PFCs in Aquifer Systems: The Influence of Landscape Attributes on the Spatial Distribution of Perfluorooctanoic acid (PFOA) and Perfluorooctanesulfonic acid (PFOS) in New Hampshire Aquifers

Ashley M. Sutton

ABSTRACT

Perfluorinated chemicals (PFCs) negatively impact the hormonal and reproductive health of organisms while persisting long term in the environment. Although there have been numerous studies examining PFC concentrations and their movement in water, these papers focus on surface water systems and exclude groundwater, the largest source of the world's drinking water. I used an exploratory and ordinary least squares regression along with a regression tree analysis comparing New Hampshire well concentrations of Perfluorooctanesulfonic acid (PFOS) and Perfluorooctanoic acid (PFOA) to 28 different variables describing anthropogenic features, the surface environment, and geologic characteristics. These analyses examined whether factors shown to influence surface water PFC concentrations translated to groundwater, and if additional variables were needed to increase model accuracy. In both the linear regressions and the regression tree, solid waste disposal facilities, industry, distance to the closest known aquifer, and markers of developed land were important to predicting PFC levels. For the exploratory regressions both PFOA and PFOS were correlated to 4 anthropogenic features, 1 surface descriptor, and 2 geologic characteristics. However, the Moran's I test for spatial autocorrelation indicated that both models had clustered residuals and were missing key explanatory variables. Although the linear regression and regression tree resulted in a similar model accuracy for PFOA with an R squared of 0.4, there was a marked improvement using the regression tree for PFOS increasing the R squared from 0.04 to 0.3. The high presence of anthropogenic features in the final models indicates that, like surface water, groundwater PFC concentrations are related to sources of pollution. Nevertheless, geologic characteristics are also critical in understanding PFC distributions in aquifers.

KEYWORDS

Endocrine disruptors, public health, groundwater, multi-variate regression, wells

INTRODUCTION

Perfluorinated chemicals (PFCs) are compounds whose exposures cause harmful impacts to the environment and human health (Caliman and Gavrilescu 2009, Webster 2010, Vélez et al. 2015). PFCs are non-flammable and highly stable man-made chemicals used in various applications for fire suppression, cleaning solvents, heat transfer fluids, and atmospheric tracers (Tsai et al. 2002, Webster 2010). Contact with dust, water, and consumer products containing PFCs, along with bioaccumulation in food sources have all been established as possible routes for PFC entrance to the human body (Kubwabo et al. 2005, Tittlemier et al. 2007, Ericson et al. 2008, Guo and Krebs 2009). Although initially thought to be biologically inert, examining freshwater fish indicates that cholesterol regulation, cellular response time, cell migration, and organization of the cytoskeleton are all impacted from PFC exposure (Collí-Dulá et al. 2016). In humans, PFCs have been indicated in decreased fertility in women, decreased sperm count in men, and alteration of reproductive hormones in males exposed in utero (Calafat et al. 2007, Vested et al. 2013, Vélez et al. 2015). Laboratory studies on rodents and monkeys also indicate significant effects on development, life span, behavior, and alterations in hormone levels (Lau et al. 2007, Johansson et al. 2008, White et al. 2009). PFCs have detectable influences on biologic systems and can result in long-term negative health effects.

With indications of global dispersal, and consequently widespread exposure, the potential health impacts from PFCs are arousing concern. Because PFCs are synthetic products of human manufacturing the main output to the environment is through production, usage, and disposal of items containing PFCs. These products can range from cosmetics to furniture foam, and when disposed then break down and enter water systems and aquatic food chains (Webster 2010, National Institute of Environmental Health 2016). As a result of their chemical stability PFCs are generally not removed during the waste water treatment process and can persist in municipal waste water for years after being removed from production due to continued consumer use (Caliman and Gavrilescu 2009, Houtz et al. 2016). This wide spread use has translated into the detection of PFCs around the world, with remote regions such as the Tibetan Plateau and the northern reaches of Canada showing evidence of PFC concentrations in local animal tissue analysis (Kubwabo et al. 2005, Shi et al. 2010). Although there has been some evidence to indicate water as a global transport mechanism, allowing for PFCs to distribute themselves so widely, understanding the process of PFC movement throughout the environment is an area that requires more examination (Taniyasu et al. 2013). However, even with the knowledge that PFCs are now globally present, the spatial pattern of PFC contamination is only beginning to be understood.

As PFCs move throughout the environment, especially through water systems, climate and human factors are influencing their dispersal patterns. Seasons and temperature have measurable effects on PFC concentrations in surface water on an annual basis, specifically in estuaries and bays (Montagner and Jardim 2011, Rocha et al. 2014). In addition, human influences such as pollution point sources and land use can cause differences in the PFC levels within a surface water system (Zushi and Masunaga 2010). The presence of military fire training areas, waste water treatment plants and industrial sites are all key determinants of PFC concentrations in surface water (Hu et al. 2016). However, though there are clear patterns of PFCs in surface water, depending on various factors of time and space, analysis into the spatial patterns of PFCs in aquifer systems is lacking. Because long-term storage holds a large amount of the world's potable water, spatial analysis of PFCs to encompass these critical below ground resources is critical.

This study examined the spatial patterns of PFC concentrations in well water throughout the watersheds of New Hampshire. Specifically, I aimed to determine how physical landscape characteristics influence the spatial pattern of PFCs in aquifer systems and how important geologic characteristics are to long-term spatial distributions of PFCs. I included anthropogenic features (population density, location of superfund sites and industry, land cover, location of airports and train stations, percentage of impervious surfaces, waste disposal sites, arterial traffic area, location of sewage treatment plants, and military fire training locations), surface environment (river and lake areas, elevation, and soil runoff potential), and geologic characteristics (aquifer extent and location, transmissivity, saturation thickness, water table elevation, quaternary deposits, location of seismic fault lines, and bedrock type). Using an exploratory regression, ordinary least squares, and regression tree analysis to compare these characteristics against the distribution of PFC concentrations in New Hampshire I aimed to determine which factors are highly correlated to PFC levels. I expect that these results will elucidate the relative importance of aquifer characteristics in their contamination.

METHODS

Southern New Hampshire Watershed

Southern New Hampshire's aquifer system is the product of millennia of tectonic and erosional forces shaping the landscape into its current form. When the North American, Eurasian, and African plates collided, during the creation of Pangea over 300 million years ago, it caused the forces pushing on the continental plate of North America to thrust and fault the bedrock of what is currently New Hampshire, compressing it and folding it over (Billings 1956). With the extreme heat and pressure caused by the convergence, the bedrock was converted to hard and erosionally resistant crystalline rock. The breakup of Pangea 150 million years later introduced volcanic activity into New Hampshire, leading to the insertion of igneous rock formations among the crystalline bedrock (Figure 1) (Billings 1956).

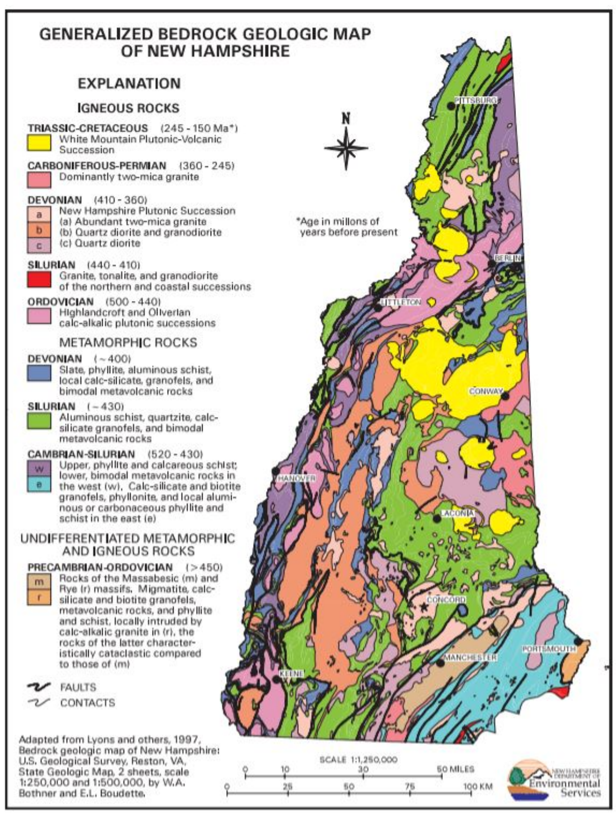

Figure 1. Generalized Bedrock Geologic map of New Hampshire
Source: New Hampshire Department of Environmental Services

Overlying the bedrock are surficial deposits (unconsolidated sediment) that were primarily created from the deglaciation of the Laurentide Ice Sheet, which retreated less than 14 thousand years ago (Goldthwait et al. 1951). Erosion by plucking, abrasion, and scour happened during the life of the glacier with additional sediment deposit and redistribution occurring as the glacier melted and retreated. The composition of the resulting surficial deposits is dependent on the parent bedrock type and the form of erosion that took place.

New Hampshire has two main types of aquifers (an area of rock that contains and can move groundwater); stratified drift and bedrock aquifers. Stratified drift aquifers are the result of surficial deposits of layered sand and gravel that determines the flow, recharge abilities, and storage of aquifers (Stekl and Flanagan 1992, Medalie and Moore 1995). The majority of these aquifers were deposited in large valleys and other lowland areas from when rising sea levels and proglacial lakes covered these regions (Randall 2001). Characteristics of each aquifer, such as yield (amount of water transmitted) and storage ability, are determined by the size of the deposit particles, with larger particle aquifers holding more water and having a higher yield (Medalie and Moore 1995). Stratified drift aquifers underlie fourteen percent of New Hampshire and are important to public, industry, and commercial water supply (Medalie and Moore 1995).

Bedrock (parent rock material) always lies beneath surficial deposits and can contain water in fractures of various size and extent creating bedrock aquifers (Medalie and Moore 1995). As stratified drift aquifers occur primarily in lowland areas and valleys, not all cities have one easily accessible. Because of this, bedrock aquifers can be the only source of aquifer water available to some communities (Medalie and Moore 1995). Due to the crystalline rock structure, water doesn't move through the rock to reach these water stores, but rather through fractures in the material (United States Geologic Survey 2016). In 2016, ninety percent of private wells in New Hampshire were connected to bedrock aquifers. Overall, between both stratified drift and bedrock aquifers "46% of New Hampshire households get water from private wells," according to the supervisor of the Department of Environmental Services drinking water division (MacKenzie 2016)

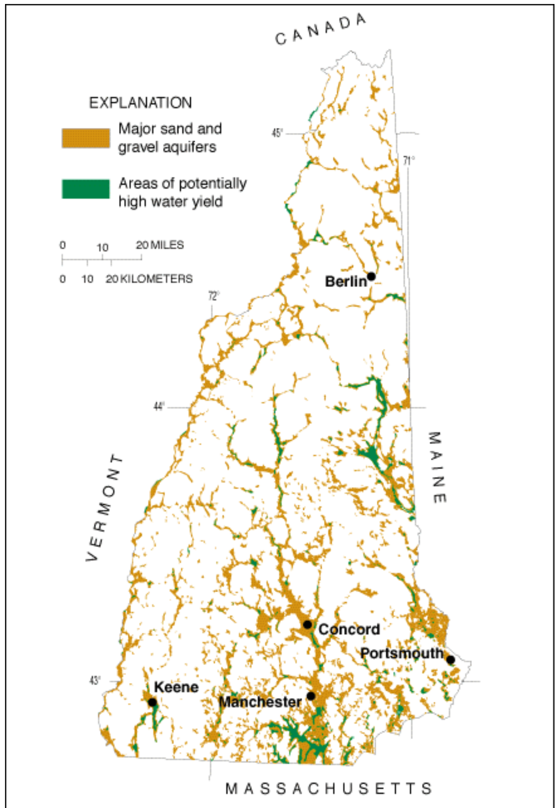

Figure 1. Location of major sand and gravel aquifers and areas of potentially high water yield in New Hampshire.

Figure 2. Sand and Gravel Aquifers of New Hampshire
Source: USGS

Data Collection Methods

Perfluorinated Compound Data

For the outcome variable of the exploratory regression, ordinary least squares, and regression tree analysis, I included data samples of PFCs from private and public wells across New Hampshire, data that was collected during the New Hampshire Department of Environmental Services' (NHDES) state-wide perfluorinated chemicals in drinking water investigation. This inquiry began in 2016 when a plastic production company notified the NHDES of elevated Perfluorooctanoic Acid (PFOA) levels in water faucets at their facility (NHDES) . From there NHDES collected water samples from wells around known point source polluters such as firefighting training facilities, landfills, industrial production sites, and areas of human health concern such as schools and private drinking wells. Samples were tested by labs certified by the Department of Defense or the National Environmental Laboratory Accreditation Program (NELAP) and used reporting limits of 5 nanograms per liter (*Laboratory Testing Guidelines for Per- and Polyfluoroalkyl Substances (PFAS)* 2016). NELAP accredited labs used either standard isotope dilution in their methods or EPA Method 537 Rev 1.1 (*Technical Advisory - Laboratory Analysis of Drinking Water Samples for Perfluorooctanoic Acid (PFOA) Using EPA Method 537 Rev. 1.1* 2016). According to the EPA's Unregulated Contaminant Monitoring Rules (UCMR) Round 3 lists a set of chemicals that must be encompassed in the analysis. This includes Perfluorooctanesulfonic acid (PFOS), Perfluorooctanoic acid (PFOA), Perfluorononanoic acid (PFNA), Perfluorohexanesulfonic acid (PFHxS), Perfluoroheptanoic acid (PFHpA), and Perfluorobutanesulfonic acid (PFBS). In addition, because Perfluorobutanoic acid (PFBA), Perfluoropentanoic acid (PFPeA), and Perfluorohexanoic acid (PFHxA) are regularly found in New Hampshire groundwater they were included in the testing (*Laboratory Testing Guidelines for Per- and Polyfluoroalkyl Substances (PFAS)* 2016). I chose to focus this analysis on two of the tested PFCs; PFOA and PFOS. This publicly available data set was one main reason for New Hampshire being chosen as the site for this study, because other data on PFCs is difficult to find at this scale.

Because the testing sites for the dependent variables were conducted at well locations, the data set is comprised of highly clustered sample locations. This is due to the fact that wells occur at a higher frequency in connection to areas of high population density. Although this sample location distribution is not desirable for spatial analysis, the location of aquifer access points via wells constrains the data collection potential.

Key Variables

This data analysis included data for 28 different variables in the categories of human factors (15), surface environmental data (5), and geologic data (8) (Table 1). The human factors included land use, population density and percent of impervious surfaces. These factors are expected to be correlated with the amount of PFCs introduced into the environment as are distance to potential point source pollutants such as military fire training locations, TRI reporting manufacturing facilities, superfund sites, total road length, sewage treatment plant locations and outfalls, airport and train station locations, and solid waste disposal sites. The surface environmental data included pathways of movement such as soil and surface water through soil runoff potential, river and lake locations, along with elevation and slope. For geologic data I used characteristics of aquifers that noted their ability to move water and contaminants along with potential connectivity (Stekl and Flanagan 1992). These variables were aquifer size, perimeter area ratio, transmissivity, and saturation thickness along with elevation of the water tables, quaternary deposits, bedrock type, and seismic fault lines. A complete list detailing data source information for each of these variables is in Appendix A.

To compile the various data layers into one table I primarily used intersect, near, reclassify, extract values to points, and join by spatial geometry in ArcGIS 10.5.1. By using these methods I made several key assumptions. The applicability of the near tool rests on the basis that Euclidean distance will account for the relationship between the PFC concentrations at each well location and distance from tested variables. This increases the chance that this model won't correctly account for more complex connections. Though this isn't ideal, as a first approximation this assumption is acceptable.

Additionally, when using both the tool intersect and extract values to points I assume that the only significant influence from each variable is directly at the specific well locations. Because many of the rasters used in this analysis were products of other government organizations and already contain some level of spatial simplification in their raster classification, I wanted to avoid further averaging of the data so as to minimize the risk of excluding existing correlations due to over simplification.

Table 1. Data Variables Included in Regression

Categories	Data Variable
Dependent Variable	PFOA Concentrations
	PFOS Concentrations
Human Influence	Distance to Military Fire Training Locations
	Distance to Superfund Sites
	Distance to TRI Reporting Facilities
	Land Cover Type
	Percent Impervious Surface
	Population Density
	Total Road Length Around Wells
	Distance to the Closest Road
	Distance to Waste Water Treatment Plants (WWTP)
	Distance to WWTP Violators
	Distance to Airports
	Distance to Train Stations
	Distance to Solid Waste Disposal Site Facilities (SWF)
	Distance to Unlined SWF
	Distance to Sewage Outfalls
Surface Environment	Elevation
	Slope
	Soil Runoff Potential
	Distance to Rivers
	Distance to Lakes
Geologic Characteristics	Aquifer Area
	Perimeter Area Ratio
	Aquifer Transmissivity
	Aquifer Saturation Thickness
	Elevation of Water Table
	Surficial Deposits
	Primary Bedrock Type
	Seismic Fault Lines

Data Analysis Methods

Exploratory Regression and Ordinary Least Squares Analysis

To analyze the spatial and statistical differences between each of the 28 variables and tested PFC levels of varying ranges I used ArcGIS 10.5.1 (Esri 2017). To begin I used the Getis-Ord Gi* Hotspot Analysis on my initial PFC concentration dataset. The analysis compares the sum of neighboring features to the sum of all features. When the local sum is significantly different from the expected sum, and the difference is too large to be attributed to chance, you get a resulting Z-value that is statistically significant. This value then indicates where features with high and low values are spatially clustered (Esri 2017). Therefore, the output identified regions of elevated PFC concentrations across the state.

Because regressions use linear relationships, it is typically assumed that the independent variables and the residuals of the model follow a normal distribution. The normality assumption is necessary to unbiasedly estimate standard errors of the coefficients of the regression model, and hence confidence. Therefore, I then evaluated my variables to determine which needed to be normalized through a Boxcox transformation or logarithmic transformation before adding them to the final database. An exploratory regression in this software was used to remove redundant variables while also determining the correlations between variables. Using this form of regression allowed for the testing of multiple models for their fit, using the variables in my database, before deciding on the best combination of factors. This model also showed the statistical significance of each correlation, the model's stationarity, and the model's bias. To incorporate the categorical data into the regression each classification was transformed into binary dummy variables.

To compare regression models I used the Akaike Information Criterion (AIC), with lower values indicating better model fit. I then used an Ordinary Least Squares (OLS) model to compute the final regression with best fit model variables from the exploratory regression. To examine whether all key factors were accounted for I used Moran's I statistic test for spatial dependence on the OLS residuals.

Classification and Regression Tree Analysis

Although the exploratory and OLS regressions determined correlations between the 28 selected variables and aquifer PFC concentration levels it does have several restrictions. Linear regressions give the best results when the input data is normal. However, if data are not completely normal, even when transformed, is not ideal for this analysis, although some research debates this assertion (Schmidt and Finan 2017). Additionally, using a linear regression required binary variable substitutes for categorical data and it doesn't handle outliers well. Therefore, although OLS can be more accurate, I used Classification and Regression Trees as an additional analysis (Briand et al. 2000). This method requires no normalization, is built to handle categorical information, and can manage for outliers (Timofeev 2004).

Because the dependent variable of the trees, PFOA and PFOS concentrations, are continuous a regression tree was used. I incorporated the same data table used in the exploratory regression into R version 3.4.3 using the rpart and rpart.plot packages (Milborrow 2017, Therneau and Atkins 2018). The tree is built through binary recursive partitioning, which is a process that divides the data into branches, an action that is repeated as the method moves up the branches and splits the data into smaller groups. The full data set is partitioned off based upon all the possible binary splits in every field, with the algorithm selecting the divergence that "minimizes the sum of the squared deviations from the mean in the two separate partitions" (Frontline Systems 2012). This rule is followed for all other divisions along the tree until the branch reaches the terminal node.

Because the tree will partition until the sum of the squared deviations from the mean is zero at the node, these algorithms can suffer from over fitting of the model. To address this issue I pruned the tree using the complexity parameter which determined the number of decision nodes for the tree. The larger the complexity parameter value the smaller the final tree. Adding this complexity parameter resulted in a regression tree that shows the most important decision nodes, which lists the variable and value that determined the split, along with the average PFC concentration value and percent of the data set that meets that concentration value on the leaf.

RESULTS

PFCs in Southern New Hampshire Drinking Water

Elevated levels of PFOAs clustered south of Manchester with the highest observed concentration of 1600 ng/L (Figure 3a). Alternatively, PFOS elevated levels occurred south of Portsmouth and low levels south of Manchester, with the highest concentration being 11000 ng/L (Figure 3b). However, the sampling data sets for both compounds is highly skewed to favor lower concentrations with PFOA having a minimum reported value of 2 ng/L and a mean of 38 ng/L and PFOS having a minimum reported value of 4 ng/L and a mean of 19 ng/L. Although higher concentrations of PFOS and PFOA are not frequent, when they do occur they can be highly elevated.

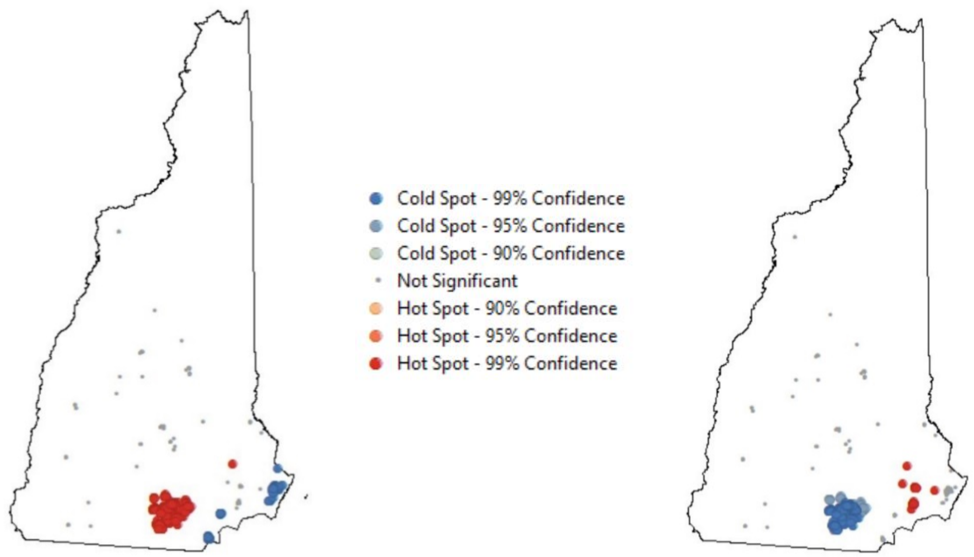

Figure 3a. PFOA Getis-Ord Hotspot Analysis

Figure 3b. PFOS Getis-Ord Hotspot Analysis

Results of the Exploratory Regression

Among the 28 variables tested, 4 sets of variables each were correlated above 0.8; Land use classed as developed and forests, metamorphic rock and igneous rock, aquifer area and perimeter area ratio, and surficial deposits of sand/gravel and ground moraine deposits. These values are considered collinear, so in the final model developed land, metamorphic rock, aquifer area, and surficial deposits of sand/gravel were used while the others were left out so as to avoid redundant variables.

The best fit exploratory linear regression model explained 39% of the variation in New Hampshire well water PFOA. The best fit model used 5 of the 28 variables, which were distance to trains, distance to waste water treatment plant violators, distance to rivers, land use (developed), and the thickness of the water table. It also had the lowest AIC score of 4901.15 (Table 2). In a correlation matrix, out of the 7 variables in the top 3 best fit models, none were highly correlated to PFOA levels, only distance to waste water treatment plant violators was moderately correlated with a value of -0.58, while distance to trains had low correlation at -0.47. However, all of the other variables; land use, distance to rivers, water table thickness, metamorphic rock type, and population density had individual correlation values between 0.1 and 0.3.

Alternatively, the best fit exploratory linear regression model was only able to explain 4% of the variation in New Hampshire well water PFOS with an AIC of 2198.93 (Table 3). The selected variables for this model were the distance from unlined landfills, total road length within a 500m radius, elevation, maximum transmissivity, distance to TRI reporting facilities, distance to waste water treatment plants, and distance to saturation thickness measurement location. All of the individual correlations for these variables were between 0 and 0.15 in the correlation matrix.

Table 2. Comparison of Models from the PFOA Exploratory Regression (*indicates that variable was present in the model)

Model Number	Adjusted R squared	AIC	Land Use (Developed)	Distance from Train	Distance from WWV	Distance from River	Water Table Thickness	Metamorphic Rock	Population Density
Model 1	0.39	4901.15	*	*	*	*	*		
Model 2	0.39	4902.89		*	*	*		*	*
Model 3	0.39	4903.43		*	*	*	*		*

Table 3. Comparison of Models from the PFOS Exploratory Regression (*indicates that variable was present in the model)

Model Number	Adjusted R squared	AIC	Distance from Unlined Landfills	Total Road Length within a 500m radius	Elevation	Maximum Transmissivity	Distance to TRI Reporting Facility	Distance to WWTP	Distance to Closest Recorded Aquifer
Model 1	0.04	2198.93	*	*	*	*	*		
Model 2	0.04	2199.16	*	*	*			*	*
Model 3	0.04	2199.27	*	*	*		*		*

PFOA Ordinary Least Squares Multiple Regression

After completing the exploratory regression for PFOA I used an ordinary least squares multiple regression on PFOA Model 1 to find specific coefficients and statistical significance.

Anthropogenic Features

Of the human factors submitted for the model, 3 variables, developed land (Figure 4a), distance from train (Figure 4b), and distance from waste water treatment plant violators (Figure 4c) were used in the final PFOA model (Table 4). Developed land has a positive coefficient which indicates a positive relationship. On the other hand, distance from the closest train has a negative coefficient, and also has the smallest value out of the 3. In contrast, distance from waste water treatment plant violators has the largest coefficient out of all of the variables in the best fit model and a negative relationship.

Table 4: OLS Results for PFOA Model 1 Variables

	Land Use (Developed)	**Distance from Train**	**Distance from WWV**	**Distance from River**	**Water Table Thickness**
Coefficient	0.262	-0.172	-1.008	0.059	-0.084
p-value	<0.001	<0.001	<0.001	<0.001	<0.001

Surface Environment

The surface environment only contributed one variable to the best fit model, distance to the closest river (Figure 4d). This variable has a positive relationship with a relatively small coefficient when compared to the other variables in the model (Table 4).

Geologic Characteristics

Finally, the geologic characteristics contributed one variable to the best fit model, water table thickness (Figure 4e). The coefficient is larger than the distance to rivers, but is still relatively small (Table 4). The relationship is also positive between PFOA levels and the thickness of the water table at the testing locations.

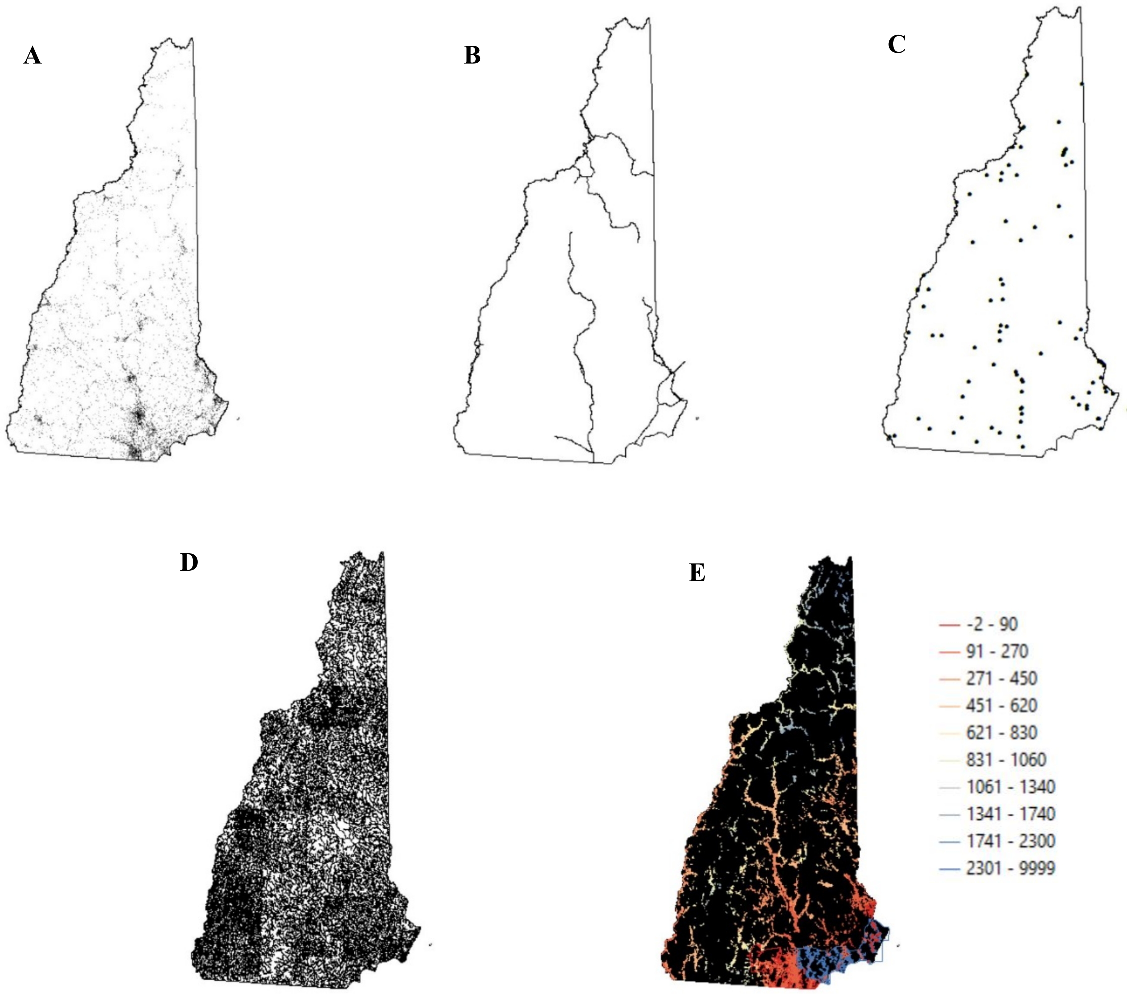

Figure 4. The 5 top model PFOA variables: (a) Developed Land Cover, (b) Railway Features, (c), Waste Water Treatment Plants with Violations, (d) New Hampshire Rivers, (e) Water Table Thickness.

Moran's I Test for Autocorrelation

The Moran's I statistic test for spatial dependence of the final model was statistically significant ($p<0.001$) and reflected a non-random distribution of residuals. The spatial distribution of high values and/or low values in the dataset is more spatially clustered than would be expected if underlying spatial processes were random. With a z-value of 13.4 the residuals were highly clustered and there was less than a 1% probaility that the clusters were caused by random chance, suggesting that there are other key factors missing from this model that would account for this clustering pattern (Figure 5). Completing a geographically weighted regression on the same variables showed no improvement to the R squared, so this particular spatial regression model was unable to improve over the OLS model for PFOA as expected.

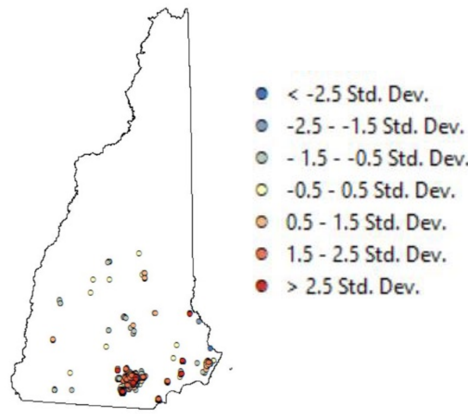

Figure 5. Standard Residuals from the best fit PFOA model

PFOS Ordinary Least Squares Multiple Regression

Like PFOA, after completing the exploratory regression for PFOS I used an ordinary least squares multiple regression on PFOS Model 1 to find specific coefficients and statistical significance.

Anthropogenic Features

Three of the five factors in the final PFOS model are linked to human development: distance from unlined landfills (Figure 6a), total road length within a 500m radius (Figure 6b), and distance to the closest TRI reporting facility (Figure 6c). Out of the three only total road length had a negative relationship with the highest coefficient out of all variables in the model (Table 5). Distance from unlined landfills and TRI facilities both have negative relationships to PFOS levels. Distance to unlined landfills had the smallest coefficient out of all five variables while industrial facilities had an intermediate coefficient value.

Table 5: OLS Results for PFOS Model 1 Variables.

	Distance from Unlined Landfills	**Total Road Length within a 500m radius**	**Elevation**	**Maximum Transmissivity**	**Distance to TRI Reporting Facility**
Coefficient	0.0005	-0.2435	-0.2079	-0.0109	0.0504
p-value	<0.001	<0.001	<0.001	0.015	<0.001

Surface Environment

Elevation was the only variable from the surface environment category to contribute to the best fit model (Figure 6d). It has a relatively high coefficient compared to the other variables and a negative relationship to PFOS levels (Table 5).

Geologic Characteristics

The last variable in the best fit model, maximum transmissivity, falls into the geologic category (Figure 6e). Though the coefficient is higher than distance from unlined landfills and TRI facilities it is still relatively small. Additionally, the relationship between maximum transmissivity of aquifers and PFOS concentrations is negative.

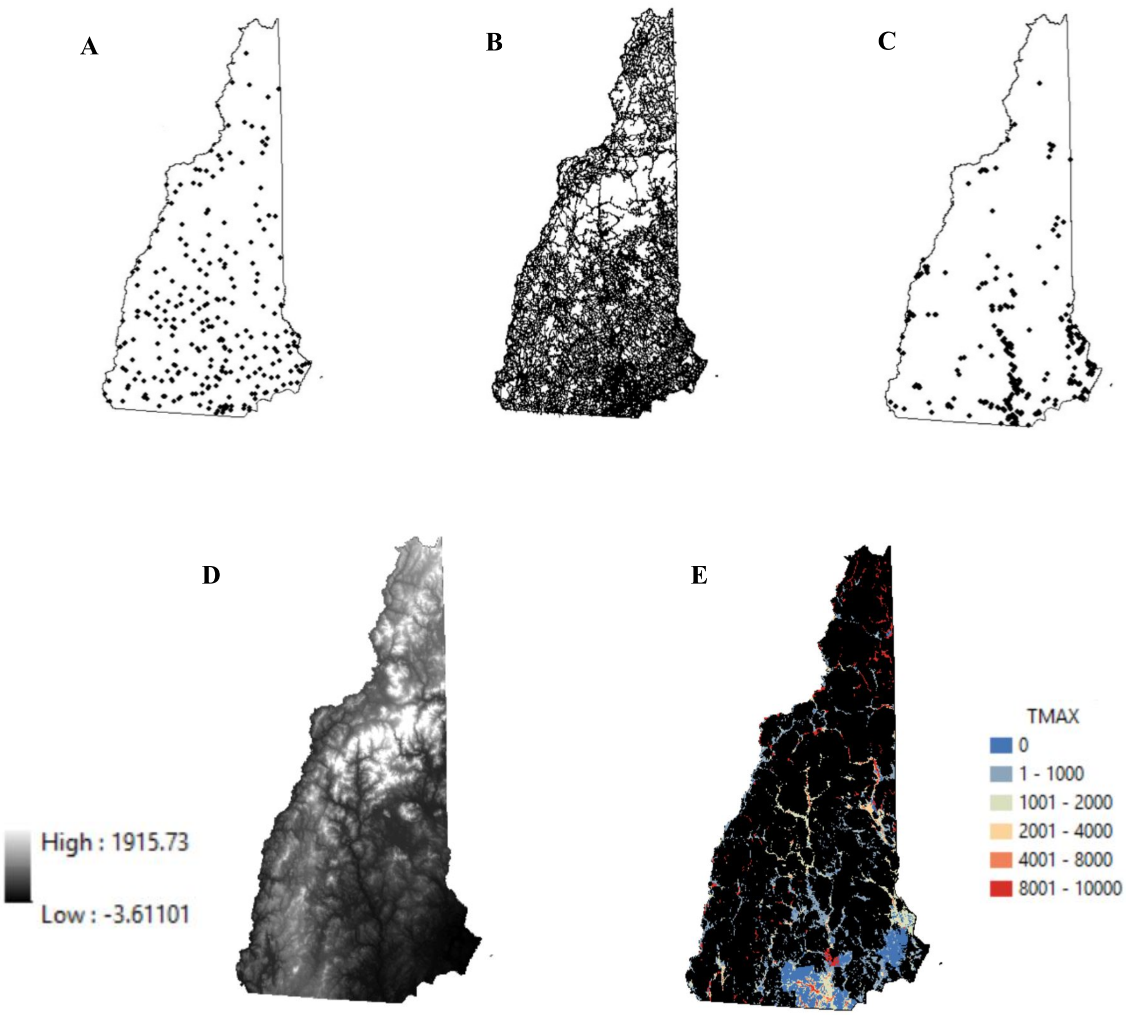

Figure 6. The 5 top model PFOS variables: (a) Unlined Landfill Locations, (b) Road Segments, (c), TRI Reporting Facilities, (d) New Hampshire Elevation, (e) Aquifer Maximum Transmissivity

Moran's I Test for Autocorrelation

The Moran's I statistic test for spatial dependence of the final model showed that there wasn't a random distribution of residuals for the PFOS best fit model with statistically significant results ($p<0.001$). The spatial distribution of high values and/or low values in the dataset is more spatially clustered than would be expected if underlying spatial processes were random. The results showed a z-score of 6.05 indicating that the residuals were highly clustered and such clustering had a less than 1% probability of occurring randomly. Similar to PFOA, this shows that there are missing variables from the best fit model that could explain the clustered pattern of the residuals (Figure 7). Given the low model fit for PFOS it was not expected that a geographically weighted regression would substantially improve this model, and that was indeed the case.

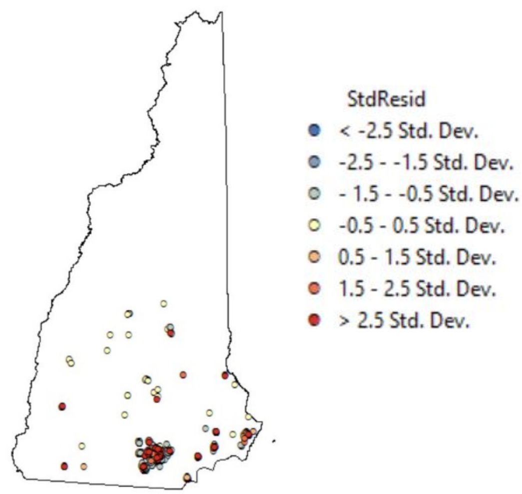

Figure 7. Standard Residuals from the best fit PFOA model

PFOA Regression Tree Analysis

The regression tree for PFOA had an R squared of 0.4 explaining 40% of the variation in PFOA. It used 4 out of the 28 variables, which were distance to waste water treatment plant violators (WWV), distance to TRI reporting facilities, distance to solid waste disposal facilities, and distance to seismic fault lines. All variables were anthropogenic features except for one geologic characteristic, distance to seismic fault lines. The first node shows a split at greater than or less than 840 m from the closest WWV. Wells less than 840 m had the highest average PFOA levels of the testing locations at 483, but accounted for a relatively small percentage of all wells with 0.7%. This is then further dissected at the next node that separates wells greater than 2547 m from a WWV from those that are less than 2547 m, but still greater than 840 m from a WWV. From there the tree splits even further and introduces new variables of interest. Several nodes occurred at specified distances to TRI reporting facilities and distance to solid waste disposal facilities while distance to seismic fault lines only had one node. Overall, testing locations that were less than 2547 m from a WWV accounted for the highest average PFOA levels.

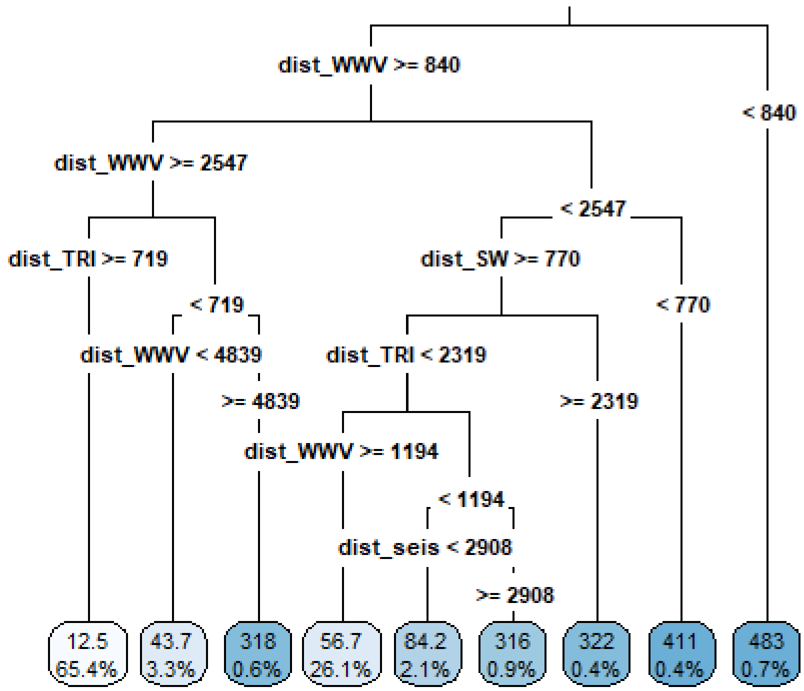

Figure 8a. Regression Tree for PFOA. The top number on the bottom circles indicates the average PFOA concentration for the specified group. The bottom number shows what percentage of the total data set meets the specified characteristics for that node. Dist_WWV notes the variable distance to waste water treatment plant violators. Dist_TRI indicates the distance to the closest TRI reporting facility, dist_SW marks the distance to solid waste disposal facilities, and dist_seis indicates the distance to the closet fault line.

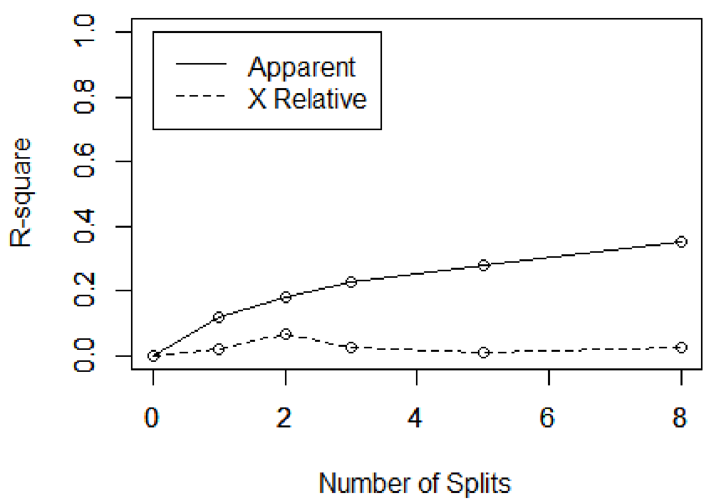

Figure 8b. R squared of PFOA Regression Tree. This shows the trend of the R squared as you increase the number of partition nodes.

PFOS Regression Tree Analysis

For PFOS the regression tree analysis was able to explain 30% of the variation in PFOS, with an R squared of 0.3. This tree used three distinct variables from the 28 put into the model, distance to the location of water table elevation measurement, percent impervious surface, and distance to solid waste disposal facilities. Distance to location of water table elevation measurement is a representative variable for distance to closest recorded aquifer boundary. Therefore, similar to PFOA, there is one geologic variable present, the distance to aquifer boundary, with the other two being anthropogenic features. This tree's first split sections off the wells that are less than 2008 m from an aquifer boundary from those that are more than 2008 m away. According to Figure 9b, after this first partition the additional nodes do not add any accuracy to the model. However, it does makes two more splits, locations were the impervious surface classification is greater than 4.5 and then where the distance to solid waste disposal facilities is larger than 195 m. Wells that were more than 2008 m from the closest aquifer boundary had the highest average PFOS levels at 2962, nonetheless they also accounted for the smallest percentage of all testing locations at 0.4%.

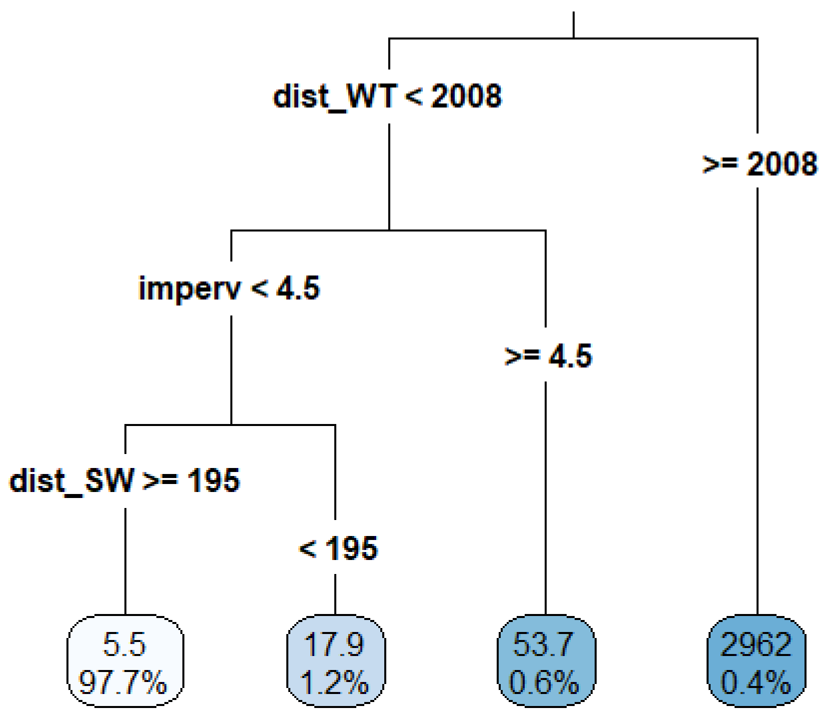

Figure 9a. Regression Tree for PFOS. The top number on the bottom circles indicates the average PFOA concentration for the specified group. The bottom number shows what percentage of the total data set meets the specified characteristics for that node. Dist_WT note the variable standing in for proximity to closest known aquifer boundary, imperv indicates the percent of impervious surface, and dist_SW marks the distance to the closest solid waste disposal facility.

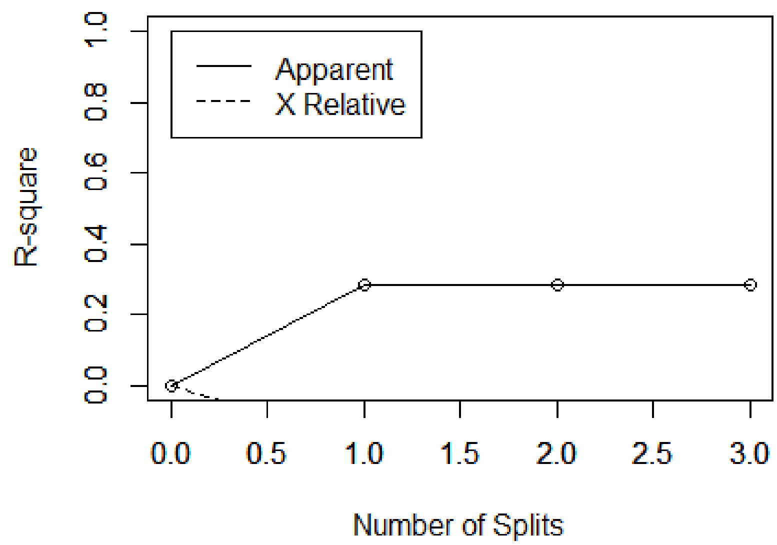

Figure 9b. R squared of PFOA Regression Tree. This shows the trend of the R squared as you increase the number of partition nodes.

DISCUSSION

Understanding how the spatial distribution of PFOA and PFOS is influenced in aquifers is critical to maintaining water quality as exposure to these chemicals is biologically harmful. For both PFOS and PFOA human activities on the landscape surface have the highest influence on the contamination levels in aquifers. In contrast to predictions, the surface environment contributed few key variables in the best fit models. However, characteristics of the aquifer landscape proved salient in modeling the PFC concentrations, indicating their importance in PFOA and PFOS long term transport. These findings had basis in previous studies which indicated that in contrast to surface water, groundwater had long term impacts from PFC exposure resulting from the interaction between soil layers and aquifer water (Shin et al. 2011).

PFOA Exploratory Regression Top Model Variables

In contrast to PFOS, the best fit linear regression models for PFOA showed moderate accuracy in determining a spatial pattern from the 28 given variables with an adjusted R^2 of 0.39.

Key variables in these models have a higher correlation to elevated levels of PFOA than the variables in the PFOS models. However, like PFOS the results for the Moran's I test for autocorrelation still had a high z-value indicating clustering of high values and/or low values in the residuals and thus, indicating missing factors from the model. In previous studies a major contributor to PFOA distribution is the location of industry (Hu et al. 2016). Similarly to PFOS, because industry was accounted for by using TRI reporting facilities the potential for extraneous manufacturing locations to cloud out higher correlation coefficients between elevated PFOA levels and PFAS manufacturers needs to be noted. This may partially explain why the adjusted R^2 didn't reach above 0.5.

Anthropogenic Features

Variables indicating human alteration of the landscape were the most prominent in the final regression model. Within the top 3 best models from the exploratory regression, the variables that were the most correlated were developed land ($p<0.001$), distance to trains ($p<0.001$), waste water treatment plant violators ($p<0.001$), and population density ($p<0.001$). The positive relationship between developed land and PFOA levels reflect that the more developed regions, which have higher population densities, also have elevated PFOA concentrations. Downtown regions have a higher mean runoff concentration of PFOA when compared to less developed areas (Xiao et al. 2012). Because the variable accounting for percent of impervious surface was not significant in the analysis, a different aspect related to higher populated and developed areas must be supporting elevated PFOA levels. High traffic locations and parking lots have elevated PFOA levels which has been further corroborated by analysis looking at PFASs in dust particles (Kim and Kannan 2007, Murakami et al. 2009). Although these data apply specifically to surface runoff, in areas of PFOA contaminated groundwater, 40% of the pollution could be traced directly back to surface water seepage (Liu et al. 2017). When considered together it presents an argument for traffic as a specific source of PFOA introduction into the environment.

Because the relationship between the distance to trains and PFOA levels is negative PFOA levels decrease the further they get from trains. Train stations have been linked with statistical significance ($p<0.05$) to elevated levels of PFOA in surface water (Yasuyuki and Shigeki 2008).

This relationship carries over to groundwater as all three of my top exploratory models used distance from trains ($p<0.001$).

In addition, waste water treatment plant violators are negatively related to PFOA levels as PFOA concentrations decrease the farther away WWTP violators are from the wells. WWTPs in general have been shown to correlate to elevated PFC levels due to industry and continued PFC product use (Eriksson et al. 2017). However, because WWTP processing does not remove PFCs and may increase the PFC load from the breakdown of precursor chemicals, WWTPs are considered point sources for this family of chemicals (Zhang et al. 2013). Therefore, violating WWTPs, discharging more pollutants than legally allowed, would be key sources for PFOA into the environment. This combined with the data on surface water seepage and sludge contamination effluent from these violating water treatment plants indicates a direct impact to groundwater quality.

Surface Environment

The main variable of importance when comparing PFOA concentrations against the existing landscape was the distance of each sample location to the closest river. The positive and statistically significant ($p<0.001$) relationship between the distance to the closest river and PFOA levels indicate that as the distance increases so do the levels of PFOA. Two main pathways account for almost all of the PFOA in groundwater**:** surface water seepage and soil leaching (Liu et al. 2017). Because soil leaching takes years, if not decades, surface water seepage accounts for the majority of the fast moving contamination (Xiao et al. 2015). Additionally, rivers can be a source for pollutants as they act as a transport pathway for water contaminants (Su et al. 2013). Though it could be argued that industry or other sources of pollution are collinear to distance from rivers, and may be a cofounding variable, the exploratory regression results negated this explanation. Collectively, regions further from river routes may convey more toxins directly to groundwater instead of being transported via surface water.

Geologic Characteristics

The presence of 2 geologic variables in the top 3 exploratory regression models speaks to the importance of an aquifer's contributions to PFOA levels. Metamorphic rock type was positively statistically significantly ($p<0.001$) related to PFOA levels. However, metamorphic rock was collinear to igneous rock potentially indicating a general increase in PFOA levels when metamorphic and igneous rock are present. There has been some indication that metamorphic rock, gneiss in particular, can develop a series of channels along a sloping surface, called swales, which help facilitate the movement of water and contaminants into groundwater stores (Carson 2017). However, the ability of fractures to transmit water varies greatly from rock that carries large quantities of water to being nearly impervious (USGS 2016). Even with intensely fractured rock in Mirror Lake, New Hampshire the most flow occurred in only one or two discrete fractures out of all the potential pathways (Paillet F. L. et al. 2006). Therefore, while the presence of metamorphic and igneous rock may indicate higher amount of fractures leading to increased PFOA levels, it is equally likely that the relative abundance of these rock types in relation to sedimentary rock has facilitated a higher correlation specific to New Hampshire.

In addition to rock type the water table thickness was important in the final models. The negative, but statistically significant relationship ($p<0.001$) shows that as the water table gets thicker the levels of PFOA decrease. Although PFOA moves into groundwater systems through liquid matrices interactions with soil, as evidenced by increased concentration with increased soil depth, the decline in aquifer concentrations moving away from known sources is majoritively due to dispersion and dilution (Xiao et al. 2015). This phenomena coupled with the importance of water table thickness may indicate that as the water table expands in vertical length the contaminant concentration is diluted.

PFOA Regression Tree Top Variables

Anthropogenic Features

In comparison with the exploratory regression models the regression tree provided a model with a similar R squared of 0.4. Additionally, the regression tree results highlighted the importance of waste water treatment plant violator locations in relation to wells in identifying points of elevated PFOA levels. However, unlike the exploratory regression the regression tree was able to demonstrate how specific distances influenced elevated PFOA levels, because this variable showed up at four different decision nodes within the tree. Wells with the highest PFOA concentrations, averaging at 483, were shown to be those less than 840 m from violating WWTPs. In conjunction, wells with moderate levels of contamination, with PFOA concentrations ranging from 411 to 56, were less than 2547 m from the same treatment plants. A majority of these wells exceed the EPA recommended exposure limit of 70 parts per trillion. Similar to the OLS PFOA model, the continued use of products containing PFOA and lack of removal in waste water processing has made treatment plants sources of pollution for PFCs overall (Zhang et al. 2013, Eriksson et al. 2017). These regression tree results point to the fact that the specific distance from treatment plants that are in violation of EPA standards is critical to predicting PFOA contamination.

Although the R squared value was similar and WWTP violators showed up in both methods those where the only two similarities between analysis. Unlike the exploratory regression, the regression tree only included four variables of importance with distance to WWTP violators being the only overlapping variable. Distance to TRI reporting facilities, distance to solid waste disposal facilities, and distance to fault lines were the other three key factors. Industry facilities, specifically those that use PFOA or other PFCs, are known sources of PFOA pollution (Hu et al. 2016). The regression tree includes two decision nodes where distance to industry locations are important with average PFOA concentrations getting higher the closer wells were to TRI facilities. Those wells that were both close to industry facilities and WWTP violators had relatively high PFOA concentration averages.

The other major anthropogenic feature included in the tree was distance to solid waste disposal facilities (SW). Similarly to PFOA pollution from waste water treatment plants, solid waste disposal facilities show high levels of PFOA contamination due to the continued use of products containing these chemicals by the general population (Eriksson et al. 2017). In solid waste disposal facilities, such as landfills, as much as 60% of all PFOA can enter leachate and move into groundwater through surface water seepage (Liu et al. 2017). The dataset that gave landfill locations provided two variables in the regression tree**:** all landfills and specifically unlined landfills. Because unlined landfills weren't used in the tree it is clear that solid waste disposal facilities are indicators of elevated PFOA concentrations whether or not they are lined.

Geologic Characteristics

The last key factor in the PFOA regression tree was distance to the closest fault line. This variable is dependent on the well locations first being close to WWTP violators and TRI reporting facilities. However, once a well is exposed to these anthropogenic pollution sources those less than 2908 m from a fault line had an average PFOA concentration of 84 while those further away were much higher at 316, both values higher than the EPA recommended 70 parts per trillion. This difference is supported by previous findings that show natural vertical faults to be conducive pathways of contamination movement upwards, but find faults with high hydraulic gradients across them act as barriers to lateral flow (Bense and Person 2006, Meyer 2016). Therefore, these faults are actually preventing movement of contaminants horizontally to other aquifers and explain why PFOA levels are higher further away from the fault lines, as PFC pathways can reach those regions more easily.

PFOS Exploratory Regression Top Model Variables

The extremely low adjusted R^2 of the exploratory regression models for PFOS demonstrate that there is little to no correlation between the 28 tested variables and PFOS levels in this model. The high z-value of the Moran's I test for autocorrelation that indicated clustering in the residuals further confirms that this model lacks predictive power. It shows that key factors are missing that would account for the clustering pattern in the residuals. Previous studies found that PFOS is highly correlated to military fire training areas (MFTAs) and major industrial sites (Hu et al. 2016). However, Hu et al. (2016) had two key differences from my model in regards to these variables. First, the difference in scale was a key component with Hu et al. (2016) being a larger, nationwide analysis with 290 site, and mine focusing only on the state of New Hampshire with 2 locations. Because of this Hu et al. (2016) was better able to identify relationships between PFOS and MFTAs. Additionally, Hu et al. (2016) included registered manufacturing locations from the EPA PFOA Stewardship program. Because New Hampshire doesn't have any member locations of the EPA PFOA program, all industry locations in this study are only identified as Toxic Release Inventory (TRI) participants. Unlike the known PFOA factories in Hu et al. (2016), TRI participants are not specifically PFOS users and therefore could be clouding out potential correlations.

Anthropogenic Features

Despite the low R^2, the best fit models indicated that unlined landfills ($p<0.001$), arterial road length ($p<0.001$), manufacturing locations ($p<0.001$), and WWTP locations ($p<0.001$) were important variables to the final analysis. In China, unlined landfills contributed to elevated PFOS levels in groundwater through soil leaching (Liu et al. 2017). The contribution from landfills was small when compared with PFOA statistics on the same process, however these differences were linked to the higher usage of PFOA in homes after PFOS was phased out of production processes in recent years (Liu et al. 2017). This contradicts my findings that there is a positive relationship between unlined landfills and elevated PFOS levels which indicate that as distance from these landfills increased the concentrations of PFOS did as well. A potentially unidentified source of pollution may be acting as a collinear variable here that would better explain this trend. WWTPs were also shown to be potential sources for PFOS and could be linked to the elevated levels of PFOS in the resulting sludge (Guo et al. 2010, Zhang et al. 2013). This sludge is used on agricultural land (24%), to cover mines and dump sites (24%), and, in some cases, to create new soils (29%) (Eriksson et al. 2017). High concentration pollution events, such as sludge contamination, on a landscape surface can interact with the soil interface to move pollutants towards groundwater reserves and create major contamination pathways (Shin et al. 2011, Xiao et al. 2015). In addition, arterial road length has been shown to be highly correlated to elevated levels of PFOS in surface runoff, indicating a tie between traffic and construction activities to increased exposure which is now seen to carry to groundwater (Zushi and Masunaga 2010).

Similar to PFOA, major PFAS production companies are predictors for elevated PFOS levels (Hu et al. 2016). The presence of TRI reporting facilities in the final analysis indicates that this pattern may be holding true for my model, however extraneous manufacturing locations could be confusing the true relationship. Because PFOA and many precursor chemicals are not required reporting compounds (PFOS was phased out by 2002) for the TRI there is no way to differentiate those facilities that are using them from the rest without their cooperation, such as in the EPA's PFOA stewardship program.

Surface Environment

Although the final PFOS models had an extremely low R^2, these previous findings support the marginal correlations in this study, and show that it could be relevant to consider the other key variables in the final analysis as potential indicators to examine. These include elevation ($p<0.001$), maximum transmissivity of the aquifer ($p<0.001$), and the distance to the closest saturation thickness measurement of the aquifer ($p<0.001$). Elevation had a negative relationship to PFOS levels as the concentrations of the pollutant decreased as the elevation increased. Groundwater flow can be complex and integrate soil and water pressures to influence the direction of water movement throughout watersheds (Winter 1999). Even testing around Mirror Lake in new Hampshire has shown water to move around obstacles and discharge in unexpected regions (Winter 1999). Although these flows can be complex, from the negative relationship my data shows between PFOs levels and elevation, water at lower elevation is collecting and/or being exposed to higher levels of PFOS pollution.

Geologic Characteristics

It's important to note that two out of the seven variables in the top 3 exploratory models deal directly with the characteristic of the aquifer itself. This indicates that, unlike surface water, the nature of an aquifer may be principal in determining PFOS distributions, and in turn exposure. Maximum transmissivity had a statistically significant negative relationship to PFOS levels as the pollutant concentrations decreased as maximum transmissivity increased. Aquifer transmissivity describes the ability of a medium to allow water to move through it, therefore my findings indicate that the better an aquifer's ability to transmit water the more likely it was to have decreased PFOS levels. This falls in line with previous studies that show declining chemical concentrations moving away from the source due to dilution and dispersion (Xiao et al. 2015). A higher aquifer transmissivity would mean these processes occur on a faster timeline.

These findings are also important to the other geologic characteristic, distance to the closest recorded aquifer, that had a positive and statistically significant relationship ($p<0.001$). As the distance to known aquifers increase PFOS levels also increased. Knowing that decreases in contamination levels are a result of dilution, the closer a body of water, like an aquifer is, the lower the concentration of PFOS.

PFOS Regression Tree Top Variables

Anthropogenic Features

Unlike PFOA which had similar R^2 values for the exploratory regression and the regression tree, PFOS had a marked improvement. The exploratory regression R^2 was very low at 0.04 which is in contrast to the regression tree with an R squared of 0.3. The model fit is still relatively low, but substantially better than the exploratory regression results. In addition the PFOS regression tree only used three variables: distance to solid waste disposal facilities, percent of impervious surface, and distance to the closest recorded aquifer. Although the exploratory regression did show distance to unlined landfills to be a key factor the regression tree didn't use distance from unlined landfills, but rather all landfills. This reflects the results of the PFOA regression tree that didn't differentiate out the unlined facilities and indicates that any solid waste disposal facility location is important in determining elevated PFOS levels in surrounding groundwater. Dissimilar to PFOA only 3.3% of PFOS in a landfill will enter soil leachate (Liu et al. 2017). Nevertheless, it is clearly enough to make a difference as wells greater than 195 m from a landfill had an average PFOS concentration of 5.5 while those closer were at 17.9, both levels that fall within the EPA exposure guidelines.

The second anthropogenic feature in the regression tree, though it could also be seen as a characterization of surface environment, was the percent of impervious surface. In this table I reclassified land cover percentages so that 1 represented 0-20% of impervious surface, 2 was 20-40%, 3 was 40-60%, 4 was 60-80% and 5 was 80-100%. Those wells in an area where the percent impervious surface value was greater than 4.5, meaning that the amount of permeable land was less than 10% of the ground cover, had at the very least double the PFOS concentration in their groundwater than those regions with a higher percentage of permeable land. Higher amounts of impervious land is correlated to higher levels of development, higher population densities, and more roads. Although none of these variables were used in the regression tree they were present in the exploratory regression. With the higher R squared from the regression tree this may indicate that percent of impervious surface is a better indicator than any of these other factors. Arterial road length and population density have been linked to elevated PFOS levels in Japan and analysis of storm water runoff indicates commercial regions to be sources of PFCs in general (Zushi and Masunaga 2010, Xiao et al. 2012). Therefore, percent impervious surface may be a key indicator value for other markers of development and anthropogenic sources of pollution around wells.

Geologic Characteristics

The most important variable in the PFOS regression tree was distance to the closest water table elevation measurement which is a proxy for the distance to the known aquifer boundary. This variable is the first decision node in the tree with all wells more than 2008 m from an aquifer boundary having the highest average PFOS concentration of 2962 while those closer to an aquifer had lower levels. It is also the most important variable to this model, as the overall fit ceases to rise with the addition of more variables after this initial partition. Because the decline in aquifer contaminant concentrations is due to dispersion and dilution, the further a well is away from the upper most aquifer the more likely surface water seepage is to maintain high concentrations (Xiao et al. 2015). This accounts for the positive relationship between distance to an aquifer and PFOS levels. Without the presence of an aquifer body to provide a mechanism for dilution of PFOS concentrations, the contamination from surface water seepage remains relatively high. Only those wells that were less than 2008 m from the closest aquifer had PFOS concentrations below the EPA exposure limit of 70 parts per trillion. Therefore, the enhanced model fit coupled with the specified limit distinctions that lead to elevated PFOS levels indicate that the regression tree was a better method for this dataset.

Limitations

As a result of the clustered nature of the available data set this study was inherently restricted. Because the tested locations were a part of a public health and environmental safety initiative the main focus of testing was on wells in population centers and rural sources were under sampled. This invariably led to a clustered data set which was not ideal for the statistical analysis as it may have introduced some preexisting bias into the results. This could explain why both PFOA and PFOS models had R squared values that maxed out at 0.4. In addition, the assumptions made to create the data table may have simplified the relationships between specific variables and PFC levels. Although care was taken to normalize and check for non-linear relationships in the exploratory and OLS regressions this may require further investigation. Furthermore, the rate of pumping from public water wells may alter the groundwater concentration predictions (Shin et al. 2011). Because the data did not come with pumping rate information this study was not able to account for this factor.

Future Directions

An expansion of several important variables in light of initial results would provide a more detailed analysis of PFOA and PFOS variance and correlations. Because this data set was so restricted based upon sampling location, testing locations outside of pre-existing wells within New Hampshire is necessary to provide a more spatially uniform data set to enhance further regression and spatial analysis. In addition, more data on soil and PFC interactions is necessary beyond broad trends to determine specific site potential for PFOA and PFOS contamination. This approach includes developing a better understanding of the influence water table depth may have on contaminant concentration. Finally, the relationship between surface water seepage and river transportation of contaminants needs more detailed examination to ascertain the specific connection between high seepage influence and decreased PFOA concentrations closer to rivers.

Broader Implications

Although these avenues for potential data would provide room for improved understanding of the relationship between PFCs and the environment, the model results are still important to the discussion around understanding what influences the health of our aquifers' water. The results support the growing number of studies finding links between point source polluters such as waste water treatment plants, landfills, and industry (Zhang et al. 2013, Hu et al. 2016, Liu et al. 2017, Eriksson et al. 2017). With the results of this analysis and understanding the interconnectedness of surface and groundwater, information pertaining to PFCs in surface water can be extrapolated to aquifer systems (Winter 1999). In addition, the combination of surface water seepage and soil leaching to create a long term contamination pathway along with the influence from bedrock type and potentially water table depth all point to the fact that understanding a multitude of factors that include aquifer characteristics is critical to understand exposure potential (Winter 1999). Knowing what aspects of an aquifer are correlated to elevated PFC levels can help inform health and safety offices about where they need to be testing in order to ensure public safety. Comprehending these factors can help predict future issues and advise management and policy decisions.

ACKNOWLEDGMENTS

This undergraduate thesis was a part of the Environmental Science thesis program in the Environmental Science, Policy, and Management Department of UC Berkeley. As such I give a huge thank you to alumni and donors who support the continuation of this program. The New Hampshire Department of Environmental Services also deserves a big thank you as they were unfailingly helpful in providing me with state data throughout this project. Tom Bruton also helped me collect critical data for my analysis along with Professor Russel Conglaton and Dr. Wally Bothner who explained how to consolidate some of my data categories. Thank you to Patina Mendez and Kurt Spreyer who walked me through the writing process and regularly gave advice on ways to expand and advance my topic. Thank you also to the amazing GSIs, Alison Ecker and Leslie McGinnis who were there to answer every question and motivate us throughout the year. Gregory Biging, my thesis mentor, also deserves my never ending gratitude for explaining and aiding me in my statistical analysis and providing much appreciated career and life advice. A couple quick shout outs to several important people; Maddie Heaps, my roommate, who listened to me talk about this paper for a year and a half; Jessica Redden who kept me optimistic and on schedule; and my project group, Annie Peacock, Bessie Liu, Sarah Nordahl, and Sophia Leiker, for editing my paper and giving invaluable critiques. Finally, thank you to my parents for their continued support and unceasing love. Without them this thesis wouldn't have been possible for so many reasons.

REFERENCES

Bense, F. V., and A. M. Person. 2006. Faults as conduit-barrier systems to fluid flow in siliciclastic sedimentary aquifers. Water Resources Research 42.

Billings, M. P. 1956. The Geology of New Hampshire. New Hampshire Department of Resources and Economic Development.

Briand, L. C., T. Langley, and I. Wieczorek. 2000. A Replicated Assessment and Comparison of Common Software Cost Modeling Techniques. Pages 377–386 ICSE ’00 Proceedings of the 22nd international conference on Software engineering. ACM New York, Limerick, Ireland.

Calafat, A. M., Z. Kuklenyik, J. A. Reidy, S. P. Caudill, J. S. Tully, and L. L. Needham. 2007. Serum Concentrations of 11 Polyfluoroalkyl Compounds in the U.S. Population: Data from the National Health and Nutrition Examination Survey (NHANES) 1999−2000. Environmental Science & Technology 41:2237–2242.

Caliman, F. A., and M. Gavrilescu. 2009. Pharmaceuticals, Personal Care Products and Endocrine Disrupting Agents in the Environment – A Review. CLEAN – Soil, Air, Water 37:277–303.

Collí-Dulá, R. C., C. J. Martyniuk, S. Streets, N. D. Denslow, and R. Lehr. 2016. Molecular impacts of perfluorinated chemicals (PFASs) in the liver and testis of male largemouth bass (Micropterus salmoides) in Minnesota Lakes. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics 19:129–139.

Ericson, I., M. Nadal, B. van Bavel, G. Lindström, and J. L. Domingo. 2008. Levels of perfluorochemicals in water samples from Catalonia, Spain: is drinking water a significant contribution to human exposure? Environmental Science and Pollution Research International 15:614–619.

Eriksson, U., P. Haglund, and A. Kärrman. 2017. Contribution of precursor compounds to the release of per- and polyfluoroalkyl substances (PFASs) from waste water treatment plants (WWTPs). Journal of Environmental Sciences (China) 61:80–90.

Frontline Systems. 2012, March 23. Regression Trees. https://www.solver.com/regression-trees.

Goldthwait, J., L. Goldthwait, and R. Goldthwait. 1951. The Geology of New Hampshire Part 1-Surficial Geology. New Hampshire State Planning and Development Commission.

Guo, R., W.-J. Sim, L. Eung-Sun, J.-H. Lee, and J.-E. Oh. 2010. Evaluation of the fate of perfluoroalkyl compounds in wastewater treatment plants. Water research 44:3476–86.

Guo, Z., and K. Krebs. 2009. Perfluorocarboxylic acid content in 116 articles of commerce. Environmental Protection Agency, Research Triangle Park, NC: U.S.

Houtz, E. F., R. Sutton, J.-S. Park, and M. Sedlak. 2016. Poly- and perfluoroalkyl substances in wastewater: Significance of unknown precursors, manufacturing shifts, and likely AFFF impacts. Water Research 95:142–149.

Hu, X. C., D. Q. Andrews, A. B. Lindstrom, T. A. Bruton, L. A. Schaider, P. Grandjean, R. Lohmann, C. C. Carignan, A. Blum, S. A. Balan, C. P. Higgins, and E. M. Sunderland. 2016. Detection of Poly- and Perfluoroalkyl Substances (PFASs) in U.S. Drinking Water Linked to Industrial Sites, Military Fire Training Areas, and Wastewater Treatment Plants. Environmental Science & Technology Letters 3:344–350.

Johansson, N., A. Fredriksson, and P. Eriksson. 2008. Neonatal exposure to perfluorooctane sulfonate (PFOS) and perfluorooctanoic acid (PFOA) causes neurobehavioural defects in adult mice. NeuroToxicology 29:160–169.

Kim, S.-K., and K. Kannan. 2007, November 13. Perfluorinated Acids in Air, Rain, Snow, Surface Runoff, and Lakes: Relative Importance of Pathways to Contamination of Urban Lakes. research-article. http://pubs.acs.org/doi/abs/10.1021/es072107t.

Kubwabo, C., B. Stewart, J. Zhu, and L. Marro. 2005. Occurrence of perfluorosulfonates and other perfluorochemicals in dust from selected homes in the city of Ottawa, Canada. Journal of environmental monitoring: JEM 7:1074–1078.

Laboratory Testing Guidelines for Per- and Polyfluoroalkyl Substances (PFAS). 2016. . New Hampshire Department of Environmental Services (NHDES).

Lau, C., K. Anitole, C. Hodes, D. Lai, A. Pfahles-Hutchens, and J. Seed. 2007. Perfluoroalkyl acids: a review of monitoring and toxicological findings. Toxicological Sciences: An Official Journal of the Society of Toxicology 99:366–394.

Liu, Z., Y. Lu, P. Wang, T. Wang, S. Liu, A. C. Johnson, A. J. Sweetman, and Y. Baninla. 2017. Pollution pathways and release estimation of perfluorooctane sulfonate (PFOS) and perfluorooctanoic acid (PFOA) in central and eastern China. Science of The Total Environment 580:1247–1256.

MacKenzie, J.-A. 2016, April 27. Granite State Analytical - News. http://www.granitestateanalytical.com/news/.

Medalie, L., and R. Moore. 1995. Ground-Water Resources in New Hampshire: Stratified-Drift Aquifers. U.S. Geological Survey.

Meyer, R. 2016, August 11. A Major Pacific Northwest Quake Looks Even Likelier, Thanks to Underwater Landslide Evidence - The Atlantic. https://www.theatlantic.com/science/archive/2016/08/a-major-earthquake-in-the-pacific-northwest-just-got-more-likely/495407/.

Montagner, C. C., and W. F. Jardim. 2011. Spatial and seasonal variations of pharmaceuticals and endocrine disruptors in the Atibaia River, São Paulo State (Brazil). Journal of the Brazilian Chemical Society 22:1452–1462.

Murakami, M., H. Shinohara, and H. Takada. 2009. Evaluation of wastewater and street runoff as sources of perfluorinated surfactants (PFSs). Chemosphere 74:487–493.

National Institute of Environmental Health. 2016. Perfluorinated Chemicals (PFCs). Pages 1–4. U.S. Department of Health and Human Services.

New Hampshire Department of Environmental Services State-wide Perfluorinated Chemicals (PFC) in Drinking Water Investigation. 2017. . Pages 1–4. New Hampshire Department of Environmental Services (NHDES).

Paillet F. L., Hess A. E., Cheng C. H., and Hardin E. 2006. Characterization of Fracture Permeability with High-Resolution Vertical Flow Measurements During Borehole Pumping. Groundwater 25:28–40.

Randall, A. D. 2001. Hydrogeologic framework of stratified-drift aquifers in the glaciated Northeastern United States. US Geological Survey Professional Paper:B1–B176.

Rocha, M. J., C. Cruzeiro, M. Reis, M. Â. Pardal, and E. Rocha. 2014. Spatial and seasonal distribution of 17 endocrine disruptor compounds in an urban estuary (Mondego River, Portugal): evaluation of the estrogenic load of the area. Environmental Monitoring and Assessment 186:3337–3350.

Schmidt, A. F., and C. Finan. 2017. Linear regression and the normality assumption. Journal of Clinical Epidemiology.

Shi, Y., Y. Pan, R. Yang, Y. Wang, and Y. Cai. 2010. Occurrence of perfluorinated compounds in fish from Qinghai-Tibetan Plateau. Environment International 36:46–50.

Shin, H.-M., V. M. Vieira, P. B. Ryan, R. Detwiler, B. Sanders, K. Steenland, and S. M. Bartell. 2011. Environmental Fate and Transport Modeling for Perfluorooctanoic Acid Emitted from the Washington Works Facility in West Virginia. Environmental Science & Technology 45:1435–1442.

Stekl, P. J., and S. M. Flanagan. 1992. Geohydrology and Water Quality of Stratified-Drift Aquifers in the Lower Merrimack and Coastal River Basins, Southeastern New Hampshire. U.S. Geological Survey.

Su, S., R. Xiao, X. Mi, X. Xu, Z. Zhang, and J. Wu. 2013. Spatial determinants of hazardous chemicals in surface water of Qiantang River, China. Ecological Indicators 24:375–381.

Taniyasu, S., N. Yamashita, H.-B. Moon, K. Y. Kwok, P. K. S. Lam, Y. Horii, G. Petrick, and K. Kannan. 2013. Does wet precipitation represent local and regional atmospheric transportation by perfluorinated alkyl substances? Environment International 55:25–32.

Technical Advisory - Laboratory Analysis of Drinking Water Samples for Perfluorooctanoic Acid (PFOA) Using EPA Method 537 Rev. 1.1. 2016. . Environmental Protection Agency.

Timofeev, R. 2004. Classification and Regression Trees (CART) Theory and Applications:40.

Tittlemier, S. A., K. Pepper, C. Seymour, J. Moisey, R. Bronson, X.-L. Cao, and R. W. Dabeka. 2007. Dietary exposure of Canadians to perfluorinated carboxylates and perfluorooctane sulfonate via consumption of meat, fish, fast foods, and food items prepared in their packaging. Journal of Agricultural and Food Chemistry 55:3203–3210.

Tsai, W.-T., H.-P. Chen, and W.-Y. Hsien. 2002. A review of uses, environmental hazards and recovery/recycle technologies of perfluorocarbons (PFCs) emissions from the semiconductor manufacturing processes. Journal of Loss Prevention in the Process Industries 15:65–75.

United States Geologic Survey. 2016. Crystalline-rock aquifers. https://pubs.usgs.gov/ha/ha730/ch_m/M-text6.html.

USGS. 2016, December 6. U.S. Geological Survey Fact Sheet 112-02 -- Fractured-Rock Aquifers--Understanding an Increasingly Important Source of Water. https://toxics.usgs.gov/pubs/FS-112-02/.

Vélez, M. P., T. E. Arbuckle, and W. D. Fraser. 2015. Maternal exposure to perfluorinated chemicals and reduced fecundity: the MIREC study. Human Reproduction (Oxford, England) 30:701–709.

Vested, A., C. H. Ramlau-Hansen, S. F. Olsen, J. P. Bonde, S. L. Kristensen, T. I. Halldorsson, G. Becher, L. S. Haug, E. H. Ernst, and G. Toft. 2013. Associations of in utero exposure to perfluorinated alkyl acids with human semen quality and reproductive hormones in adult men. Environmental Health Perspectives 121:453–458.

Webster, G. 2010. Potential human health effects of perfluorinated chemicals (PFCs). Pages 1–10. National Collaborating Center for Environmental Health.

White, S. S., K. Kato, L. T. Jia, B. J. Basden, A. M. Calafat, E. P. Hines, J. P. Stanko, C. J. Wolf, B. D. Abbott, and S. E. Fenton. 2009. Effects of perfluorooctanoic acid on mouse mammary gland development and differentiation resulting from cross-foster and restricted gestational exposures. Reproductive Toxicology (Elmsford, N.Y.) 27:289–298.

Winter, T. C. 1999. Relation of streams, lakes, and wetlands to groundwater flow systems. Hydrogeology Journal 7:28–45.

Xiao, F., M. F. Simcik, and J. S. Gulliver. 2012. Perfluoroalkyl acids in urban stormwater runoff: Influence of land use. Water Research 46:6601–6608.

Xiao, F., M. F. Simcik, T. R. Halbach, and J. S. Gulliver. 2015. Perfluorooctane sulfonate (PFOS) and perfluorooctanoate (PFOA) in soils and groundwater of a U.S. metropolitan area: Migration and implications for human exposure. Water Research 72:64–74.

Yasuyuki, Z., and M. Shigeki. 2008. Train Station as an Indicator of the Nonpoint Source of Perfluorinated Compounds in Urban River Basin. Organohalogen Compounds 70:002466.

Zhang, W., Y. Zhang, S. Taniyasu, L. W. Y. Yeung, P. K. S. Lam, J. Wang, X. Li, N. Yamashita, and J. Dai. 2013. Distribution and fate of perfluoroalkyl substances in municipal wastewater treatment plants in economically developed areas of China. Environmental Pollution 176:10–17.

Zushi, Y., and S. Masunaga. 2010. Novel Source Apportionment Method Based on GIS Receptor Model for Aquatic Perfluorinated Compound (PFC) Pollution- A Case Study in the Basin of Tokyo, Japan.

APPENDIX A: Data Source Information

Categories	Downloaded Data	Data Source	Website	Date of Access	Format	Cell Size	Projection/Datum	Variable(s) Created	Variable Range
Anthropogenic Features	PFC Testing Values	New Hampshire Department of Environmental Services	Available via Electronic Correspondence	1/6/2017	Excel Table	N/A	GCS North American 1983	Well Locations	N/A
								PFOS Tested Concentrations	13-110000 ng/L
								PFOA Tested Concentrations	0-1600 ng/L
	Superfunds Sites	EPA Facility Registration Service (FRS)	https://catalog.data.gov/dataset/epa-facility-registry-service-frs-facility-interests-dataset-download	11/2/2017	Vector	N/A	GCS North American 1983	Distance to Superfund Site	261-23606 m
	Industry Sites						GCS North American 1983	Distane to TRI Repoting Facility	78-17730 m
	Military Fire Training Locations	Google Earth	Available via Electronic Correspondence	1/23/2018	Excel Table	N/A	WGS 1984	Distance from Military Fire Training Locations	1338-86178 m
	Land Cover	Multi-Resolution Land Characteristics Consortium (MRLC)	https://www.mrlc.gov/nlcd11_data.php	11/28/2017	Raster	30, 30	NAD 1983 Albers	Land Cover Type [1]	Categorical Data
					Raster	30, 30		Percent Impervious Surface[2]	Categorical Data

Anthropogenic Features	Census Data	New Hampshire Office of Strategic Initiatives	https://www.nh.gov/osi/data-center/census/	1/21/2018	Vector	N/A	GCS North American 1983	Population Density	0-0.351 (Total Population of Census Block normalized over Census Block Size)
	Transportation	United States Geologic Survey	https://viewer.nationalmap.gov/basic/#productSearch	11/1/2017	Vector	N/A	GCS North American 1983	Distance to Closest Road	0-770 m
								Total Road Length in a 500m radius around well	0-3 m
								Distance to Closest Airport	484-15218 m
								Distance to Closest Train Track	15-34665 m
	Sewage Treatment Plant Locations	EPA Facility Registration Service (FRS)	https://catalog.data.gov/dataset/epa-facility-registry-service-frs-wastewater-treatment-plants	11/2/2017	Vector	N/A	WGS 1984 Web Mercator Auxiliary Sphere	Distance to closest Waste Water Treatment Plants (WWTP)	59-20247 m
								Distance to closest WWTP violators	217-24427 m

Anthropogenic Features	Solid Waste Disposal Site Locations	New Hampshire Department of Environmental Services	Available via Electronic Correspondence	11/25/2017	Vector	N/A	NAD 1983 StatePlane New Hampshire FIPS 2800 Feet	Distance from Solid Waste Disposal Facilites (SWF)	28-5626 m
								Distance from closest unlined SWF	53-9988 m
								Distance from closest Sewage Outfall	1481-68139 m
Surface Environment	Soil Type	New Hampshire's Statewide Clearing House	http://www.granit.unh.edu/data/downloadfreedata/category/databycategory.html	10/17/2017	Vector	N/A	NAD 1983 StatePlane New Hampshire FIPS 2800 Feet	Soil Runoff Potential[3]	Categorical Data
								Slope	7-65 degrees
	Digital Elevation Model	United States Geologic Survey	https://viewer.nationalmap.gov/basic/#cart	11/1/2017	Raster	9.2592593e-005, 9.2592593e-005	GCS North American 1983	Elevation	(-3)-1280 Meters
	Hydrology	United States Geologic Survey	https://viewer.nationalmap.gov/basic/#cart	11/1/2017	Vector	N/A	GCS North American 1983	Distance to Closest River	0-1488 m
								Distance to Closest Lake	0-1788 m

Geologic Characteristics	Aquifer Characteristics	New Hampshire Department of Environmental Services	Available via Electronic Correspondence	11/2/2017	Vector	N/A	NAD 1983 StatePlane New Hampshire FIPS 2800 Feet	Aquifer Area	1-2819946000 m squared
								Perimeter Area Ratio	0.0003-0.012
								Aquifer Transmissivity	0-94416
								Aquifer Saturation Thickness	0-96 L
	Water Table Characteristics	New Hampshire Department of Environmental Services	Available via Electronic Correspondence	11/2/2017	Vector	N/A	NAD 1983 StatePlane New Hampshire FIPS 2800 Feet	Water Table Elevation	(-2)-9430 Meters
	Quaternary Deposits	United States Geologic Survey	https://pubs.usgs.gov/imap/i-2789/	10/19/2017	Vector	N/A	NAD 1983 Lambert Azimuthal Equal Area	Surficial Deposit Type[4]	Categorical Data
	Bedrock Type	United States Geologic Survey	https://mrdata.usgs.gov/geology/state/state.php?state=NH	10/19/2017	Vector	N/A	GCS North American 1927	Primary Bedrock Type[5]	Categorical Data
	Seismic Fault Lines	New Hampshire Department of Environmental Services	Available via Electronic Correspondence	11/2/2017	Vector	N/A	NAD 1983 StatePlane New Hampshire FIPS 2800 Feet	Distance to Closest Seismic Fault	7-23740 m

Table A1. The source information of regression variables. This table includes all of the collection information, formatting, and resulting variables from each source.

[1]- Appendix B, [2]- Appendix C, [3]- Appendix D, [4]- Appendix E, [5]- Appendix F

APPENDIX B: Land Cover Types

MRLC Classification Number	Reclassification Categories
11	Water
12	Ice
31	Bare
21, 22, 23, 24	Developed
52, 71, 81, 82	Low Vegetation
41, 42, 43	Forests
90, 95	Wetlands

Table B1. Land Cover Categories. This table demonstrates the numeric categories of the MRLC land cover data and the groups that I separated them into for the dummy columns used in the exploratory regression.

APPENDIX C: Percent Impervious Surface Categories

Percent of Impervious Surface
0-19 %
20-39 %
40-59 %
60-79 %
80-100 %

Table C1. Percent Impervious Surfaces. This table represents the grouping cutoffs for the percent of impervious surface from the MRLC data. These categories were used in the dummy columns for the exploratory regression.

APPENDIX D: Soil Runoff Potential

Group A. Soils having a high infiltration rate (low runoff potential) when thoroughly wet. These consist mainly of deep, well drained to excessively drained sands or gravelly sands. These soils have a high rate of water transmission.
Group B. Soils having a moderate infiltration rate when thoroughly wet. These consist chiefly of moderately deep or deep, moderately well drained or well drained soils that have moderately fine texture to moderately coarse texture. These soils have a moderate rate of water transmission.
Group C. Soils having a slow infiltration rate when thoroughly wet. These consist chiefly of soils having a layer that impedes the downward movement of water or soils of moderately fine texture or fine texture. These soils have a slow rate of water transmission.
Group D. Soils having a very slow infiltration rate (high runoff potential) when thoroughly wet. These consist chiefly of clays that have a high shrink-swell potential, soils that have a high water table, soils that have a claypan or clay layer at or near the surface, and soils that are shallow over nearly impervious material. These soils have a very slow rate of water transmission.

Table D1. Soil Runoff Potential Key. These descriptions come from the New Hampshire government's classifications of soil groups for their potential to release water after rainfall. The categories were used in the dummy columns for categorical data in the exploratory analysis.

APPENDIX E: Surficial Deposit Types

Surficial Deposit Type
Clay and silt
Glaciolacustrine kame-delta deposits
Glaciomarine deposits and till
Ice-contact sand and gravel
Marine kame-delta deposits
Outwash sand and gravel
Sand and gravel
Sandy till--Ground-moraine deposits

Table E1. Surficial Deposit Categories. These groups represent the different surficial deposits above bedrock in New Hampshire. These were converted into dummy columns for the exploratory regression.

APPENDIX F: Bedrock Types

Rocktype	Reclassification
Granite	Igneous
Granitic gneiss	Metamorphic
Granodiorite	Igneous
Granofels	Metamorphic
Meta-argillite	Sedimentary
Metasedimentary rock	Metamorphic
Mica schist	Metamorphic
Migmatite	Metamorphic
Pelitic schist	Metamorphic
Phyllite	Metamorphic
Quartz diorite	Igneous
Quartzite	Metamorphic
Tonalite	Igneous

Table F1. Bedrock Types and Reclassifications. These are the main bedrock types in New Hampshire and their parent rock group. I used the parent rock group to break this information into dummy columns for the dummy variables in the exploratory regression.

APPENDIX G: Data Transformation

Variable	Transformation	Boxcox Lambda
PFOA	Log	-
PFOS	Log	-
Distance to Military Fire Training Locations	Boxcox	0.1
Percent Impervious Surface	None	-
Population Density	Log	-
Distance to Closest Road	Boxcox	0.1
Distance to Closest Airport	Boxcox	0.4
Distance to Train Track	Boxcox	0.05
Distance to Waste Water Treatment Plant	Boxcox	0.1
Distance to Waste Water Treatment Plant Violators	Log	-
Distance to Sewage Outfalls	Boxcox	-0.4
Distance to Solid Waste Facility Locations (SWF)	Boxcox	0.6
Distance to Unlined SWF	Boxcox	0.6
Slope	Log	-
Soil Runoff Potential	None	-
Total Road Length in a 500m radius around well	Log	-
Elevation	Boxcox	-0.1
Distance to Closest River	Boxcox	0.2
Distance to Closest Lake	Log	-
Aquifer Area	Log	-
Perimeter-Area Ratio	Log	-
Maximum Transmissivity	Log	-
Saturation Thickness	Log	-
Water Table Elevation	Log	-
Distance to Closest Seismic Fault Line	Log	-
Distance to Superfund Site	Boxcox	0.3
Distance to TRI Reporting Facility	Boxcox	0.05

Table G1. Variable data transformation. This table shows if a variable underwent a transformation to increase normality before being used in the exploratory regression. If a variable used a Boxcox transformation the lambda value of the data set is listed in the third column.