2012; Barringer et al., 2011).

Stratigraphy likely plays an important role in influencing As concentrations in groundwater within the Coastal Plain. In the southern portion of the Virginia Coastal Plain, the Potomac Aquifer is not divided by confining units, as in the Northern Neck and Middle Peninsula, but is rather a single, thick aquifer (McFarland, 2013). In many places, the Potomac Aquifer is overlain by the Aquia Aquifer, which contains abundant glauconite. The close proximity of these aquifers and lack of confining units could facilitate mixing between the two aquifers. The Potomac Aquifer is lithologically similar to the Aquia and Piney Point aquifers (both contain glauconite and calcitic shell material).

Although we were not able to obtain information on screened intervals for wells in the Eastern Shore, they are likely open to the Yorktown-Eastover Aquifer, as the deeper aquifers contain saltwater (McFarland and Bruce, 2006). Both datasets in this study contained samples with As concentrations that exceed 5 ppb along the Eastern Shore (Figure 13). The source for As on the Eastern Shore is currently unclear, as there are potentially natural sources, including glauconite, but the Eastern Shore also has extensive poultry production. Poultry litter, applied to agricultural fields, may be a potential source for As in shallow groundwater (Denver et al., 2004). Denver and others (2004) completed a study of water quality along the Delmarva Peninsula (Eastern Shore of VA, Maryland, and Delaware). Results showed that As was detected (detection limit was 0.10 ppb) in approximately 50% of wells sampled in the surficial aquifer in agricultural areas (Yorktown-Eastover Aquifer in VA; Pocomoke Aquifer in MD); however, the As concentrations were generally below the EPA MCL of 10 ppb.

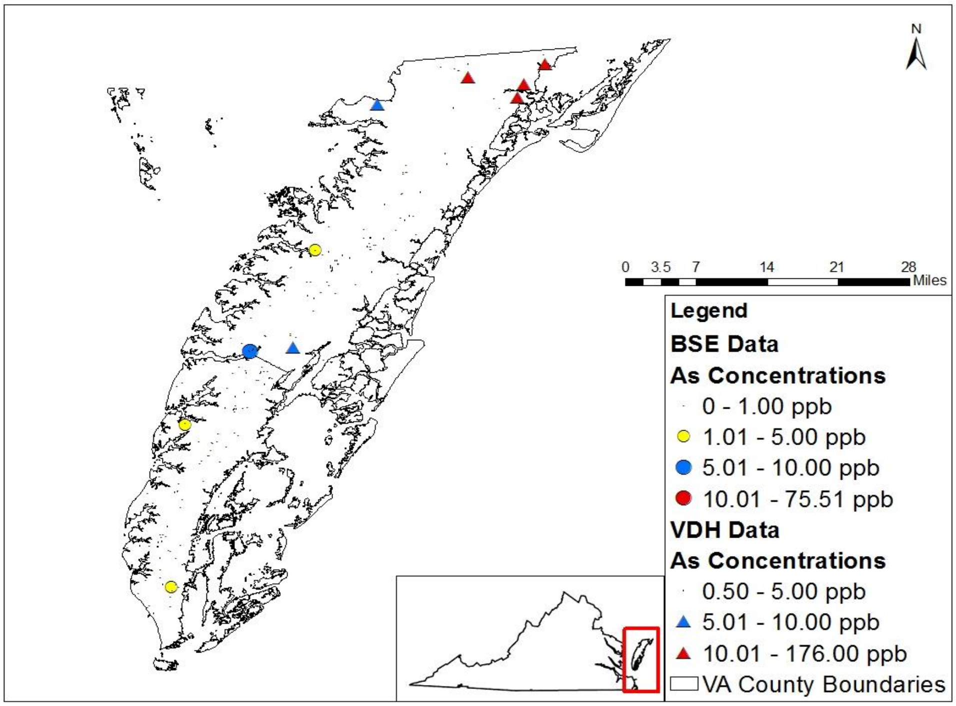


Figure 13. Spatial locations of As in groundwater from the VDH and BSE datasets in the Eastern Shore of Virginia.

Data Gaps

Our model indicated that two geologic units in Virginia have a higher probability of observing elevated As concentrations in groundwater. However, it should be noted that our datasets were based on available samples and do not have consistent sample coverage throughout the state. Thus, it is possible that areas with poor spatial distribution of samples, including counties in western, far southwestern, and southside Virginia, may have geologic units that are associated with elevated As concentrations. Additional information is needed to evaluate As associations with geologic units in these areas.

Chapter 5: Conclusions

The logistic regression model constructed in this study indicated that Triassic-aged sedimentary rocks and Devonian-aged shales/sandstones are associated with elevated As in groundwater in Virginia. From the model output and investigation of groundwater in areas of the Coastal Plain we interpreted that elevated As concentrations in groundwater may be observed in three areas (Table 10):

1) In the Culpeper Basin, elevated As concentrations in groundwater are associated with Triassic-aged sedimentary rocks. Comparison with studies of other Mesozoic basins (e.g. Newark, Gettysburg) suggests that the original source of As is from pyrite in the black shales, but that that the release of As to groundwater is due to desorption from oxides in the mudstones.

2) In the northwestern region of Virginia, elevated As concentrations in groundwater are associated with the Hampshire Formation, a Devonian shale. The source of As is likely sulfides in the shale; however, it is possible that iron oxides, resulting from sulfide oxidation, represent another source of As. The mobilization mechanism is currently unclear. While the Devonian-aged shales/sandstones extend farther along the Valley and Ridge Province, few As samples were collected in these areas and the association with elevated As could not be determined and;

3) The Virginia Coastal Plain, which had approximately 30 groundwater samples with elevated As, with most occurrences in the Northern Neck. Wells with elevated As in the Northern Neck are not limited to one or two aquifers, but include all four major aquifers in this region: Piney Point, Potomac, Yorktown-Eastover, and Aquia aquifers. These aquifers are made up of similar sedimentary material, including glauconite, which is potential source for As in groundwater. In the Eastern Shore, several wells with elevated As are open to the Yorktown-Eastover Aquifer.

However, the intense poultry production and widespread use of litter as fertilizer on the Eastern Shore may also influence As concentrations. At this time, there is not enough information from this study to differentiate between anthropogenic or natural sources in this area.

While the model was successful for evaluating potential geogenic sources of As to groundwater, the model fit analysis indicated that it should not be used for predictive purposes using our datasets. The poor model fit is a direct result of the distribution of As concentrations in our datasets, with most (95-99%) of the samples having concentrations less than or equal to the threshold value of 5 ppb. If another dataset on As concentrations in groundwater in Virginia becomes available, it could be included to further examine the predictability of our model.

Table 10. Sources and release mechanisms for arsenic in Virginia groundwater.

Region	Source	Possible Mobilization Mechanism	Examples of Similar Mechanisms and Sources
Culpeper Basin, Virginia	Triassic-aged sedimentary rocks and Jurassic Diabase	Sulfide oxidation; reductive dissolution of iron oxides; desorption from iron oxides	Serfes et al., 2010; Blake and Peters, 2014; Peters and Burkert, 2008.
Northwestern Virginia	Devonian Shale (Hampshire formation)	Sulfide oxidation; reductive dissolution of iron oxides; desorption from iron oxides	MGS, 2014; Tuttle et al., 2009
Coastal Plain/ Eastern Shore	Tertiary gravel and sands with glauconite; Land application of poultry litter	Release from glauconite; leaching from poultry litter	Drummond and Bolton, 2010; Denver et al. 2004; Pearcy et al., 2011

Suggestions for Further Research

Groundwater quality in the Culpeper Basin, northwestern portion of Virginia, and Coastal Plain of Virginia should be investigated further to examine the relationships between well depths, geologic units, and groundwater geochemistry, as these likely influence As concentrations in groundwater. Depths of wells from which samples were collected should also be documented and verified. While the model indicated two geologic units that were more likely to contain elevated As in groundwater, our datasets are limited by the spatial distribution. It is possible there are other geologic units within Virginia that may release As into groundwater; further study is needed to address this possibility.

References

Amini, M., Abbaspour, K.C., Berg, M., Winkel, L., Hug, S.J., Hoehn, E., Yang, H., and Johnson, C.A. 2008. Statistical Modeling of Global Geogenic Arsenic Contamination in Groundwater. Environmental Science and Technology, Volume 42, 3669-3675.

Ayotte, J.D., Nolan, B.T., Nuckols, J.R., Cantor, K., Robinson, G.R., and Baris, D. 2006. Modeling the Probability of Arsenic in Groundwater in New England as a Tool for Exposure Assessment. Environmental Science and Technology, Volume 40, Issue 11: 3578-3585.

Barringer, J.L., Mumford, A., Young, L.Y., Reilly, P.A., Bonin, J.L. and Rosman, R. 2010. Pathways for arsenic from sediments to groundwater to streams: Biogeochemical processes in the Inner Coastal Plain, New Jersey, USA. Water Research, Volume 44: 5532-5544.

Barringer, J.L., Reilly, P.A., Eberl, D.D., Blum, A.E., Bonin, J.L., Rosman, R., Hirst, B., Alebus, M., Cenno, K. and Gosrka, M. 2011. Arsenic in sediments, groundwater, and streamwater of a glauconitic Coastal Plain terrain, New Jersey, USA—Chemical ‘‘fingerprints’’ for geogenic and anthropogenic sources. Applied Geochemistry, Volume 26: 763-776.

Bednar, A.J., Garbarino, J.R., Ranville, J.F., and Wildeman, T.R. 2002. Preserving the Distribution of Inorganic Arsenic Species in Groundwater and Acid Mine Drainage Samples. Environmental Science and Technology, Volume 36, Issue 10: 2213–2218.

Berg, M., Trang, P.T.K., Stengel, C., Bushman, J., Viet, P.H., Dan, N.V., Giger, W., and Stüben, D. 2008. Hydrological and sedimentary controls leading to arsenic contamination of groundwater in the Hanoi area, Vietnam: The impact of iron-arsenic ratios, peat, river bank deposits, and excessive groundwater abstraction. Chemical Geology, Volume 249: 91-112.

Bhattacharya, P., A.B. Mukherjee, J. Bundschuh, R. Zevenhoven, and R. Loeppert. 2007. Arsenic in Soil and Groundwater Environment Biogeochemical Interactions, Health Effects and Remediation. Elsevier, Amsterdam, The Netherlands. Chapter 1.

Blake, J.M. and Peters, S.C. 2014. The occurrence and dominant controls on arsenic in the Newark and Gettysburg Basins. Science of the Total Environment, Volume 505: 1340-1349.

Bolton, D.W., Quinn, H.A., and Andreasen, D.C. 2014. Compilation and Review of GIS-Based Ground-Water Quality Data from Multiple Databases for Garrett County, Maryland. Administrative Report No. 14-02-02.1. Maryland Geological Survey.

Bräuner, E. V., R.B. Nordsborg, Z.J. Andersen, A. Tjønneland, S. Loft, and O. Raaschou-Nielsen. 2014. Long-term Exposure to Low-Level Arsenic in Drinking Water and Diabetes Incidence: A Prospective Study of the Diet, Cancer and Health Cohort. Environmental Health Perspectives, Volume 122, Issue 10: 1059-1065.

Burgess, W.G. and Pinto, L. 2005. Preliminary observations on the release of arsenic to groundwater in the presence of hydrocarbon contaminants in UK aquifers. Mineralogical Magazine, Volume 69, Issue 5; 887-896.

Cai, Y., Cabrera, J.C., Georgiadis, M., Jayachandran, K. 2002. Assessment of arsenic mobility in the soils of some golf courses in South Florida. The Science of the Total Environment, Volume 291: 123-134.

Chapman, M.L., Cravotta, C.A., Szabo, Z., and Lindsey, B.L. 2013. Naturally Occurring Contaminants in the Piedmont and Blue Ridge Crystalline-Rock Aquifers and Piedmont Early Mesozoic Basin Siliciclastic-Rock Aquifers, Eastern United States, 1994–2008

Cleaves, E.T., Edwards, J., Jr., Glaser, J.D., 1968, Geologic Map of Maryland: Maryland Geological Survey, Baltimore, Maryland, scale 1:250,000.

Cozzarelli, I.M., Schreiber, M.E., Erickson, M.L., Ziegler, B.A., 2016. Arsenic cycling in hydrocarbon plumes: Secondary effects of natural attenuation. Groundwater, Volume 1: 35-45.

Cule, E. and Iorio, M. 2012. A Semi-Automatic Method to Guide the Choice of Ridge Parameter in Ridge Regression. Annals of Applied Statistics.

deLemos, J., Bostick, B., Renshaw, C., Sturup, S., and Feng, X. 2006. Landfill-Stimulated Iron Reduction and Arsenic Release at the Coakley Superfund Site (NH). Environmental Science and Technology, Volume 40, 67-73.

Delmarva Poultry Industry Inc., Look what the poultry industry is doing for Delmarva, 2015. https://dpichicken.org/newsletters/docs/DPI%20IN%20ACTION%20MAY%2015.pdf.

Denver, J.M., Ator, S.W., Debrewer, L.M., Ferrari, M.J., Barbaro, J.R., Hancock, T.C., Brayton, M.J., and Nardi, M.R. 2004. Water Quality in the Delmarva Peninsula, Delaware, Maryland, and Virginia, 1999–2001. Reston, Virginia. United States Geological Survey.

Dooley, J.H. 2001. Baseline Concentrations of Arsenic, Beryllium and Associated Elements in Glauconite and Glauconitic Soils in the New Jersey Coastal Plain. The New Jersey Geological Survey Investigation Report. http://www.state.nj.us/dep/dsr/soils/baseline.pdf

Dove, P.M. and Rimstidt, J.D. 1985. The solubility and stability of scorodite, $FeAsO_4.2H_2O$. American Mineralogist, Volume 70: 838-844.

Drummond, D.D., and Bolton, D.W. 2010. Arsenic in groundwater in the Coastal Plain Aquifers of Maryland. Maryland Geological Survey Report of Investigations No. 78.

Dubey, B., Spalvins, E., Townsend, T., and Solo-Gabriele, H. 2009. Comparison of Metals Leaching from CCA- and ACQ-Treated Wood in Simulated Construction and Demolition Debris Landfills. Journal of Environmental Engineering, Volume 135, Issue10; 910−917.

ESRI, 2013. ArcGIS Desktop: Release 10.2. Redlands, CA: Environmental Systems Research Institute.

Ettinger, A.S., Zota, A.R., Amarasiriwardena, C.J., Hopkins, M.R., Schwartz, J., Hu, H., and Wright, R.O. 2009. Maternal arsenic exposure and impaired glucose tolerance during pregnancy. Environmental Health Perspective, Volume 117, Issue 7: 1059-1064.

Fajčíková, K., Cvečková, V., Stewart, A., and Rapant, S. 2014. Health risk estimates for groundwater and soil contamination in the Slovak Republic: a convenient tool for identification and mapping of risk areas. Environmental Geochemistry and Health, Volume 36: 973-986.

Farzan, S.F., Chen, Y., Rees J.R., Zens, M.S., Karagas, M.R. 2015. Risk of death from cardiovascular disease associated with low-level arsenic exposure among long-term smokers in a US population-based study. Toxicology and Applied Pharmacology, Volume 287, Issue 2: 93-97.

Fei, D.L., Koestler, D.C., Li, Z., Giambelli, C., Sanchez-Mejias, A., Gosse, J.A., Marsit, C.J., Karagas, M.R., Robbins, D.J., 2013. Association between In Utero arsenic exposure, placental gene expression, and infant birth weight: a US birth cohort study. Environmental Heath, Volume 12: Issue 58.

Garbarino J.R., Bednar, A.J., Rutherford, D.W., Beyer, R.S., and Wershaw, R.L. 2003. Environmental fate of roxarsone in poultry litter. I. Degradation of roxarsone during composting. Environmental Science and Technology, Volume 37, Issue 8: 1509-1514.

Garelick, H., Jones, H., Dybowska, A., and Valsami-Jones, E. 2008. Reviews of Enviornmental Contamination. Volume 197.

Gong, G. and O’Bryant, S.E. 2012. Low-level arsenic exposure, AS3MT gene polymorphism and cardiovascular diseases in rural Texas counties. Environmental Research, Volume 113: 52-57.

Gress, 2015. Cleaning-induced arsenic mobilization and chromium oxidation from CCA-wood deck: Potential risk to children. Environment International, Volume 82: 35-40.

Harvey, M.C., Schreiber, M.E., Rimstidt, J.D., and Griffith, M.M. 2006. Scorodite Dissolution Kinetics: Implications for Arsenic Release. Environmental Science and Technology, Volume 40, Issue 21: 67-0-6714.

Hossain, M. and Piantanakulchai, M. 2013. Groundwater arsenic contamination risk prediction using GIS and classification tree method. Engineering Geology, Volume 156: 37-45.

Ismael, S.I. and Kharbish, S. 2013. Removing of As (V) from Aqueous Solution Using Natural and Pretreated Glauconite and Halloysite. Carpathian Journal of Earth and Environmental Sciences, Volume 8, Issue 1: 187-198.

Kent, D.B. and Fox, P.M. 2004. The influence of groundwater chemistry on arsenic concentrations and speciation in a quartz sand and gravel aquifer. Geochemical Transactions, Volume 5, Issue 1: 1-12.

Kent, D.V., Olsen, P.E., and Witte, W.K. 1995. Late Triassic-earliest Jurassic geomagnetic polarity sequence and paleolatitudes from drill cores in the Newark rift basin, eastern North America. Journal of Geophysical Research, Volume 100: 14,965-14,998.

Korte, N. 1991. Naturally occurring arsenic in groundwaters of the midwestern United States. Environmental Geology and Water Sciences, Volume 18, Issue 2: 137-141.

Larkins, P., Reed, Bradley, and Gilpin R. Robinson, Jr., 20061008, Digital Data Set of Orchards Where Arsenical Pesticides Were Likely Used in Clarke and Frederick Counties, Virginia and Berkeley and Jefferson Counties, West Virginia: U.S. Geological Survey Open-File Report 2006-1330.

Lan, C.C, Yu, H.S., and Ko, Y.C. 2011. Chronic arsenic exposure and its adverse health effects in Taiwan: A paradigm for management of a global environmental problem. Kaohsiung Journal of Medical Sciences, Volume 27; 411-416.

Lee, K.Y. 1977. Triassic Stratigraphy in the Northern Part of the Culpeper Basin, Virginia and Maryland. United States Geological Survey Bulletin 1422-c.

Li, Y. and Chen, T. 2005. Concentrations of additive arsenic in Beijing pig feeds and the residues in pig manure. Resources, Conservation and Recycling, Volume 45: 356–367.

Lui, C.W., Wang, S.W., Jang, C.S., and li, K.H. 2006. Occurrence of Arsenic in Ground Water in the Choushui River Alluvial Fan, Taiwan. Journal of Environmental Quality, Volume 1: 68-75.

Ma, L.Q., Hanis, W., and Sartain, J. 2000. Environmental Impacts of Lead Pellets at Shooting Ranges & Arsenical Herbicides On Golf Courses in Florida. State University System of Florida Florida Center for Solid and Hazardous Waste Management, Report #00-03. http://soils.ifas.ufl.edu/lqma/Publication/MA-00-R.pdf.

Mandal, B.K. and Suzuki, K.T. 2002. Arsenic round the world: a review. Talanta, Volume 58: 201-235.

McFarland, E.R. 2013. Sediment Distribution and Hydrologic Conditions of the Potomac Aquifer in Virginia and Parts of Maryland and North Carolina. Scientific Investigations Report 2013–5116. http://pubs.usgs.gov/sir/2013/5116/pdf/sir2013-5116.pdf.

McFarland, E.R. and Bruce, T.S. 2006. The Virginia Coastal Plain Hydrogeologic Framework. Professional Paper 1731. http://pubs.usgs.gov/pp/2006/1731/PP1731.pdf

Mercer, T.G. and Frostick, L.E. 2014. Evaluating the potential for environmental pollution from chromated copper arsenate (CCA)-treated wood waste: A new mass balance approach. Journal of hazardous Material, Volume 276: 10-18.

Michael, H. 2013. An Arsenic Forecast for China. Science, Volume 341: 852-853.

Mumford, A.C., Barringer, J.L., Benzel, W.M., Reilly, P.A., and Young, L.Y. 2012. Microbial transformations of arsenic: Mobilization from glauconitic sediments to water. Water Research, Volume 46: Pages 2859-2868.

National Research Council. 2001. Arsenic in Drinking Water. The National Academies Press, Washington D.C.

New Jersey Department of Environmental Protection. 2004. Dep Adopts New Mercury, Arsenic Standards: New Rules Are Strictest in the Nation. News Release. http://www.nj.gov/dep/newsrel/2004/04_0130.htm.

Nickson, R., McArthur, J., Burgess, W., Ahmed, K.M., Ravenscroft, P., Mizanur Rahman, M. 1998. Arsenic poisoning of Bangladesh groundwater. Nature, Volume 395: 338.

Nordstrom, D. K. 2002. Worldwide Occurrences of Arsenic in Ground Water. Science, Volume 296, Issue 5576: 2143-2145.

Nriagu, J.O., J.M. Pacyna. 1988. Quantitative assessment of worldwide contamination of air, water, and soil by trace metals. Nature, Volume 333: 134–139.

O'Day, P.A., Vlassopoulos, D., Meng, X., Benning, L.G. 2005. Advances in Arsenic Research; ACS Symposium Series; American Chemical Society: Washington, DC.

Olsen, P.E., Kent, D.V., Cornet, B., Witte, W.K., and Schlische, R.W. 1996. High-resolution stratigraphy of the Newark rift basin (early Mesozoic, eastern North America). GSA Bulletin, Volume 108: 40-77.

Oyewumi,O. Madeline E. Schreiber, M.E. 2012. Release of arsenic and other trace elements from poultry litter: Insights from a field experiment on the Delmarva Peninsula, Delaware. Applied Geochemistry, Volume 27: 1979-1990.

Paikaray, S. 2012. Environmental hazards of arsenic associated with black shales: a review on geochemistry, enrichment and leaching mechanism. Reviews in Environmental Science and Technology, Volume 11: 289–303.

Pearcy, C.A., Chevis, D.A., Haug, T.J., Jeffries, H.A., Yang, N., Tang, J., Grimm, D.A., and Johannesson, K.H. Evidence of microbially mediated arsenic mobilization from sediments of the Aquia aquifer, Maryland, USA. Applied Geochemistry, Volume 26: 575-586.

Penga, Q., Harlow, S.D., and Parka, S.K. 2015. Urinary arsenic and insulin resistance in US adolescents. International Journal of Hygiene and Environmental Health, Volume 218: 407–413.

Peryea, F. J. and Creger, T. L. 1994. Vertical distribution of lead and arsenic in soils contaminated with lead arsenate pesticide residues. Water, Air, and Soil Pollution, Volume 78: 297-306.

Peters, S.C. and Burkert, L. 2008. The occurrence and geochemistry of arsenic in groundwaters of the Newark basin of Pennsylvania. Applied Geochemistry, Volume 23, Issue 1: 85-98.

Pippin, C.E. 2005. Distribution of Total Arsenic in Groundwater in the North Carolina Piedmont Radium, Radon, and Uranium. NGWA Naturally Occurring Contaminants Conference: Arsenic, February 2005.

Polizzotto, M.L., Harvey, C.F., Sutton, S.R., and Fendorf, S. 2005. Processes conducive to the release and transport of arsenic into aquifers of Bangladesh. Proceedings of the National Academy of Sciences in the United States, Volume 102, Issue 52: 18819–18823.

Powell, J.T., Townsend, T.G., and Tolymat, T.M. 2015. Does Disposing of Construction and Demolition Debris in Unlined Landfills Impact Groundwater Quality? Evidence from 91 Landfill Sites in Florida. Environmental Science and Technology, Volume 49, Issue 15: 9029-9036.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.Rproject.org/.

Reedy, R.C., Scanlon, B.R., Nicot, J.P., and Techovsky, J.A. 2007. Unsaturated Zone Arsenic Distribution and Implications for Groundwater Contamination. Environmental Science and Technology, Volume 41: 6914-6919.

Robinson, G.R. Jr., Ayotte, J.D. 2006. The influence of geology and land use on arsenic in stream sediments and ground waters in New England, USA. Applied Geochemistry, Volume 21: 1482–1497.

Robinson, G. R. Jr., P. Larkins, C.J. Boughton, B.W. Reed, and P.L. Sibrell. 2007. Assessment of Contamination from Arsenical Pesticide use on Orchards in the Great Valley Region, Virginia and West Virginia, USA. Journal of Environmental Quality, Volume 36, Issue 3: 654-663.

Ruiz-Navarro, M.L., Navarro-Alarcón, M., Lopez G-de la Serrana, H. Pérez-Valero, V., and M.C López-Martinez, M.C. 1998. Urine arsenic concentrations in healthy adults as indicators of environmental contamination: Relation with some pathologies. Science of the Total Environment, Volume 216, Issue 1-2: 55-61.

Rutherford, D.W., Bednar, A.J., Garbarino, J.R., Needhman, R., Staver, K.W., Wershaw, R.L. 2003. Environmental fate of roxarsone in poultry litter. Part II. Mobility of arsenic in soils amended with poultry litter. Environmental Science and Technology, Volume 37: 1515-1520.

Schooley, T., Weaver, M., Mullins, D., and Eick, M. 2008. The History of Lead Arsenate Use in Apple Production: Comparison of its Impact in Virginia with Other States. Journal of Pesticide Safety Education, Volume 10: 22-53.

Schreiber, M.E., Simo, J.A., and Freiberg, P.G. 2000. Stratigraphic and geochemical controls on naturally occurring arsenic in groundwater, eastern Wisconsin, USA. Hydrogeology Journal, Volume 8, Issue 2: 161-176.

Schrenk, M.O., Edwards, K.J., Goodman, R.M., Hamers, R.J., and Banfield, J.F. 1998. Distribution of *Thiobacillus ferrooxidans* and *Leptospirillum ferrooxidans*: Implications for Generation of Acid Mine Drainage. Science, Volume 279, Issue 5356: 1519-1522

Senior, L.A., Sloto, R.A., 2006. Arsenic, Boron, and Fluoride Concentrations in Ground Water in and Near Diabase Intrusions, Newark Basin, Southeastern Pennsylvania. USGS Scientific Investigations Report 2006–5261.

Serfes, M., 2005. Arsenic occurrence, sources, mobilization, transport and prediction in the major bedrock aquifers of the Newark Basin. Ph.D. Thesis, Rutgers University.

Serfes, M.E., Herman, G.C., Spayd, S.E., and Reinfelde, J. 2010. Sources, Mobilization and Transport of Arsenic in Ground Water in the Passaic and Lockatong Formations of the Newark Basin, New Jersey. N.J. Geological Survey Bulletin 77, Chapter E of Contributions to the Geology and Hydrogeology of the Newark Basin.

Shiber, J.G. 2005. Arsenic in domestic well water and health in central Appalachia, USA. Water, Air, and Soil Pollution Volume 160: 327–341.

Silbergeld, E.K. and Nachman, K. 2008. The Environmental and Public Health Risks Associated with Arsenical Use in Animal Feeds. Annals of the New York Academy of Sciences, Volume 1140: 346-357.

Singh, N., Kumar, D., and Sahu. A. P. 2007. Arsenic in the Environment: Effects on Human Health and Possible Prevention. Journal of Environmental Biology. Volume 28; Issue 2: 359-365.

Singh, S.K. and Vedwan, N. 2015. Mapping composite vulnerability to groundwater arsenic contamination: an analytical framework and a case study in India. Natural Hazards, Volume 75, Issue 2: 1883-1908.

Smedley, P.L. and Kinniburgh, D.G. 2002. A review of the source, behaviour and distribution of arsenic in natural waters. Applied Geochemistry, Volume 17: 517-568.

Smith, A.H., Lingas, E.O., and Mahfuzar Rahman, M. 2000. Contamination of drinking-water by arsenic in Bangladesh: a public health emergency. Bulletin of the World Health Organization, Volume 78, Issue 9: 1093-1103.

Solo-Gabriele, H. and Townsend, T. 1999. Disposal practices and management alternatives for CCA-treated wood waste. Waste Management and Research, Volume 17: 378-389.

Thompson, G.R. and Hower, J. 1975. The Mineralogy of Glauconite. Clays and Clay Minerals, Volume 23: 289 300.

Tibsharani, R. 1996. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society, Volume 58, Issue 1: 267-288.

Tuttle, M.L.W., Breit, G.N., and Goldhaber, M.B. 2009. Weathering of the New Albany Shale, Kentucky: II. Redistribution of minor and trace elements. Applied Geochemistry, Volume 24: 1565-1578.

United States Geological Survey. 1988. Hydrogeologic Framework of the Virginia Coastal Plain. U.S. Geological Survey Professional Paper 1404-C. http://pubs.usgs.gov/pp/pp1404-C/pdf/pp_1404-c.pdf (accessed December 2, 2014).

United States Geological Survey. 1992. The Regional Aquifer System Underlying the Northern Atlantic Coastal Plain in Parts of North Carolina, Virginia, Maryland, Delaware, New Jersey, and New York – Summary. USGS Professional Paper 1404-A. http://pubs.usgs.gov/pp/1404a/report.pdf (accessed November 30, 2014).

USEPA. 2003. Comprehensive Environmental Response, Compensation, and Liability Information System database; U.S. EPA: Washington, DC (accessed March 3, 2016)

Wasserman, G.A., Liu, X., Lolcono, N.J., Kline, J., Factor-Livtak, P., vanGeen, A., Mey, J.L., Levy, D., Abramson, R., Schwartz, A., and Graziano, J.H. A cross-sectional study of well water arsenic and child IQ in Maine schoolchildren. Environmental Health, Volume 13, Issue 23.

Welch, A.H., Westjohn, D.B., Helsel, D.R., and Wanty, R.B. 2000. Arsenic in ground water of the United States-- occurrence and geochemistry: Ground Water v.38 no.4, p.589-604.

Whitmore, T.J., Riediniger-Whitmore, M.A., Smoak, J.M., Kolasa, K.V., Goddard, E.A., Bindler, R. 2008.Arsenic contamination of lake sediments in Florida: evidence of herbicide mobility from watershed soils. Journal of Paleolimnology, Volume 40, Issue 3: 869-884.

Williams, M. 2001. Arsenic in mine waters: an international study. Environmental Geology, Volume 40: 267-278.

Yang, N., Winkel, L.H.E., and Johannesson, K.H. 2014. Predicting Geogenic Arsenic Contamination in Shallow Groundwater of South Louisiana, United States. Environmental Science and Technology, Volume 48: 5660-5666.

Yang, Q., Jung, H.B., Marvinney, R.G., Culbertson, C.W., and Zheng, Y. 2012. Can Arsenic Occurrence Rates in Bedrock Aquifers Be Predicted? Environmental Science and Technology, Volume 46: 2080-2087.

Appendices

Appendix A. Drinking Water Sampling Procedures

Taken from the EPA Quick Guide to Drinking Water Sampling, April 2005.

Metals to for Sample Collection:

Antimony, Arsenic, Barium, Beryllium, Cadmium, Calcium, Chromium V, Magnesium, Manganese, Mercury, Nickel, Selenium, Sodium, Silver, Thallium, Lead, Copper, Zinc, and others

Bottles to Use:

Plastic or glass bottles may be used but plastic is preferred. Preservative to Use Nitric Acid (HNO3) to pH

Holding Time:

28 days for mercury, 6 months for other metals

Sampling Instructions:

1. Prepare a Sampling and Analysis Plan (SAP) which describes the sampling locations, numbers and types of samples to be collected, and the quality control requirements of the project.

2. Check with the laboratory before collecting samples to ensure that sampling equipment, preservatives, and procedures for sample collection are acceptable. It is best to obtain sampling supplies directly from the laboratory performing the analyses. Gather all equipment and supplies necessary for the project.

3. The acids and bases used in preservation of many types of samples described in this document are dangerous and must be handled with care. Always wear gloves and eye protection when handling preservatives. When opening a preservative bottle, particularly a glass ampoule, break open the ampoule away from yourself and others. Have acid/base neutralization supplies (baking soda) on hand in the event of a spill.

4. Collect samples in an area free of excessive dust, rain, snow or other sources of contamination.

5. Select a faucet for sampling which is free of contaminating devices such as screens, aeration devices, hoses, purification devices or swiveled faucets. Check the faucet to be sure it is clean. If the faucet is in a state of disrepair, select another sampling location.

6. Collect samples from faucets which are high enough to put a bottle underneath, generally the bath tub or kitchen sink, without contacting the mouth of the container with the faucet. If you are collecting a first-flush sample for lead/copper, allow the water to run just a bit before collecting the sample but do not flush the lines as you want to collect a sample which has been in contact with the distribution system pipes for at least six hours.

8. If you are collecting other types of samples, open the faucet and thoroughly flush. Generally 2 to 3 minutes will suffice, however longer times may be needed, especially in the case of lead distribution lines. Generally, the water temperature will stabilize which indicates flushing is completed. Once the lines are flushed, adjust the flow so it does not splash against the walls of the bathtub, sink or other surfaces.

9. Follow the collection instructions provided for the analytes of interest described on the following pages. Wear eye protection and gloves if you are handling containers with acidic/basic preservatives and when you are collecting samples.

10. Use a sample tag to record the site location, name of the sampler, date and time of collection, method of collection, type of analysis to be completed, and preservative in use. Attach the sample tag to the bottle.

11. Fill out the chain of custody form with the sample collection information.

12. Deliver or ship samples to the laboratory to ensure that holding times are met.

13. Return empty preservative containers to the laboratory for proper disposal.

Appendix B. ArcGIS Procedures

ArcGIS 10.2 software was used in all geospatial mapping and analyses.

Each dataset was received as an excel file. The excel file was loaded into an ArcMap document and converted to a table. The x-y data was then displayed which converted the table to a shapefile. The datum NAD83 UTM zone 18 was used to display the data as well as each of the GIS layers. Each sample location was mapped spatially and color gradation was chosen to represent As concentration ranges in groundwater. From the mapped sample locations, it was observed that elevated As concentrations exist in the northern Virginia region with sporadic elevated As throughout the state.

Each of the layers as described in the body of the thesis (Geologic Units, Lithology, Physiographic Province, Land Use, Soil Series and Particle Size) were added to ArcMap as a shapefile with the exception of land use, which was added as in a raster format. Boundary data for counties in Virginia was obtained from the United States Census Bureau.

In order to create a complete data set, where the As sampling locations are associated with each of the attributes in each of the layers, a spatial join was created between the As locations and layer attributes. This spatial join was created by choosing As to be located within its associated feature. For example, if an As sample was located within the Manassas Soil Series, the spatial join will associate the As sample with the Manassas Soil Series.

The completed data set was then exported as an excel file where the data were formatted to be input into the statistical software R. This process was repeated for each dataset to obtain a complete data set.

Appendix C. Statistical Software R Code

```
###### Create variables for all categorical variable ########

mydata <-(VDH)

head(mydata)

dim(mydata)

colnames(mydata)

colnames(mydata)<-c("AS", "LATITUDE", "LONGITUDE", "Land_Use", "PROVINCE","Geologic.Unit",
"Lithology",  "SoilSeries","SoilParticleSize")

#add one fake line

fake <- c(rep(1,3),rep("a",6))

fake_res <- rep(0:1,times=c(609,dim(mydata)[1]-608))

mydata[,4:9] <- sapply(mydata[,4:9] ,as.character)

mydata_new <- rbind(mydata,fake)

mydata_new[,4:9] <- sapply(mydata_new[,4:9],as.factor)

mydata_new[,1:3] <- sapply(mydata_new[,1:3],as.numeric)

tail(mydata_new)

str(mydata_new)

dim(mydata_new)

A<-model.matrix(fake_res~AS + Geologic.Unit + Lithology + PROVINCE + Land_Use + SoilSeries +
SoilParticleSize,data=mydata_new)

colnames(A)

dim(A)

mydata_all<- A[1:(dim(A)[1]-1),2:dim(A)[2]]

head(mydata_all)

tail(mydata_all)

dim(mydata_all)

write.table(mydata_all, "BSEcategorized.csv", col.names=TRUE,row.names=FALSE,sep=",")

####### Logistic LASSO Regression – to Select Variables ########

mydata <-(BSEcategorized)

dim(mydata)

mydata$As_binary <- as.numeric(mydata$AS>5)

library(glmnet)

myLASSOLogit<-glmnet(as.matrix(mydata[,-c(1:1,dim(mydata)[2])]),mydata$As_binary, family =c ("binomial"),
alpha=1)
```

```
print(myLASSOLogit)
plot(myLASSOLogit)
#plot variable coefficients vs. shrinkage parameter lambda.
plot(myLASSOLogit,xvar="lambda")
grid()
#cross validation can also be used to select lambda.
cv.myLASSOLogit<- cv.glmnet(as.matrix(mydata[,-c(1:3,dim(mydata)[2])]), mydata$As_binary, family=c("binomial"), alpha=1)
plot(cv.myLASSOLogit)
best_lambda <- cv.myLASSOLogit$lambda.min
best_lambda
log(best_lambda)
#myLASSOcoef <- coef(myLASSOLogit, s=best_lambda)
myLASSOcoef <- coef(myLASSOLogit, s=exp(-4.954366))
sig_LASSO <- myLASSOcoef[which(abs(myLASSOcoef[,1])>0),]
length(sig_LASSO)
sig_LASSO[order(abs(sig_LASSO))]
####### Logistic Ridge Regression – to find Significance for each regressor ########## - library(ridge)
mydata <- (BSEcategorized)
dim(mydata)
mydata$As_binary <- as.numeric(mydata$AS>5)
mylogitRidge1 <- logisticRidge(As_binary ~ Geologic.Unit_Tm +Geologic.Unit_D +Geologic.Unit_M+ Geologic.Unit_Tr +Geologic.Unit_DS, data = mydata[,-c(1:3)], lambda = "automatic")
summary(mylogitRidge1)
print(mylogitRidge1)
plot(mylogitRidge1)
mylogitRidge1_est <- pvals(mylogitRidge1)
index1 <- which(mylogitRidge1_est$pval[ ,1]<0.1)
round(data.frame(coef=mylogitRidge1_est$coef[index1,1],
expCoef=exp(mylogitRidge1_est$coef[index1,1]),SE=mylogitRidge1_est$se[index1,1],tstat=mylogitRidge1_est$tstat[index1,1], p_value=mylogitRidge1_est$pval[index1,1]),digit=6)
######Rerun Ridge Model with only significant regresors to confirm significance######
mylogitRidge1 <- logisticRidge(As_binary ~ Geologic.Unit_Tm +Geologic.Unit_D +Geologic.Unit_M+ Geologic.Unit_Tr +Geologic.Unit_DS, data = mydata[,-c(1:3)], lambda = "automatic")
summary(mylogitRidge1)
```

```
print(mylogitRidge1)

plot(mylogitRidge1)

mylogitRidge1_est <- pvals(mylogitRidge1)

index1 <- which(mylogitRidge1_est$pval[ ,1]<0.1)

round(data.frame(coef=mylogitRidge1_est$coef[index1,1], expCoef=exp(mylogitRidge1_est$coef[index1,1]),
SE=mylogitRidge1_est$se[index1,1],tstat=mylogitRidge1_est$tstat[index1,1],
p_value=mylogitRidge1_est$pval[index1,1]),digit=6)

######Use VDH Data to Test BSE Model ########## - library(ridge)

train_temp <- (BSEcategorized)

train <- train_temp[ ,c("AS", "Geologic.UnitTm", "Geologic.UnitD", "Geologic.UnitM"," Geologic.UnitTr",
"Geologic.UnitDS")]

train$As_binary <- as.numeric(train$AS>5)

trainX <- as.matrix(train[ ,c("Geologic.UnitTm","Geologic.UnitD","Geologic.UnitM","Geologic.UnitTr",
"Geologic.UnitDS")])

test_temp <- (VDH_final_categorized2)

test <- test_temp[ ,c("AS", "Geologic.Unit_Tm", "Geologic.Unit_D", "Geologic.Unit_M", "Geologic.Unit_Tr",
"Geologic.Unit_DS")]

test$As_binary <- as.numeric(test$AS>5)

testX <- as.matrix(test[ ,c("Geologic.Unit_Tm", "Geologic.Unit_D", "Geologic.Unit_M", "Geologic.Unit_Tr",
"Geologic.Unit_DS")])

colnames(testX) <- c("Geologic.UnitTm", "Geologic.UnitD", "Geologic.UnitM", "Geologic.UnitTr",
"Geologic.UnitDS")

mylogitRidge <- logisticRidge(As_binary ~Geologic.UnitTm + Geologic.UnitD +Geologic.UnitM + Geologic.UnitTr
+ Geologic.UnitDS, data = train, lambda = “automatic”)

mylogitRidge_est <- pvals(mylogitRidge)

index <- which(mylogitRidge_est$pval[ ,1]<0.1)

round(data.frame(coef=mylogitRidge_est$coef[index,1], expCoef=exp(mylogitRidge_est$coef[index,1]),
SE=mylogitRidge_est$se[index,1],tstat=mylogitRidge_est$tstat[index,1],
p_value = mylogitRidge_est$pval[index,1]), digit=6)

mypred <- predict(mylogitRidge, newdata = as.data.frame(testX), type = c("response"))

which(mypred>0.5)

P_all <- sum(test$As_binary==1)
N_all <- sum(test$As_binary==0)
P_pred <- length(which(mypred>=0.5))
N_pred <- length(which(mypred<0.5))

TP <- 0
FN <- P_all
FP <- 0
TN <- N_all
matrix(c(TP, FN, FP, TN), nrow=2, byrow=TRUE)
```

Appendix D. Additional Figures

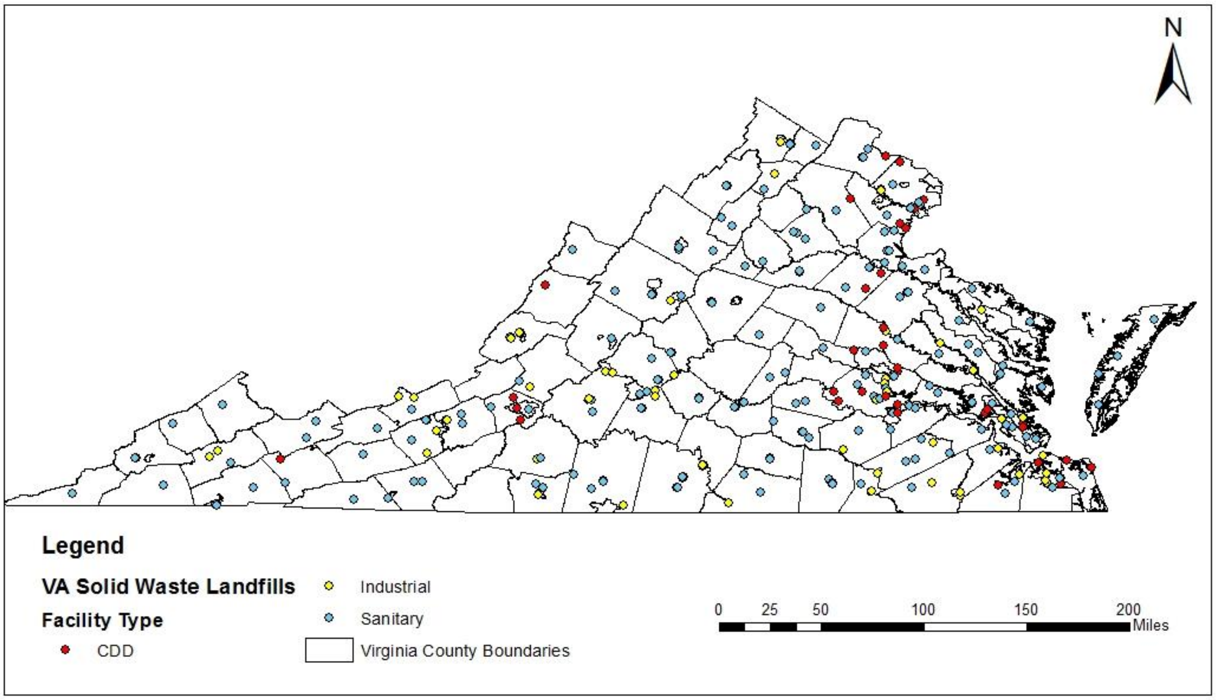


Figure D1. Landfill locations in Virginia.

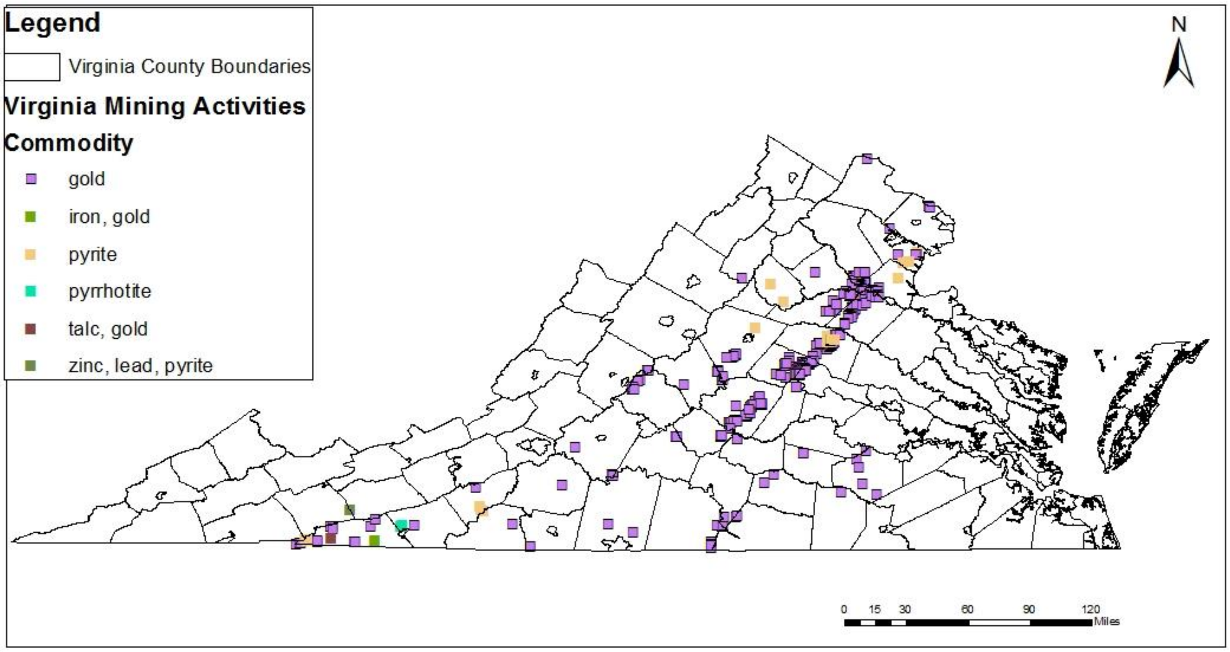


Figure D2. Mining activities in Virginia.

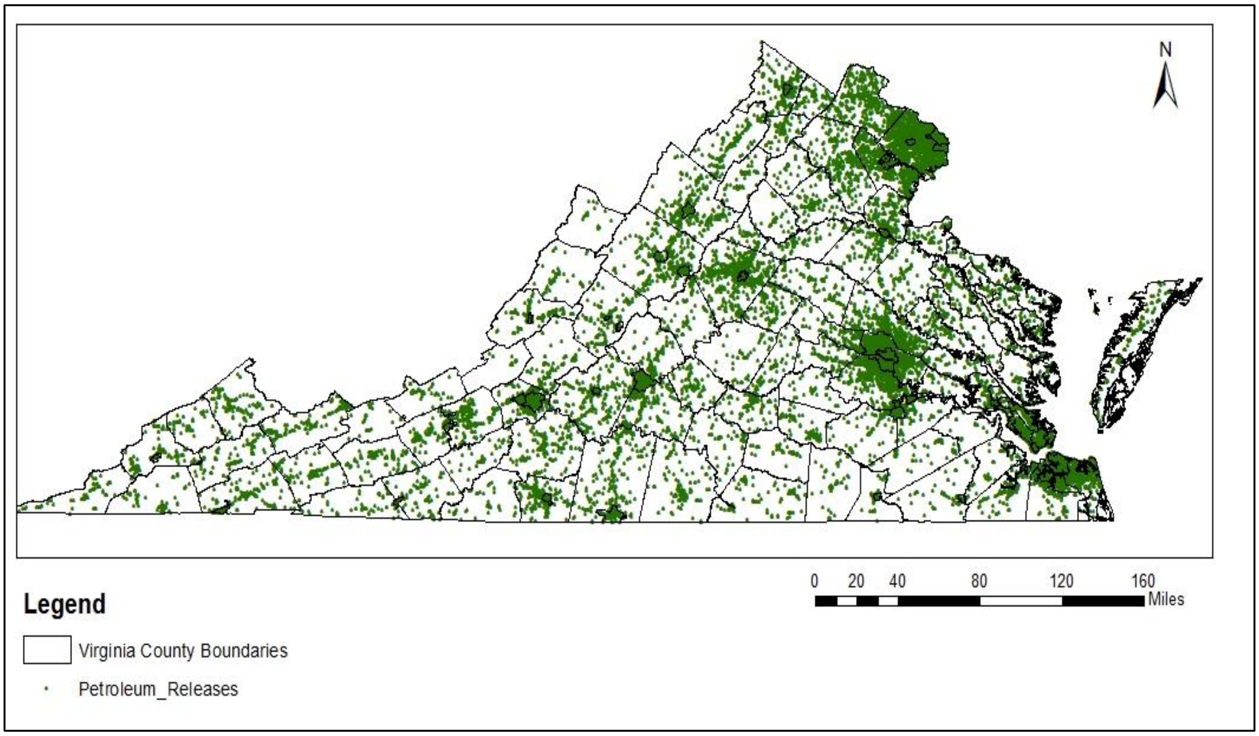


Figure D3. Documented petroleum releases in Virginia since the late 1970's.

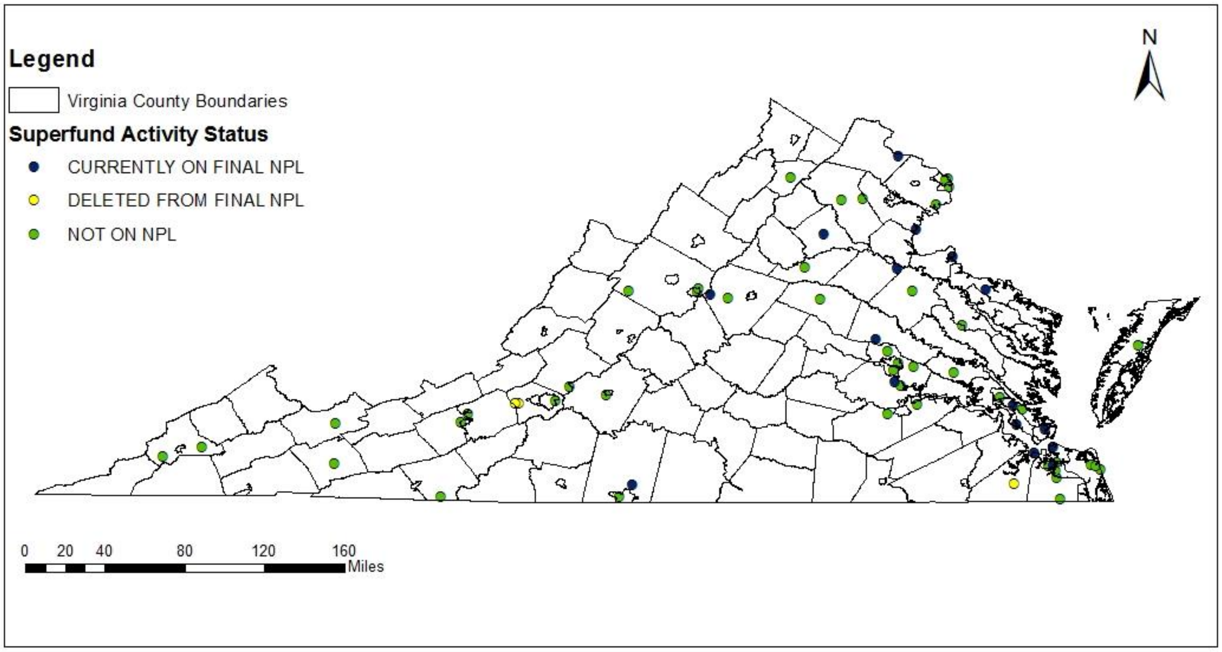


Figure D4. Superfund sites in Virginia, including sites both off and on the National Priorities List (NPL).

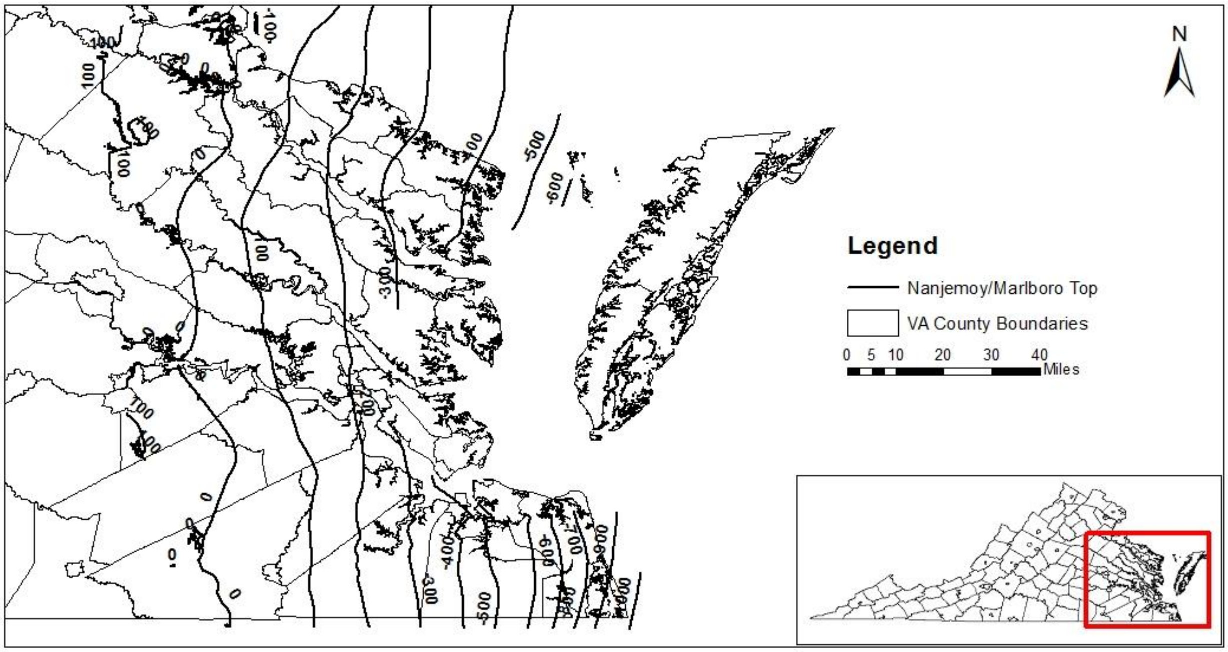


Figure D5. Surface contours for the Nanjemoy/Marlboro confining unit.

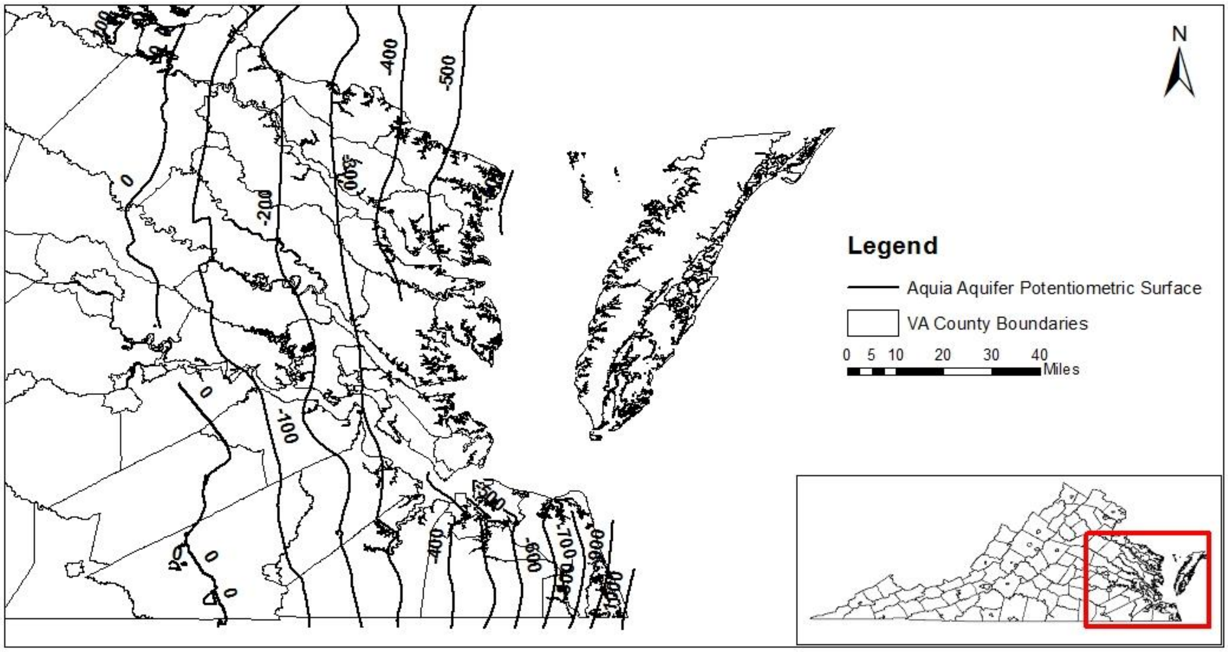


Figure D6. Surface contours for the Aquia Aquifer.

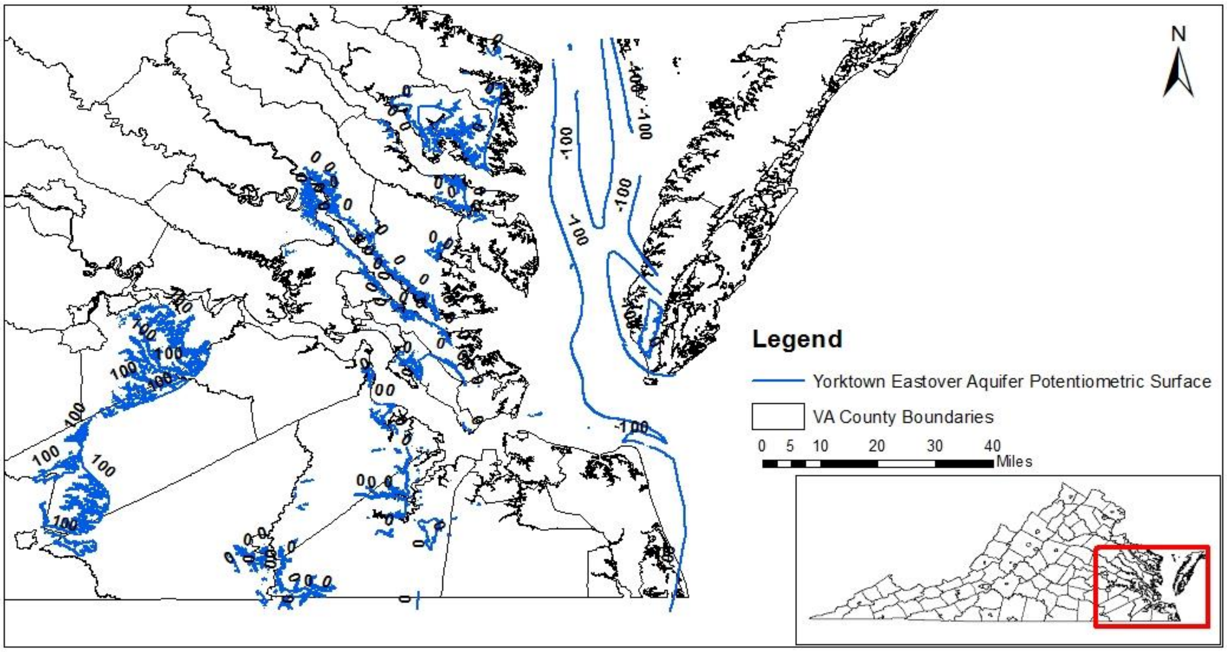


Figure D7. Surface contours for the Yorktown-Eastover Aquifer.

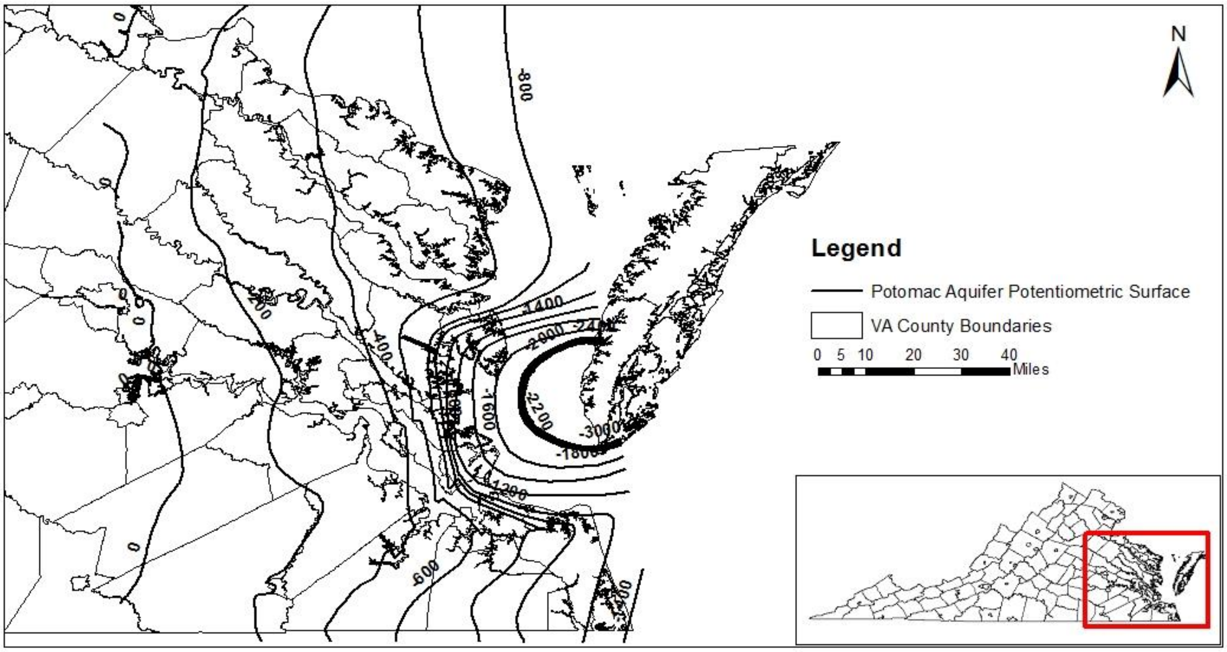


Figure D8. Surface contours for the Potomac Aquifer.

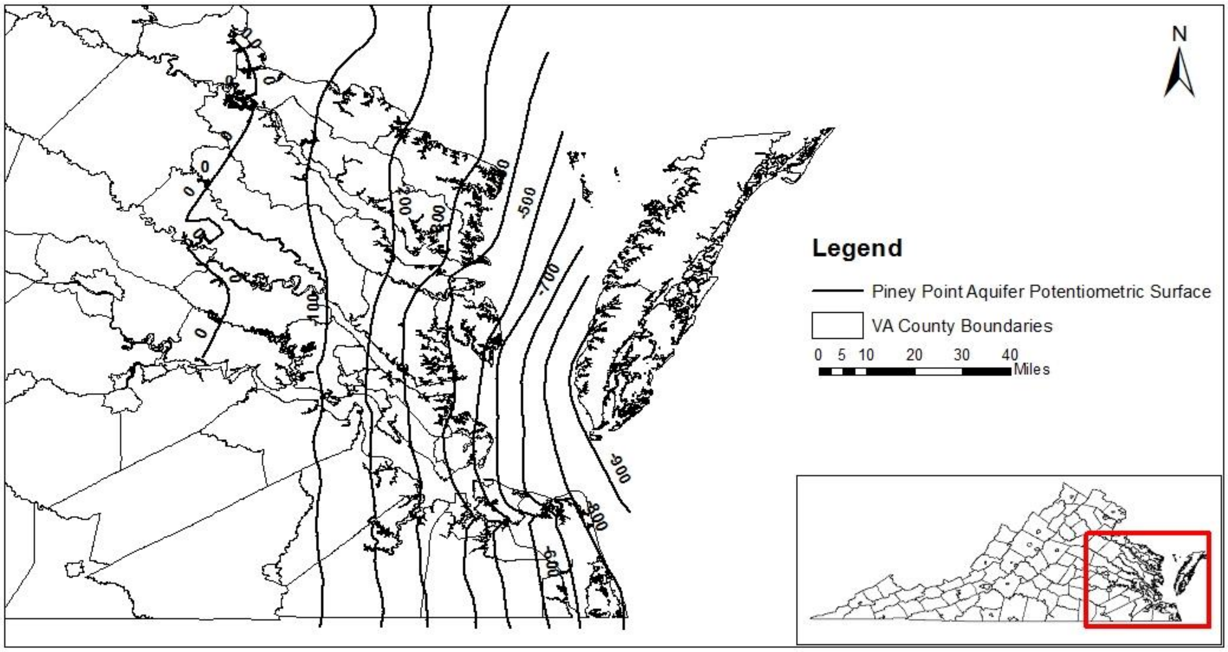


Figure D9. Surface contours for the Piney Point Aquifer.

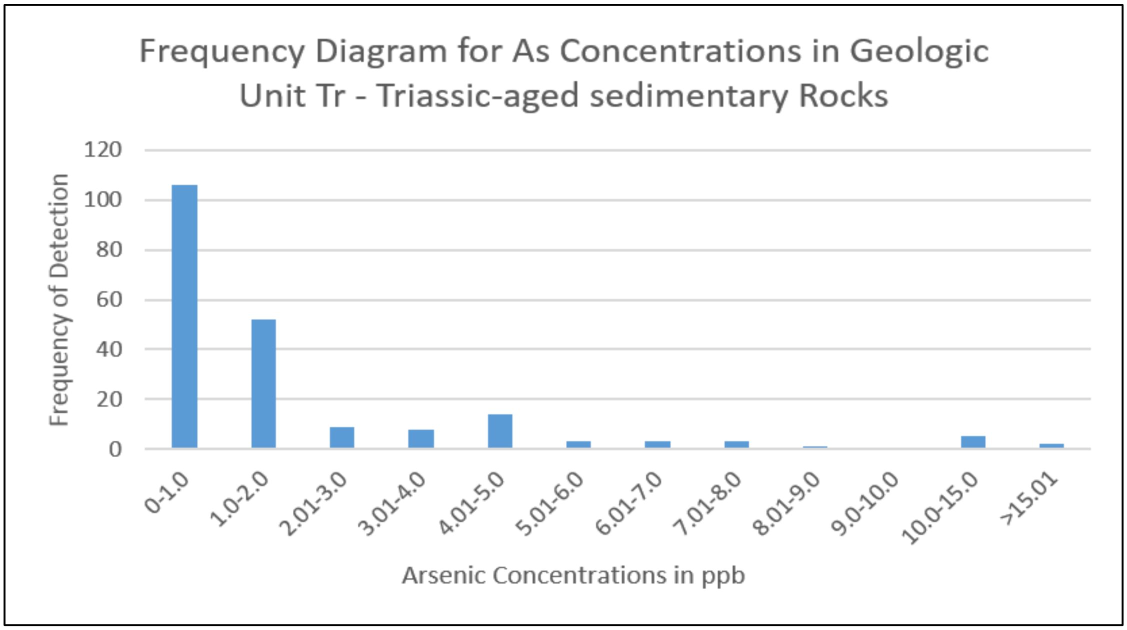


Figure D10. Frequency diagram for As concentrations in geologic unit Tr - Triassic-aged sedimentary rocks. Data is from the BSE and VDH datasets.

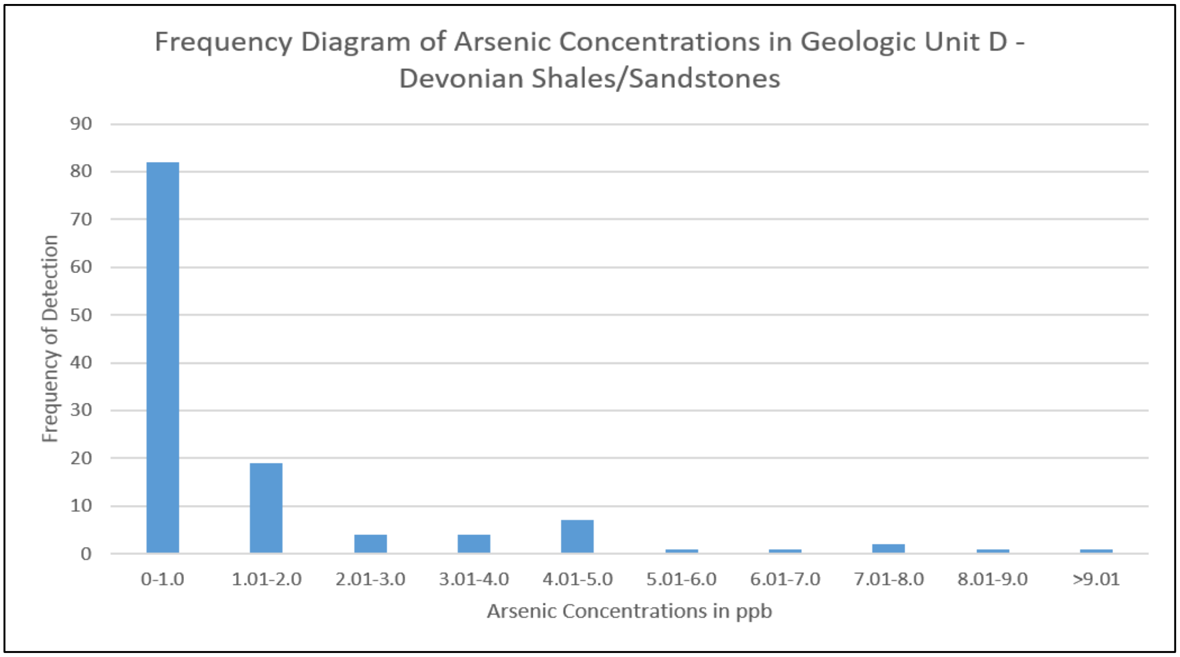


Figure D11. Figure D10. Frequency diagram for As concentrations in geologic unit D - Devonian-aged shales/sandstones. Data is from the BSE and VDH datasets.

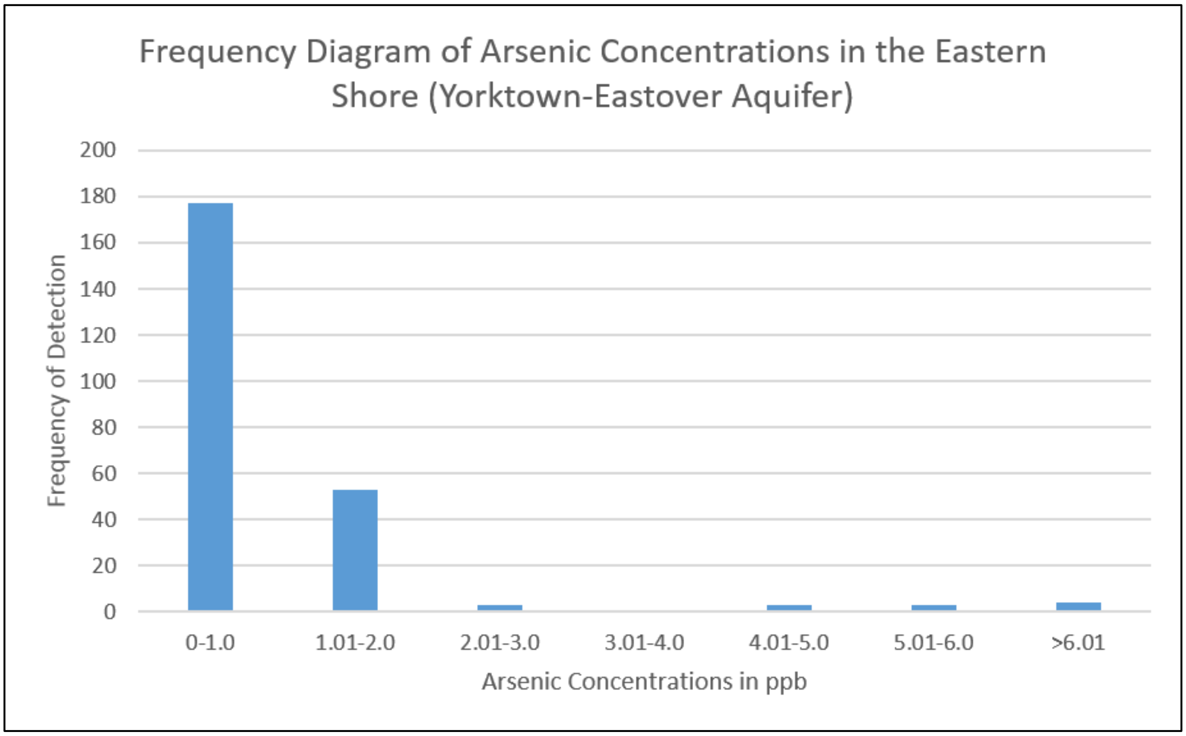


Figure D12. Figure D10. Frequency diagram for As concentrations in the Eastern Shore (Yorktown-Eastover Aquifer). Data is from the BSE and VDH datasets.