THESIS

FORECASTING GROUNDWATER CONTAMINANT PLUME DEVELOPMENT USING STATISTICAL AND MACHINE LEARNING METHODS

Submitted by

Elizabeth McConnell

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2022

Master's Committee:

Advisor: Jens Blotevogel

Kayvan Karimi Askarani
Jay Ham
Joseph Scalia

ABSTRACT

FORECASTING GROUNDWATER CONTAMINANT PLUME DEVELOPMENT USING STATISTICAL AND MACHINE LEARNING METHODS

A persistent challenge in predicting the fate and transport of groundwater contaminants is the inherent geologic heterogeneity of the subsurface. Contaminant movement has been primarily modeled by simplifying the geology and accepting assumptions to solve the advection-dispersion-reaction equation. With the large groundwater quality datasets that have been collected for decades at legacy contaminated sites, there is an emerging potential to use data-driven machine learning algorithms to model contaminant plume development and improve site management. However, spatial and temporal data density and quality requirements for accurate plume forecasting have yet to be determined.

In this study, extensive historical datasets from groundwater monitoring well samples were initially used with the intent to increase our understanding of complex interrelations between groundwater quality parameters and to build a suitable model for estimating the time to site closure. After correlation analyses applied to the entire datasets did not reveal compelling correlation coefficients, likely due to poor data quality from integrated well samples, the initial task was reversed to determine how many data are needed for accurate groundwater plume forecasting. A reactive transport model for a focus area downgradient of a zero-valent iron permeable reactive barrier was developed to generate a detailed, synthetic carbon tetrachloride concentration dataset that was input to two forecasting models, Prophet and the damped Holt's

method. By increasing the temporal sampling schedule from the industry norm of quarterly to monthly, the plume development forecasts improved such that times to site closure were accurately predicted. For wells with declining contaminant concentrations, the damped Holt's method achieved more accurate forecasts than Prophet. However, only Prophet allows for the inclusion of exogenous regressors such as temporal concentration changes in upgradient wells, enabling the predictions of future declining trends in wells with still increasing contaminant concentrations.

The value of machine learning models for contaminant fate and transport prediction is increasingly apparent, but changes in groundwater sampling will be required to take full advantage of data-driven contaminant plume forecasting. As the quantity and quality of data collection increases, aided by sensors and automated sampling, these tools will become an integral part of contaminated site management. Spatial high-resolution data, for instance from multi-level samplers, have previously transformed our understanding of contaminant fate and transport in the subsurface, and improved our ability to manage sites. The collection of temporal high-resolution data will similarly revolutionize our ability to forecast contaminant plume behavior.

ACKNOWLEDGEMENTS

The path to getting this degree has been long and circuitous, but I am very thankful to have the support of friends and family that made this accomplishment possible. The spirit of self-improvement and curiosity was instilled in me by my parents from a young age, and that motivated me to pursue a master's degree seven years after finishing my bachelor's degree. The professors from my undergrad experience at The University of Texas provided excellent formative education, and I want to give special thanks to Jay Banner and John Abbott for their mentorship and writing me letters of recommendation for my application to Colorado State University. Additionally, I would like to thank my former colleague and other recommendation writer, Alan Posnick, who also taught me so much about paddling and paleontology. Everyone I worked with at Texas Commission on Environmental Quality created a supportive environment where I absorbed great knowledge about Texas rules and regulations and made lasting connections. Without encouragement and support from my community and Austin, I would not have had the confidence to take this chance and embark on a new adventure.

I want to thank all of my professors at CSU, who were able to pivot to online instruction halfway through my second semester. I know it was an immense challenge for everyone involved, and I think they did an impressive job even as a global pandemic broke out. Special thanks goes to those on my committee, Jens Blotevogel, Kayvan Karimi Askarani, Jay Ham, and Joe Scalia. All of your guidance through my graduate program has made me a better student and I am grateful that you were each able to take on a role in my master's program. Kristina Quinn and the CSU Writes program gave me the space and tools to write a compelling manuscript and thesis.

Finally, I would like to thank anyone who aided me in my job search as this long program came to a close. From borrowing an FE study guide to forwarding my email on to a colleague, I know that I would not have had as much luck getting exceptional job offers without the help of my network. On that note, I am incredibly lucky to have widespread support and that has enabled me to choose a new challenge in my career, as a remediation engineer in New York City.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER ONE. INTRODUCTION

1.1. Problem Statement

A target of the UN Sustainable Development Goals is "By 2030, improve water quality by reducing pollution, eliminating dumping and minimizing release of hazardous chemicals and materials, halving the proportion of untreated wastewater and substantially increasing recycling and safe reuse globally." To reach this multi-disciplinary objective, attention must be paid to the clean-up of contaminated water in addition to the reduction of pollution released into the environment. Groundwater accounts for 30.1% of Earth's fresh water (Figure 1) (Gleick, Pacific Institute for Studies in Development, and Stockholm Environment Institute. 1993), and 99% of all liquid freshwater (J. Cherry 2020). This is a widely used resource for municipal and rural drinking water, but in many developed areas the shallow groundwater has been contaminated as a result of industrialization and the lagging regulation of chemicals. Modern data collection and analysis is crucial to understand how this vital resource is impacted and ways that it will be restored to productive reuse – either anthropogenically or through natural processes. Only through rigorous and transparent analysis of groundwater data can we begin to follow through with the mission of the US Environmental Protection Agency (EPA), "to protect human health and the environment."

Since the emergence of contaminant hydrology as a new discipline, the struggle to delineate and track movement of contaminant plumes in soil and groundwater has made vast strides. However, the heterogeneity of the subsurface challenges an accurate understanding. The advection-dispersion-reaction (ADR) equation was derived in the 1970s and provides a mathematical explanation for contaminant transport, but the determining mechanisms are often convoluted by

complex processes (J. A. Cherry and Freeze 1979). This poses a challenge not only in the delineation of groundwater plumes, but also the understanding of contaminant transport in the subsurface. Analysis of contaminant plumes is dominated by process-based analytical and numerical models that solve the ADR equation. Notably, tracer studies at highly characterized research sites have highlighted limitations in the predictive power of these models. Tracer studies have compared model outcomes with ground-truthed samples, with mixed results (Freyberg 1986; Pickens and Grisak 1981; Stephens et al. 1998). These issues are due to unrealistic assumptions inherent in the governing equation, lack of accounting for dual-domain mass transfer, and the role of preferential flow paths (Zheng, Bianchi, and Gorelick 2011).

Advances in the understanding of chemical reactions that govern contaminant breakdown have been made both in the field and in the lab, but the diversity of parameters that affect these reactions in the field make it difficult to apply general understandings to individual sites (Elder, Benson, and Eykholt 2002; Arnold and Roberts 2000; Fetter 1999). However, data collected from contaminated sites are becoming more comprehensive, and with that shift, empirical data-driven models may offer advanced opportunities in creating a better understanding of contaminant fate and transport at individual sites. Additionally, emerging contaminants have the potential to affect previously unregulated sites or change the regulatory targets at multi-contaminant sites. New methods of understanding the structure of the subsurface and predicting fate and transport of contaminants are needed.

Machine learning models have potential to transform earth sciences as has occurred in many other areas. Methods such as Positive Matrix Factorization (PMF) and linear discriminant analysis that are based on linear algebra have been applied to remediation site data. For example, PMF was used to identify areas of a plume in different stages of attenuation at a large

site in New Jersey (Capozzi et al. 2018). Linear discriminant analysis was used to show 1,4-dioxane attenuation at sites documented in California's Geotracker public records system (Adamson et al. 2015). Matrix factorization has proved to be a very useful tool in forensics to determine plume sources (Vesselinov, Alexandrov, and O'Malley 2018; Sun et al. 2020).

A variety of other statistical machine learning techniques have been applied to remediation data, including decision trees and neural networks. Decision trees have been used with text mined from reports of gasoline station sites to predict remediation costs (Farrell et al. 2007). Neural networks have also been applied to this type of data, notably to predict parameters, such as permeability and distribution coefficients, that are inputs to reactive transport models (Santos et al. 2021). Already, through these studies, there is evidence that machine learning techniques are applicable to data in the field of contaminant hydrology. Additionally, there are many more areas of machine learning should be explored to understand how they can improve our understanding of complex environmental systems. For instance, pattern recognition could be applied to imagery of fluoresced soil samples and recommendation systems could be created as decision tools for selecting remedies.

More specifically, the time series forecasting methods that have been developed in the field of econometrics that incorporate machine learning have recently been gaining popularity applied to earth sciences. Time series data of contaminant concentration in monitoring wells are foundational to contaminated site investigation and monitoring. The interpretation of these data is routinely done by non-parametric trend analysis such as the Mann-Kendall method that is not appropriate to predict future concentrations (ITRC 2013). An open source model named Prophet (Taylor and Letham 2018) has been demonstrated in applications to earth science, for example to predict air pollution (Nath et al. 2021; Shen, Valagolam, and McCalla 2020),

groundwater levels (Aguilera et al. 2019; Fernández-Ayuso et al. 2019), surface water pollution, and streamflow values (G. Papacharalampous and Tyralis 2020; Dabrowski et al. 2020; G. A. Papacharalampous and Tyralis 2018). Prophet is a well-suited model for predicting these values in some cases, such as applied to temperature, precipitation, groundwater level, and air pollution data, but in other cases using streamflow data there have been mixed results. The use of forecasting algorithms and numerical models in tandem may be beneficial, and indeed some have argued that an iterative process using forecasting prior to intensive numerical model calibration may improve upon predictive performance of numerical models (White 2017).

Numerical models are a category of code that use simplified representations of subsurface geology, then mechanistically apply a transport model and reactions. The time series models focused on here in this study use historical data to train a model that then predicts future values. For these types of models, there is not a mechanistic interpretation for either the historical data or the prediction. This can provide an advantage because the complexity of realistically representing the many processes contributing to fate and transport becomes impractical for large sites over long time scales (National Research Council 2013). However, a major disadvantage of this approach is that any mechanistic interpretation is based on the data and a general understanding of the physical and chemical processes at play, not explained by mathematics that represent those processes in the model as would be done in a numerical flow model.

1.2.Legacy Site Historical Datasets

Prior to creation of the EPA in 1970 and the implementation of the Resource Conservation and Recovery Act (RCRA) in 1976 and the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) in 1980, much of the waste from industrial operations was released

directly into the environment. Large facilities with historical releases were later identified and entered into state and federal regulatory programs. This necessitated characterization and remediation efforts carried out using boreholes and groundwater monitoring wells sampled on a site-specific basis. Identifying trends in contaminant concentrations and areas that required corrective measures based on this data guided site management. The long-term nature of environmental contamination means that many of these sites continue to be monitored and have archives of sampling data that span decades. The data compiled for this project consists of historical sampling data from twelve large legacy sites in North America. The largest of these sites had over one thousand monitoring wells sampled over the duration of remediation work, which has been ongoing since the late 1980s. Changing technology and sampling plans created gaps in the data that we have attempted to remedy or removed when necessary. The methods of collecting this data have largely remained the same; monitoring wells are sampled and the samples are sent to a lab for chemical analysis. Some parameters were tested in the field using probes or field kits (e.g. Hach sets). The results of these analyses are subject to biases related to monitoring well mixing, sampling errors, or quality control errors.

1.2.1. Analytical Data

In many cases the analytes are not detected and the result is reported as the detection limit of the analytical method. The *method detection limit* is the lowest concentration that can be identified with usually 99% confidence as different than the method blank (ITRC 2013). The *reporting limit* or *quantitation limit* is the lowest concentration of analyte that is within quality control guidelines when compared with known spiked concentrations. If the analyte is between the method detection limit and the reporting limit, it is flagged with a "J" and is considered present but not at a quantifiable level. This is a common laboratory analytical flag, but there are many

other flags that are identified in quality assurance/quality control of the laboratory analytical data can obscure the interpretation of the results. These flags are used to identify problems in the quality standards of the labs and are reported with the results of the analytical testing. Quality assurance and quality control are strictly regulated within labs accredited through the National Environmental Laboratory Accreditation Program (NELAP). This accreditation is required for analytical samples that are submitted to regulatory agencies to ensure that the results meet standards set in the governing regulations.

1.2.2. Field Data

Field data are generally collected using probes, rapid tests, or observations. Probes have commonly been used to determine depth to groundwater, oxidation-reduction potential, pH, dissolved oxygen, and temperature. The pH of groundwater is susceptible to changes when exposed to the surface because carbon dioxide that is dissolved in groundwater will off-gas into the atmosphere due to a lower partial pressure, and this will increase the pH of the groundwater sample (J. A. Cherry and Freeze 1979). The composition of groundwater depends on many factors, including soil organic matter, infiltration, depth to water table, temperature, and soil structure. For geochemical parameters, field rapid tests are commonly used to measure nitrate and nitrite, iron, and other constituents that provide information about the makeup of the groundwater. These kits are useful for field characterization but the results are not of the same quality as the samples that are analyzed in a lab. Some rely on colorimetric judgement, which is not standardized from person to person. In some cases these results are more valuable as qualitative indicators rather than intended to be recorded and used in a quantitative study.

Another kind of data that is collected in the field consists of observations of soil borings during site characterization that are recorded and serve to create cross-sections of the subsurface.

Although these observations are done by trained geologists, they can often be incorrect as a result of small variations in composition that may lead to different designations of Unified Soil Classification System (USCS) types. This is a common issue that may lead to problems in the creation and interpretation of a conceptual site model.

Many of the field recorded values from site characterization and sampling are not as reliable as the analytical values that are measured in a lab. This is not to say that these values are not important for a more complete understanding of the site, but the abstraction of the values themselves from the rest of the site data, as is done in preliminary investigations in this study, makes it difficult to use them in a quantitative manner.

1.3. Advances in Site Management

The short history of contaminated site remediation has seen several large shifts in the way that characterization and cleanup is approached. Because the field of contaminant hydrology largely grew out of the environmental movement and subsequent passage of regulations, there is a shorter history to the field than many other scientific disciplines. In many ways the basic tools of investigation are the same – classifying soils, laboratory analysis of samples, and applying principles of subsurface flow to contaminant plumes. However, the instruments and methods to collect those samples have improved with technology, and insights into flaws with early methodologies have led to more sophisticated ways of taking samples and measuring contaminants.

Wells construction for irrigation and drinking water wells have been well studied and understood but monitoring well construction has different objectives. The methods of construction have been refined several times to arrive at modern best practices. One area of concern for collecting representative groundwater samples is the length of well screen and the portion of the aquifer

that the well is screened in. These are important considerations because the contaminants of concern are often present in very small quantities and proper detection without dilution is a priority. The use of multi-level samplers led to much greater understanding of the fine resolution at which contaminant distribution may be widely variable in a complicated heterogeneous subsurface (J. A. Cherry and Freeze 1979; Nielsen 2005). If a well screen is very long, the sample may be diluted due to large contributions of groundwater from areas outside of the target stratum. The resulting composite sample over a large interval does not actually represent the heterogeneous conditions of the groundwater plume. This can be detrimental when planning effective remedies, particularly where to place permeable reactive barriers, since the design targets areas of the highest concentration of contaminants (Nielsen 2005). In addition, understanding what interval to place the screen is critical because a screen that spans an aquitard may lead to contamination of a previously unimpacted groundwater zone. In fact, many cases of incorrect well installation have led to contamination problems that did not formerly exist. Despite the long-known advantages of multi-level sampling wells and the dangers of conventional monitoring well installation with long screened intervals, the standard for most contaminated sites across the country remains the conventional method.

Beyond the design and construction of effective monitoring wells, the understanding of how contaminants themselves degrade has had a similar transformation. Since the inception of remediation at legacy sites, a paradigm shift has emphasized the importance of natural attenuation by microbes rather than the formerly predominant method of pumping and treating groundwater (Wiedemeier et al. 2007; Lu et al. 1999). Many of the modern remediation techniques involve changing the geochemical environment of the subsurface and/or adding microbes and nutrients to encourage microbially-mediated breakdown of common contaminants

(Fennell, Gossett, and Zinder 1997; Smatlak, Gossett, and Zinder 1996). Additionally, increased understanding of the concept of back diffusion from low-permeability zones has highlighted the problems with pump-and-treat methods (Chapman and Parker 2005; Seyedabbasi et al. 2012). This phenomenon is widely understood to contribute to rebounding in measured concentrations after treatment (Wilking, Rodriguez, and Illangasekare 2013).

Tools to improve operation efficiency, including genetic algorithms, reactive transport modeling software, and optimization problems, have been presented but no one method is consistently used. More commonly, basic guidance on screening out contaminants or pathways and models that take into account contaminant transport to develop a complete conceptual site model are employed. One such model is the 14-compartment model, which isolates the different possible media reservoirs and connects them through possible transport pathways to identify on which areas to focus (T. Sale and Newell 2011).

One method of optimization that has been proposed for long term monitoring is the use of genetic algorithms that identify, based on past data, well locations that may be redundant for understanding plume behavior and therefore could be abandoned (P. Reed, Minsker, and Valocchi 2000). Additionally, these algorithms can be combined with human subject matter experts to support decision making while giving power to users familiar with important regulatory or judgmental information (Babbar-Sebens and Minsker 2008). Genetic algorithm optimization problems are the foundation for a variety of decision support frameworks including the Monitoring and Remediation Optimization System (MAROS) and Adaptive Strategies for Sampling in Space and Time (ASSIST) (Aziz et al. 2003; Patrick Reed, Minsker, and Goldberg 2001). Although this area of research has yielded a wealth of site management tools, additional algorithms should be explored that may yield different answers.

Numerical models with parameters meant to simulate subsurface conditions have been developed in a variety of user-friendly formats to aid in site management. Stemming from the derivation of the advection-dispersion equation, computational simulations of geologic layout combined with details about groundwater movement that solve this equation have led to massive growth in our understanding of how fluids move underground. The most common framework is the numerical finite-difference model MODFLOW, created by the USGS and released as open-source software that has been adapted for many research applications. Some notable examples of basic models that include reactive considerations are REMchlor and PREMchlor (Liang et al. 2010). At smaller contaminated sites these models are unlikely to be used due to the amount of information that is required to be input before the models is able to run. However, at larger sites these tools may be among those accessed to create a detailed conceptual site model to produce an approximate understanding of where the plume may go next.

1.4. Machine Learning Overview

Machine learning is a field of study encompassing many algorithms that use existing data to make predictions (Alpaydin 2014). Machine learning algorithms are described by whether they are trained based on "right" answers provided by the user (supervised) or whether they only use input data to find relationships (unsupervised). There are also different categories of algorithms based on what they do, for example, classification, regression, and logistical regression. One algorithm may have several applications depending on the need, such as clustering, dimensionality reduction, and prediction. This document is not meant to provide detailed explanations of every form of machine learning, however, a brief overview of several relevant methods is presented to aid in understanding the basis of the following work.

1.4.1. Dimensionality Reduction

Dimensionality reduction is a sub-group of machine learning algorithms that simplify high dimensional data and enable visualization of the relationships between different parameters. This is a fundamental tool in understanding relationships between variables in large datasets with many different variables. Principal Component Analysis (PCA) is the dimensionality reduction algorithm that was used to analyze data in this study.

PCA is a technique that creates variables that are linear functions of the original variables and uses either an eigenvector or singular value decomposition solution to determine which variables account for the maximum variance in the dataset (Jollife and Cadima 2016). The new variables that account for the largest percentages of variance are the principal components. The amount that each variable contributes to the new variables are called the PC scores. Although this method based on simple linear algebra has been well known since the early 1900s, computers now allow for the application of PCA to large multi-dimensional datasets (Hotelling 1933). PCA has been used to understand complex data in many fields including genetics, atmospheric sciences, and paleontology. Other methods based on linear algebra include single value decomposition, linear discriminant analysis, and matrix factorization (Alpaydin 2014). These methods are similarly used to reduce large multidimensional data into interpretable dimensions.

1.4.2. Time Series Forecasting

Time series data are foundational to contaminated site monitoring. The interpretation of this data is often based on non-parametric trend analysis such as the Mann-Kendall method that is not appropriate to predict future concentrations (ITRC 2013). Advanced forecasting models have been developed based on demand for accurate predictions in business analytics and marketing. A standout model called Prophet that uses interpretable components and expands on a

decomposable time series model was released as an open source project by Facebook in 2018 (Taylor and Letham 2018). The creators of this method argued that in many industries there is a need for "forecasting at scale", in which many analysts that are familiar with their field but may not be experts in forecasting models should be able to make accurate forecasts using an interpretable and tune-able model. The model is designed to use an "analyst in the loop" workflow, where an analyst can easily create forecasts, detect anomalies, and tune parameters to correct them, all while efficiently forecasting with many datasets. This model has increasingly been applied to hydrology, atmospheric sciences, and other areas of earth science.

The basis of the Prophet model is a combination of the three components trend, seasonality, and holidays.

$$y(t) = g(t) + s(t) + h(t) + Et$$

In this equation, g(t) is the trend function, s(t) is a seasonal component, and h(t) represents holidays. Et is the error, which is assumed to be normally distributed. The trend term is based on an ecosystem carrying capacity equation and has mostly been used to model growth, though in this case it will be used to model degradation. The rate of growth can be adjusted at *changepoints*, which are automatically detected in historical data but can also be changed by hand. When the model makes a prediction, the trend term will include changepoints that reflect the frequency and magnitude of rate changes in the historical data. Endogenous variables are the past values of the variable to be predicted. Exogenous variables are external values that have some relationship that can improve the prediction of the target variable.

The seasonal component of the model is based on a smooth standard Fourier series. Seasonal patterns can be programmed by the analyst and can include daily, weekly, monthly, quarterly, or

annual seasonality. The holiday component is meant to account for events that can cause shocks to the projection but are not seasonal and require user knowledge. Common holidays for some countries are included, but the user can also program a list of past and future events that will function as holidays. In this analysis the holiday component is representative of remediation treatment events.

Each component can be individually inspected as a way of interpreting how each contributes to the full forecast. To determine the accuracy of the forecasts, mean absolute percentage error (MAPE) can be calculated as well as a method of simulated historical forecasts, which creates forecasts within the existing data timeline that can be compared to the real data. The ability to catch problems in the forecasts and fix them through tuning parameters is an important feature of the prophet model. Automatic anomaly detection is highly encouraged to highlight any unreasonable forecasts.

These forecasting methods are included in the fable package for the R programing language, with an additional package for prophet (Hyndman and Athanasopoulos 2021). The time series forecasting analysis in this study was done with this package. The availability of multiple time series models in this package creates a simple programming environment to test these models with consistent syntax rather than using different packages and programming languages. In addition to the models that were selected for detailed analysis, basic models including naïve and mean were used and tested for effectiveness, but ultimately performed much worse than those selected for further study.

Numerical models attempt to mimic the processes going on in the subsurface, then mechanistically apply a transport model and reactions. Prophet uses the historical data to train a model that then predicts future values in that time series. There is not a mechanistic

interpretation for either the historical data or the prediction. This provides an advantage because the complexity of realistically representing the many processes contributing to fate and transport becomes impractical for large sites over long time scales (National Research Council 2013). A disadvantage of this approach is that any mechanistic interpretation is based on the data and a general understanding of the physical and chemical processes at play, not explained by mathematics that represent those processes in the model.

The use of machine learning has revolutionized many fields and could have the potential to make similar advances with respect to contaminated site management (Suthersan et al. 2016). Challenges inherent in machine learning include data quality issues and having enough data. Using historical datasets, these challenges are amplified by lack of control over the sampling practices in the past. Nevertheless, sampling data from legacy sites should be explored using this lens to ensure that we can get the most out of the data we already have.

1.5. Reductive Dechlorination of Chlorinated Volatile Organic Compounds

Chlorinated volatile organic compounds are among the most recalcitrant organic contaminants in the world (*Alternatives for Managing the Nation's Complex Contaminated Groundwater Sites* 2013). Chlorinated ethenes and methanes are the focus in this document. These chemicals have been commonly used as solvents in dry cleaning and electronics manufacturing. Therefore, they have been rigorously studied and their pathways of dechlorination are well understood (Vogel et al. 1987). The geochemical conditions of the subsurface control which kind of reactions occurs and how fast (Fetter 1999). Reactions that break down chlorinated ethenes and methanes include substitution, dehydrohalogenation, oxidation, reduction, and elimination (Fetter 1999). The chlorinated ethene series is also referred to with the suffix -ethylene in previous studies, though they refer to the same chemicals.

Besides abiotic reduction by reduced metal species, both chlorinated ethenes and methanes may undergo biologically-mediated dechlorination in environmentally relevant conditions – though high concentrations of contaminants may be toxic to bacteria and reduce the effectiveness of this pathway (Mikesell and Boyd 1990; Harker and Kim 1990). The parallel pathways of PCE through either triple-bonded ethynes (a.k.a. acetylenes) or double-bonded ethenes (a.k.a. ethylenes) are shown in Figure 1. The triple-bonded pathway primarily occurs through reaction with zero-valent iron the permeable reactive barrier is composed of, while the double-bonded pathway is largely mediated through ubiquitous soil bacteria (Arnold and Roberts 2000).

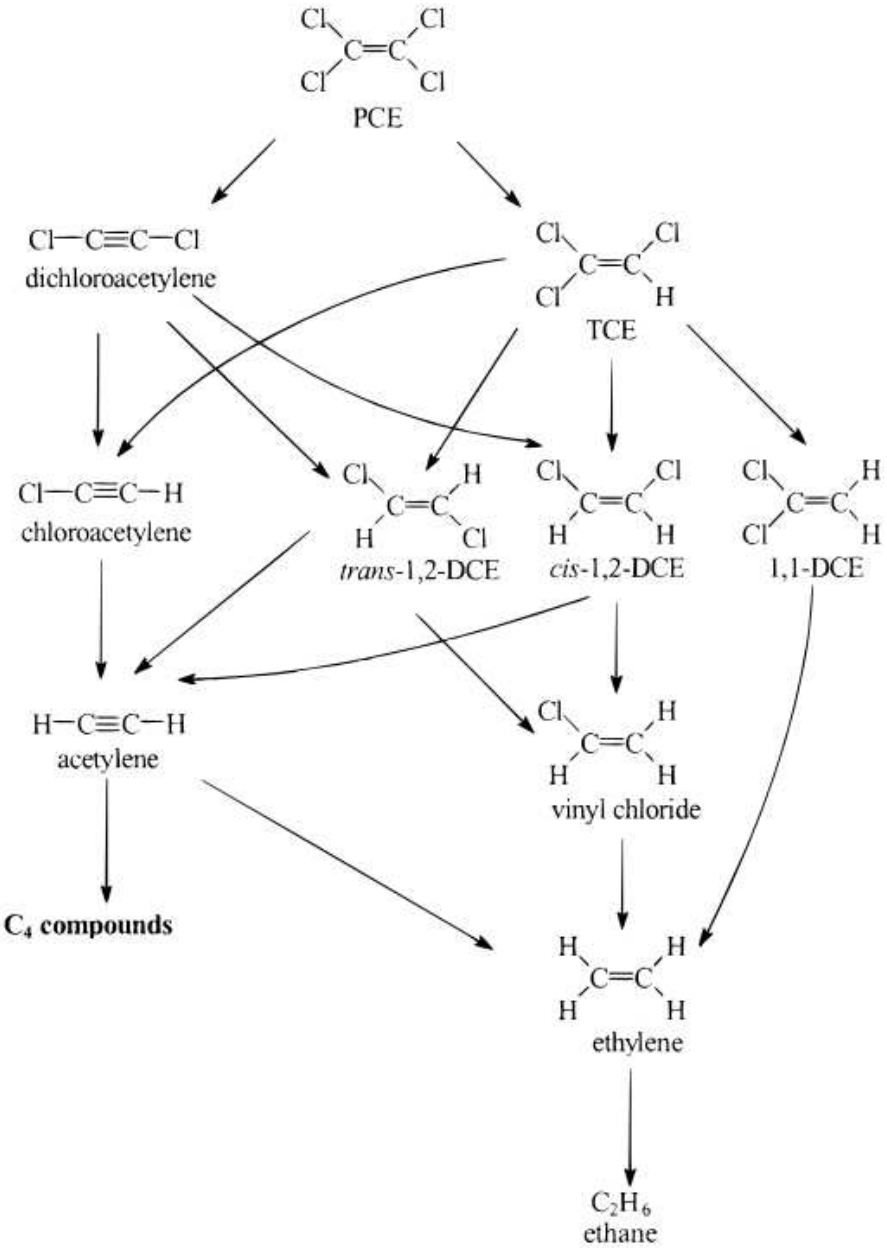

Figure 1: Reductive dechlorination pathways for perchloroethene (PCE) and its transformation products. The microbially mediated pathway is on the right, while the abiotic pathway is shown on the left. (Arnold and Roberts 2000)

1.6. Objectives

The first objective of this study was to explore the application of predictive machine learning methods to historical monitoring well data from legacy sites to improve efficiency of site management. The dimensionality reduction technique PCA, as well as basic correlation analysis were applied to the real-world historical dataset.

The second objective of this study was to reverse the initial task by creating a realistic synthetic dataset and exploring the question of how many data are needed for accurate groundwater plume forecasting. The modeling algorithms Prophet, damped Holt's method, and neural network autoregression were tested on the synthetic dataset. Although the objectives are exploratory in nature, it is important to test the applicability of machine learning models as it continues to impact modern data analytics.

The work included here uses a data-driven approach of providing available contaminant concentrations to train forecasting models without solving governing transport equations. Predictive forecasts using both a modified exponential smoothing model and Prophet have been compared to the output of a realistic synthetic dataset generated for a site with historical contamination. The effects of modifying training dataset size and temporal data density are interrogated to understand potential improvements in model performance. Additionally, we use Prophet to evaluate the benefit of incorporating training data from multiple observation wells as regressors to improve the performance of the prediction at downgradient wells. Forecasting in this manner will introduce a new tool for developing and refining conceptual site models and site management plans, and the computational efficiency of these methods creates opportunities to streamline decision-making and reduce monitoring/remediation costs at a wide variety of sites. With increasing quality and quantity of data, investigating the applications of these tools

will improve our understanding of contaminant transport and lead to practical uses that have the ability to improve modern site management.

1.7. Publications

As a result of my MS research, I submitted one first-author manuscript for peer review to a Special Issue of Groundwater Monitoring and Remediation on "Environmental Data Management and Analysis":

McConnell, L.; Karimi Askarani, K.; Cognac, K.E.; Mack, E.E.; Bartlett, C.; Ronayne, M.J.; Blotevogel, J.: Forecasting Groundwater Contaminant Plume Development Using Statistical and Machine Learning Methods. *Groundwater Monitoring & Remediation* **2022**, in review.

CHAPTER TWO. METHODS

2.1. Correlation Analysis and Site Description

The first objective in this investigation was finding which parameters were highly correlated in the large historical groundwater dataset as potential predictors. As discussed in chapter one, chlorinated volatile organic compounds were chosen as a focus of the study due to their widespread presence and well understood degradation pathways (Matheson and Tratnyek 1994). Simultaneous measurements of other parameters that could be correlated and thus potentially used as proxies were of particular interest.

The dataset contained measurements of both chemical and field parameters; however, the samples were taken as directed by the project managers overseeing the sites at the time. Many factors including cost, weather, and judgement dictated what was sampled for and when. This led to inconsistencies in the dataset that made consistent analysis of correlations difficult. For example, in some cases redox and chemical concentration were sampled concurrently many times, but dissolved oxygen was infrequently sampled. The correlation between redox and chemical concentration would then be based on different observations than the dissolved oxygen and chemical concentration. Additionally, there is a range of reliability in the field sampled results based on sampling methods and protocols. Dissolved oxygen is a parameter that is often incorrectly measured due to air exposure of the sample. Because of this, a binary designation of oxygen presence or absence was also included as a possible predictor. Another complication is that some chemicals (particularly trichloroethene) are both a parent and a daughter product, i.e., they were released at the site and are a degradation product of another chemical also released at the site. To try to address this, in addition to testing correlations between the raw chemical concentrations, the ratio of daughter to parent product for the chemicals in the tetrachloroethene

degradation series was also used. All chemical concentrations were converted to micrograms per liter. The distributions of chemical concentrations in general were skewed low, therefore, data were log-transformed prior to analysis.

The possible predictors were Total Iron, Dissolved Iron, Ratio of Dissolved to Total Iron, Dissolved Oxygen, Presence of Oxygen, Oxidation-Reduction Potential, pH, Sulfate, Manganese, Redox Couples. The target values included are tetrachloroethene (PCE) concentration, trichloroethene (TCE) concentration, *cis*-1,2 dichloroethene (*cis*-1,2 DCE) concentration , vinyl chloride (VC) Concentration, ratio of PCE plus TCE Concentration to *cis*-1,2 DCE plus VC concentration, psuedo first-order rate of PCE, psuedo first-order rate of TCE, psuedo first-order rate of 1,1-cDCE, psuedo first-order rate of VC, carbon tetrachloride (CT), and psuedo first-order rate of CT. The pseudo first-order rate constants were calculated by subtracting the subsequent concentration from the original concentration and dividing by the time in days between samples. This value is useful for estimating the time to reach a cleanup goal at a particular point in space (Newell et al. 2002).

Preparing a model to analyze site data requires thoughtful investigation of the controlling factors in a dataset. This includes handling non-detect values, joining and reshaping data, and exploratory analysis. Data cleaning of any field or analytical data was done prior to analysis. Concentration data was extracted for the selected chemicals from the raw dataset. Non-detect values were removed from analysis. If multiple results were reported for the same well and sample date, only the maximum value was used. Both concentration and rate values were log-transformed prior to analysis. Pearson and spearman correlation coefficients were calculated using the cor function in baseR and visualized using the corrplot package. The correlation

coefficient used is pearson unless noted to be the spearman correlation coefficient, which is nonparametric and preferred in some instances.

One field site that was chosen for more detailed study is located in California, U.S.A. (Figure 2). Three aquifers characteristic of fluvial deposition were identified in site characterization. Groundwater flows to the north and northeast toward a tidal river outlet into a bay. The surficial aquifer is dominated by fine-grained sands approximately five to 15 feet below ground surface. The upper aquifer is dominated by sand, silty sand, and gravelly sand approximately 35 to 50 feet below ground surface. The lower aquifer is dominated by sand, silty sand, gravelly sand, and sandy gravels approximately 50 to 110 feet below ground surface. The three layers are separated by aquitards that limit the movement of water between them. However, there is some communication between the three aquifers in portions of the site. All aquifers are considered part of a larger groundwater bearing unit that overlies a regional aquitard (Hocking et al. 2006).

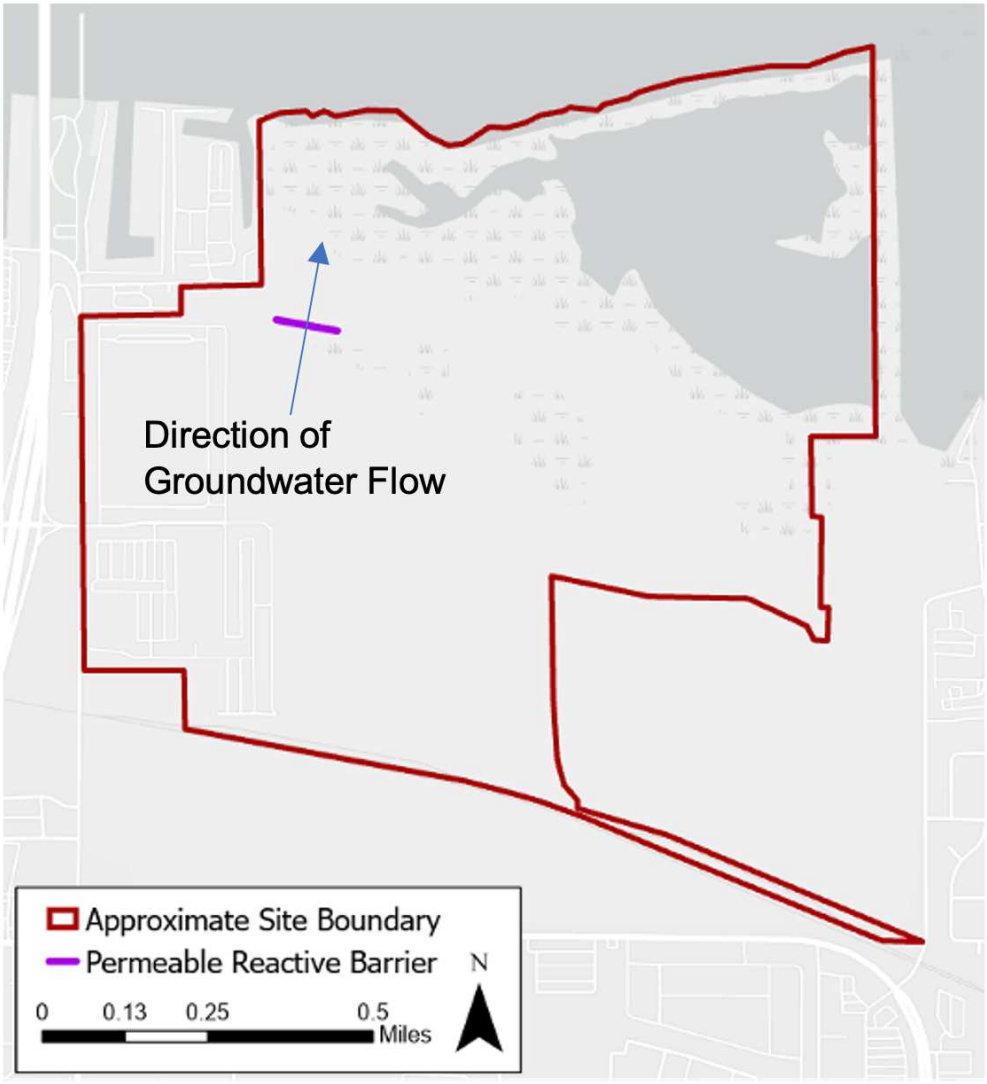

Figure 2: Map of approximate site boundaries and location of the permeable reactive barrier installed at the site.

The data at this site includes individual chemical observations back to the 1990s. There have been several remediation treatments applied to different areas of this site. One that this study will focus on is a permeable reactive barrier installed in the lower two aquifer units. Figure 3 shows the individual observations of carbon tetrachloride from wells on either side of the permeable reactive barrier.

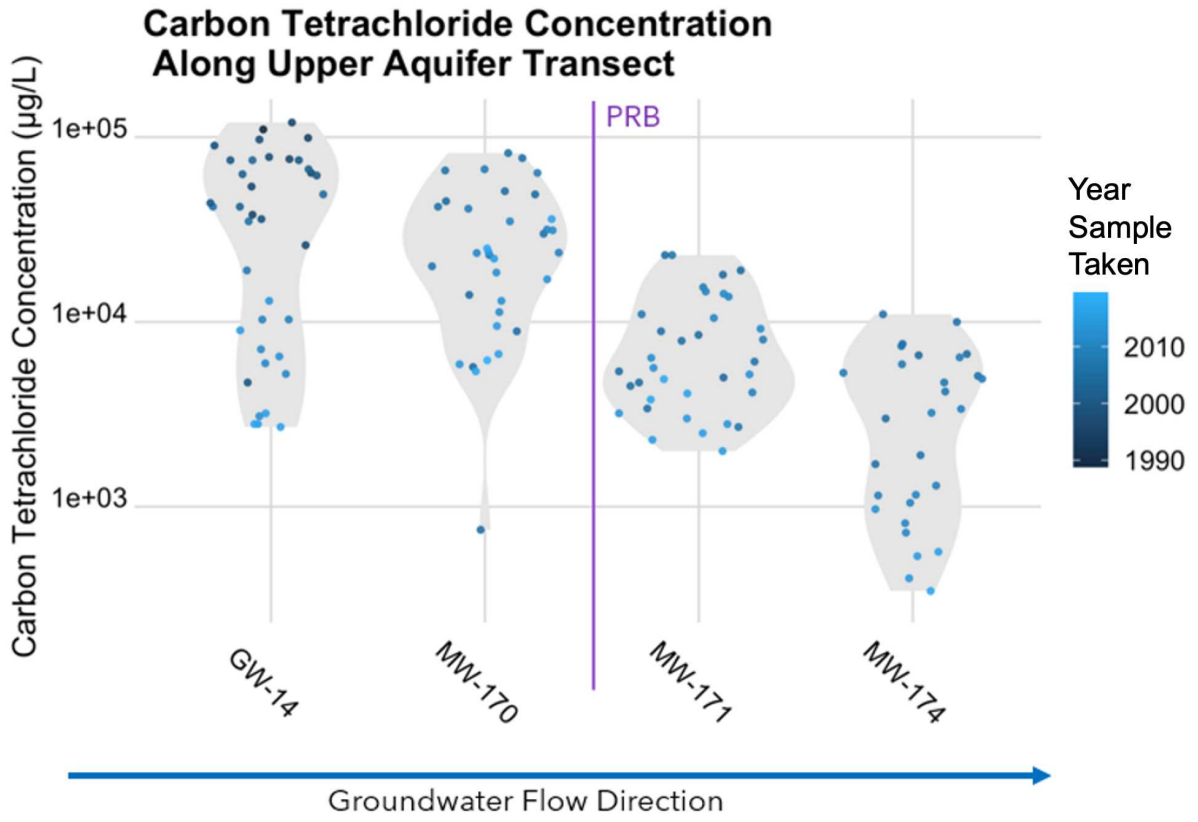

Figure 3: Concentrations of carbon tetrachloride before and after groundwater passes through the permeable reactive barrier. The wells to the left of the purple line are upgradient of the PRB, while those to the right are downgradient. The dots represent individual observations, colored by year taken.

2.2. Dimensionality Reduction Analysis

As another way of investigating the relationships between some of the parameters sampled from the same wells, dimensionality reduction techniques were used on the real-world data. The primary dimensionality reduction method that was employed was principal component analysis. As discussed previously, PCA is able to show the dimensions of the data that account for the most variability in the multivariate data. The R prcomp function was used to do the analysis and the factoextra package was used to visualize the results.

2.3. Synthetic Dataset Creation

Due to the inability to find strong correlations with parameters in the site data, we reversed our task by creating a realistic synthetic dataset using a numerical model and exploring the question of how many data are needed for accurate groundwater plume forecasting. With the understanding that chemical and geochemical samples taken from monitoring wells with well screens that allow mixing may be inaccurate, we decided to create a representative synthetic dataset to explore this question instead. This synthetic dataset was not created by me, but by my collaborators (K.E. Cognac and M.J. Ronayne, Dept. of Geosciences, Colorado State University). Nevertheless, this work is described here for a complete report and interpretation of this study's findings.

An existing groundwater flow model from the site investigation was adapted to generate detailed, synthetic concentration data sets that were input to machine learning algorithms to test input data requirements. The groundwater model was initially designed during normal site operations to evaluate the efficacy of pump-and-treat scenarios for containment of site contaminants (DERS, 1997). For this study, the spatial extent of the modeling domain was focused on the plume area rather than the full extent of the site. Additionally, the layout was rotated to gain better alignment of the grid with the primary transport direction and perpendicular PRB orientation. The flow model was converted from an older version (MODFLOW-96) to MODFLOW-2005 for manipulation (Harbaugh 2005). Reactive transport of carbon tetrachloride was performed using a reactive transport modeling add-on to MODFLOW, MT3DMS (Zheng 2010).

The revised groundwater flow and reactive transport model consisted of three layers, with 507 rows and 300 columns. Grid spacing in the horizontal direction ranged from 0.3 m near the PRB to a maximum of 3 m. Thickness of the layers ranged from 2 meters to 20 meters, reflecting

spatially variable aquifer thicknesses. Confining units between the aquifers were simulated using MODFLOW's quasi-3D approach of a vertical leakance parameter, as was done with the original flow model. Constant head values representing a river were assigned to the northern boundary cells. The southern boundary cells were also assigned constant heads – the head values were extracted from the steady-state solution of the original model. The eastern, western, and lower boundaries parallel to flow were assigned no-flow conditions (Figure 4). Parameters for hydraulic values were extracted from the original model, with horizontal hydraulic conductivity (K) values ranging from 26 to 100 m day^{-1} for different aquifers. Leakance values, necessary to simulate vertical flow through confining units, were calculated using a constant vertical hydraulic conductivity of $8.6 \cdot 10^{-5}$ m day^{-1}. Porosity for the geological unit was assigned a constant value of 0.3.

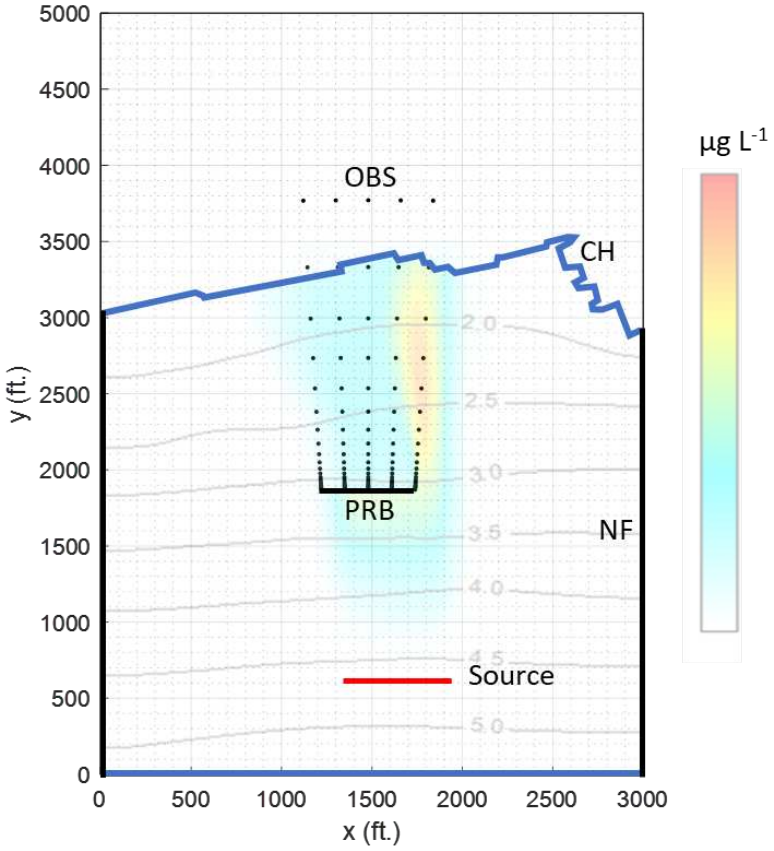

Figure 4: Groundwater flow and transport model diagram. This figure shows the layout (layer 2) delineating no flow (NF) and constant head (CH) boundary condition, observation point (OBS), source zone, and permeable reactive barrier (PRB) positions with simulated carbon tetrachloride concentrations (µg L^{-1}).

The permeable reactive barrier (PRB) installed at the site is composed of zero-valent iron filings injected into the subsurface. These types of barriers work by providing an electron donor to enhance the reductive dehalogenation of contaminants of concern as contaminated water travels through them. They have been the focus of numerous studies, several of which have sought to understand the effect they have on effective degradation rates (Henderson and Demond 2007). This effective degradation rate is one parameter in the numerical MODFLOW model that was modified to simulate the plume at this site. To calculate the effective degradation rate, several scientific studies were consulted to select reasonable values of surface area, density, and rate

constant. Values for rate constants of 0.012 L m^{-2} h^{-1} and 0.045 min^{-1} have been reported previously (Scherer et al. 1998; Matheson and Tratnyek 1994). The rate constant is written in terms of surface area since that varies greatly between reactive agents. A value of 1.0 m^2/g that is typical for zero-valent iron was applied to this site. A value of 2.9 kg/L was used as the density of zero valent iron filings (Scherer et al. 1998). A first calculation using the rate constant described by Scherer and co-workers was done, with the effective degradation rate calculated at 8352 per day. Another calculation done using the rate constant described by Matheson and Tratnyek (1994), yielded an effective degradation rate of 64.8 per day. These values are representative of barriers with fresh zero-valent iron that has not been passivated. Using data from the field site, the average effective degradation constant for the permeable reactive barrier was 2.57 per day. This value calculated from the real site was used in the model for the cells that represented the permeable reactive barrier because it best approximated the plume at the site.

The flow at the site was simulated under steady-state conditions, consistent with the original model and considered appropriate for the long-term simulation with no large changes in storage. Transient simulations of carbon tetrachloride (CT) transport and degradation used groundwater velocity vectors calculated from the steady-state hydraulic heads. CT was incorporated in the model through constant concentration cells assigned in a source area (Figure 4). In order to approximate historical conditions at the site, constant concentration cells in layers 1 and 2 were active during an initial source pulse lasting 5.5 years. The constant concentration cells were deactivated, effectively removing the source, for the subsequent 20 years of the simulation. The time steps for the transport model ranged from 0.1 to 20 days. CT degradation within PRB cells was simulated as a first-order decay reaction, meaning that the degradation followed a linear relationship. A rate constant of 2.57 day^{-1} (half-life = 0.270 days) was estimated from an

independent analysis of the percent reduction of CT due to the PRB as determined from field data, as discussed previously. Model-simulated concentrations of CT during the 25.5-year transient simulation were extracted for a series of observation points arranged into 5 transects extending downgradient from the PRB for use in the forecasting simulation (Figure 5). Synthetic Gaussian noise proportional to concentration values was added to the simulated CT concentrations prior to training with machine learning algorithms to account for spatiotemporal heterogeneities and other complexity not explicitly represented in the numerical model (Shuryak 2017). Addition of the noise was completed by drawing numbers from a normal distribution with a mean of zero and spatially scaled standard deviation set to 40% of the peak concentration at that location. Any negative concentration values that were generated using this method were subsequently set to zero.

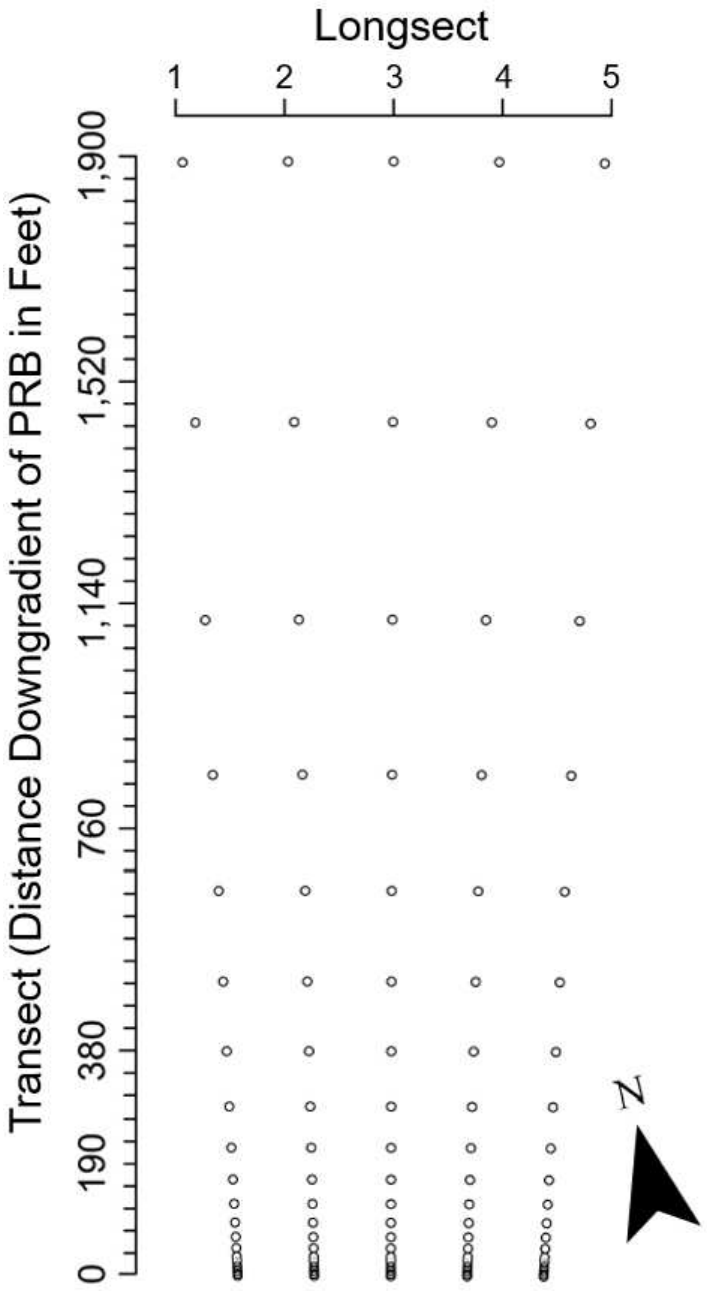

Figure 5: Naming conventions of longsects and transects of the synthetic monitoring well network. Transects are identified by the distance downgradient of the modeled permeable reactive barrier. They are spaced with exponentially increasing distance from the PRB, at 13, 17, 22, 29, 37, 48, 63, 82, 106, 138, 179, 233, 303, 394, 512, 665, 865, 1125, 1462, and 1900 feet downgradient.

2.4. Time Series Forecasting Models

Two forecasting models, Prophet and the damped Holt's exponential smoothing model, were used to predict CT concentrations using the synthetic dataset. Exponential smoothing models are a type of time series forecasting method that predict future values based on averages of past values, with exponentially less weight given to older observations. Methods based on this framework have been studied for many years and are quick and computationally efficient and have performed well in a wide variety of applications (Hyndman and Athanasopoulos 2021). Holt's method is a variation of an exponential smoothing model incorporating trend and seasonality (Holt 2004). The trend component in this method is based on a linear formula, which tends to overestimate values far into the future. The variation of this method that accounts for this overestimation includes a dampening parameter that reduces the magnitude of the trend component until a point where it no longer has an effect (Gardner and Mckenzie 1985). This damped Holt's exponential smoothing model was selected for analysis in this study based on performance with eight observation points distributed within the synthetic plume that were analyzed using a variety of other basic forecasting methods.

Prophet is another type of time series forecasting model that uses interpretable components and expands on a generalized additive model. It is fairly recent in its provenance – it was released as an open source project by Facebook in 2018 (Taylor and Letham 2018). The creators of this method argued that in marketing, as well as many industries there is a need for "forecasting at scale", in which many analysts that are familiar with their field but may not be experts in forecasting models should be able to make accurate forecasts using an easily interpretable and tunable model. The model is designed to use an "analyst in the loop" workflow, where an analyst can easily create forecasts, detect anomalies, and tune parameters

to correct them, all while efficiently forecasting with many datasets. In this case represented by individual monitoring wells at a contaminated site. This model has increasingly been applied to hydrology, atmospheric sciences, and other areas of earth science, primarily in research contexts.

The synthetic data described above were used to train these models. In this analysis, different simulated sampling intervals and amounts of univariate training data were considered to better understand how limitations of real-world data might impact the usability of these forecasting methods. The input data for each forecasting model was obtained from 100 individual sampling points (or synthetic monitoring wells) in the synthetic plume. The 100 chosen points comprise the lower layer of the numerical model, which is the layer with the most real-world data points. A value of 0.96 for phi as the damping parameter was used, as determined by iteratively tuning models between the recommended reasonable values of 0.8 to 0.98 (Hyndman and Athanasopoulos 2021) on eight wells distributed through the plume. The Prophet model was customized using the logistic growth setting of the model specified for the trend, with a floor of 0 and a capacity of the maximum value of the training dataset plus one (in this case 86.8). The change point prior scale of 0.5 was similarly determined by iterating through recommended reasonable values of between 0.001 and 0.5 (Prophet 2021) for the same eight representative wells.

2.5. Accuracy Measures

Predicted CT concentrations generated by each forecasting model were compared to the synthetic data at various locations (Figure 13) and different times in the same location. Model prediction errors were quantified using the root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{N}(y_i - x_i)^2}$$

and mean absolute error (MAE):

$$MAE = \frac{1}{N}\sum_{N}|y_i - x_i|$$

In the preceding equations, N represents the total number of observations (concentration values) for a particular location in the synthetic dataset, x_i is the i^{th} observation in the test set, and y_i is the predicted equivalent of the i^{th} observation. RMSE is the mean of the squared prediction errors, while MAE is the mean of the absolute values of the prediction errors. Both measures are in mg/L, which is the same units of the time series. In other cases where different units are being compared scaled error metrics would be advantageous (Taylor and Letham 2018), but because the units of measurement in our study are the same and some values are close to zero, we chose to use RMSE and MAE to quantify the error. Smaller values for both metrics represent lower error and therefore better performance of the models. The RMSE is more sensitive to outlier errors and as such tends to be higher than the MAE. These error metrics were calculated based on the forecasts over the interval of the test period compared to the original smooth synthetic data extracted from the numerical groundwater transport model.

CHAPTER THREE. RESULTS AND DISCUSSION

3.1. Correlation Analysis at Different Scales

Initially, lab and field historical groundwater datasets from large legacy sites across North America were used in analysis to determine if parameters measured on a standard basis (e.g., oxidation-reduction potential, electron acceptor concentrations) were correlated with contaminant concentrations or contaminant parent-to-daughter ratios. Parent-to-daughter ratios were chosen because reductive dechlorination kinetics have been reported to linearly correlate with the redox potential, i.e., the lower the ORP is the faster reductive dechlorination proceeds (Valiev et al. 2008). No strong correlations were found between chlorinated compound concentrations and values of pH, redox potential, and dissolved oxygen (DO) (Figure 6). Due to the notorious inaccuracies of DO measurements taken from field samples, dissolved oxygen was subsequently recoded to a dummy variable of "oxygen presence" with a value of 1 for a reported concentration greater than 1 mg/L and a value of 0 otherwise. Additionally, the ratio of parent to daughter product between carbon tetrachloride and chloroform was included as a descriptor of dechlorination in the analysis.

After correlation analysis with data combined from multiple large sites did not yield satisfactory correlations, a targeted area of enhanced remediation – a permeable reactive barrier (PRB) – was chosen for detailed study. A PRB is an engineered barrier, sometimes a backfilled trench but in this case a zone of injected zero-valent iron filings, that is intended to intersect a plume where the major contaminant of concern is reactive. In this case the major contaminant of concern in this plume was carbon tetrachloride (DERS 1997). Several wells installed on either side of the PRB had been sampled on a regular schedule since the time that the barrier was installed, and a few had been sampled since far before the construction of the PRB began. Although data were

available for indicators of dechlorination and parent and daughter products, no strong correlations were found to relate the electrochemical indicators to the redox processes that degrade chlorinated organic compounds. The clear correlations that we previously expected in such a strongly biased environment downgradient of a PRB were not found.

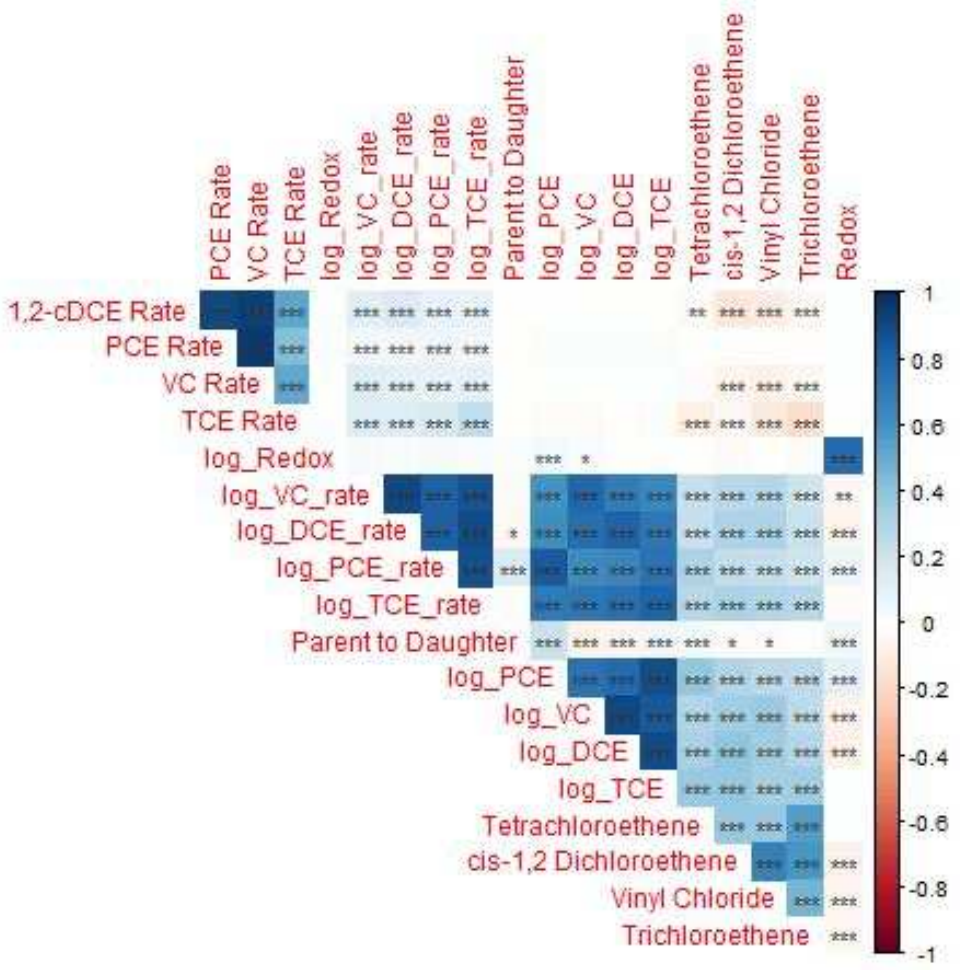

Figure 6: Correlation plot of relevant Tetrachloroethene degradation series parameters and oxidation-reduction potential. The colors represent the Pearson correlation coefficient and one, two and three asterisks represent p-values of 0.05, 0.01, and 0.001, respectively, at a 95% confidence interval (Friendly 2002). Specific values of the coefficient are not included but rather represented by color for ease of reading.

3.1.1. Entire Dataset

Initial correlation analysis was intended to validate the most available data and explore the possibly of using large datasets from historical sites to investigate common contaminant degradation processes at field sites. Using the most possible simultaneous measurements from individual wells across all field sites, correlation analysis was done on relevant parameters. Redox did not correlate with any other parameter (except log redox) more than +/- 0.2 (Figure 6). These unsupportive results could be due to diverse sites in terms of release histories, geographical location, and heterogeneity of the aquifer / groundwater samples.

3.1.2. Site-wide

The spatial extent of the analysis was narrowed to within a single site. This was done to eliminate any distorting effects from sites with different major contaminants of concern or in different geochemical environments. In this analysis, redox is positively correlated with parent to daughter ratio, which is an expected result because a higher redox value should cause higher persistence of the parent compound within a reductive transformation pathway. However, the correlation is not very strong (0.07) (Figure 7). This correlation analysis was done using a non-parametric spearman correlation coefficient. The correlation between redox and pH was 0 for this site, which differs from the pattern seen with other data of consistently negative correlations, consistent with redox chemistry that proceeds with higher pH at lower redox potentials and vice versa (J. A. Cherry and Freeze 1979).

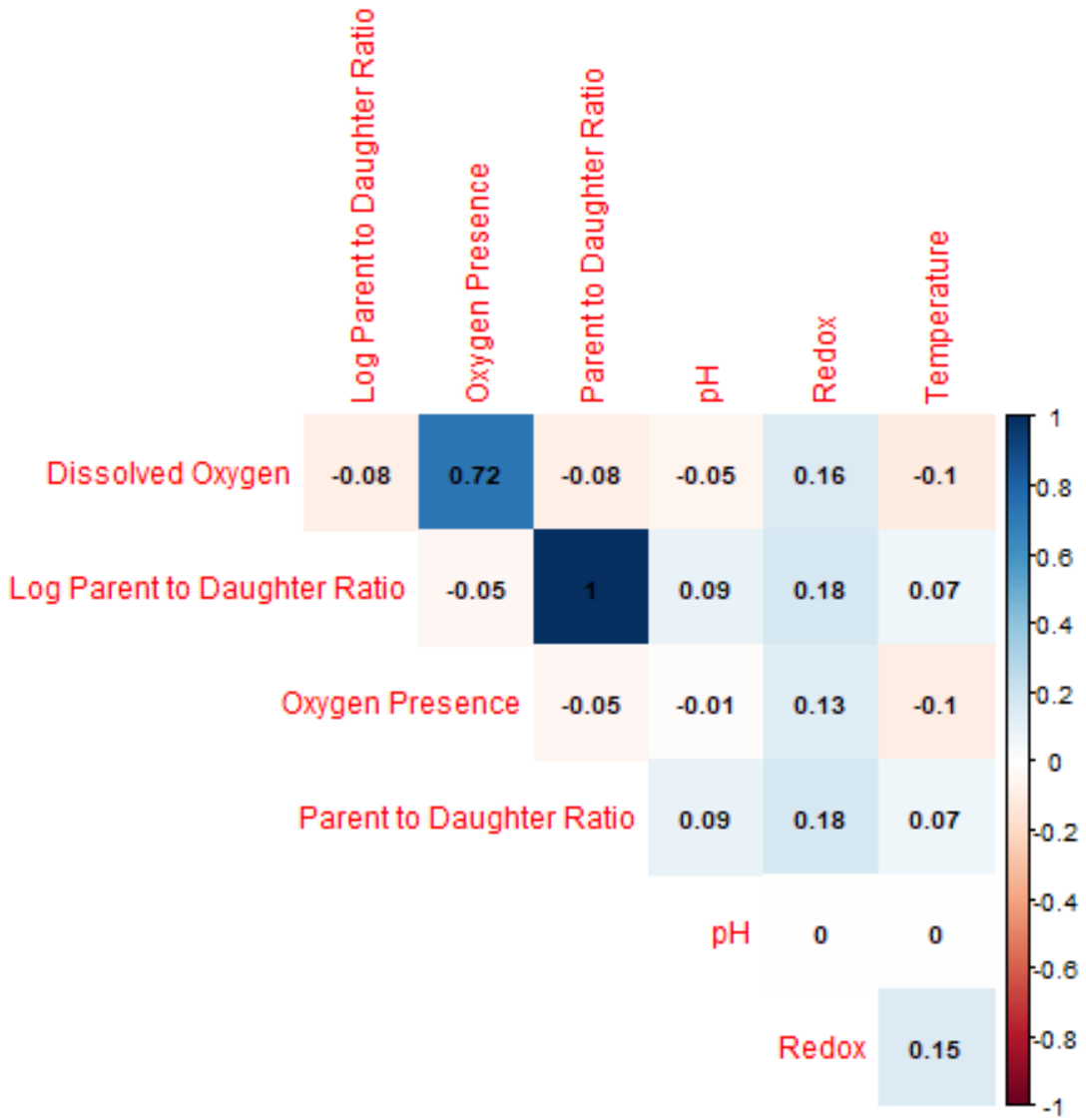

Figure 7: Spearman correlation analysis at a large former industrial site.

3.1.3. Active Remediation Areas

After testing whether correlations were observable between relevant parameters in the entire dataset and the data from within one site only and finding unconvincing results, the spatial area included in the correlation analysis was once again narrowed to only the area of a treatment to increase reductive dechlorination, the permeable reactive barrier. Since the effect of zero-valent

iron on the dechlorination of contaminants has been proven both in mechanistic laboratory studies and in the field, it was hypothesized that those relationships would be evident based on the historical data from this portion of the site. The wells immediately downgradient of the permeable reactive barrier that were screened in the same unit as the barrier were analyzed together to see if the relationship was apparent for this area. Redox and pH did have a slight negative correlation (-0.22) (Figure 8). However, the results did not show the expected positive correlation between redox values and the ratio of parent to daughter compound (Figure 8). In fact, a slight negative correlation was found between those parameters (-0.23).

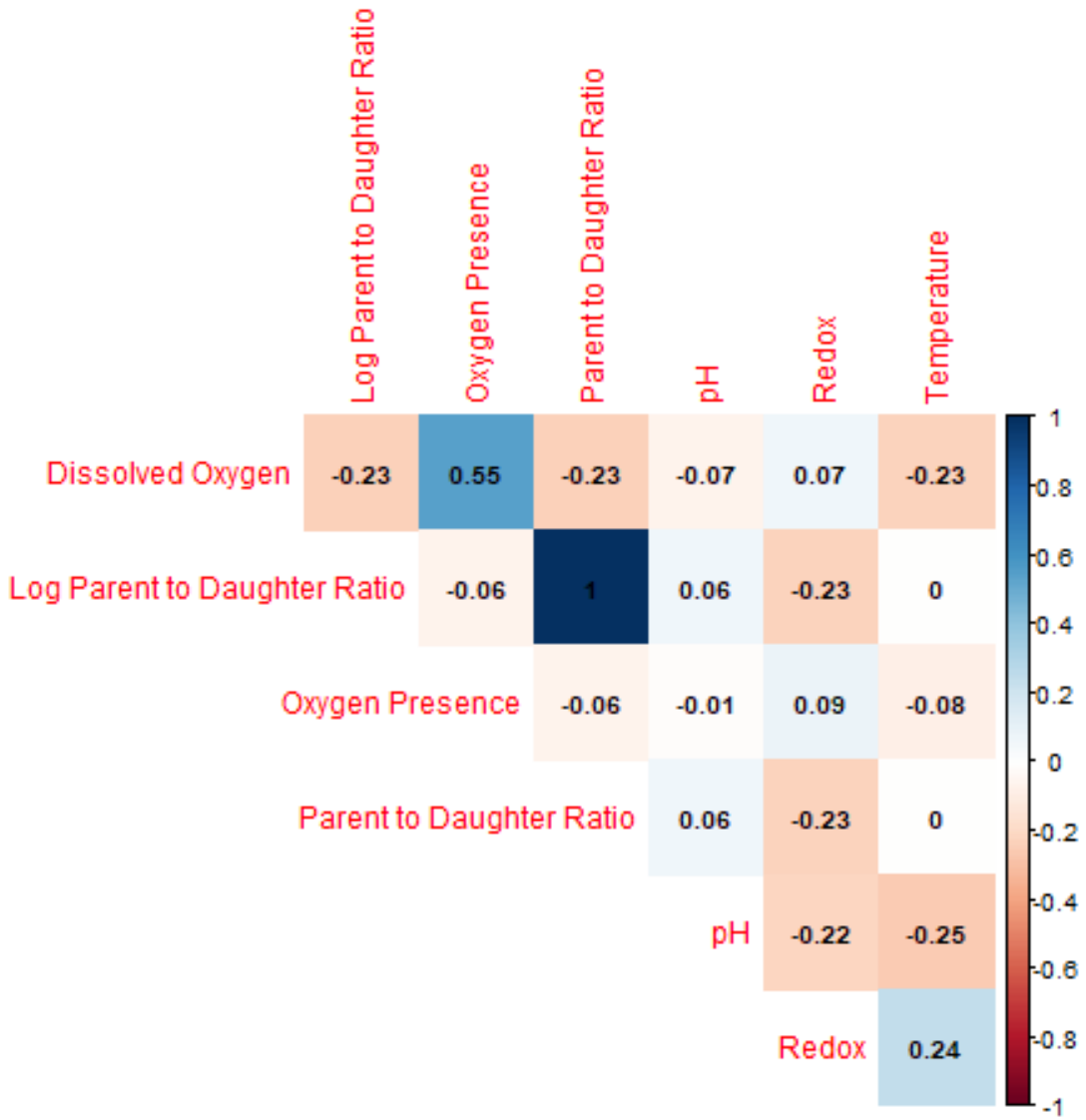

Figure 8. Spearman correlation analysis of parameters in wells downgradient of the PRB.

This result was unexpected because the decline in concentrations of carbon tetrachloride in the wells downgradient of the permeable reactive barrier is apparent in the data. The plot of concentrations of carbon tetrachloride for the upgradient (MW-184) and downgradient (MW-185) of PRB wells along the same flow path shows the sharp decline in the downgradient well compared to the gradual decline in the upgradient well (Figure 9). Additionally, Figure 3 shows

another visualization of the decline of the concentration of carbon tetrachloride across the PRB. The two wells that are downgradient of the PRB have much lower concentrations after the time that the PRB is installed. The observations are colored by date, and it is generally evident that more recent samples have the lowest concentrations (Figure 3). This figure also gives an idea of the number of observations that exist for one monitoring well at this site.

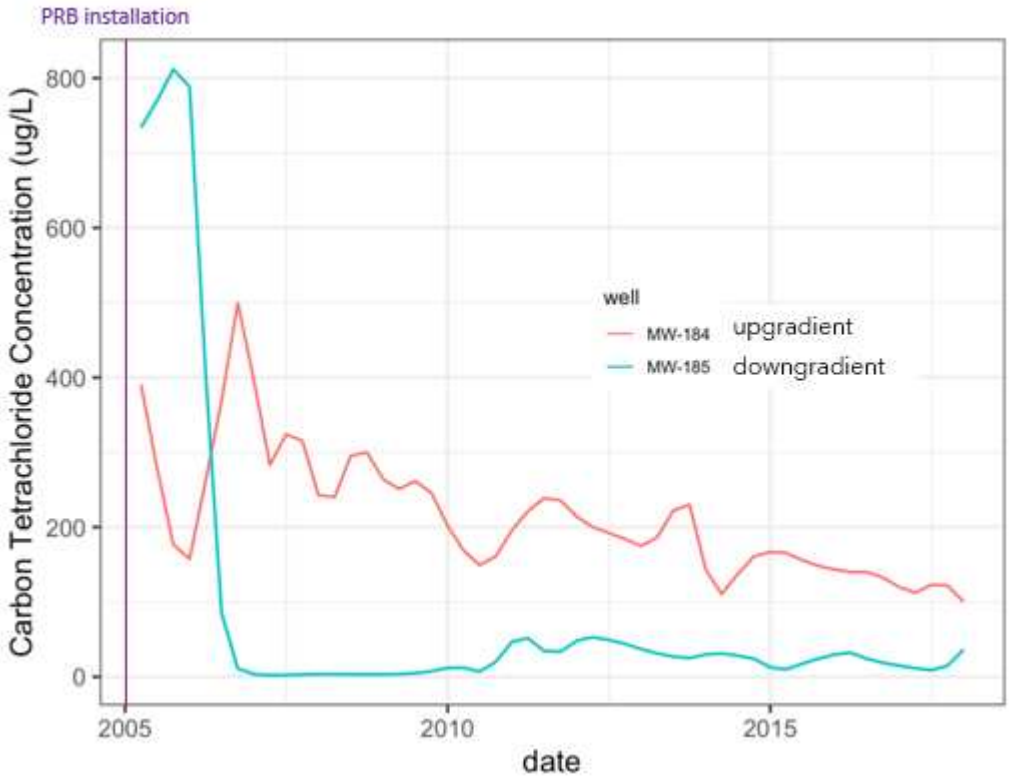

Figure 9: Time-series plot showing carbon tetrachloride concentrations at one well upgradient and one well downgradient of the permeable reactive barrier.

The site-wide correlation did not reveal a compelling pattern, however, if the spatial scale of the correlation analysis is once again narrowed to just the downgradient well, a positive correlation as would be expected is apparent (Figure 10). This analysis uses the least number of observations in any of the analyses done up to this point, yet is the most compelling evidence of the relationship that exists between the reductive dechlorination of contaminants and redox values.

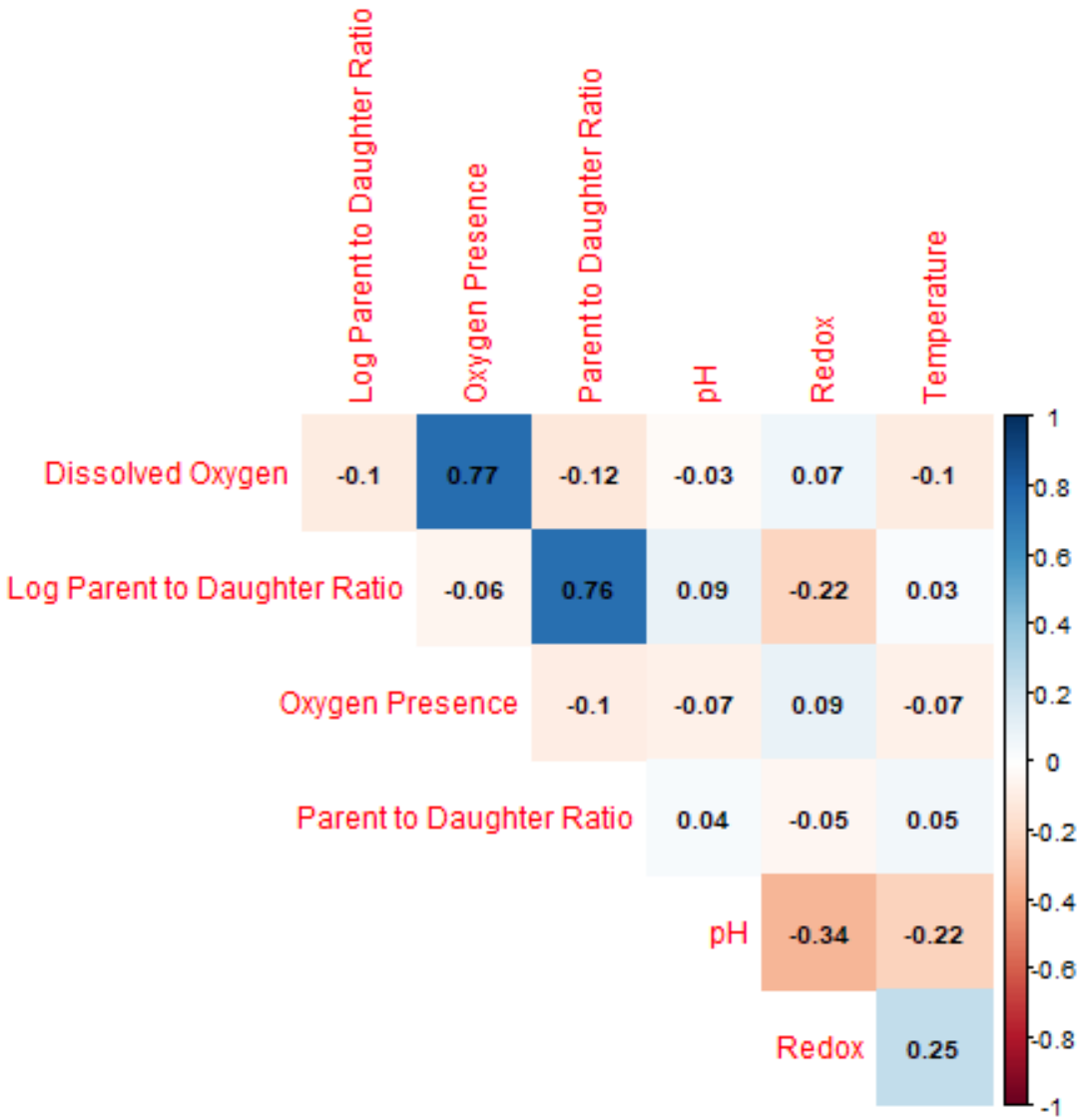

Figure 10: Pearson correlation plot of relevant parameters at the well downgradient of the permeable reactive barrier.

Ultimately, these data are not compelling for our purposes due to the few data points that compose the dataset showing the correlation that we expected to find across all data points in this study. While correlation analysis is valuable to measure the effect of a remedy at the well-specific level, we did not find evidence that this relationship is present across combined multiple well, simultaneously collected data. Because the mechanism of reductive dechlorination has been

demonstrated, the relationship between redox and decreasing concentrations would not be expected to conflict with field data. Differences in sample collection, measurement, and reporting could contribute to this complication.

3.2. Dimensionality Reduction

Correlation analysis using Pearson and Spearman correlation coefficients measures the linear relationship between two variables. Other dimensionality reduction methods are able to detect a nonlinear relationship between many variables. Principal component analysis was applied to a combination of analytical and field data from individual wells. Again, a relationship between the chemical products of reductive dechlorination and geochemical indicators was interrogated. The orientation of the redox parameter in relation to the top two principal components and the chlorinated hydrocarbons shows that there is a relationship, particularly between the more chlorinated species, although the first and second dimensions of the dataset do not account for most of the variance of the data (27.8 and 8.4, respectively) (Figure 11).

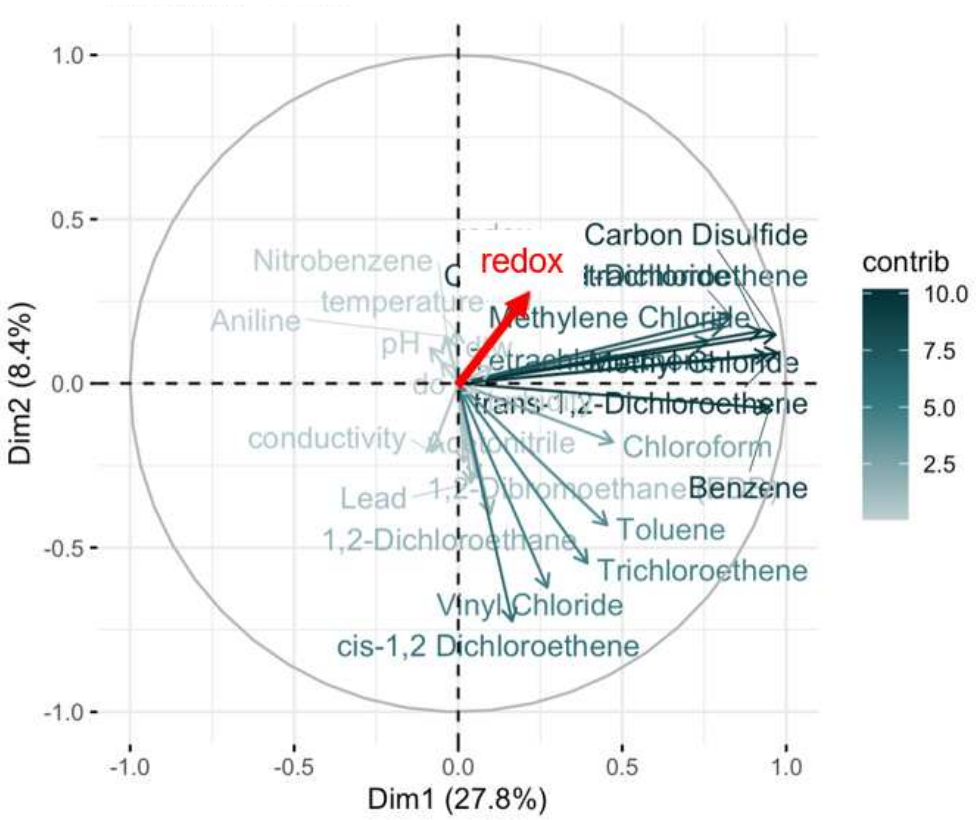

Figure 11. PCA plot showing that redox points in the same direction of some of the higher chlorinated constituents, indicating a relationship between redox and the attenuation to daughter products.

3.3. Forecasting Model

To address the question of what spatial and temporal data resolution is required to accurately forecast groundwater contaminant behavior, portions of the synthetic dataset were used to train forecasting models and to compare the predictions against test data to investigate what model performed best. Additionally, the models were trained with data at different simulated sampling intervals to determine the effect of decreasing temporal density of data on the accuracy of forecasts. The amount of training data was also varied to determine the effect that this would have on the model performance. Finally, data from upgradient wells were used as regressors in the forecasting models to understand if including data from different spatial arrangements could substantially improve the forecasting performance.

3.3.1. Model Performance

Time series plots for eight representative synthetic wells in the dataset downgradient of the PRB are shown in Figure 12. The synthetic observation well concentration forecasts are closest to the test values in areas near the PRB (e.g., transect 17 ft downgradient of the PRB). Both the RMSE and MAE were smallest for the transect 17 ft downgradient of the PRB, except for the forecast using the damped Holt's method in transect 303, longsect 2, which had values of RMSE and MAE of 0.21 mg/L and 0.17 mg/L, respectively. For the observation point in transect 17 and longsect 2, RMSE and MAE were 0.06 mg/L and 0.04 mg/L, respectively, using the damped Holt's method, and 0.54 mg/L and 0.51 mg/L, respectively, using Prophet. For wells in the area near the PRB, the downward trend is correctly predicted by both methods. For the synthetic wells in transect 303, the test data in longsect 2 is well predicted by Holt's method, while the pattern of the data in longsect 4 is not correctly identified by either method. The predicted concentrations for the synthetic wells in transect 865 and 1462 are inaccurate because the training data covers only a period of increasing trend. The highest MAE for the eight wells was 10.4, for transect 303, longsect 4, using the Prophet method. The highest RMSE in these wells was 11.4, for transect 865, longsect 4. This highlights the limits of long-term forecasting when training data are from an individual monitoring well where concentrations may still be rising. Many legacy sites that are in the monitoring stage are comprised of wells with consistently declining concentrations due to the nature of historical sources. However, with less frequent sampling, natural variability can often cause results that are not representative of the prevailing trend behavior in the subsurface (McHugh et al. 2011). For wells with increasing concentrations of contaminants, these forecasting models that are not constrained by data from other spatial locations would be inappropriate due to ongoing source loading. Additionally, if new releases or

significant changes in groundwater flow occur, these forecasting algorithms may not effectively take into account and therefore forecast trend changes.

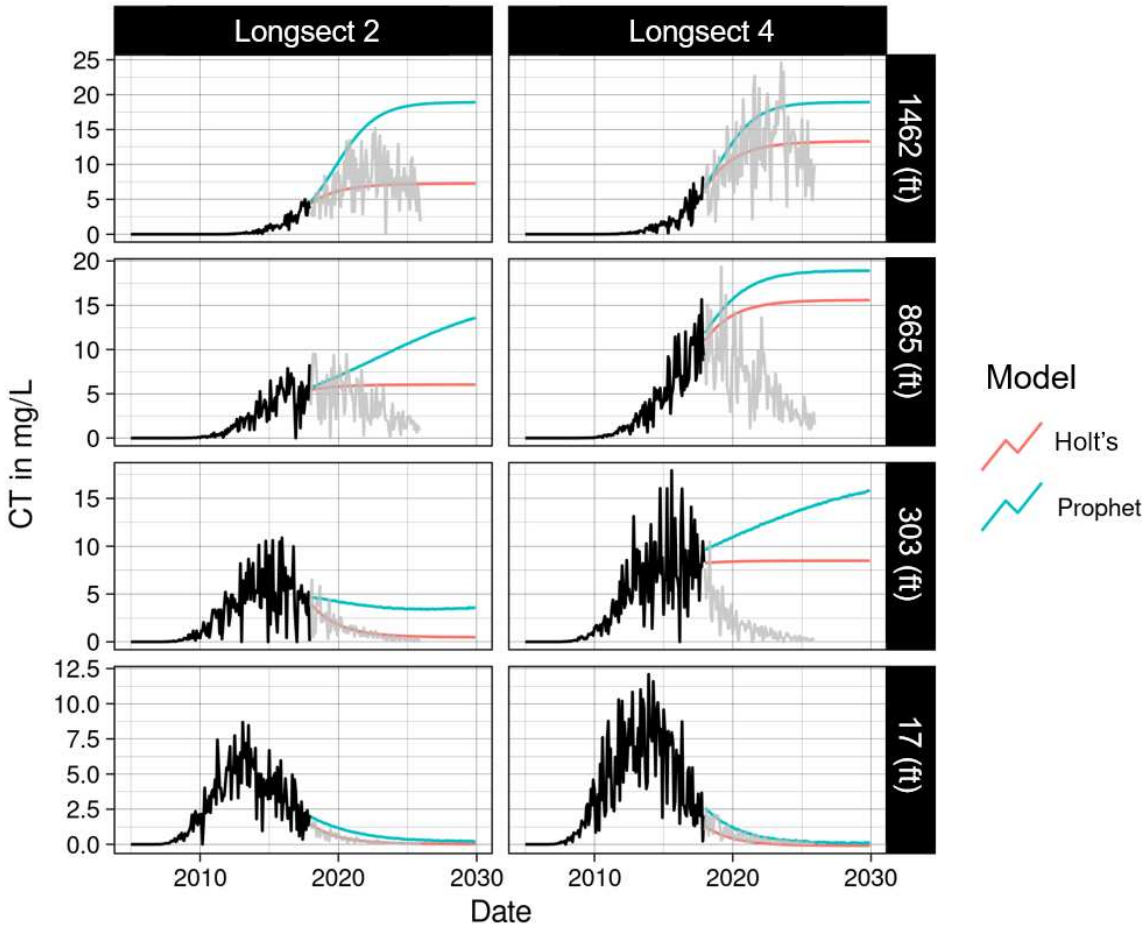

Figure 12. Forecasts of carbon tetrachloride (CT) concentrations at eight simulated wells in longsects 2 and 4, at 17, 303, 865, and 1462 feet downgradient from the permeable reactive barrier. Both Prophet and the damped Holt's method models were trained with data up to the end of 2017. The black line represents the synthetic observation training data, whereas the gray line represents the synthetic observation test data.

Considering predictions based on training data through 2017, the mean RMSE was 13.4 for Prophet and 4.6 for the damped Holt's model, and the mean MAE was 11.6 for Prophet and 4.2 for the damped Holt's model. Prophet RMSE ranged from 0.2 to 69.2 and the damped Holt's model RMSE ranged from 0.05 to 40.3. Prophet MAE ranged from 0.2 to 65.1 and the damped Holt's model MAE ranged from 0.04 to 37.4. Forecasting accuracy improved using training data from 2005 through 2020, with a mean RMSE of 5.6 for Prophet and 2.7 for the damped Holt's model, and a mean MAE of 4.7 for Prophet and 2.4 for the damped Holt's model. Prophet RMSE ranged from 0.03 to 54.8 and the damped Holt's model RMSE ranged from 0.01 to 58.6. Prophet MAE ranged from 0.03 to 49.1 and the damped Holt's model MAE ranged from 0.01 to 54.2.

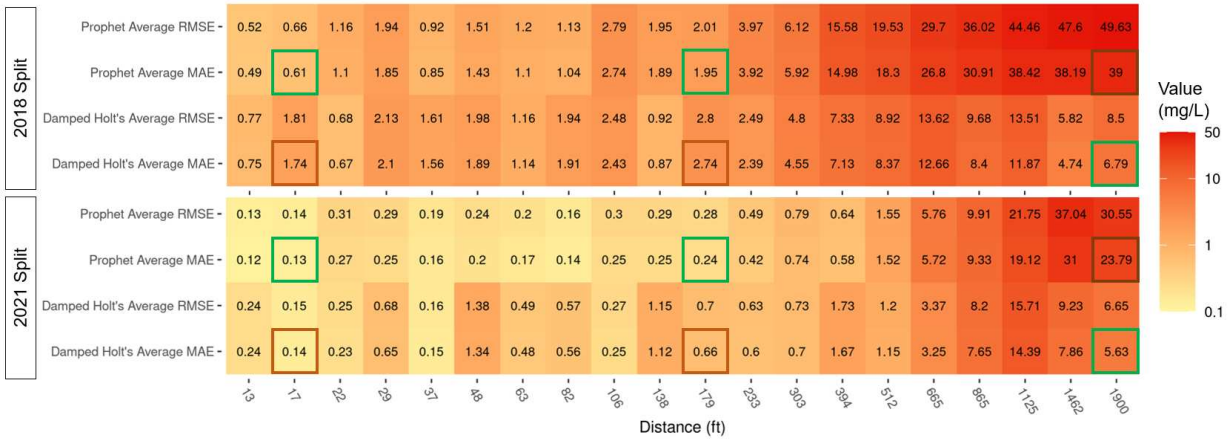

Figure 13. Transect averages of accuracy measures for forecasts trained on data up to 2018 (top) and 2021 (bottom). Highlighted are the mean absolute error at different transects.

These values vary spatially, affected by factors including maximum concentration and distance from the source. Errors are smaller with increasing amounts of training data, the data trained up to 2021. In this analysis, prophet performs better than the damped Holt's model for the closer wells, but worse for the wells furthest downgradient of the PRB.

Maps of the forecast carbon tetrachloride concentrations in comparison to the output of the numerical model in the year 2025 are shown in Figure 14. The predicted concentrations

generated by the damped Holt's method tends to slightly overestimate concentrations in the upgradient southern portion of the well field, while this forecast underestimates the carbon tetrachloride concentrations further downgradient. Prophet largely overestimates the values of the concentrations compared to the synthetic dataset.

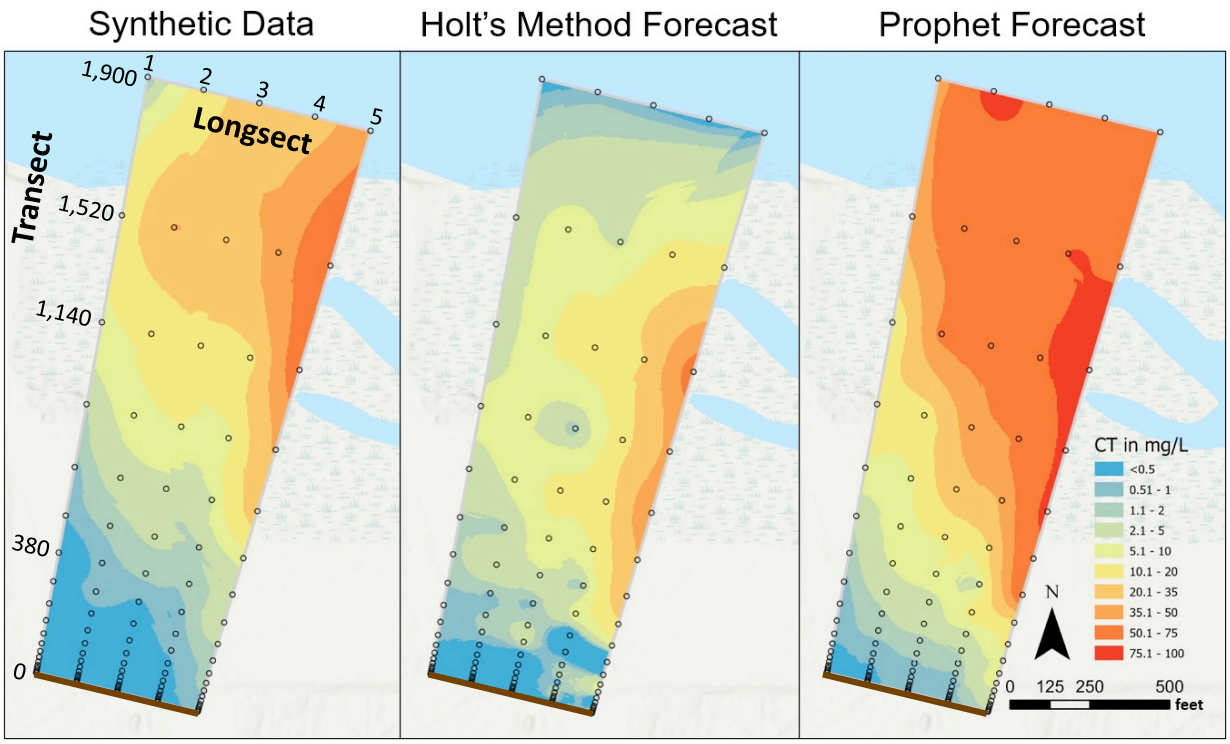

Figure 14. Maps of predicted carbon tetrachloride concentrations. Groundwater plumes of carbon tetrachloride downgradient of the PRB (brown line at the southern end of the predicted plume) in 2025 based on the synthetic data (left), damped Holt's method forecast with training data through 2017 (middle), and Prophet forecast with training data through 2017 (right). The direction of groundwater flow is north-northeast, perpendicular to the PRB. The map of the synthetic data (left) includes naming conventions of longsects and transects of the observation point network.

3.3.2. Effects of Increased Sampling Intervals

At typical remediation sites, groundwater sampling frequency is determined by many factors, including cost, risk, regulatory requirements, and degree of site characterization. Quarterly sampling is a common sampling schedule required by regulatory agencies, and has been used as a benchmark for cost and time required to calculate an effective attenuation rate (McHugh et al. 2016). Sites where the effectiveness of a remedy or nearby potential receptors are being closely monitored may be sampled on a more frequent basis. At older or smaller sites where there is little risk, sampling schedules are more likely to be semi-annual or annual. Typically, the sampling schedule is more frequent when monitoring wells are initially installed, then decreases as a better

understanding of contaminant fate and transport evolves and risk diminishes. To understand the impact of a less frequent sampling schedule on forecast accuracy, two different temporal densities, monthly and quarterly, up to the year 2018 were used for model training and compared to the test data until 2030 (Figure 15). At the synthetic well in longsect 3 in the transect 303 ft downgradient from the PRB, damped Holt's method performs satisfactory compared to the test data, using monthly monitoring data – with an RMSE of 0.19 mg/L, and MAE of 0.15 mg/L. Prophet performs worse using monthly training data at this location (RMSE = 2.9 mg/L; MAE = 2.8 mg/L), however, it still approximately replicates the shape and direction of the trend. When the same forecast models are trained with quarterly data, predictive capability is significantly reduced (Figure 15) and the Prophet model does not correctly forecast the trend change. The damped Holt's method replicates the trend direction, however, this is likely an artifact, as the exponential smoothing places higher weight on more recent samples and the quarterly sample variability includes low values close to the training data cutoff date (Figure 15). In addition, the damped Holt's method does not correctly predict the magnitude of the change. Therefore, while quarterly or semi-annual groundwater monitoring are often sufficient to evaluate risks for downgradient receptors, accurate plume forecasting may require more frequent groundwater sampling schedules.

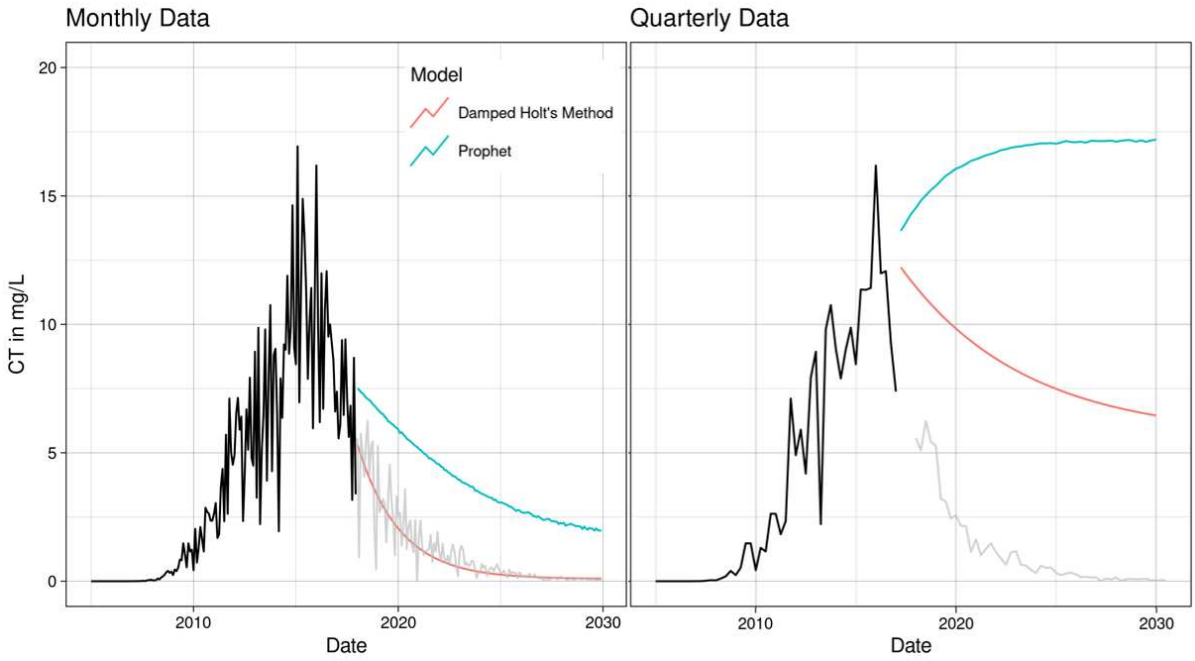

Figure 15. Forecasts of the synthetic data with monthly and quarterly sampling frequencies. Synthetic well in longsect 3, transect 303 ft downgradient of the PRB with training data through 2017, using monthly data on the left and quarterly data on the right. The change in data density from monthly to quarterly sampling programs results in poor forecasts by both models.

3.3.3. Comparing Different Training and Testing Data Splits

The amount of training data that is fed into the forecasting model is critical to the forecast's accuracy. As shown and discussed previously with wells at different distances from the source area, the point at which the concentrations begin to decline is a factor in the ultimate accuracy of the prediction. Additionally, this is illustrated for a longsect 3 observation point in Figure 16. With training data up to the end of 2017, both models perform poorly (Figure 16). With a few years' more training data (training data split in 2021), the models are still not sufficiently recognizing the change in the trend. Interestingly, while the forecast should improve with a later training split, the damped Holt's method forecast worsens, supporting our conclusion above that the forecast is heavily biased by more recent data points. The case with the most training data, with the training split in 2024, both models provide an acceptable forecast until the end of the

dataset in 2030. The RMSE using Prophet was 12.3, 9.96, and 0.42 for 2018, 2021, and 2024 splits, respectively. The RMSE using Damped Holt's Method was 3.23, 6.34, and 0.22 for 2018, 2021, and 2024 splits, respectively. The MAE using Prophet was 11.4, 9.5, and 0.38 for 2018, 2021, and 2024 splits, respectively. The MAE using Damped Holt's Method was 2.6, 6.09, and 0.18 for 2018, 2021, and 2024 splits, respectively. Additionally, the models performed best for short predictions and declined in accuracy for far ahead forecasts. In conclusion, a robust training data set needs to be collected before reliable forecasts can be achieved.

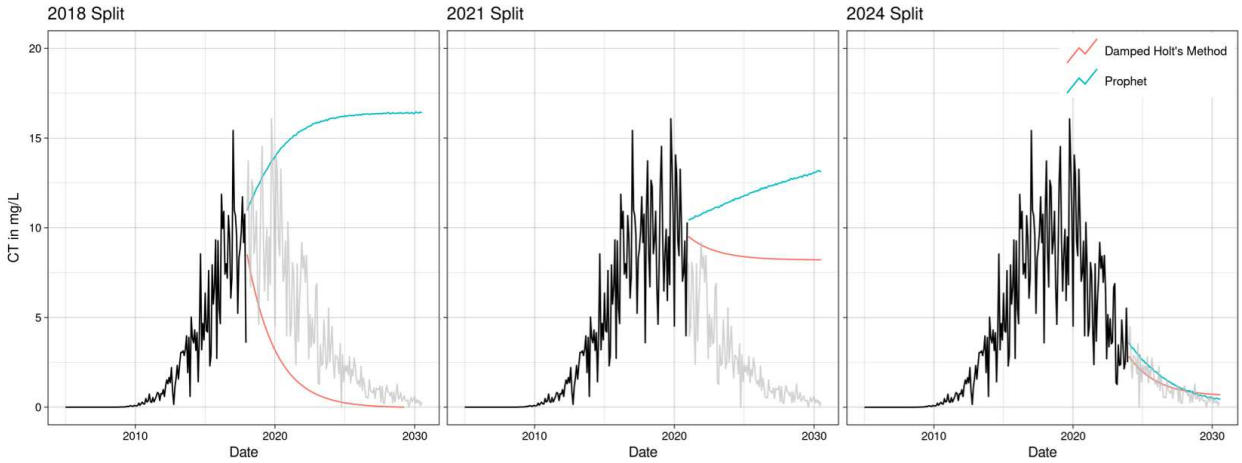

Figure 16. Forecasts of the synthetic well in longsect 3, transect 865 ft downgradient of the PRB with the training and testing data split at three different dates, 2018, 2021, and 2024. The black line is the training data, the blue line is the Prophet forecast, the red line is the damped Holt's method forecast, and the grey line is the test data.

3.3.4. Including Upgradient Well Concentrations

The analysis discussed up to this point was done using univariate data. Groundwater wells are not installed in isolation; rather, they are installed as a network. This is done because monitoring wells can only measure point concentration in a continuous aquifer. Contaminants may be measured in all the wells, but at different times based on advection, diffusion, and reactions. Upgradient well data may contain a valuable record of what patterns may appear in downgradient data, as shown in Figure 12. Thus, we chose to test whether including this

upgradient well data as a nonlinear regressor could improve the forecast at downgradient wells. The damped Holt's method does not support using exogenous regressors; thus, only the results using Prophet are presented here.

Figure 17 shows that a forecast using training data up to 2018 with one upgradient regressor well produces a better forecast than the historical data up to 2018 alone. The RMSE of the prediction using one upgradient well as a regressor decreases from 8.8 to 5.6. One important improvement is that the trend is changed from increasing to decreasing with the additional information from another spatial location. When four upgradient regressors in adjacent longsects are added to the model, the RMSE decreases to 4.1 from the original 8.8. This is an incremental improvement when using additional upgradient regressor data. However, relatively greater benefits were observed when changing from quarterly to monthly data (Figure 15) or increasing the amount of training data (Figure 16). The spatial relationship of the regressor wells may play an important role in the improvement of forecasts on the target well. Unfortunately, the monitoring well network is typically determined by other factors, and while they may be placed to gain the most insight into the subsurface plume behavior, they are not amenable to relocation for optimization after their installation.

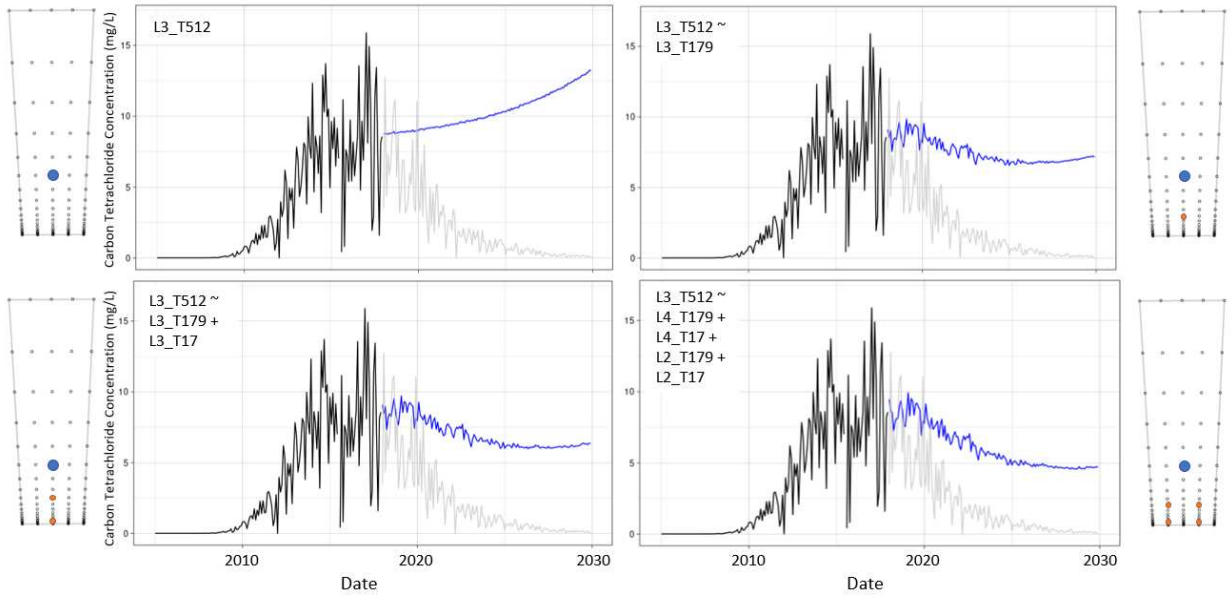

Figure 17. Forecasts showing historical and predicted data using the Prophet model with historical data only, one upgradient regressor well, two upgradient regressor wells, and four upgradient regressor wells. All models were trained with data up to the end of 2017. The large blue dot on the diagram of the synthetic monitoring well network represents the well being forecast and the orange dots represent the regressor well.

CHAPTER FOUR. SUMMARY AND CONCLUSIONS

Many large historical datasets compiled from contaminated sites with data from samples taken over 20 or 30 years have been collected. While not without unique value, the utility of these data for the purpose of contaminant plume forecasting may be limited due to improper or outdated groundwater sampling design and techniques. In this project we were unable to find meaningful correlations or patterns using basic correlation analysis and dimensionality reduction techniques. However, data collection is becoming increasingly sophisticated; as the quality and quantity of environmental data improve, data-driven analysis will have greater potential to address relevant questions beyond compliance monitoring.

Using two different models with univariate data, this analysis has shown that the damped Holt's modified exponential smoothing model largely outperformed Prophet in forecasting plume behavior. However, in wells with consistently increasing contaminant concentrations, a future declining trend was unable to be accurately predicted without external regressors. In those cases, only when upgradient well data are included in the analyses, can decline curve forecasts be achieved. Overall, satisfactory predictions of time to closure are possible if the collected data used for model training are of sufficient spatial and temporal density. Nevertheless, specific model biases should be well-understood and established during sensitivity analysis as has been done in our study.

Spatially high-resolution data, for example from multi-level samplers, have previously transformed our understanding of contaminant fate and transport in the subsurface, and improved our ability to manage sites. Collecting temporally high-resolution data will similarly revolutionize our ability to forecast contaminant concentrations and time to site closure. It is becoming increasingly clear that this revolution is imminent with the rapid emergence of sensor-

based collection of better and cheaper data. The applicability of sensors to high-resolution monitoring of contaminated sites has been recently demonstrated (Blotevogel et al. 2021; Askarani and Sale 2020; Sale et al. 2019). This analysis documents the results of the performance of only select forecasting algorithms – more research in this area will be necessary to take advantage of the full potential of advanced statistical and machine learning algorithms. Additionally, dashboards with real-time data are becoming more common, and implementing forecasting capabilities automated for many wells will ensure that predictive capabilities are available to site managers (Askarani and Sale 2020). Building trust that predictions are reliable, both at a regulatory and operational level is crucial to the adoption of forecasting models across the field.

CHAPTER FIVE. RECOMMENDATIONS AND FUTURE WORK

The practice of environmental remediation is currently dictated by guidance and regulations published by state and federal governments. However, the science of contaminant hydrology has advanced beyond the basic requirements of regulations. Where regulatory compliance is the primary driver, collection of data above and beyond the requirements is disincentivized because of additional costs and potential to discover contaminants that require further delineation or remediation. This leads to a data-poor environment for most contaminated sites. However, to use the statistical and machine learning tools now available, more and higher quality data are critically required. This is a paradox highlighted by this work, and a crucial challenge that prevents the applicability of advanced data analysis in the practice areas of contaminant hydrology.

Additional research is necessary to understand better ways to use the existing data from legacy sites. Environmental consulting firms and site owners hold a large amount of data for the many sites that are managed by them. Internal software programs allow this data to be shared within these companies; however, due to liabilities and competitive business strategies, this data is not freely available. Building relationships with these companies to facilitate sharing of data to get better insights, as has been done in this project, is a strategy that has synergistic benefits in both the private and research communities. Data mining using text mining and accessing the publicly available documents from contaminated sites is one way to collect this previously recorded data. Data management and warehousing is an important and often neglected area that will become more important as the amount of data available increases. There may be other questions that are better suited for methods such as dimensionality reduction to answer. A key challenge is asking the correct questions that we can use these innovative tools and the trove of existing data to

answer. However, a move to incorporate these technologies into the field of contaminant hydrology should reinforce rather than replace the vast understanding of the subsurface that has been gained in the field. With the first generation of scientists that worked in this nascent field nearing retirement, transferring that knowledge through education and mentorship is extremely important (Hadley and Newell 2012). John Cherry, a forefather of the field, has endeavored to create a free database, called the Groundwater Project, that will help preserve his legacy and allow learning scientist to access educational material without the barriers that have historically been in place (J. Cherry 2020).

In addition to using the data that exists, new guidelines regarding how much data should be collected are warranted. As such, the Interstate Technology and Regulatory Council (ITRC) has convened a work group to discuss and recommend changes to these guidelines. Whether by a regulatory or research body, these guidelines could help streamline the kinds and amounts of data collected for various sites and thus enable better use of statistical analysis. Research and private consulting have both made great strides in understanding the best practices for characterization and cleanup of contaminated sites; however, federal and state regulatory agencies have lagged in providing up to date, consistent, and clear guidance that would streamline remediation operations.

Microcontroller-connected autosamplers could improve the quality of the samples we collect and reduce the amount of field work required to collect those samples. Sensors that take real time measurements of geochemical conditions are already being used on a large scale (T. Sale et al. 2021). A key challenge for the future will be incorporating machine learning to analysis of this data to address questions such as: Does the seasonal nature of the environment significantly affect the breakdown of contaminants in the subsurface? In what ways does the impact of rising

sea levels influence the geochemical environment in coastal remediation sites? Certainly, the area of climate change impacts to the subsurface will need to be further explored in the coming years. Research not only in the field of contaminant hydrology, but also in computer science, public administration, and social science can aid in creating a better working relationship that would ultimately allow for the highest and best use of contaminated properties. However, communication between practitioners of these fields and relationship building is necessary to truly move the field forward in a synergistic manner. When silos in these areas are broken down such that sites are managed in a collaborative way, the understanding and possibilities for reuse of these contaminated sites will drastically improve.

REFERENCES

Adamson, David T., R. Hunter Anderson, Shaily Mahendra, and Charles J. Newell. 2015. "Evidence of 1,4-Dioxane Attenuation at Groundwater Sites Contaminated with Chlorinated Solvents and 1,4-Dioxane." *Environmental Science and Technology* 49 (11): 6510–18. https://doi.org/10.1021/acs.est.5b00964.

Aguilera, H., C. Guardiola-Albert, N. Naranjo-Fernández, and C. Kohfahl. 2019. "Towards Flexible Groundwater-Level Prediction for Adaptive Water Management: Using Facebook's Prophet Forecasting Approach." *Hydrological Sciences Journal* 64 (12): 1504–18. https://doi.org/10.1080/02626667.2019.1651933.

Alpaydin, Ethem. 2014. *Introduction to Machine Learning*. *The MIT Press*.

Alternatives for Managing the Nation's Complex Contaminated Groundwater Sites. 2013. *Alternatives for Managing the Nation's Complex Contaminated Groundwater Sites*. National Academies Press. https://doi.org/10.17226/14668.

Arnold, William A., and A. Lynn Roberts. 2000. "Pathways and Kinetics of Chlorinated Ethylene and Chlorinated Acetylene Reaction with Fe(0) Particles." *Environmental Science and Technology* 34 (9): 1794–1805. https://doi.org/10.1021/es990884q.

Aziz, Julia J., Meng Ling, Hanadi S. Rifai, Charles J. Newell, and James R. Gonzales. 2003. "MAROS: A Decision Support System for Optimizing Monitoring Plans." *Ground Water* 41 (3): 355–67. https://doi.org/10.1111/j.1745-6584.2003.tb02605.x.

Babbar-Sebens, Meghna, and Barbara Minsker. 2008. "Standard Interactive Genetic Algorithm—Comprehensive Optimization Framework for Groundwater Monitoring

Design." *Journal of Water Resources Planning and Management* 134 (6): 538–47. https://doi.org/10.1061/(asce)0733-9496(2008)134:6(538).

Blotevogel, Jens, Kayvan Karimi Askarani, Andrea Hanson, Sam Gallo, Brian Carling, Carol Mowder, Jim Spain, Andrew Hartten, and Tom Sale. 2021. "Real-Time Remediation Performance Monitoring with ORP Sensors." *Groundwater Monitoring & Remediation* 41 (3): 27–28. https://doi.org/10.1111/GWMR.12479.

Capozzi, Staci L., Lisa A. Rodenburg, Valdis Krumins, Donna E. Fennell, and E. Erin Mack. 2018. "Using Positive Matrix Factorization to Investigate Microbial Dehalogenation of Chlorinated Benzenes in Groundwater at a Historically Contaminated Site." *Chemosphere* 211 (November): 515–23. https://doi.org/10.1016/j.chemosphere.2018.07.180.

Chapman, Steven W., and Beth L. Parker. 2005. "Plume Persistence Due to Aquitard Back Diffusion Following Dense Nonaqueous Phase Liquid Source Removal or Isolation." *Water Resources Research* 41 (12): 1–16. https://doi.org/10.1029/2005WR004224.

Cherry, John. 2020. "The Groundwater Project: Democratizing Groundwater Knowledge." *Groundwater* 58 (5): 682–83. https://doi.org/10.1111/GWAT.13029.

Cherry, John A., and R. Allen Freeze. 1979. *Groundwater*. Prentice-Hall.

Dabrowski, Joel Janek, Ashfaqur Rahman, Daniel Edward Pagendam, and Andrew George. 2020. "Enforcing Mean Reversion in State Space Models for Prawn Pond Water Quality Forecasting." *Computers and Electronics in Agriculture* 168 (January). https://doi.org/10.1016/j.compag.2019.105120.

DERS. 1997. “Groundwater Modeling Report, DuPont Antioch Works, Antioch, California. Project No. 4012-02, March 1997.”

Elder, Carl R., Craig H. Benson, and Gerald R. Eykholt. 2002. “Effects of Heterogeneity on Influent and Effluent Concentrations from Horizontal Permeable Reactive Barriers.” *Water Resources Research* 38 (8): 27-1-27–19. https://doi.org/10.1029/2001wr001259.

Farrell, Dara M., Barbara S. Minsker, David Tcheng, Duanne Searsmith, Jane Bohn, and Dennis Beckman. 2007. “Data Mining to Improve Management and Reduce Costs of Environmental Remediation.” *Journal of Hydroinformatics* 9 (2): 107–21. https://doi.org/10.2166/hydro.2007.004.

Fennell, Donna E., James M. Gossett, and Stephen H. Zinder. 1997. “Comparison of Butyric Acid, Ethanol, Lactic Acid, and Propionic Acid as Hydrogen Donors for the Reductive Dechlorination of Tetrachloroethene.” *Environmental Science and Technology* 31 (3): 918–26. https://doi.org/10.1021/es960756r.

Fernández-Ayuso, Ana, Héctor Aguilera, Carolina Guardiola-Albert, Miguel Rodríguez-Rodríguez, Javier Heredia, and Nuria Naranjo-Fernández. 2019. “Unraveling the Hydrological Behavior of a Coastal Pond in Doñana National Park (Southwest Spain).” *Groundwater* 57 (6): 895–906. https://doi.org/10.1111/gwat.12906.

Fetter, C. W. 1999. *Contaminant Hydrogeology*. Second. Prentice-Hall.

Freyberg, David L. 1986. “A Natural Gradient Experiment on Solute Transport in a Sand Aquifer 2. Spatial Moments and the Advection and Dispersion of Nonreactive Tracers.” *WATER RESOURCES RESEARCH*. Vol. 22.

Friendly, Michael. 2002. “Corrgrams: Exploratory Displays for Correlatigon Matrices.” *American Statistician* 56 (4): 316–24. https://doi.org/10.1198/000313002533.

Gardner, Everette S, and Ed Mckenzie. 1985. “Forecasting Trends in Time Series” 31 (10): 1237–46.

Gleick, Peter H., Environment Pacific Institute for Studies in Development, and Stockholm Environment Institute. 1993. “Water in Crisis : A Guide to the World’s Fresh Water Resources,” 473.

Hadley, Paul W., and Charles J. Newell. 2012. “Groundwater Remediation: The Next 30 Years.” *Groundwater* 50 (5): 669–78. https://doi.org/10.1111/J.1745-6584.2012.00942.X.

Harbaugh, Arlen W. 2005. “MODFLOW-2005 : The U.S. Geological Survey Modular Ground-Water Model--the Ground-Water Flow Process.” *Techniques and Methods*. https://doi.org/10.3133/TM6A16.

Harker, A. R., and Y. Kim. 1990. “Trichloroethylene Degradation by Two Independent Aromatic-Degrading Pathways in Alcaligenes Eutrophus JMP134.” *Applied and Environmental Microbiology* 56 (4): 1179–81. https://doi.org/10.1128/aem.56.4.1179-1181.1990.

Henderson, Andrew D., and Avery H. Demond. 2007. “Long-Term Performance of Zero-Valent Iron Permeable Reactive Barriers: A Critical Review.” *Environmental Engineering Science*. https://doi.org/10.1089/ees.2006.0071.

Hocking, Grant, Melanie A. Thurman, Chad N. Givins, and Chris M. Lacko. 2006. *Installation of a Full-Scale Permeable Reactive Barrier at a Former Manufacturing Site in California*. Battelle Press.

Holt, Charles C. 2004. "Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages." *International Journal of Forecasting* 20 (1): 5–10. https://doi.org/10.1016/J.IJFORECAST.2003.09.015.

Hotelling, H. 1933. "Analysis of a Complex of Statistical Variables into Principal Components." *Journal of Educational Psychology* 24 (6): 417–41. https://doi.org/10.1037/h0071325.

Hyndman, Rob J (Monash University), and George (Monash University) Athanasopoulos. 2021. "Forecasting: Principles and Practice (3rd Ed)." 2021. https://otexts.com/fpp3/index.html.

ITRC. 2013. "Guidance Document Groundwater Statistics and Monitoring Compliance Statistical Tools for the Project Life Cycle Prepared by The Interstate Technology & Regulatory Council Groundwater Statistics and Monitoring Compliance Team."

Jollife, Ian T., and Jorge Cadima. 2016. "Principal Component Analysis: A Review and Recent Developments." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. Royal Society of London. https://doi.org/10.1098/rsta.2015.0202.

Karimi Askarani, Kayvan, and Thomas Clay Sale. 2020. "Thermal Estimation of Natural Source Zone Depletion Rates without Background Correction." *Water Research* 169 (February). https://doi.org/10.1016/j.watres.2019.115245.

Liang, Hailian, Ronald Falta, Charles Newell, Shahla Farhat, P.S. Rao, and Nandita Basu. 2010. "PREMChlor: Probabilistic Remediation Evaluation Model for Chlorinated Solvents."

Lu, Guoping, T. Prabhakar Clement, Chunmiao Zheng, and Todd H. Wiedemeier. 1999. "Natural Attenuation of BTEX Compounds: Model Development and Field-Scale Application." *Ground Water* 37 (5): 707–17. https://doi.org/10.1111/j.1745-6584.1999.tb01163.x.

Matheson, Leah J, and Paul G Tratnyek. 1994. "Reductive Dehalogenation of Chlorinated Methanes by Iron Metal." *Environ. Sel. Technol*. Vol. 28.

McHugh, Thomas E., L M Beckley, C Y Liu, and C J Newell. 2011. "Factors Influencing Variability in Groundwater Monitoring Data Sets." *Ground Water Monitoring and Remediation* 31 (2): 92–101. https://doi.org/10.1111/j.1745-6592.2011.01337.x.

McHugh, Thomas E., Poonam R. Kulkarni, and Charles J. Newell. 2016. "Time vs. Money: A Quantitative Evaluation of Monitoring Frequency vs. Monitoring Duration." *Groundwater* 54 (5): 692–98. https://doi.org/10.1111/gwat.12407.

Mikesell, M. D., and S. A. Boyd. 1990. "Dechlorination of Chloroform by Methanosarcina Strains." *Applied and Environmental Microbiology* 56 (4): 1198–1201. https://doi.org/10.1128/aem.56.4.1198-1201.1990.

Nath, Pritthijit, Pratik Saha, Asif Iqbal Middya, and Sarbani Roy. 2021. "Long-Term Time-Series Pollution Forecast Using Statistical and Deep Learning Methods." *Neural Computing and Applications*. https://doi.org/10.1007/s00521-021-05901-2.

National Research Council. 2013. *Alternatives for Managing the Nation's Complex Contaminated Groundwater Sites*. *Alternatives for Managing the Nation's Complex Contaminated Groundwater Sites*. https://doi.org/10.17226/14668.

Newell, Charles J, Hanadi S. Rifai, John T Wilson, John A Connor, Julia A. Aziz, and Monica P. Suarez. 2002. "Calculation and Use of First-Order Rate Constants for Monitored Natural Attenuation Studies." *Environmental Protection*, 1–28.

Nielsen, David M., ed. 2005. "Practical Handbook of Environmental Site Characterization and Ground-Water Monitoring," September. https://doi.org/10.1201/9781420032246.

Papacharalampous, Georgia A., and Hristos Tyralis. 2018. "Evaluation of Random Forests and Prophet for Daily Streamflow Forecasting." *Advances in Geosciences* 45 (August): 201–8. https://doi.org/10.5194/adgeo-45-201-2018.

Papacharalampous, Georgia, and Hristos Tyralis. 2020. "Hydrological Time Series Forecasting Using Simple Combinations: Big Data Testing and Investigations on One-Year Ahead River Flow Predictability." *Journal of Hydrology* 590 (November). https://doi.org/10.1016/j.jhydrol.2020.125205.

Pickens, John F, and Gerald E Grisak. 1981. "Scale-Dependent Dispersion in a Stratified Granular Aquifer." *WATER RESOURCES RESEARCH*. Vol. 17.

R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Reed, P., B. Minsker, and A. J. Valocchi. 2000. “Cost-Effective Long-Term Groundwater Monitoring Design Using a Genetic Algorithm and Global Mass Interpolation.” *Water Resources Research* 36 (12): 3731–41. https://doi.org/10.1029/2000WR900232.

Reed, Patrick, Barbara S Minsker, and David E Goldberg. 2001. “A Multiobjective Approach to Cost Effective Long-Term Groundwater Monitoring Using an Elitist Nondominated Sorted Genetic Algorithm with Historical Data.”

Sale, Thomas C., Jay M. Ham, William Samuel Gallo, Kayvan Karimi Askarani, Zachary Scott Ferrie, and Joseph Scalia. 2019. Devices and methods for measuring temperature, oxidation reduction potential, and water-level within a subsurface formation, issued October 14, 2019. https://patents.google.com/patent/US10901117B2/en.

Sale, Tom, Sam Gallo, Kayvan Karimi Askarani, Maria Irianni-Renno, Mark Lyverse, Harley Hopkins, Jens Blotevogel, and Scott Burge. 2021. “Real-Time Soil and Groundwater Monitoring via Spatial and Temporal Resolution of Biogeochemical Potentials.” *Journal of Hazardous Materials* 408 (April): 124403. https://doi.org/10.1016/J.JHAZMAT.2020.124403.

Sale, Tom, and Charles Newell. 2011. “A Guide for Selecting Remedies for Subsurface Releases of Chlorinated Solvents.”

Santos, Javier E., Ying Yin, Honggeun Jo, Wen Pan, Qinjun Kang, Hari S. Viswanathan, Maša Prodanović, Michael J. Pyrcz, and Nicholas Lubbers. 2021. “Computationally Efficient Multiscale Neural Networks Applied to Fluid Flow in Complex 3D Porous Media.” *Transport in Porous Media 2021*, May, 1–32. https://doi.org/10.1007/S11242-021-01617-Y.

Scherer, Michelle M., Barbara A. Balko, David A. Gallagher, and Paul G. Tratnyek. 1998. "Correlation Analysis of Rate Constants for Dechlorination by Zero- Valent Iron." *Environmental Science and Technology* 32 (19): 3026–33. https://doi.org/10.1021/es9802551.

Seyedabbasi, Mir Ahmad, Charles J. Newell, David T. Adamson, and Thomas C. Sale. 2012. "Relative Contribution of DNAPL Dissolution and Matrix Diffusion to the Long-Term Persistence of Chlorinated Solvent Source Zones." *Journal of Contaminant Hydrology* 134–135 (June): 69–81. https://doi.org/10.1016/J.JCONHYD.2012.03.010.

Shen, Justin, Davesh Valagolam, and Serena McCalla. 2020. "Prophet Forecasting Model: A Machine Learning Approach to Predict the Concentration of Air Pollutants (PM2.5, PM10, O3, NO2, SO2, CO) in Seoul, South Korea." *PeerJ* 8. https://doi.org/10.7717/peerj.9961.

Shuryak, Igor. 2017. "Advantages of Synthetic Noise and Machine Learning for Analyzing Radioecological Data Sets." *PLOS ONE* 12 (1): e0170007. https://doi.org/10.1371/JOURNAL.PONE.0170007.

Smatlak, Concordia R., James M. Gossett, and Stephen H. Zinder. 1996. "Comparative Kinetics of Hydrogen Utilization for Reductive Dechlorination of Tetrachloroethene and Methanogenesis in an Anaerobic Enrichment Culture." *Environmental Science and Technology* 30 (9): 2850–58. https://doi.org/10.1021/es9602455.

Stephens, Daniel B, Kuo-Chin A Hsu Mark Prieksat, Mark D Ankeny Neil Blandford, Tracy L Roth, James A Kelsey Julia R Whitworth, Kuo-Chin Hsu, Mark A Prieksat Mark D Ankeny, Neil Blandford, Tracy L Roth James A Kelsey Daniel B Stephens, and Julia R

Whitworth. 1998. “A Comparison of Estimated and Calculated Effective Porosity.” *Hydrogeology Journal*. Vol. 6. Springer-Verlag.

Sun, Xiang, Haoqi Wang, Zhigang Guo, Peili Lu, Fuzhong Song, Li Liu, Jiaxin Liu, Neil L. Rose, and Fengwen Wang. 2020. “Positive Matrix Factorization on Source Apportionment for Typical Pollutants in Different Environmental Media: A Review.” *Environmental Science: Processes & Impacts*. https://doi.org/10.1039/c9em00529c.

Suthersan, Suthan, Margaret Gentile, Caitlin Bell, Joseph Quinnan, and John Horst. 2016. “Big Data and Environmental Remediation: Gaining Predictive Insights.” *Groundwater Monitoring and Remediation* 36 (2): 21–31. https://doi.org/10.1111/gwmr.12156.

Taylor, Sean J., and Benjamin Letham. 2018. “Forecasting at Scale.” *American Statistician* 72 (1): 37–45. https://doi.org/10.1080/00031305.2017.1380080.

Valiev, Marat, Eric J Bylaska, Michel Dupuis, and Paul G Tratnyek. 2008. “Combined Quantum Mechanical and Molecular Mechanics Studies of the Electron-Transfer Reactions Involving Carbon Tetrachloride in Solution.” https://doi.org/10.1021/jp7104709.

Vesselinov, Velimir V., Boian S. Alexandrov, and Daniel O’Malley. 2018. “Contaminant Source Identification Using Semi-Supervised Machine Learning.” *Journal of Contaminant Hydrology*. https://doi.org/10.1016/j.jconhyd.2017.11.002.

Vogel, Timothy M, Craig S. Criddle, Perry L. McCarty, and Joseph C. Arcos. 1987. “Transformations of Halogenated Aliphatic Compounds: Oxidation, Reduction, Substitution, and Dehydrohalogenation Reactions Occur Abiotically or in Microbial and Mammalian Systems.” *Environmental Science and Technology* 21 (8): 722–36. https://doi.org/10.1021/es00162a001.

White, Jeremy T. 2017. “Forecast First: An Argument for Groundwater Modeling in Reverse.” *Groundwater* 55 (5): 660–64. https://doi.org/10.1111/GWAT.12558.

Wiedemeier, Todd H., Hanadi S Rifai, Charles J Newell, and John T Wilson. 2007. *Natural Attenuation of Fuels and Chlorinated Solvents in the Subsurface*. *Natural Attenuation of Fuels and Chlorinated Solvents in the Subsurface*. https://doi.org/10.1002/9780470172964.

Wilking, Bart T., Derrick R. Rodriguez, and Tissa H. Illangasekare. 2013. “Experimental Study of the Effects of DNAPL Distribution on Mass Rebound.” *Groundwater* 51 (2): 229–36. https://doi.org/10.1111/J.1745-6584.2012.00962.X.

Zheng, Chunmiao. 2010. “MT3DMS v5.3 Supplemental User’s Guide.”

Zheng, Chunmiao, Marco Bianchi, and Steven M. Gorelick. 2011. “Lessons Learned from 25 Years of Research at the MADE Site.” *Ground Water*. https://doi.org/10.1111/j.1745-6584.2010.00753.x.