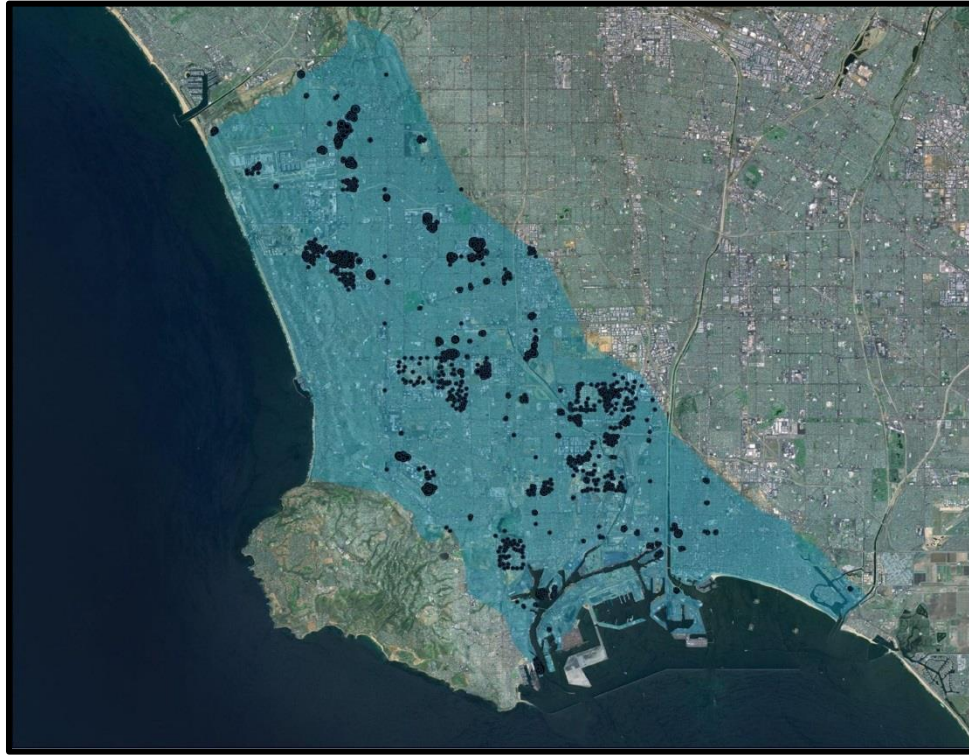


Mass Quantification of PCE in Los Angeles Groundwater from the GeoTracker Database



Client: Los Angeles Regional Water Quality Control Board

Advisor: Travis Brooks

Group Members: Samantha Burke, Grace Chu, Johanna Heyer, Janna Lee, Xina Tang, Tiffany Tran, Lawrence Zhang

**UCLA Institute of the Environment and Sustainability
Senior Practicum
June 2013**

Contents

I. Introduction	3
II. Background	4
Contaminant: Tetrachloroethene.....	4
Case Study Area: Torrance, CA.....	5
GeoTracker GAMA	7
Sampling and Analytical Chemistry	7
Interpolation by Kriging: Conceptual and Mathematical Foundations.....	8
III. Methodology	10
Development of Conceptual Model	10
Limiting and Formatting Data	12
Kriging Concentration	15
Integration in ArcGIS	23
IV. Results.....	27
V. Discussion.....	30
Sources of Uncertainty.....	30
Recommendations and Considerations for Large-Scale Implementation.....	31
Works Cited	31
Appendix: Description of steps taken in ArcGIS	33

I. Introduction

In an average year, groundwater meets nearly half of California's water demand and provides water to 80 percent of the state's residents for drinking, domestic use, and irrigation (SWRCB 2012). In southern California, groundwater provides 40 percent of the region's freshwater supply (WRDSC 2011). Increased statewide monitoring efforts beginning in the 1950s have contributed to a growing awareness of the contamination of this underground resource with an ever-developing catalogue of pollutants. In Los Angeles, the aerospace industry and manufacturing boom of the 1950s contributed large quantities of chlorinated solvents into local aquifers. This class of synthetic contaminants has been linked to health effects ranging from organ damage to cancer (Moran *et al.* 2007).

Currently, groundwater is regulated by measuring the concentration of contaminants against health and environmental standards. This approach allows regulatory agencies to detect exceedances, but it is limited in its capacity to predict how much contaminant input will cause an exceedance. Surface water is inherently easier to monitor and regulate because it is more accessible and because the volume of surface water bodies is more easily measured. Section 303 (d) of the Clean Water Act introduced the regulatory approach of defining a Total Maximum Daily Load (TMDL) for surface water bodies. This approach involves defining the total mass flow of a contaminant that a given waterbody can accommodate without exceeding a specific concentration standard. The Los Angeles Regional Water Quality Control Board (LARWQCB) is interested in the possibility of applying a similar mass-based and waterbody-specific regulatory approach to groundwater. This would require a method of quantifying the total mass of a contaminant currently present in a given aquifer or groundwater basin from monitoring well concentration data. This report details a set of approaches and methods explored for determining the mass of a contaminant in a waterbody from concentration measurements at sampling wells.

The State Water Resources Control Board (SWRCB) is responsible for protecting the quality of both surface and groundwater in California. The LARWQCB is charged with advancing these goals with respect to local conditions in Los Angeles. The state's current groundwater monitoring program, Groundwater Ambient Monitoring and Assessment (GAMA), was implemented in 2000. The Groundwater Quality Monitoring Act of 2001 (California Assembly Bill 599) updated the program and was passed with two primary goals: increased groundwater quality monitoring, and increased availability of monitoring data to the public. The public access component of GAMA is the GeoTracker database, which makes statewide groundwater quality data information available to the public as downloadable spreadsheets, reports, and an interactive map. Our client asked us to develop a method for quantifying the mass of a contaminant in a groundwater basin using the GeoTracker database. The initial study area was the West Coast Basin, a sub-basin of the Los Angeles Coastal Plains Basin, but the methodology was developed on a small case study in Torrance and then scaled up in a simplified form.

II. Background

Contaminant: Tetrachloroethene

LARWQCB initially proposed the quantification of all volatile organic compounds (VOCs) present in the study area. Because of the focus on methodology, this set of compounds was scoped to just one VOC present in Los Angeles groundwater, tetrachloroethene (PCE).¹ PCE is a chlorinated solvent and the parent compound of three degradation products: trichloroethylene (TCE), dichloroethylene (DCE), and vinyl chloride (ATSDR 1997). Although TCE and vinyl chloride are both associated with greater health risks than PCE, we decided to consider the parent compound rather than the degradation products as it is a precursor for other compounds.

PCE is an effective man-made chemical most commonly used in dry cleaners and metal degreasing. Eighty percent of the dry cleaning industry relies on PCE because it is safe to use on fabrics while it effectively removes fats, oils and greases (USEPA, 1994). As an effective degreaser, many commercial and industrial companies also rely on PCE to degrease metal parts because it is noncorrosive. In the past, environmental laws were not as strict as they are today and dry cleaners, commercial and industrial companies were allowed to discard waste by pouring wastewater down sewers, into trashcans, and directly on the ground. The discarded waste eventually enters the water table and contaminates the groundwater.

PCE's high density causes it to sink when it infiltrates an aquifer. It can remain there for decades due to its low degradation rate in groundwater and its inability to evaporate (ATSDR 1997). At high concentrations following a large spill, PCE is generally a dense non-aqueous phase liquid; however, for this report it is assumed to be in the dissolved phase because there are no new PCE plumes in the study area and concentrations are low enough that non-aqueous phase is unlikely.

The maximum contaminant level (MCL) for PCE in drinking water is 0.005 mg/L or 5 ppb to protect human health while considering the feasibility in cost and ability for public water systems to remove and treat PCE (USEPA 2012). Exposure to water with PCE below the MCL has not been shown to cause acute health effects. However exposure to PCE concentrations that exceed the MCL over many years have been linked to liver and kidney damage (ATSDR 1997). Human exposure to PCE also occurs by inhalation. Since PCE is volatile, PCE migrates from contaminated groundwater through the soil and into buildings. Health effects associated with the inhalation of low concentrations of PCE includes dizziness and headaches while inhalation of high concentrations of PCE depresses the central nervous system (ATSDR 1997).

¹ Tetrachloroethene is the International Union of Pure and Applied Chemistry name for the compound, but it is also known as tetrachloroethylene and perchloroethylene, from which the acronym PCE is derived.

Case Study Area: Torrance, CA

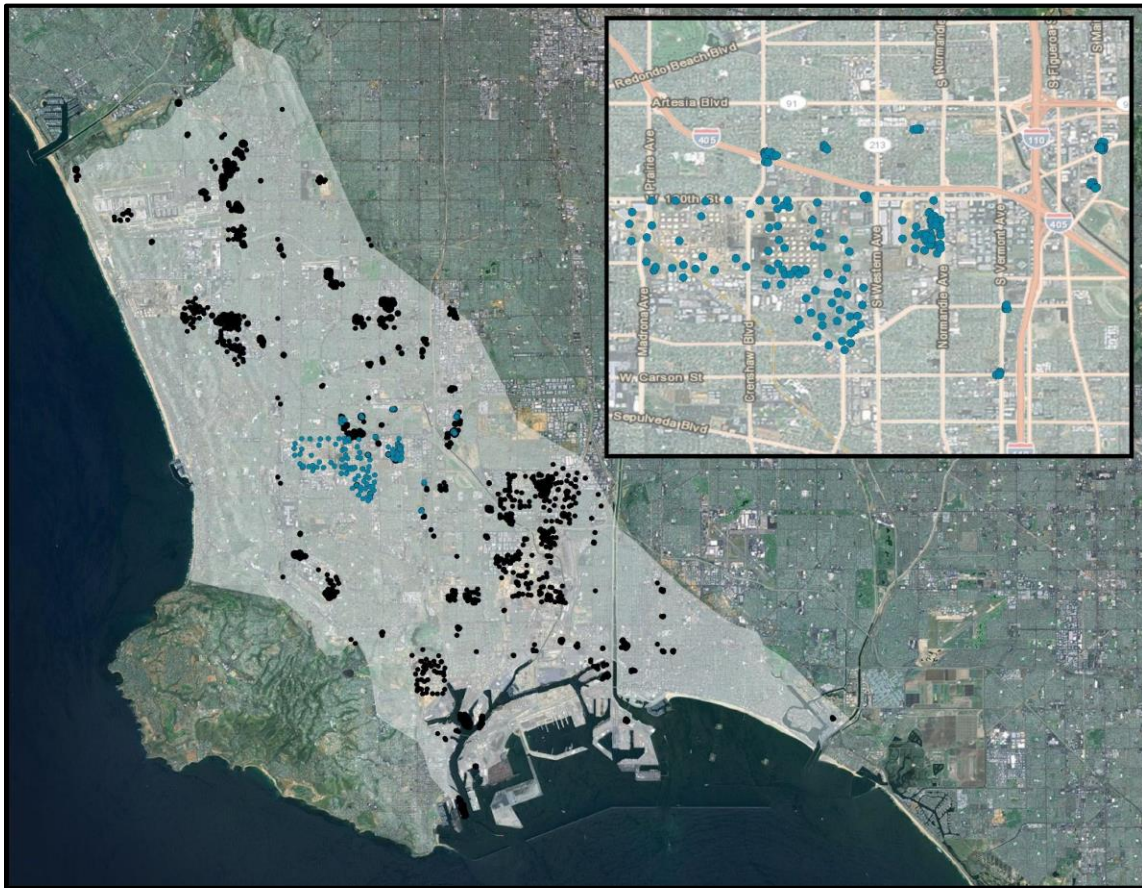


Figure 1 Monitoring well network across the West Coast Basin. Case study area in Torrance, CA is highlighted in blue and shown in inset.

The client initially proposed quantifying mass for the entire Los Angeles Coastal Plains basin. This was scoped first to the West Coast Basin, a sub-basin of the Coastal Plains, and then to a case study area within the West Coast basin to allow for exploration of methodology with a smaller dataset (Figure 1). Narrowing the project to a case study enabled more detailed considerations of factors beyond the concentration data, such as local hydrology and contamination history. The case study area covers several city blocks in Torrance, CA. It contains ten contamination sites, including an aerospace company, gas stations, a dry cleaner, and an oil refinery (Table 1).

The city of Torrance has an abundance of private dry-cleaners, a large active oil refinery, and an industrial area. It was an important center of the aerospace industry beginning in the 1950s and many of the old and extant aerospace buildings are likely to contain masses of machinery where PCE is used extensively. Torrance is home to two superfund sites, the Del Amo Hazardous Waste Site and Montrose Chemical Corporation.

There are four shallow aquifers in Torrance: Gage, Gardena, Lynwood, and Silverado. The Gardena and Gage aquifers lie adjacent to each other at depths of around 200 feet. The Gage aquifer has a largely sand and gravel composition (Reichard *et al.* 2003). The Gardena aquifer consists of coarse deposits.

The Lynwood and Silverado Aquifers belong to the upper San Pedro aquifer system, and are at depths of approximately 400 feet. The Lynwood aquifer lies directly above the Silverado aquifer, and both contain sand and gravel similar to the Gage aquifer, but with marine deposits such as shell and wood fragments embedded as well as the common silt and clay embeddings that often come with sand and gravel (Kirk 2006).

Table 1 Global IDs, names, and addresses of the contamination sites in the Torrance case study area.

Global ID	Site Name	Address
SL0603776467	Boeing C-6 Facility	19503 S. Normandie Ave. Los Angeles
SL184361419	Bee Chemical Co. (Former)	1500 178th St W. Gardena
SL372452438	Mobil - Torrance Refinery	3700 W. 190th Street. Torrance, CA
SLT43716714	Bushee Cleaners	2131 W. 182 ND Street. Torrance
T0603701494	Mobil 18-D9E	18200 Crenshaw Blvd. Torrance
T0603703643	Chevron #9-3874 (Former)	111 Victoria Street. Carson
T0603704674	Mobil #18-MAP	20802 Vermont Ave S. Torrance
T0603719669	Nissan North America, Inc.	125 Griffith. Carson
T0603759521	Shell	911 Carson Street. W. Torrance
T0603792935	Unocal #6075	1875 190th Street W. Torrance

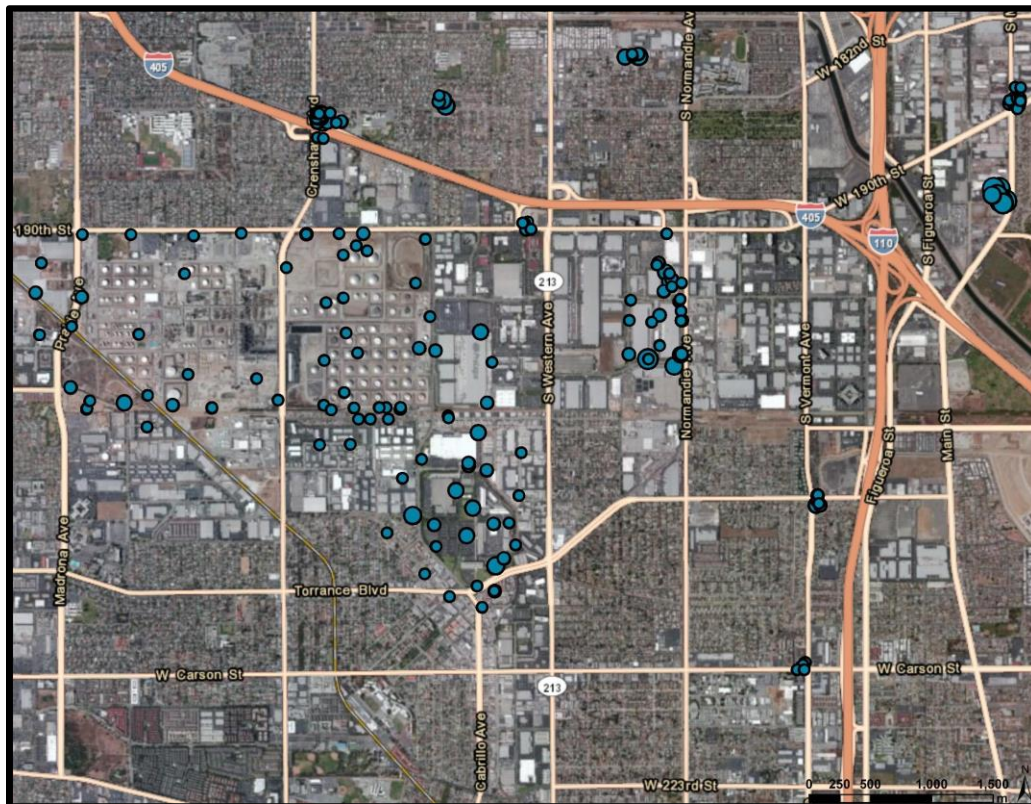


Figure 2 Wells in case study area. Wells sized according to measured PCE concentration.

GeoTracker GAMA

The GeoTracker database was designed to make statewide groundwater quality data accessible to the public. It coordinates data from different types of wells, such as drinking and monitoring wells, and different agencies across California. It has three major components: an online interactive mapping feature, downloadable water quality data with geographic coordinates, and downloadable reports from monitoring sites. This project relied on both the water quality data and the monitoring site reports. The interactive map feature was used to identify a suitable case study area, but it was not used to quantify mass.

Water quality data can be limited by water basin, contaminant, and year of collection and then downloaded as an excel spreadsheet. The options for years of collection are past one, three, and ten years. Each well in the dataset is identified by geographic coordinates as well as a “Global ID” or “Well ID” (both names are used for the same ID in different contexts) that is specific to the monitoring site and a “Well Name” that is specific to the well within the site. The other columns in the dataset that were used for this project are “Date,” which indicates the date the sample was collected, and “Result,” which gives the detected concentration of the selected contaminant in the sample. The “Qualifier” column was also useful in identifying samples in which no concentration was detected. It shows “=” for detected concentrations, and “ND” for Non-Detect results where there was no contaminant detected.

The interactive map reflects the data limiting (basin, contaminant, years), and displays monitoring locations as “clusters” of monitoring sites. Clicking on a cluster displays all the wells within the cluster. This cluster dataset can be downloaded to Excel, but the only concentration results included are the maximum detected concentration for each well. Each well has “Native Data” which leads to more detailed monitoring data for that specific well, which can be downloaded to Excel, and “View Graph,” which displays a plot of concentration over time for the selected contaminant. Clicking directly on a well name leads to the Analytical Results for that well, including all contaminants analyzed.

Monitoring site reports are organized into seven tabs: Summary, Cleanup Action Report, Regulatory Activities, Environmental Data (ESI), Site Maps/Documents, Community Involvement, and Related Cases. The tabs used for this project are Environmental Data (ESI) and Site Maps/Documents. The ESI tab contains Individual Well Analytical Data, which includes the length of well screens, through which water enters the well, and depth to water data. Laboratory Analytical Reports (EDF) are also included under the ESI tab. This tab also contains individual monitoring well reports, which include the method detection limit, which was used to assign concentration values to Non-Detect results.

Sampling and Analytical Chemistry

Because GAMA coordinates data from different agencies across the state with different monitoring goals and different laboratory equipment and protocol, the uncertainty inherent in analytical chemistry is an important consideration in interpreting GeoTracker data. Samples are taken periodically from well locations by an agency and then analyzed in a laboratory separate from the collecting agency within several days of

the sample collection date. All the wells in a given contamination site are sampled by the same agency and analyzed at the same laboratory, unless an agency switches to a new laboratory for analysis. Adjacent contamination sites, though, are often monitored by different agencies and samples analyzed in the same laboratory may be analyzed with different equipment or method.

The goal of analytical chemistry is to identify unknown compounds in the sample and quantify the concentration of the compounds in the sample (Rong 2002). VOCs are separated, identified and quantified through a gas chromatograph and detector. Gas chromatography separates different types of volatile compounds to their molecular level based on their retention time, which is the time it takes for the compound to be detected after it has been injected. On the other hand, detectors including flame ionization detectors, photo ionization detectors and mass spectrometers are used to identify and measure the quantity of the compound in the sample (Rong 2002). The certainty and the meaning of the results are particular to both the instrument and the method used to analyze each sample. In a contaminated area, the concentration of a contaminant is very rarely absolute zero, but many samples return a result of non-detect. A non-detect result does not mean that the contaminant being analyzed is absent, but rather that its concentration is below the Method Detection Limit (MDL). In this project, an understanding of the uncertainties of analytical chemistry and MDLs were important in assigning concentration values to samples in which no PCE was detected.

Interpolation by Kriging: Conceptual and Mathematical Foundations

Groundwater quality is monitored in terms of the concentration at discrete measuring points. Interpolation between these points yields a continuous surface of concentration predictions over an area. Within Geographic Information System (GIS) software, there are several available interpolation models; these fall into the two broad categories of deterministic and geostatistical interpolation (ESRI 2008). Inverse Distance Weighting (IDW) is a deterministic interpolation method resting on the principle that points near one another are similar. Each prediction point uses a weighted average of the points around it. The weight for each known point is based on the inverse of its distance from the prediction point, meaning that known points influence a prediction more strongly when they are closer to the prediction point (ESRI 2008).

Geostatistical interpolation is similar to the deterministic IDW method in that it assumes points close to one another are similar, but it uses geostatistics to determine the weighting of measured points in predictions. The family of geostatistical interpolation methods is also called kriging and was developed for ore reserve estimation in the mining industry, although it now has many other applications. Kriging produces an uncertainty surface in addition to a prediction surface by incorporating geostatistics and considering the entire spatial distribution around prediction points (ESRI 2008; Cameron and Hunter 2002). Both kriging and IDW have been used extensively in modeling groundwater parameters (Rejith *et al.* 2009; Machiwal *et al.* 2011; Nas & Berkay, 2010; Jang *et al.*, 2012).

Within the family of kriging there are several different models including simple kriging, ordinary kriging, cokriging, universal, and multivariate indicator kriging (MVIK). The choice between these models depends on the dataset and the context of the project. MVIK is often useful in modeling water quality data because it predicts the

probability of exceeding a certain value; this prediction can be used to predict the likelihood of exceeding regulatory or health-based standards for water quality. For this project, though, MVIK was inappropriate because mass prediction requires an estimate of the concentration across a surface, not an estimate of how many points exceed concentration-based standards. Cokriging enables the simultaneous interpolation of multiple parameters. This technique was considered but not fully explored for this project. Ordinary kriging was selected as the most appropriate technique based on the data and the intended output.

Ordinary kriging is the most flexible type of kriging because, unlike simple kriging, it does not assume that there is a known mean to the dataset. Ordinary kriging does assume a *fixed* mean for the entire dataset, but it does not assume that this mean is known (ESRI 2013). Ordinary kriging relies on the mathematical concepts of variance and covariance. Variance and covariance are inverses of each other and are therefore practically equivalent; this project used only variance. The mathematical model used in ordinary kriging to assign weights to measured points in predictions is the variogram: a plot of variance as a function of the separation distance between points (Figure 3). All possible pairs of points in the dataset are grouped into intervals (called lags) by the separation distance between the two points in the pair. The variance between the two points is the dependent variable. The lag size and the number of lags in the variogram depend on how regularly and how densely spaced the measured points are. If the points are distributed on a grid, then pairs of points can be grouped into narrow lags with a specific separation distance. However, if points are unevenly distributed, each separation distance may be unique and pairs of points need to be grouped into intervals of separation distances. If points are very densely arranged, the model can have more lags because there are many unique separation distances. However, if the points are sparse, they must be grouped into fewer lags.

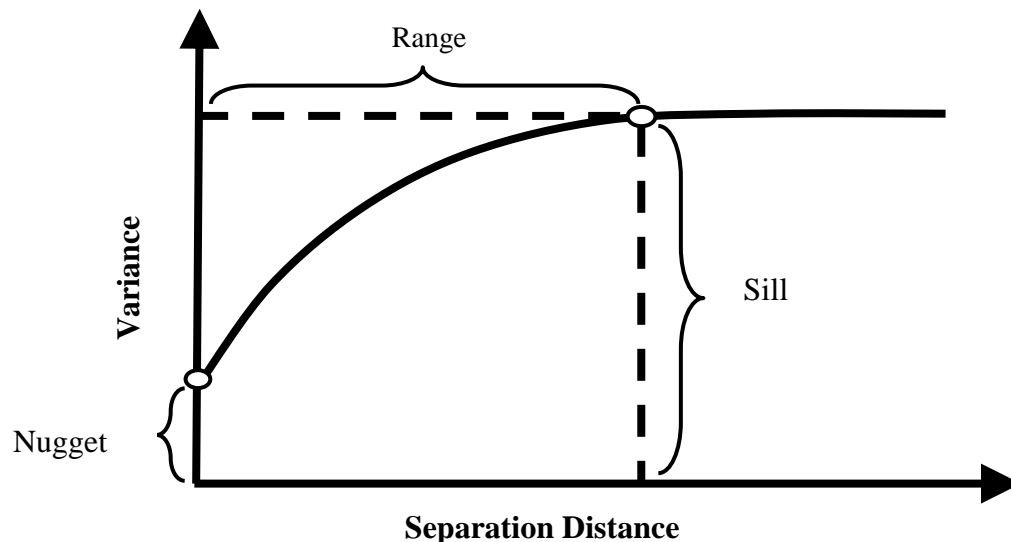


Figure 3 Model variogram showing the key parameters: range, sill, and nugget.

By convention, the variogram is often replaced by the semivariogram, which is simply the variogram with all the variance values divided in half. The key parameters of the variogram are the partial sill, the major range, and the nugget (Figure 3). The sill is the maximum variance between pairs of points in the dataset; visually it is the value at

which the variogram levels off. Beyond a certain distance variance ceases to increase. The range is the separation distance at which the sill is reached. The nugget is the variance at very close or coincident points; graphically it is the y intercept.

III. Methodology

Development of Conceptual Model

Initially the conceptual model focused heavily on factors affecting the distribution and volume of water of the shallow aquifer. These factors included precipitation, pump and recharge rates, local aquifer structure, and contamination and cleanup history (Figure 4). Some of this data was difficult to find at the local scale and there was not time to incorporate these factors into the final conceptual model. Furthermore, the intent was to create a model reliant only on data available through the GeoTracker database.

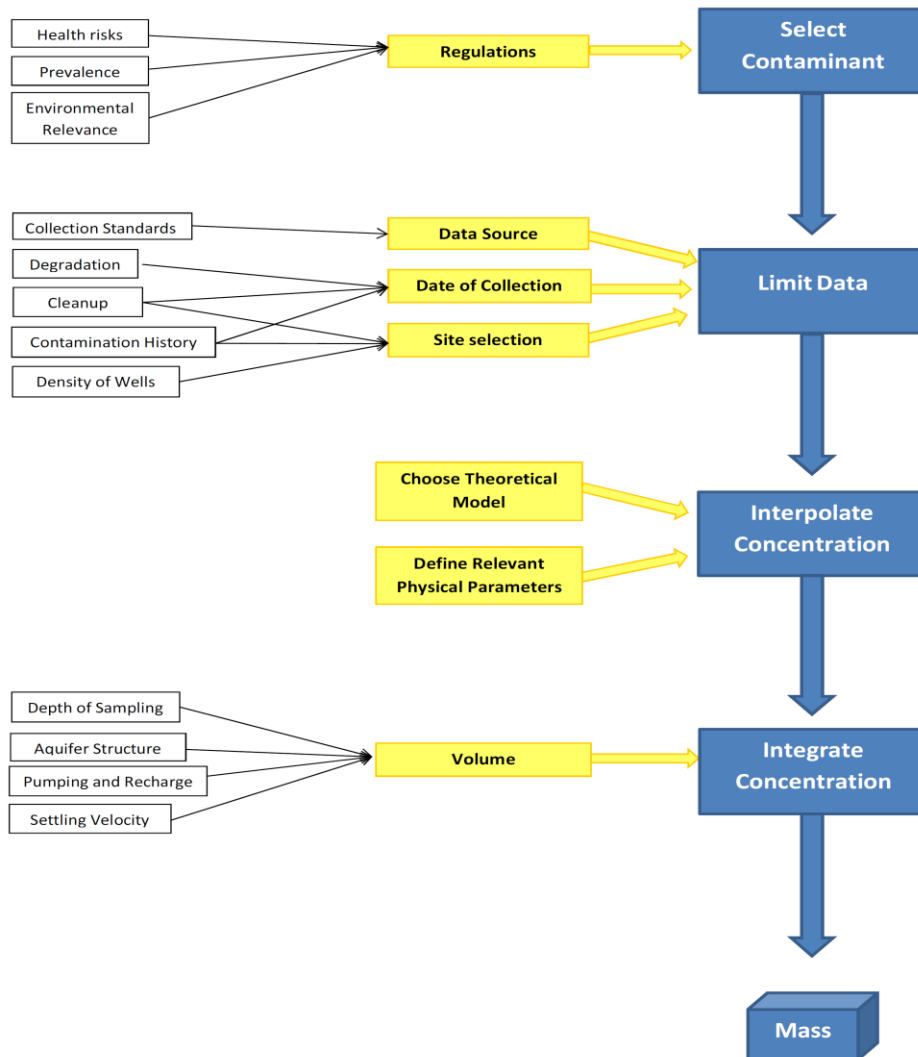


Figure 4 An early version of the conceptual model for this project. Several of the parameters shown here were determined to be irrelevant or beyond the scope of the project.

The final methodology for the case study area is summarized in Figure 5.

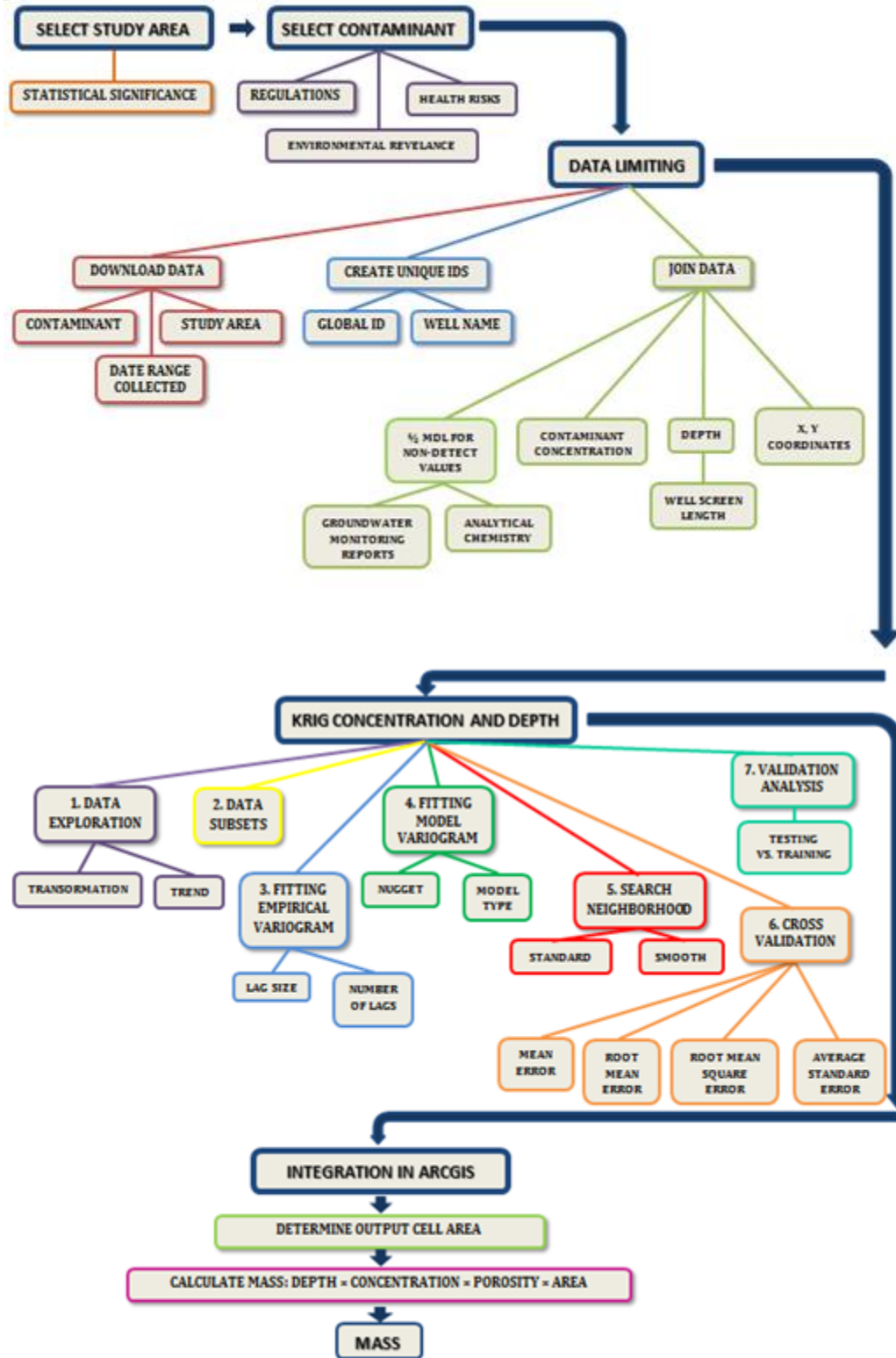


Figure 5 Model of specific methodology for this project, including case study.

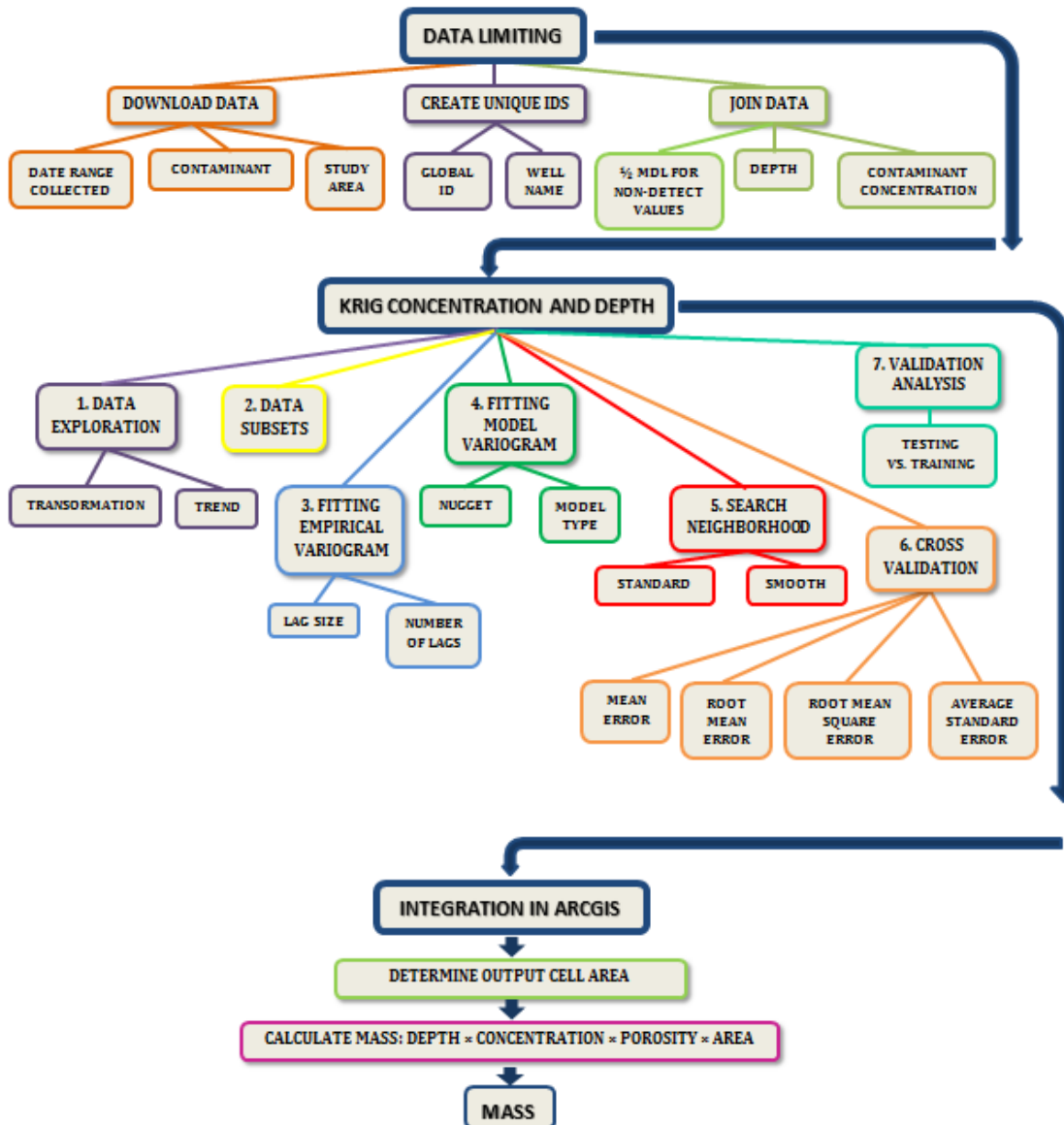


Figure 6 Conceptual Model for kriging mass using data from GeoTracker database.

The simplified conceptual model shown in Figure 6 outlines the process of obtaining a mass estimate for a contaminant across a groundwater basin in GeoTracker. This model assumes that the contaminant of interest is already known, the scale of the desired prediction is an entire groundwater basin, and the only data source is the GeoTracker database. The steps represented in the model are elaborated in the following section.

Limiting and Formatting Data

The GeoTracker database limits data according to the following categories: (1) well type, including Environmental Monitoring Wells and Supply Wells; (2) geographic extent, defined by Counties, Groundwater Basins, Regional Board Boundaries, Assembly Districts, Senate Districts, Hydrogeologic Vulnerable Areas, and GAMA Study Areas; (3) contaminant, grouped by *Wells With Results*, *All Wells That Have Been Sampled*, or

Wells With Results Above Comparison Concentration; and (3) date range. This project considered data from all available well types. The geographic extent was limited by the category of “Groundwater Basin” to “COASTAL PLAIN OF LOS ANGELES – WEST COAST.” In terms of contaminant, the dataset was limited to wells with any result for PCE (the option in the dropdown menu in GeoTracker is “Wells With Results” as opposed to) groups data into categories of the past one, three, and ten years. Originally, we proposed to limit the data to the most recent year available, 2012, and average all measurements during this year. The date range was expanded to the past three years because some wells had sparse or missing PCE data for 2012 and excluding earlier years would have limited the number of data points available. Furthermore, uncertainties in analytical chemistry can cause anomalous results that may be corrected by averaging (Figure 7). PCE has a half-life of about one to two years in groundwater or longer under certain conditions, so in the absence of remediation efforts concentration does not fall rapidly (SWRCB 2009). Still, the assumption of no degradation is not valid. The ten-year average was rejected because remediation efforts in combination with natural degradation have significantly reduced the concentration at some contamination sites and using a ten-year average would therefore overestimate current contamination levels (Figure 8). Using a three year average is not ideal, but it limits the uncertainty and the data lost by considering only one year of measurements as well as avoiding some of the overestimation resulting from assuming a constant concentration over ten years.

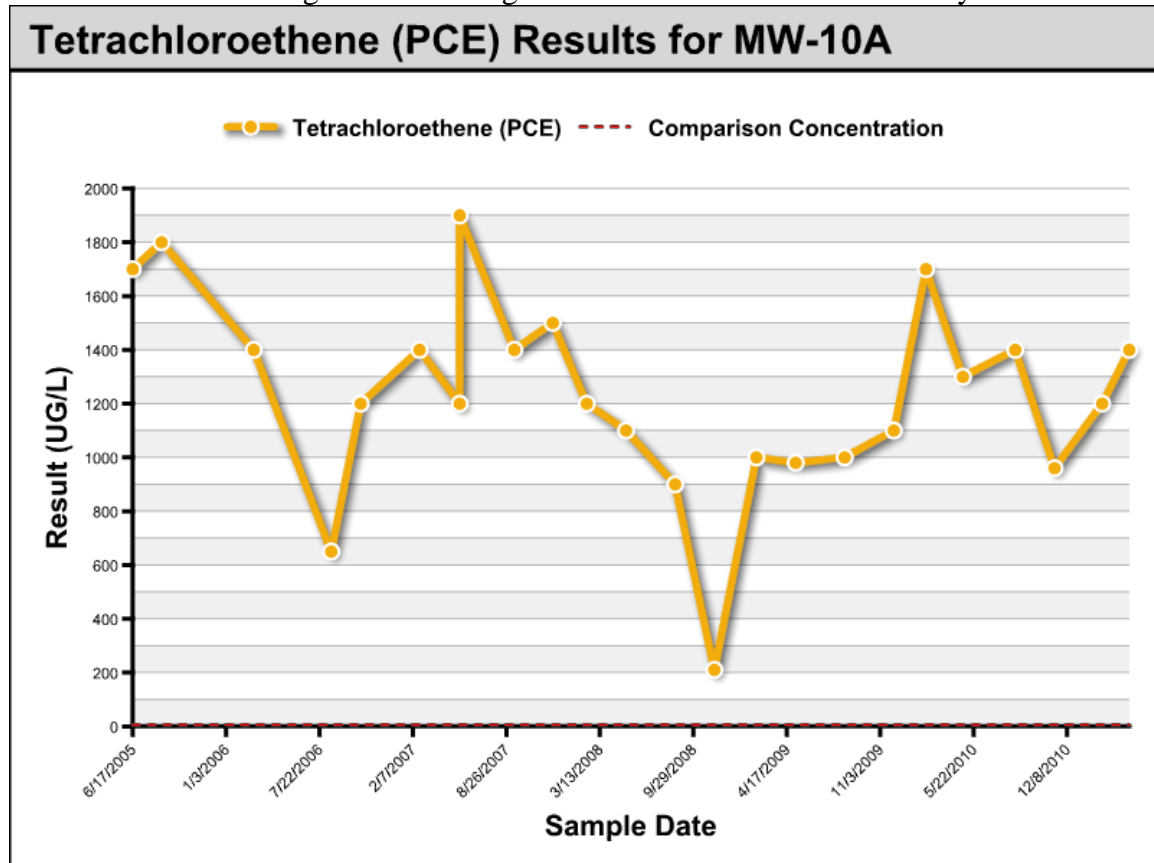


Figure 7 Measured PCE concentrations in MW-10A of the Honeywell Aerospace contamination site from 2005 to 2010, showing anomalous spikes and dips in concentration measurements.

After limiting and downloading concentration data through the GeoTracker interface, the dataset was formatted further according to the specific goals of this project. As discussed in the background section, a consideration of analytical chemistry techniques was essential to the interpretation of non-detect results. Each instrument and analytical method has a different Method Detection Limit (MDL), described as “the minimum result which can be reliably discriminated from a blank (for example, with a 99% confidence level) (Federal Advisory Committee, 2007).” Out of the many detection limits included in monitoring reports, such as Reporting Limit and Practical Quantitation Limit, MDL is the relevant measure because it describes the limits of the analytical instrument and method.

In the Torrance case study area and in the West Coast Basin, a majority of concentration values were reported as non-detect. Several treatments of non-detect values were considered. The option of discarding all non-detect values in the dataset was rejected because it would have left too few points in the dataset. The GeoTracker database assigns a zero value to non-detect measurements. This assumption would likely lead to an underestimation of the total amount of PCE in the groundwater because in an area contaminated with PCE, it is unlikely that a non-detect result actually represents a concentration of zero. It is more likely that the concentration is simply too low to be detected by the instrument and method used in analysis. As a compromise, non-detect measurements were assigned a concentration equal to half the MDL. This approach is based on the EPA’s non-detection policy elaborated for Region 3 (Mid-Atlantic) groundwater quality risk assessment (USEPA 1991). The MDLs in the Torrance case study area ranged from 0.14 µg/L to 130 µg/L; the large range means that using half the MDL is not ideal. For instance, a non-detect measurement with an MDL of 130 µg/L would be assigned a concentration result of 65 µg/L, which is thirteen times higher than the Maximum Contaminant Level and also higher than the concentrations of many of the samples with detected concentrations.

MDLs are not currently part of the downloadable data available through GeoTracker. The MDL for each measurement was found in quarterly monitoring reports, in the ESI tab of site reports. A final possibility for treating non-detect values, which circumvents the inconvenience of manually inputting MDLs, is to assign a uniform value to all non-detect measurements. For the West Coast Basin it was impractical to look up the MDL for each measurement, so a result equal to half of the lowest MDL in the dataset was assigned to all non-detect results.

Another piece of data that was manually obtained and joined to the downloaded dataset was well screen length. As will be discussed later in the methodology, well screen length was used in the case study area as a surrogate for the depth across which PCE is distributed. Well screen length data is available in the ESI tab, but the data for many wells have not been entered and had to be determined from well bore diagrams available in the Site Maps/Documents tab.

Joining data for wells from different datasets requires some unique identifier for each well. The combination of latitude and longitude uniquely identifies a monitoring location, but it is inconveniently long when put into one column together. Wells are also assigned IDs within GeoTracker, but neither of the two types of ID uniquely identifies a well. The first, called alternately a “Global ID” or a “Well ID” depending on the context, identifies the contamination site. Within contamination sites, individual wells are

distinguished by the “Well Name.” Well Names are not unique to wells either, because there may be wells with the same Well Name in different contamination site. A new field was added to the dataset called “UNIQID,” which is a concatenation of the Global ID and the Well Name.

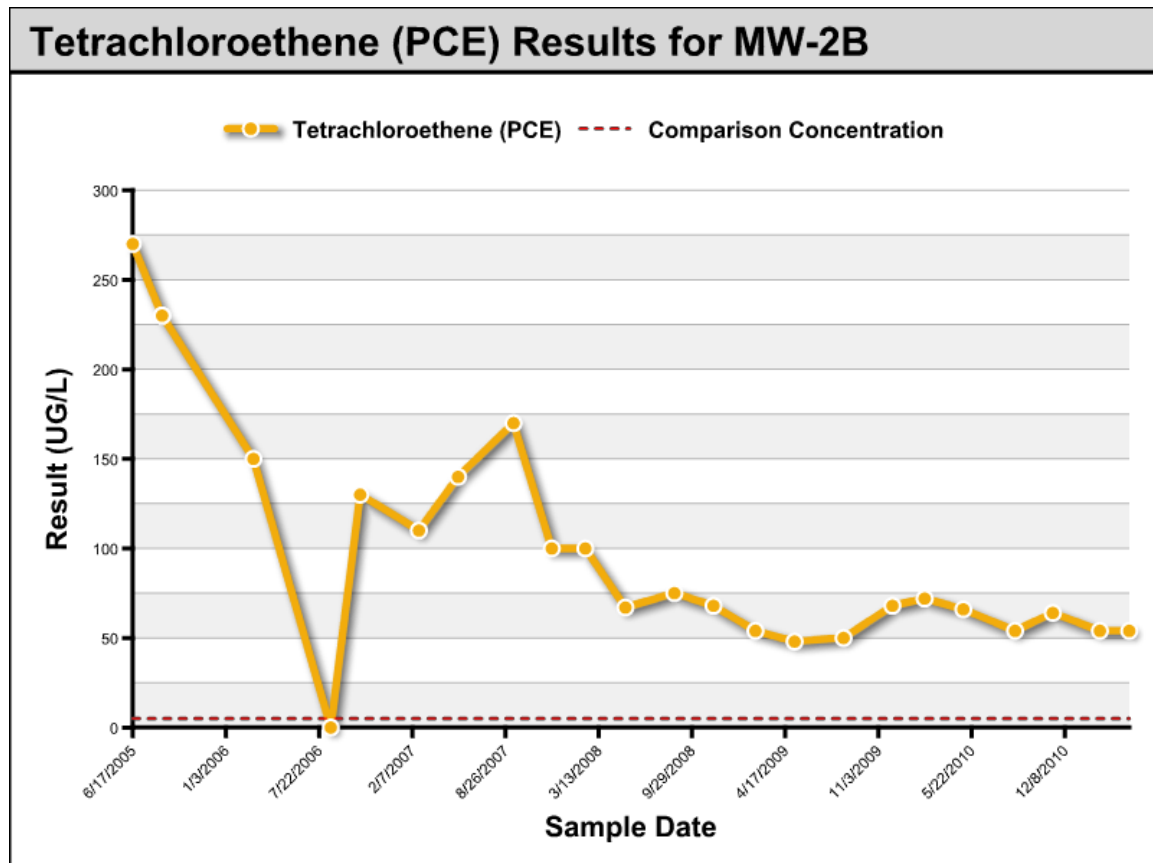


Figure 8 Measured PCE concentrations in MW-2B of the Honeywell Aerospace contamination site from 2005 to 2010, showing the natural degradation of PCE over time.

Kriging Concentration

Kriging was ultimately performed in ArcGIS, but it was also explored in R. Ordinary kriging was used because there is not a known mean to the data. The major steps in ordinary kriging are: (1) exploring the data for transformations and trend, (2) constructing the empirical Semivariogram, (3) fitting the model semivariogram, and (4) defining the search neighborhood. The kriging model is checked at two stages: cross validation assesses the fit of the data to the model semivariogram. Validation analysis compares values predicted by kriging to measured values. The kriging methodology in R was explored based on a series of guides and tutorials created by Nicholas Christou and available at http://www.stat.ucla.edu/~nchristo/statistics_c173_c273/. This methodology was only taken to the step of fitting the model semivariogram because the techniques elaborated through Christou’s tutorials only allow a global search neighborhood, which was not suitable for the dataset. Data exploration in R informed the data exploration that was part of the kriging process within ArcGIS.

1. Exploring Data for Transformations and Trend

The first step in kriging is to explore the dataset to determine any data transformations and trend removal that should be applied. Since kriging works best for normally distributed data, data transformations can improve the fit of the model. The transformations available in ArcGIS are box-cox, arcsine, and log. The concentration data in both the case study area and the West Coast Basin are highly skew toward low concentrations (Figure 9). A log transformation yields a distribution closer to normal, but it is still skew. This suggests that a robust estimator will be more effective than a classical because the robust estimator is better able to accommodate extreme values. Estimator type, which is a parameter in fitting a model semivariogram, is not part of the ArcGIS interface, but it is in R.

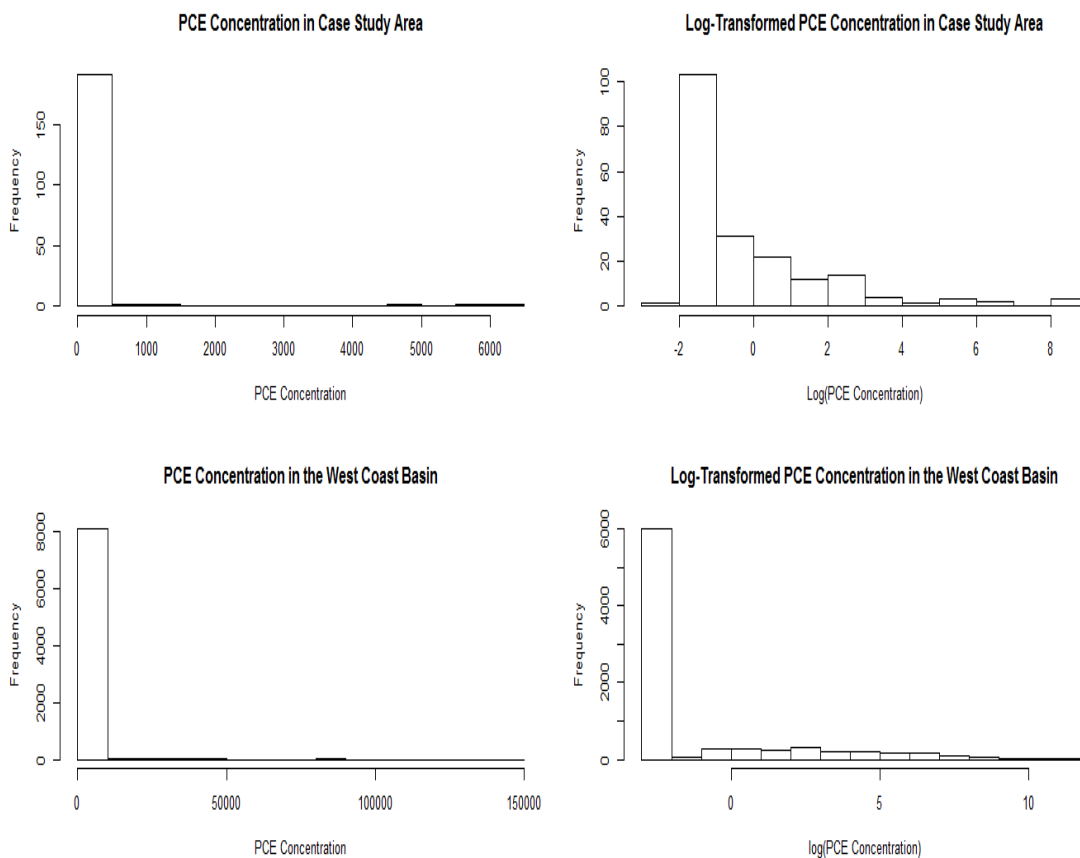


Figure 9 (Clockwise from upper left) PCE concentration in the case study area; log-transformed PCE concentration in the case study area; PCE Concentration in the West Coast Basin; log-transformed PCE Concentration in the West Coast Basin.

Another aspect of the data to explore is the presence of directional trend. Trend analysis is possible in R and ArcGIS and was performed in both. Trend should not be removed unless there is a real-world basis to explain why the data has a certain pattern along a direction. For instance, if there were a strong directional flow of groundwater across the dataset, this would be a reasonable justification for trend removal. There is, however, no strong directional flow across the West Coast Basin, or even across most

individual contamination sites. Some site reports in GeoTracker included rose diagrams for groundwater flow direction, and none showed a strong and constant flow in one direction.

The Trend Analysis tool in ArcGIS can help determine if a trend exists in the data. The shape of the green and blue curve through the projected plane indicates if a trend exists. A flat curve means that no trend exists while a U-shaped curve indicates that there is a second order trend. If a trend exists, one can decide to accept the trend or remove the trend during kriging.

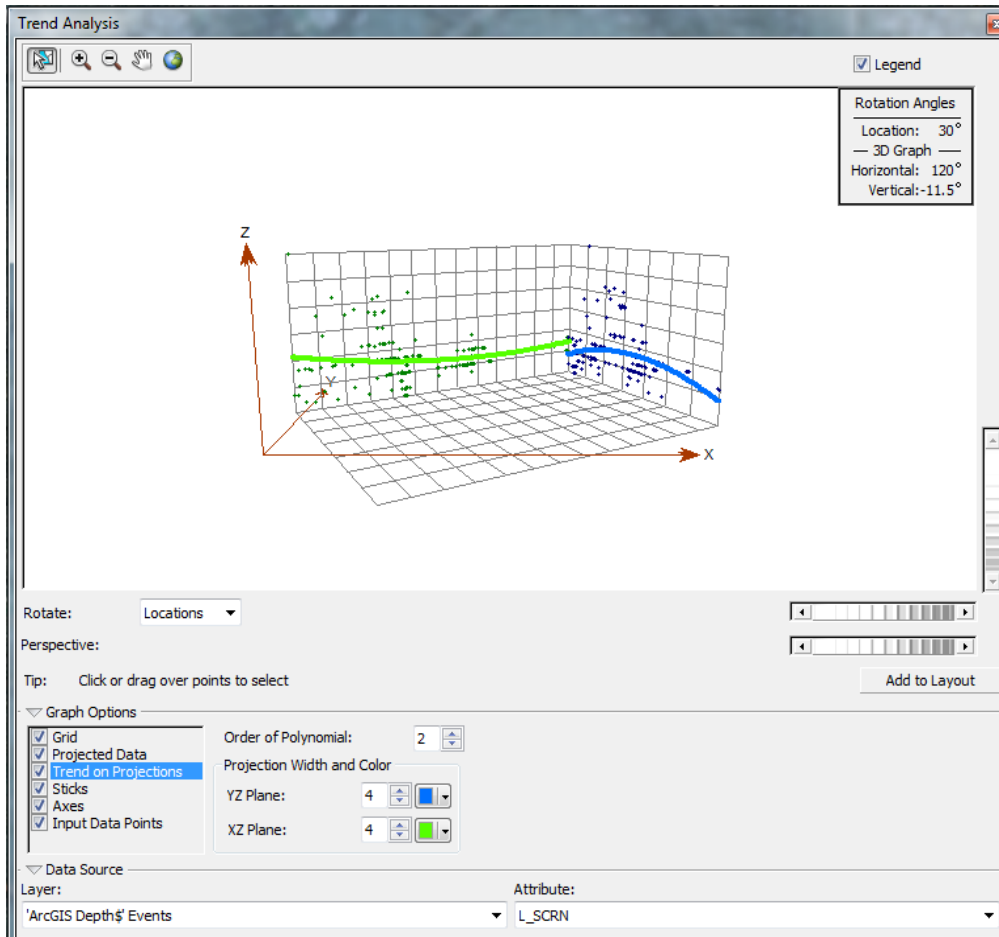


Figure 10 Trend analysis tool in ArcGIS, showing the trend in well screen length data.

2. Constructing the Empirical Semivariogram: Lags, Directionality, and Outliers

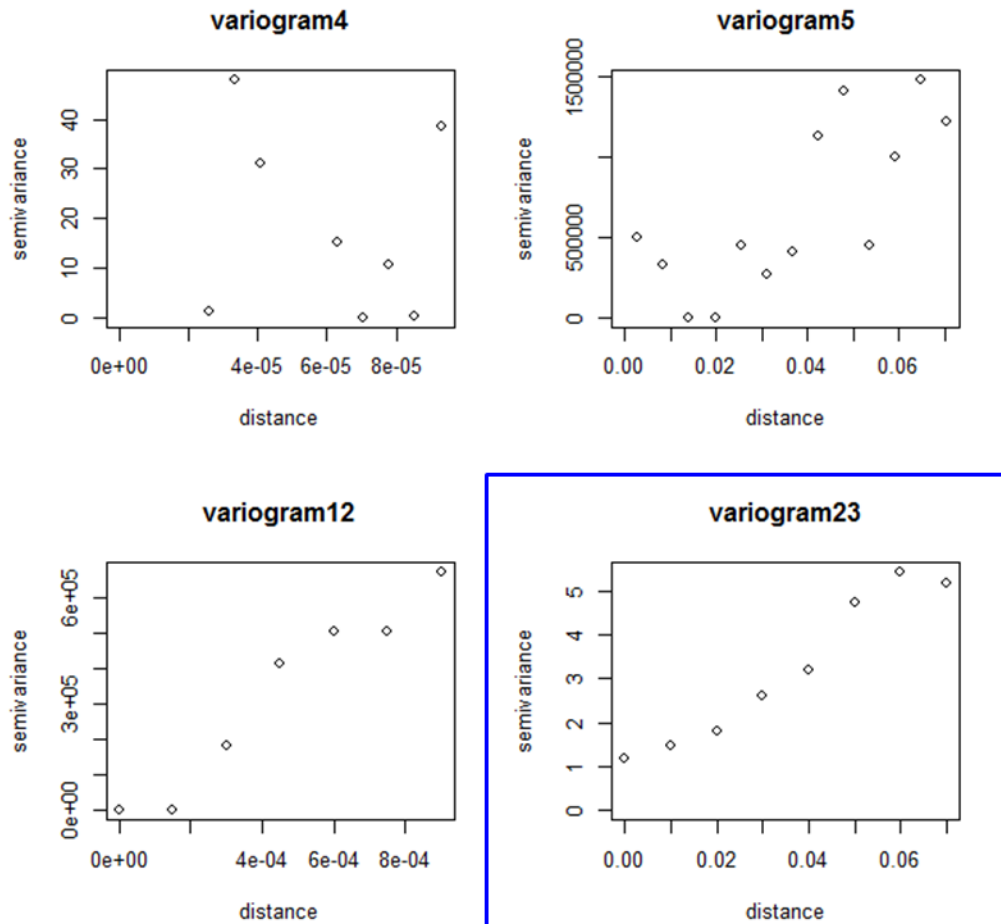


Figure 11 Four plots of the same semivariogram, with varying maximum distance and lag size in R. Points represent average semivariance of each lag.

Before constructing a model semivariogram, it is necessary to construct the empirical semivariogram by varying the size and number of lags as well as the direction of variance (for a review of lags, see background section). In R, the size of lags and the maximum separation distance are varied; in ArcGIS, the size and number of lags are varied. To convert between maximum separation distance and number of lags, divide by lag size. The effects of varying the size of lags and the maximum separation distance can be seen in Figure 11.

The other factor in constructing an empirical semivariogram is directionality. Variograms can be omnidirectional, grouping point pairs only by the separation between them, or directional. A directional variogram considers only point pairs along a certain axis (with a defined angular tolerance). The data for this project was not directional, so an omnidirectional variogram was used.

ArcGIS uses slightly different equations to construct the empirical semivariogram than R, so the parameters cannot be translated between the two programs. In ArcGIS there are two steps that deal with constructing the empirical semivariogram. First, in

Geostatistical Analyst's Data Exploration functions, there is a semivariogram tool (Figure 12).

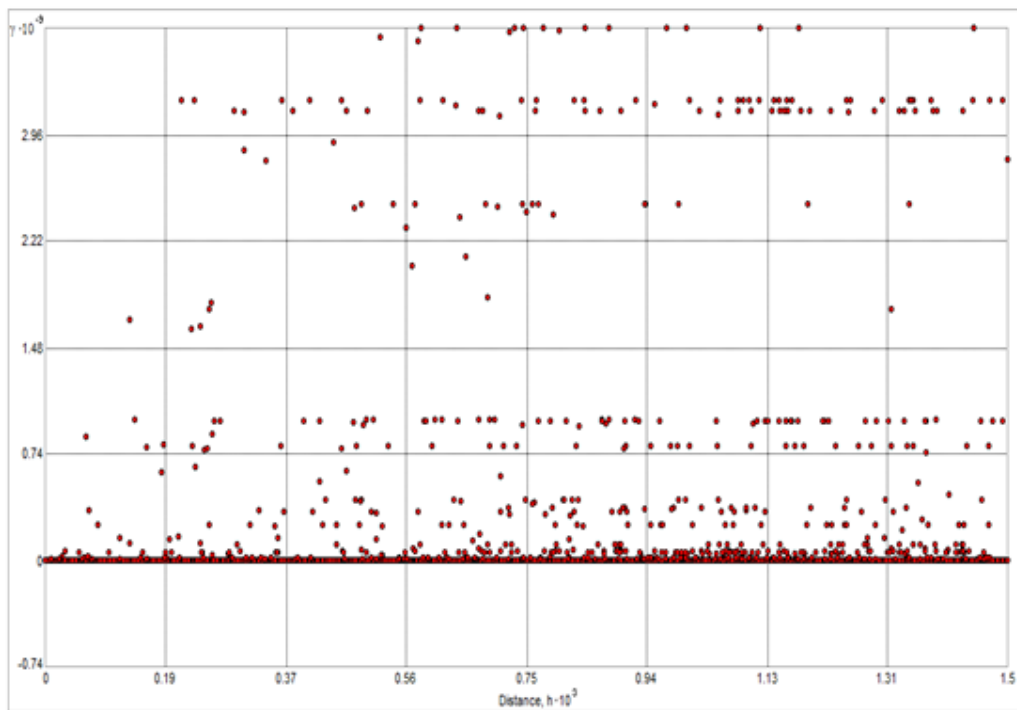


Figure 12 Empirical semivariogram for PCE concentration in the West Coast Basin, ArcGIS Geostatistical Analyst, Data Exploration, Semivariogram tool.

Second, there is a step in kriging through Geostatistical Wizard where the lag size and number are set, which shows a different form of the semivariogram, with both binned and averaged variances for each lag size (Figure 15). The maximum distance should be reflective of the maximum physical separation of two points for which the points can reasonably be expected to be related.

A final consideration in constructing the empirical semivariogram is the presence of extreme values in the dataset. These values are not outliers because they are not due to errors in measurement, but represent real measured values. One way to deal with the error these values introduce into the model is to exclude them from the data while the model is developed, but add them back in for the predictions (Krause 2012). In the case study area there is one site with PCE concentrations far higher than all the other sites. This site was excluded from the dataset during model semivariogram development (Figure 13). An extreme value near zero separation distance is still visible; this was also excluded from the dataset during model development (Figure 13).

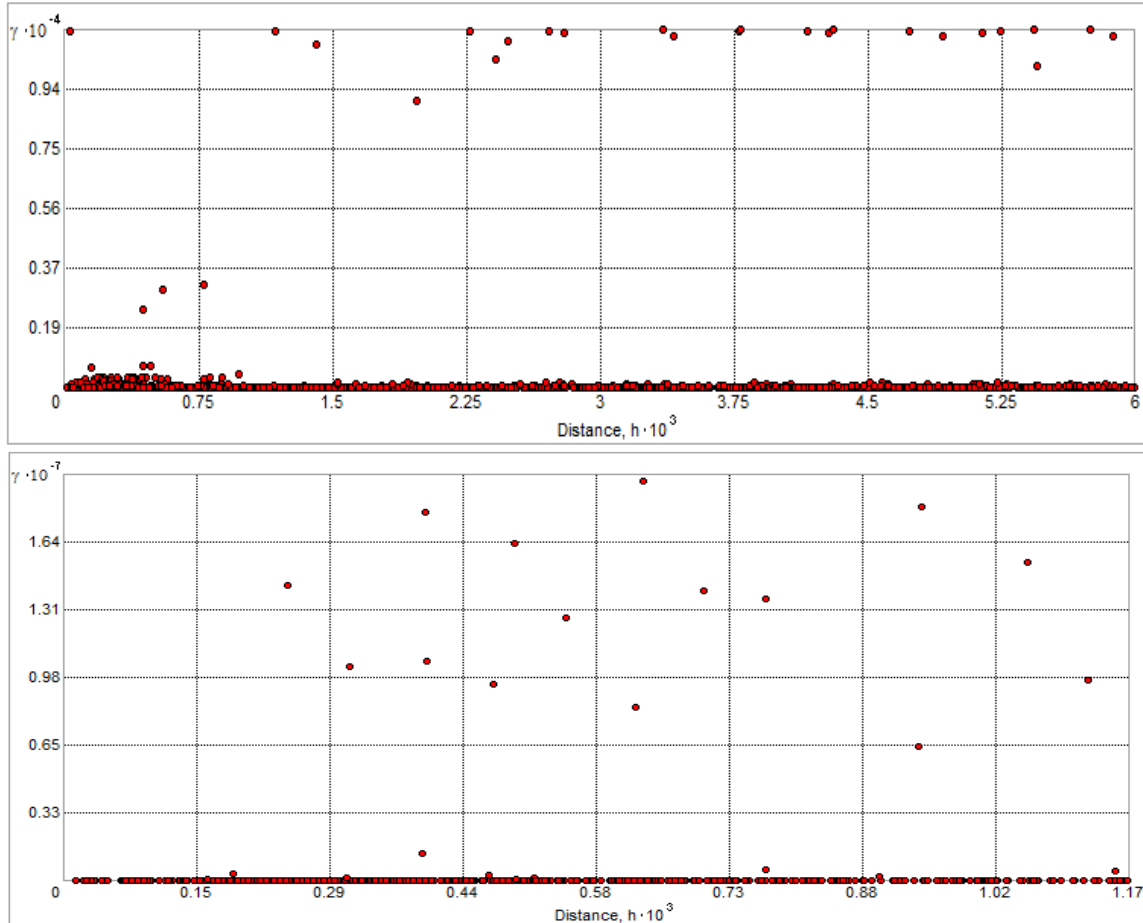


Figure 13 Top: semivariogram for PCE concentration in case study area with site of extreme values excluded. Bottom: semivariogram for PCE concentration in case study area with site of extreme values excluded and an additional extreme value near zero separation distance excluded.

3. Fitting a Model Semivariogram and Cross Validation

In fitting a model to the empirical semivariogram, the first choice to be made is the model type. Variogram types available in ArcGIS are: stable, exponential, Gaussian, spherical, circular, tetraspherical, pentaspherical, J-Bessel, K-Bessel, and Hole Effect. The most common model types are Gaussian, spherical, and exponential. The type of model can be guessed visually and then checked via cross validation. Both ArcGIS and R have optimization functions to determine the best-fit parameters for a given variogram model. In R, the function `variofit` puts initial parameters, estimated by eye, through a loss function, which is optimized to obtain the parameters for the model semivariogram. In addition to the model semivariogram, the weighting type must be defined before optimization. The three types of weighting are Cressie, Equal, and Npairs. Npairs is the default for the `variofit` function.

Table 1. Loss function values for different model types (Gaussian, Exponential, Spherical) and weights (Cressie, Equal, Npairs) applied to the PCE concentration data in the case study area.

	Cressie	Equal	Npairs
Gaussian	67.4935204423278	0.557370025158011	770.230099026166
Exponential	718.794718484257	2.70781599637019	6131.32826941955
Spherical	972.208781908626	2.44126260304492	7938.48323426944

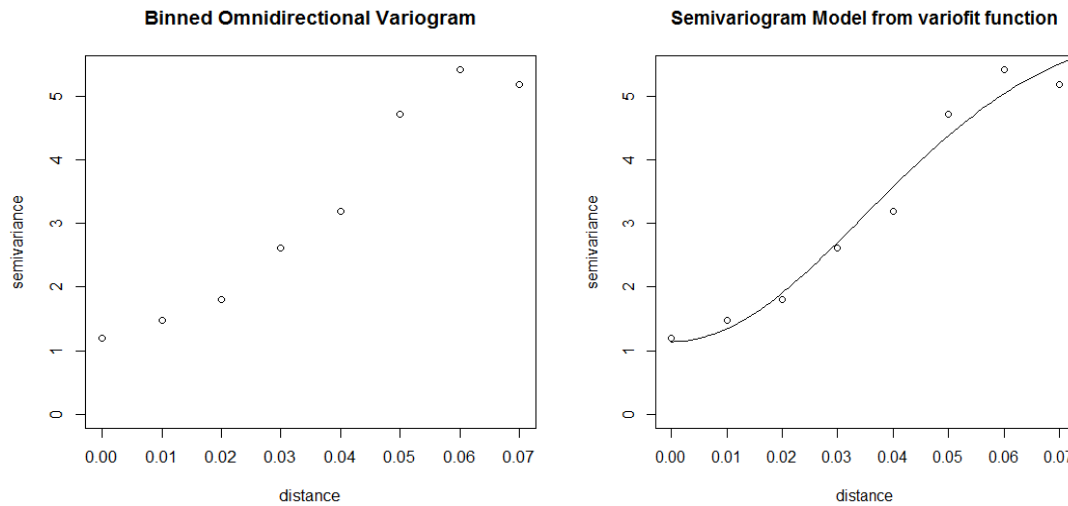


Figure 14 Binned omnidirectional empirical semivariogram with log transformation, bin size of 0.01 and bin number of 8, robust estimator (left). Empirical semivariogram with model semivariogram obtained via the variofit function in R.

Table 2. Parameters from variofit function in R, using Gaussian model and equal weights.

Parameter	Value
<i>Lag Size</i>	0.01
<i>Number of Lags (max.dist/lag size)</i>	8
<i>Nugget</i>	1.2
<i>Partial Sill</i>	5
<i>Range</i>	0.05
<i>Kappa (smoothing)</i>	0.5

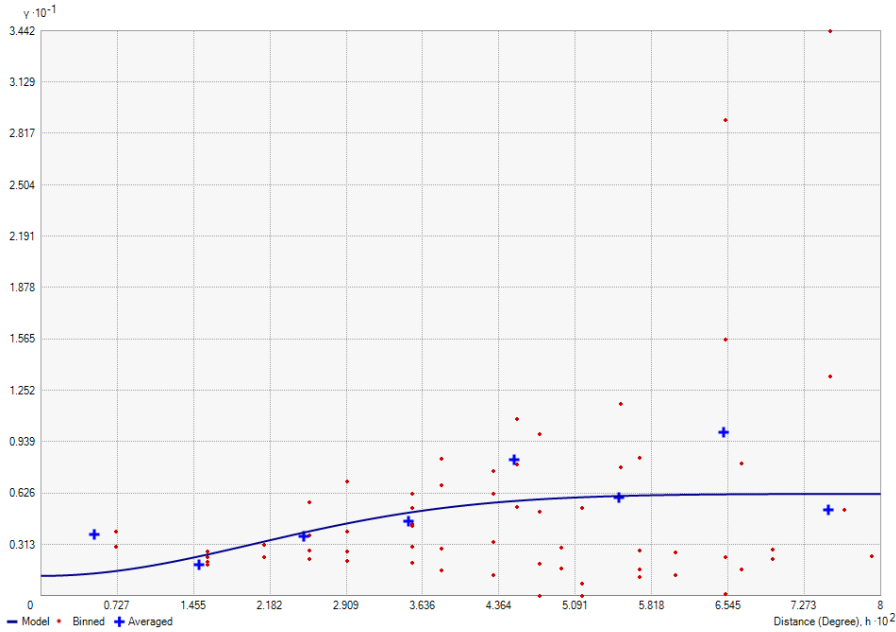


Figure 15 Empirical semivariogram (binned variance shown as red dots, averaged variance shown as blue crosses) and model semivariogram with parameters taken from best-fit model determined in R by the variofit function. Due to slight differences in the empirical and model equations, the semivariograms do not match despite the shared parameters.

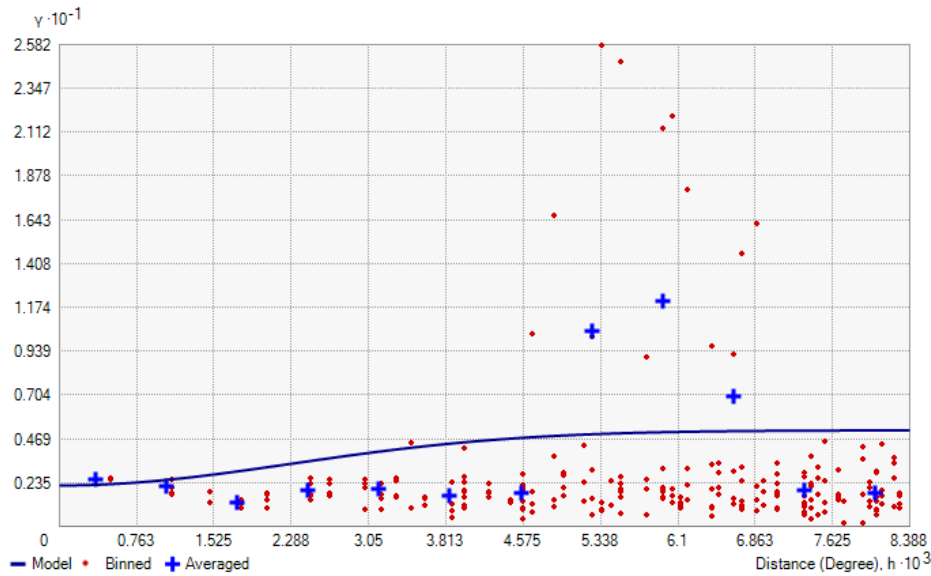


Figure 16 Empirical semivariogram (binned variance shown as red dots, averaged variance shown as blue crosses) and model semivariogram with parameters optimized in ArcGIS for a Gaussian variogram model. The major difference between this semivariogram and the one in Figure 12 is the maximum distance, which is an order of magnitude greater.

4. Defining the Search Neighborhood

The final step in generating a kriged prediction surface is defining the search neighborhood. This limits the radius and number of points that influence each prediction

point. The major and minor semi-axes define the vertical and horizontal reach of the search area. This step was not performed in R but it would be possible in future projects. The default in ArcGIS is for these parameters to be copied from the variogram, but they can also be set manually. If there is no directionality in the data, the search neighborhood should be circular and the major and minor semi-axes should be equal. For this project both axes were fixed at 0.0025. Within the search neighborhood, the minimum and maximum number of neighbors must also be set. For this project, the minimum number of neighbors was set to 0; this means that outside the search neighborhood radius from measured points, no concentration is predicted. Finally, the type of neighborhood—standard or smooth—also affects the radius and the predictions. A smooth neighborhood is better for data with lots of variation in values close to one another, which made it more appropriate for this project. Within the smooth neighborhood, a smoothing factor must be set between 0 and 1. The higher the smoothing factor, the larger the search radius. For this project, it was set at 0.4.

Validation Analysis

Due to time and data constraints, validation analysis was not performed. Validation analysis required subsetting data into test and training points. The data was not dense enough to accommodate the exclusion of training points. When the data for the West Coast Basin was randomly subsetted and then kriged, entire contamination sites were predicted to have zero PCE concentration because there were so few points in the site that they were all excluded randomly from the dataset. If future projects are able to more fully develop an effective kriging model, validation analysis will be essential to determining the confidence in predictions.

Integration in ArcGIS

The predicted concentration layer of PCE was integrated over the volume of water to obtain the mass of PCE. The concentration surface is two-dimensional and was assumed to be fixed over a given depth; variations in concentration with depth were not considered. To calculate the volume of water in the study area, the prediction cell area, depth, and porosity of aquifer media were multiplied according to the following equation:

Volume = Cell Area (m²) * Depth (ft.) * Porosity of aquifer media * Conversion factors.

The cell area refers to the area of the output cells of the raster used to sum mass. Because the cell area listed in the raster properties is in units of projected coordinates, the ruler tool was used to measure the area of cells. Some variation in the measurements was accounted for by averaging five measurements, which yielded a cell area of 334.528495 m².

Defining the depth of the saturated zone across which PCE is distributed is difficult because local hydrology, rate of pumping and recharge of groundwater, the sinking rate, and the residence time in the aquifer all affect the depth distribution of a contaminant. The initial assumption that monitoring well screens covered 80 – 90% of the saturated zone and that PCE was approximately uniformly distributed across the

saturated zone was rejected. Two definitions of depth were used during volume integration. First, the well screen length was assumed to be an approximate measure of the depth across which contaminants can be expected to be distributed. The function of well screens is to keep sand the neighboring media like sand and gravel from collapsing into the borehole, and it also supports the borehole walls and allows water to flow into the well (Brush 1971). The well screen lengths in the case study area ranged from five feet to 96.04 feet, but the majority were either 20 or 30 feet long (Figure 17).

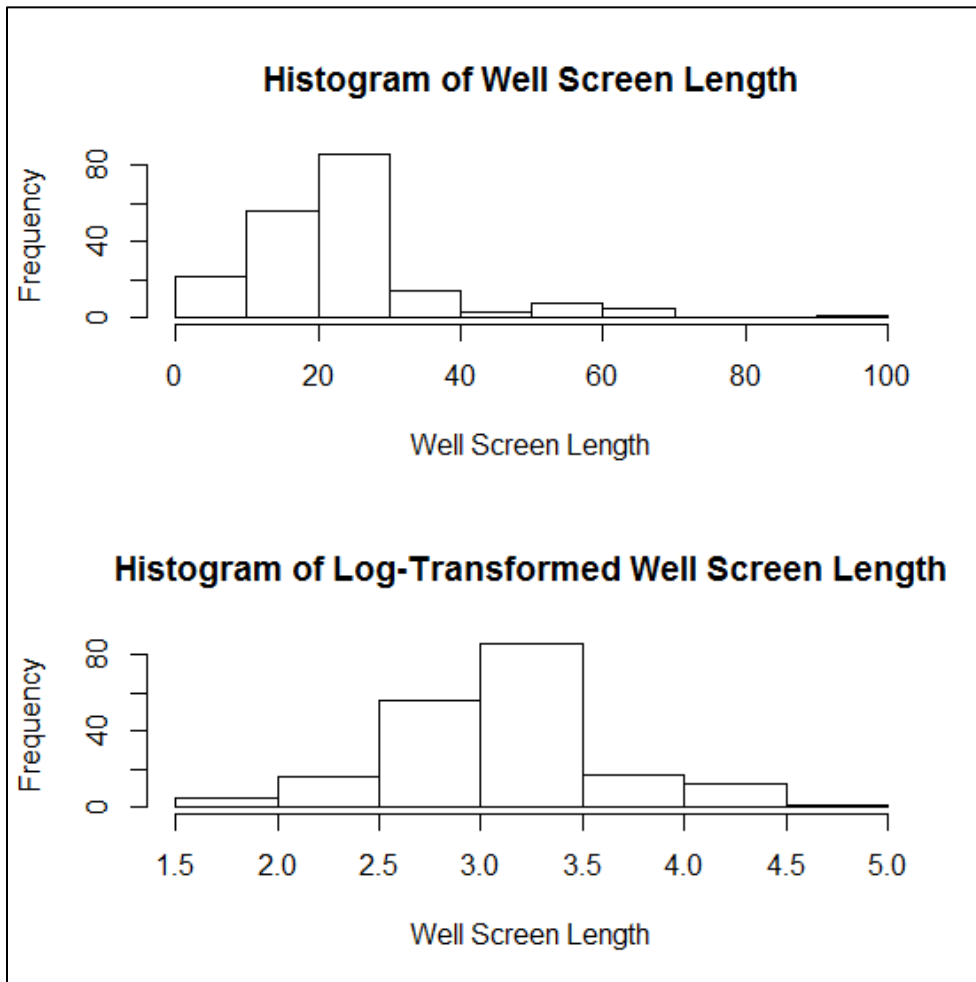


Figure 17 Well Screen Length data for the Torrance case study area.

Well screen length was interpolated by ordinary kriging (Figure 18). The concentration raster and depth raster were multiplied together along with cell area, porosity, and conversion factors to predict mass.

To determine which kriging methods was the most appropriate for the interpolation of groundwater depth, three kriging methods were used to produce a continuous depth profile for the aquifer: Simple, Ordinary and Universal. Different assumptions of the groundwater depth are made for each of the kriging. Interpolating groundwater depth by simple or ordinary kriging, assumes that the groundwater depth is a stationary variable. On the other hand, universal kriging assumes that a trend exists in the varying depths. Since we were unsure as to what kriging model fits well to the length of

well screen data, we decided to perform a test for each kriging model while adjusting the parameters.

After interpolating the well screen length by all three kriging models, the most suitable kriging model was determined by cross validation. In cross validation, four statistics were evaluated for each kriging method. These statistics are standardized mean, root mean squared (RMS), root mean squared standardized (RMSS), average standard error and standardized mean. The most suitable model should have a standardized mean closest to 0, a smallest RMS value, an average standardized error that is close to the RMS prediction error and a RMS standardized value close to 1. Models that produce a RMS that is close to the average standard error is preferred more than the model that has a lower average standard error and a lower RMS.

Based on these statistics, simple kriging seems to be the most suitable model as it has the smallest RMS value, a RMS standardized error close to 1 and an average standardized error that is close to the root mean square. It requires a fixed and known mean, though, so ordinary kriging was selected because it does not require a known mean and its standardized mean error was closest to 0.

Table 2 Summary of cross validation error statistics for different kriging methods on well screen length with parameters automatically calculated in ArcGIS.

	Simple Kriging	Ordinary Kriging	Universal Kriging
Root-mean-square	11.43666	12.62842	13.24804
Standardized mean	-0.009752081	0.004051191	0.02345767
Root-mean-square standardized	0.9909258	1.505904	1.511183
Average standard error	11.40605	7.70138	30.8332

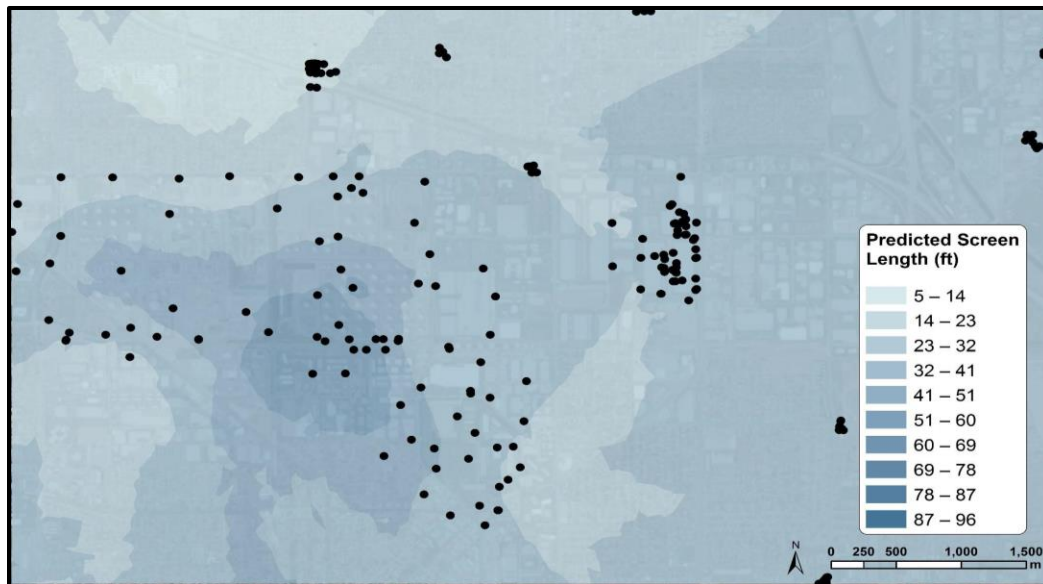


Figure 18 Monitoring well screen length predicted by ordinary kriging in case study area. Well screen length was used as a surrogate for the depth across which most PCE is concentrated in the aquifer. Monitoring wells are shown in black.

A second model for estimating the volume assumed that the depth across which PCE was approximately uniformly distributed was fixed at twenty feet. In this scenario, the kriged concentration surface was multiplied by porosity, cell area conversion factors, and a fixed depth value of 20 feet.

Porosity is included in the volume calculation because this project considered only contaminants in groundwater, not those sorbed to soil. Furthermore, concentration measurements are of contaminants present in groundwater, not in soil. Multiplying the volume of the aquifer by porosity gives the volume of water as opposed to the volume of water and soil combined. In the Site Maps/Documents tab, each well has a Geobore diagram (although many are missing) that includes verbal descriptions of the aquifer soil media at different depths for each well. The soil media are classified according to the Unified Soil Classification System (USCS), which classifies soil based on the media's texture and grain size (Stevens 1982). Porosity was initially assumed to be fixed at thirty percent, an assumption which was supported by a closer analysis of the aquifer media profiles available through GeoTracker. We compiled an Excel spreadsheet showing the depth media profile of each well represented as one foot intervals, with each interval assigned a USCS symbol to match the media type (Figure 19). A literature review was conducted to assign each USCS symbol a numeric porosity value (Morris 1967). After comparing the porosities of different media types present in the case study area, we concluded that the initial assumption of 30 percent porosity was reasonable (Table 3).

SL184361419MW-25	SL184361419MW-26	SL184361419MW-27	SL184361419MW-28
CL	SM	SM	SM
CL	SM	SM	SM
CL	SM	SM	SM
SM	ML	SC	SP
SM	ML	SC	SP
SM	ML	SC	SP
SM	ML	SC	SP
SM	ML	SC	SP
CL	SP	CL	SP
CL	SP	CL	SP
CL	SP	CL	SP
CL	SP	CL	SP
CL	SP	CL	SP
SM	SC	SC	SP
SM	SC	SC	SP
SM	SC	SC	SP
SM	SC	SC	SP
SM	SC	SC	SP
SM	SM	SC	SP

Figure 19 Aquifer media profiles of four wells in the case study area.

Table 3 Soil type and corresponding porosity values assigned according to Morris 1967.

Symbol	Soil Type	Porosity Values (%)
SM	Silty sand	29-49
SP	Poorly-graded sand	31-46
SP-SM	Silty sand/ Poorly-graded sand	26-46
SW	Well-graded sand, fine to coarse sand	31-46
CL	Clay	34-57
ML	Silt	34-61
SC	Clayey sand	26-53
SM/ML	Silty sand/ silt	34-49

IV. Results

PCE concentration was predicted in the case study area by ordinary kriging (Figure 20) and inverse distance weighting (Figure 21) in ArcGIS. The prediction surfaces are similar. The surface produced by ordinary kriging is preferable because it does not suggest that monitoring wells are sources of contamination. In the IDW prediction surface, high concentrations are predicted around wells with high measured concentrations and then fall off in a circle around individual wells. The parameters for the ordinary kriging were, as follows:

- log transformation
- smooth search neighborhood with a smoothing factor of 0.4
- major and minor semi-axes = 0.025
- number of lags = 13
- lag size = 9×10^{-5}
- nugget = 1.6463277694940552
- measurement error = 100%
- model type = stable
- range = 0.00024177790776960882
- sill = 0.9703698926393521

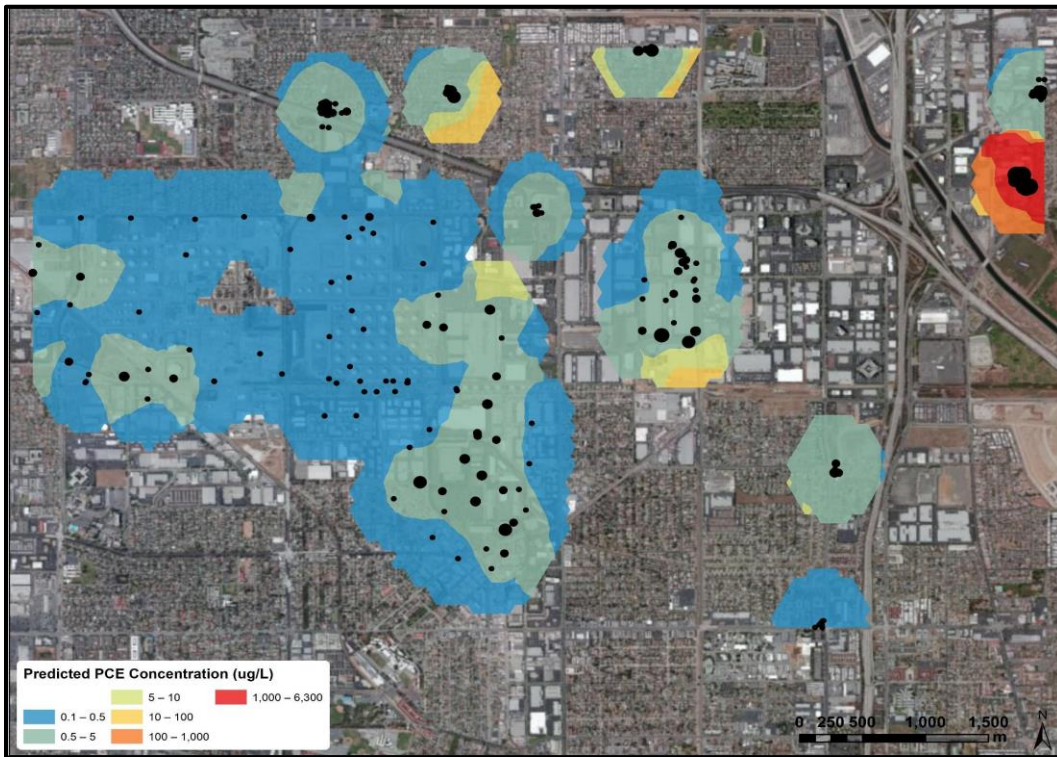


Figure 20 PCE concentration predicted by ordinary kriging in case study area. Monitoring wells are shown in black and sized proportionally to measured concentration.

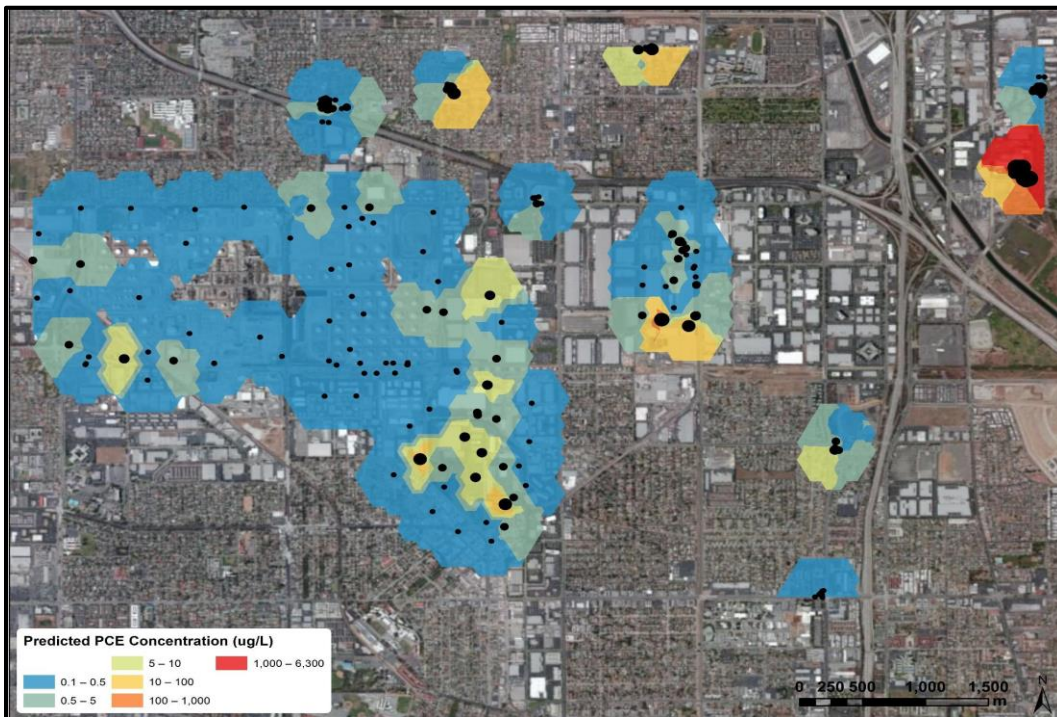


Figure 21 PCE concentration predicted by inverse distance weighting in case study area. Monitoring wells are shown in black and sized proportionally to measured concentration.

In the case study area, depth was estimated in two ways: first it was assumed that well screen length was a reasonable surrogate for depth and well screen length was predicted by ordinary kriging across the study area (Figure 18); second, a fixed depth of twenty feet was assumed. Mass was calculated and summed for the case study area according to the following equation:

$$\text{Mass} = \text{concentration} * \text{depth (ft)} * 1000 * 334.528495(\text{m}^2) * 0.3048 (\text{m/ft}) * \text{porosity}.$$

When the kriged well screen length was used to define volume, the sum of mass predictions was 1014 kilograms. When depth was assumed to be fixed at 20 feet, the sum of mass predictions was 900 kg.

Concentration in the West Coast Basin was predicted by ordinary kriging (Figure 22). Depth was assumed to be fixed at 20 feet. There was not time to collect all the well screen length data for kriging, so depth was assumed to be fixed at twenty feet. The summed mass prediction for the West Coast Basin was 4383 kg.

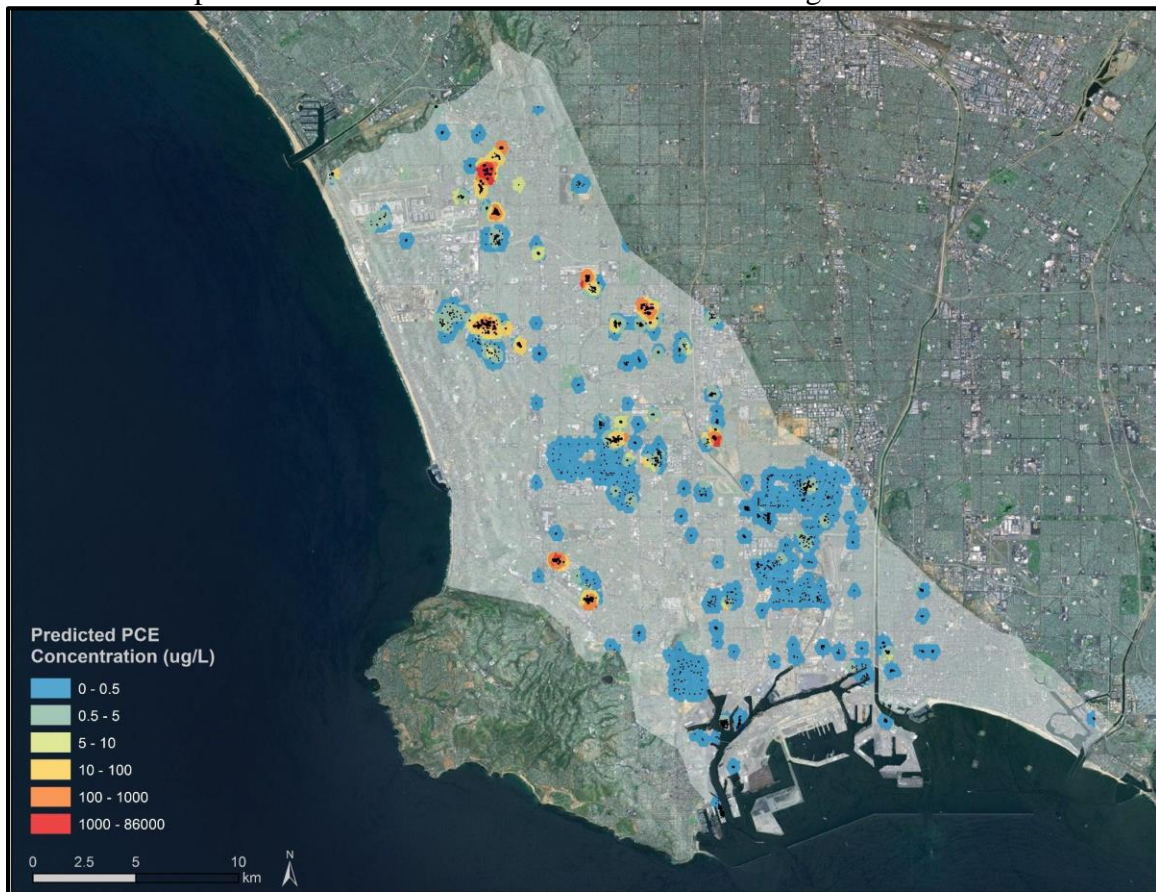


Figure 22 PCE concentration predicted by ordinary kriging in the West Coast Basin. Wells are shown in black.

V. Discussion

Sources of Uncertainty

The mass predictions from this methodology are approximate and likely overestimate the total mass. Kriging concentration introduces uncertainty because the kriging model is not ideal for the GeoTracker data. Kriging is ideal for densely distributed data on a grid. The monitoring well network supplying data to the GeoTracker database, on the other hand, is concentrated only in areas where there are known contamination sites. Wells are distributed in clusters and many sites have five or fewer monitoring wells; the sparseness and the distribution pattern of monitoring wells limit the accuracy of interpolation models (not just kriging).

The treatment of non-detect values using MDLs should be developed further. As discussed in the methodology section, MDLs in the case study area varied from 0.14 to 130 $\mu\text{g/L}$. A majority of the detected concentrations (66 of 114) were below 65 $\mu\text{g/L}$, which is the value assigned to the non-detect measurement with an MDL of 130 $\mu\text{g/L}$. Helsel (2005) provides a discussion of alternate methods for treating non-detect data.

Another limitation is that the data in GeoTracker does not distinguish between samples taken from physically distinct aquifers. Limiting the search neighborhood to local rather than global helps mediate this problem, but it cannot account for physical barriers in aquifer structure between monitoring points that are within each other's search neighborhood. Because kriging, like any interpolation model, assumes that all points in a search neighborhood are related, if two points that are close to one another are in fact separated in two distinct aquifers, assuming their values affect one another is incorrect. Some way of grouping the GeoTracker data into distinct aquifers and then kriging aquifers individually would help resolve this problem.

Another generalization in the model is that concentration is vertically uniform; the GeoTracker data does not show the changing distribution of a contaminant with depth. Several wells do have sampling points at several depths, but a great majority of the wells do not. In the Torrance case study area there were only two wells that had measurements at multiple depths. This inability to account for variations in concentration with depth introduces huge uncertainty because it assumes that whatever concentration was measured at the sampling point is uniform for the entire assumed depth either fixed (20 feet in this study) or modeled (kriging well screen length). One possible approach to accounting for variations in concentration with depth would be to assume a normal distribution around the sampling point. Another would be to conduct a study of all the wells in a larger study area that have measurements at multiple depths and construct a characteristic mass distribution curve over depth and apply this curve to all the single measured concentration points.

A case study considering the vertical distribution of PCE would also inform a more accurate estimate of depth for the volume integration step. Both the assumption of twenty feet and the assumption of interpolated well screen length as an approximation of contaminant distribution introduce uncertainty into the model. Both of these approximations are based on the assumption that well screen length is somehow reflective of the depth over which PCE is distributed. This makes sense intuitively because wells are designed to monitor contaminants and are therefore placed in parts of

the aquifer where contaminants are likely to be found; however, a review of the literature on sinking and dissolution of PCE may yield a more suitable means of modeling depth.

Recommendations and Considerations for Large-Scale Implementation

The methodology developed at the case study level included two components that were impractical to scale up to the basin-wide level: the well screen length and the MDLs had to be obtained individually and manually joined to the concentration data downloaded from GeoTracker. Much of the well screen length data was missing entirely or had to be looked up in geobore diagrams. MDL data was grouped into quarterly monitoring reports and the team had to sort through a whole series of reports in each quarter to find the relevant measurements. If the MDL and well screen length components of the methodology are included in future conceptual models, the procedure would be streamlined by the inclusion of this data in the downloadable spreadsheet. MDLs, or some other relevant detection limit will most likely be key to any mass prediction model.

Despite the uncertainties and challenges outlined above, we believe that this approach should be explored further. Restructuring the GeoTracker database to include even one representative detection limit and any relevant depth data would make the exploration of models on a large scale more efficient. Reevaluating and developing the treatment of non-detect results would strengthen the interpolation model. Exploring the vertical distributions of contaminants and how they can be estimated from the single-depth monitoring points of most wells, perhaps through a case study, will also be key to strengthening the model. California has seen the advantage of a mass-based regulatory approach in the management of surface waters. A successful mass estimation model would open up new possibilities for monitoring and regulating groundwater quality in California.

Works Cited

Agency for Toxic Substance and Disease Registry (ATSDR). 1997. Toxicological profile for tetrachloroethylene. ATSDR, Atlanta, GA.

Brush, R. E. 1971. Wells construction: hand dug and hand drilled. Washington, D.C.

Cameron, K., and P. Hunter., 2002. Using spatial models and kriging techniques to optimize long-term ground-water monitoring networks: a case study. *Environmetrics* **13**:629–656.

State Water Resources Control Board (SWRCB). 2012. GAMA program fact sheet. SWRCB, Sacramento, CA.

State Water Resources Control Board (SWRCB). 2013. Geotracker GAMA. SWRCB, Sacramento, CA.

Christou, Nicolas. 2013. Statistics C173/ 273: Applied Geostatistics. UCLA Department of Statistics, Los Angeles. Accessed May, 2013:
<http://www.stat.ucla.edu/~nchristo/statistics_c173_c273/>

Environmental Systems Research Institute, Inc. (ESRI). 2012. Understanding ordinary kriging. ESRI, Redlands, CA.

Federal Advisory Committee. 2007. Report of the Federal Advisory Committee on detection and quantitation approaches and uses in Clean Water Act programs.

Helsel, D.R. 2005. More than obvious: better methods for interpreting nondetect data. *Environmental Science & Technology* **39**:419–423.

Jang, C. S., C. W. Liu, S. K. Chen, and W. S. Lin. 2012. Using a mass balance model to evaluate groundwater budget of seawater intruded island aquifers. *Journal of the American Water Resources Association* **48**:61–73.

Kirk, T. 2006. Installing a water supply well in the Central and West Coast Basins – using available geologic information to predict how deep to drill. *WRD Technical Bulletin* **7**: 1–2.

Krause E. 2012. Dealing with extreme values in kriging. Environmental Systems Research Institute. Redlands, CA. Available from <http://blogs.esri.com/esri/arcgis/2012/05/07/dealing-with-extreme-values-in-kriging/> (accessed May 2013)

Machiwal, D. and M. K. Jha. 2011. Hydrologic time series analysis: theory and practice. Capital Publishing Company, India.

Moran, M. J., J. S. Zogorski, and P. J. Squillace. 2007. Chlorinated solvents in groundwater of the United States. *Environmental Science & Technology* **44**:74–81.

Morris, D. A. and A. I. Johnson. 1967. Summary of hydrologic and physical properties of rock and soil materials as analyzed by the Hydrologic Laboratory of the U.S. Geological Survey, U.S. Geological Survey Water-Supply Paper 1839-D, 42p.

Nas, B., and A. Berkday. 2010. Groundwater quality mapping in urban groundwater using GIS. *Environmental Monitoring Assessment* **160**:215–27.

Reichard, E.G., M. Land, S.M. Crawford, T. Johnson, R.R. Everett, T.V. Kulshan, D.J. Ponti, K.J. Halford, T.A. Johnson, K.S. Paybins, and T. Nishikawa. 2003. Geohydrology, geochemistry, and ground-water simulation-optimization of the Central and West Coast Basins, Los Angeles County, California. Water-Resources Investigations Report 03-4065. US Geological Survey, Sacramento, California, United States.

Rejith, P. G., S. P. Jeeva, H. Vijith, M. Sowmya, and A. A. Hatha. 2009. Determination of groundwater quality index of a highland village of Kerla (India) using Geographical Information System. *Journal of Environmental Health* **10**:51–8.

Rong, Y. 2002. Laboratory detection limits. *Contaminated Soil Sediment & Water Magazine*.

Stevens, J. 1982. Unified Soil Classification System. *American Society of Civil Engineers* **52**:61–62.

State Water Resources Control Board (SWRCB). 2009. Groundwater Information Sheet Tetrachloroethylene (PCE). SWRCB, Sacramento, CA.

United States Environmental Protection Agency (USEPA). 1994. Chemicals in the environment: perchloroethylene. USEPA, Washington, D.C.

United States Environmental Protection Agency (USEPA). 1991. Guidance on handling chemical concentration data near the detection limit in risk. USEPA, Washington, D.C.

United States Environmental Protection Agency (USEPA). 1997. Non-detect policy. USEPA, Washington, D.C.

Water Replenishment District of Southern California (WRDSC). 2011. The water replenishment district. WRDSC, Lakewood, CA.

Appendix: Description of steps taken in ArcGIS

Performed in ArcMap 10.1

1. Enable extensions: Geostatistical Analyst, Spatial Analyst. Add desired background imagery.
2. Add excel spreadsheet “Basin” to ArcMap
3. Add excel spreadsheet “Master Dataset - Torrance Cluster 3Yr” to ArcMap, using sheet “For ArcGIS”
4. Display XY data
 - a. X = LON
 - b. Y = LAT
 - c. Z = 3YR_AVG (for Cluster), = CON (for Basin)
 - d. Coordinate System: Geographic, WGS 1984

5. Create new geodatabase “Cluster” and export data from Cluster displayed XY data to this geodatabase, save as “Cluster”
6. Create new geodatabase “Basin” and export data from Basin displayed XY data to this geodatabase, save as “Basin”
7. Create Training and Testing Subsets (Geostatistical Analyst)
 - a. Geostatistical Analyst → Subset Features
 - i. Cluster
 1. Training: ClusterTrain
 2. Testing: ClusterTest
 3. Set to 70% Training, 30% Testing
 - ii. Basin
 1. Training: BasinTrain
 2. Testing: BasinTest
 3. Set to 70% Training, 30% Testing
 - b. Group the new layers with their sources (ie. cluster test and training go with cluster, basin test and training go with basin group)
8. Data Exploration
 - a. Separate extreme data
 - i. Export data again (from initial XY Cluster data), save to a new geodatabase named “Custom”, name file “Custom”
 - ii. Make outlier visible by using graduated size in symbology
 1. Properties-->Symbology-->graduated size, set min size to 5, max to 30
 - iii. Using Editor tool on Custom, select the one outlier site and delete it
 - b. Data Exploration (Use Custom)
 - i. Histogram of Custom
 1. Determine Transformation Type that gives closest to normal distribution (log)
 - ii. Trend Analysis of Custom
 1. See if data produces a flat surface
 - a. In this case, it does
 2. This will determine trend removal
 - iii. Semivariogram of Custom
 1. Determine lag size, lag number, partial sill, range
 2. (For Custom, there really is no good fit because there aren't enough points)
 3. Remove one more outlier (the very high value near 0 separation) to create a new subset, Custom3.

Semivariogram is still not dense enough, but you can approximate parameters:

- a. lag size = 0.00009, # lags = 13

9. Krig Cluster Concentration

a. Input Data

- i. Source dataset = ClusterTrain
- ii. Data field = YR_AVG
- iii. For coincident data, use mean

b. Kriging Type, Transformation, and Trend Removal

- i. Type = Ordinary
- ii. Transformation = log
- iii. Trend Removal = none

c. Semivariogram Modeling

- i. Use parameters from Custom3 Semivariogram (excluding outliers from empirical semivariogram model)
 1. lag size = 0.00009, # lags = 13
- ii. Fit variogram by eye (optimization function will not work here because of the lag size/# lags taken from the Custom3 semivariogram)
 1. Nugget = 0.97
 2. Model Type = Gaussian
 3. Major range (auto-calculated) = 0.00034275337953335456
 4. Anisotropy = false
 5. Partial Sill (auto-calculated) = 1.5355435370208586

d. Search Neighborhood

- i. Smooth
- ii. Smoothing Factor = 0.4
- iii. Copy from Variogram = False
- iv. Major Semiaxis = Minor Semiaxis = 0.0025

e. Validation Analysis

- i. Input geostatistical layer = kriged layer
- ii. Input point observation locations = ClusterTest
- iii. Field to validate on = YR_AVG

10. Kriging depth

- a. Ordinary kriging
- b. Log transformation
- c. model = stable

11. Map Algebra to get mass (using kriged depth layer)

- a. Export kriged concentration layer to raster
- b. Export depth to raster as well
- c. Toolbox-->Spatial Analyst-->Map Algebra-->Raster calculator

- d. Finding area of the cells in the raster
 - Zoom in to scale of one cell
 - Use the ruler tool, set it to area and meters
 - Area1 = 333.245125
 - Area2 = 334.741093
 - Area3 = 335.090416
 - Area4 = 334.961607
 - Area5 = 334.604234
 - Average Area = 334.528495 m²
- e. Use formula: $\text{Int}(\text{concentration raster} * \text{depth raster} * 30 * 0.3048 * 334.528495) = \text{Int}(\text{concentration raster} * \text{depth raster} * 3058.92855828)$
 - i. Use 30 (= porosity * 100 so as to not lose decimal places by using the Int, or integer, operation)
 - ii. Units: $(\text{ug} * \text{ft} / \text{L}) * (30) * (0.3048 \text{m} / \text{ft}) * (334.528495 \text{m}^2)$
 - iii. To convert to g/L, will have to multiply the sum by (0.01 to account for multiplying porosity by 100) $(1000 \text{ L} / \text{m}^3) (0.000001 \text{ } \mu\text{g} / \text{L}) = 0.00001$
- f. Build Raster table: Toolbox → Data Management Tools → Raster → Raster Properties → Build Raster Attribute Table
- g. Open the raster's attribute table, create new field, type = long integer, Title = MASS
- h. Right click on MASS field, select field calculator and enter expression: Count*Value
- i. Sum MASS (By right clicking column and selecting statistics)
SUM(MASS) = 1013871787793 ug = 1013.871787793 kg

Assume 20 ft. depth across:

- j. Use formula: $\text{Int}(\text{concentration raster} * 20 \text{ ft.} * 30 * 0.3048 * 334.528495) = \text{Int}(\text{concentration raster} * 611785.711656)$
 - i. Use 30 (= porosity * 100 so as to not lose decimal places by using the Int, or integer, operation)
 - ii. Units: $(\text{ug} * \text{ft} / \text{L}) * (30) * (0.3048 \text{m} / \text{ft}) * (334.528495 \text{m}^2)$
 - iii. To convert to g/L, will have to multiply the sum by (0.01 to account for multiplying porosity by 100) $(1000 \text{ L} / \text{m}^3) (0.000001 \text{ } \mu\text{g} / \text{L}) = 0.00001$
- k. Build Raster table: Toolbox → Data Management Tools → Raster → Raster Properties → Build Raster Attribute Table
- l. Open the raster's attribute table, create new field, type = long integer, Title = MASS

- m. Right click on MASS field, select field calculator and enter expression:
Count*Value* 0.0001 [* 0.0001 to accommodate large values that didn't fit]
- n. Sum MASS (By right clicking column and selecting statistics)
SUM(MASS) = 90046173, divide by 0.0001 ug = 900.46173 kg

Mass in BASIN:

Methodology is the same as cluster with the assumption of 20 feet depth.

SUM(MASS) = 438346471, then divide by 0.0001
= 4383464710000 ug = 4383.46471 kg