NAVFAC
Naval Facilities Engineering Command
EXWC Engineering & Expeditionary Warfare Center

ENVIRONMENTAL RESTORATION

Environmental Statistics

Purpose

Statistical methods play a key role in all tasks concerning data: collection, analysis, interpretation and visualization. Therefore, the use of statistics is ubiquitous throughout an environmental project life cycle and supports both the objective of advancing scientific understanding and the management goal of making decisions. For example, during remediation and monitoring stages, trend analysis is often used to evaluate whether chemical concentrations in groundwater are increasing or decreasing over time. Regardless of the media (i.e., groundwater, soil, etc.), chemical concentration data sets are inherently variable, and statistical methods provide the tools needed to understand the behavior and patterns of these chemicals over time and space.

This fact sheet provides a condensed selection and discussion of important statistical methods used for the analysis of groundwater and soil data at Navy and other federal sites. Statistical analysis of groundwater and soil data can present challenges during different stages of the project cycle, such as planning, implementation, and decision making. Challenges can be related to:

- The inherent nature of environmental data, such as censored values (i.e., less-than or non-detect values), small sample sizes, unrepresentative or biased samples, skewed data distributions, measurement error, etc.
- The selection and application of statistical methods to answer questions that are specific to the site and/or the objectives of the investigation.

This fact sheet is intended to elucidate some of these challenges and to promote the proper use, application and interpretation of statistical methods. The following topics are discussed:

- When is it appropriate to use statistics?
- Which statistical methods can be used for groundwater and soil data sets?

While this fact sheet provides definitions of statistical concepts and their application in environmental projects, it is not intended to replace other detailed statistical texts, guidance documents or software. Additional resources on environmental statistics are provided at the end of this fact sheet.

When is it Appropriate to Use Statistics?

Before conducting any statistical evaluation, data must be reviewed to ensure that the right type, quality and quantity of data are available. Systematic planning helps to ensure that the data collected are of sufficient quality to be used in statistical evaluations. This planning process is based on guidance for establishing data quality objectives (DQOs) as provided in the Uniform Federal Policy for Quality Assurance Project Plans (UFP-QAPP), along with additional U.S. Environmental Protection Agency's (USEPA's) resources (USEPA, 2005, 2006a and 2006b). The following considerations can be used to help determine whether the data are usable for statistical evaluations:

Sample Design. Data should be collected using systematic planning and incorporate a statistical (or probabilistic) sampling design. In particular, when the study objectives involve estimation or decision making, some form of probability sampling (e.g., simple random sampling, stratified sampling, etc.) should be selected to ensure that the data are representative of the target population, e.g., site and background (USEPA, 2000). Data collected using judgmental sampling, where the site expert designates where and when samples are collected, should not be used for statistical evaluations because the results will be biased and lead to erroneous conclusions.

Sample Size. To conduct defensible statistical evaluations, a sufficient sample size (number of usable data points) needs to be collected from the population of interest using appropriate DQO processes. As a general rule, for both

Distribution Statement A: Approved for public release; distribution is unlimited.

September 2019

parametric tests (i.e., tests that assume a data distribution, for example, the normal distribution) and most nonparametric tests (i.e., tests that do not assume a data distribution), the larger the sample size, the greater the power and the smaller the decision error risk (false positives and false negatives). To minimize decision error risk, minimum sample sizes should be determined before data collection, if possible, as part of the design process. However, more often than not, the data have already been collected without using a DQO process, or due to resource constraints, it may not be possible to collect as many samples as determined by using a DQO-based sample size formula (ITRC, 2013). Under these circumstances, guidance documents (e.g., Navy, 2002; Navy 2004; ITRC, 2013; USEPA, 2009) recommend a minimum of 10 independent observations for most statistical tests.

Historical Data. Historical or pre-existing data may be used for statistical evaluation, but the historical data must be checked for sufficient quality compared to more recent data. For example, changes in sample collection and analytical methods over time can introduce bias and higher levels of uncertainty in chemical measurements. These biases can affect the results of a statistical evaluation, leading to erroneous conclusions. Exploratory data analysis (EDA) methods can also be used to assess the usability of historical data. EDA is described in more detail below under the "Commonly Used Statistical Methods" section.

Handling Non-Detect Data. Non-detect (ND) data complicate statistical evaluations, especially when a large portion of the data are ND. There are two main approaches for handling ND data: (i) the substitution method, where NDs are replaced with zero, half of the reporting limit (RL), or full RL, or (ii) the use of robust statistical methods that can account for ND data, such as, Kaplan-Meier (KM), regression on order statistics (ROS), and maximum likelihood estimation (MLE). Substitution methods are generally not recommended because it can lead to biased estimates. The recommended approach for managing ND and estimated ("J-flagged") data (Navy, 2002; Navy 2004; USEPA, 2009) can be summarized as follows:

- If ND data are infrequent (<15%), half of the RL can be used in place of these data without significantly biasing the statistics.
- If ND data are 15% to 50%, parametric methods that explicitly handle NDs (e.g., MLE) or non-parametric methods insensitive to the presence of NDs should be used (e.g., KM and ROS).
- If ND data are >50%, use a non-parametric method.
- Where available, estimated results less than the RL (i.e., "J-flagged" data), but greater than the method detection limit should be considered detections for the purposes of statistical analysis.

Checking Statistical Assumptions. Any formal statistical test makes a series of assumptions about the underlying population from which the sample data were collected. These assumptions will vary depending on the objective of the statistical evaluation (e.g., trend analysis versus background comparisons, etc.) because assumptions are test specific. EDA/graphical methods and formal statistical tests should be used to verify these statistical assumptions so that accurate and defensible conclusions are being made about the data. Common statistical assumptions and the tests that can be conducted to verify them include:

- Normality – Shapiro Wilk (smaller sample sizes) or Lilliefors (larger sample sizes).
- Equality of variance (i.e., between different populations) – Levene's test.
- Temporal independence – sample autocorrelation function or rank von Neumann ratio test.
- Temporal stability – regression or Mann-Kendall.
- Spatial variability – parametric one-way analysis of variance (ANOVA) or nonparametric Kruskall-Wallis.
- Outliers – Dixon's or Rosner's test.

An introduction to EDA/graphical methods, regression, and Mann-Kendall are included in this fact sheet. A more comprehensive list of test-specific assumptions and associated methods/tests to verify them can be found in Appendix F of ITRC's Groundwater Statistics and Monitoring Compliance guidance document (ITRC, 2013) and various other guidance documents.

Which Statistical Methods Can be Used for the Analysis of Groundwater and Soil Data?

Commonly Used Statistical Methods and Tests

The following sections provide descriptions of some of the most commonly used statistical methods for analyzing groundwater and soil data sets. Table 1 provides a list of common study questions and indicates which statistical methods can be used to answer them. Guidance documents that contain the technical details for these and numerous other statistical methods/tests can be found in NAVFAC

(2002, 2004, 2010, 2012); ITRC (2013); USEPA (2009), and the many references therein. Note that a comprehensive list of software packages that implement these statistical methods are available in ITRC's Groundwater Statistics and Monitoring Compliance guidance document (ITRC, 2013) and its companion web-based guidance document Geospatial Analysis for Optimization at Environmental Sites (ITRC, 2016).

Exploratory Data Analysis/Graphical Methods

EDA consists of descriptive and graphical methods, such as summary statistics, time series plots, scatter plots, histograms, box plots, and probability plots. EDA can be used to:

- Check data quality.
- Qualitatively identify patterns, trends and relationships.
- Provide information on range, spread and shape of a contaminant distribution.
- Detect outliers and anomalies.
- Select the appropriate statistical method, e.g., parametric versus nonparametric.
- Test underlying assumptions of a statistical test, e.g., normality, equality of variance, etc.

Three commonly used statistical graphical methods used for EDA are described below. Note that many statistical software packages provide options to construct censored versions of these plots to accommodate data sets with ND values.

Table 1. Statistical Methods Used to Answer Common Questions at a Site

Study Question	GW	Soil	EDA/Graphical Methods	Regression	MK/TS	Statistical Limits	Hypothesis Tests	Geospatial Methods
What are the background concentrations?	✓	✓	✓			✓		
Are onsite concentrations greater than background concentrations?	✓	✓	✓			✓	✓	
Are concentrations above or below a criterion?	✓	✓					✓	
When will contaminant concentrations reach a criterion?	✓			✓				
Is there a trend over time in contaminant concentrations?	✓		✓	✓	✓			
Is there seasonality in the concentrations over time?	✓		✓	✓	✓			
What are the contaminant attenuation rates in wells?	✓			✓				
How do contaminant concentrations change with distance from the source area?	✓	✓	✓	✓			✓	✓
Is the sampling frequency appropriate (temporal optimization)?	✓		✓		✓			
Is the spatial coverage of the monitoring network appropriate (spatial optimization)?	✓				✓			✓
What is the extent of the contamination?	✓	✓	✓					✓
What is the footprint and depths for the treatment zone?	✓	✓						✓
What is the mass of contamination?	✓	✓						✓
What is the site-wide exposure concentration?	✓	✓				✓		✓

Box Plot. A box plot summarizes the data via percentiles: the 25th percentile (bottom of box), the 50th percentile or median (middle of box) and the 75th percentile (top of box) (see Figure 1). The length of the box represents the interquartile range (IQR) and the lines (or whiskers) that extend from the box most commonly represent 1.5 times the IQR. Any data that plot outside whiskers are considered potential outliers. The box plot gives an idea of the distribution of the data set, specifically, the range (minimum and maximum values), the variation/spread (height of box), the symmetry (sizes of box halves and whiskers), and the skewness (relative size of the box halves). Hence, they are often used to assess whether the data are normally distributed, to identify outliers, and to compare the distributions of two or more data sets. For example, plotting side-by-side box plots can be used to compare the distributions of two or more data sets, i.e., upgradient versus downgradient wells.

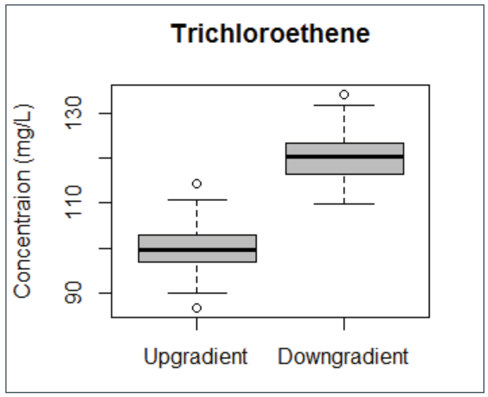

Figure 1. Side-by-Side Box Plots ***(Source: Geosyntec)***

Histogram. The histogram is a bar chart that is also used to visually inspect the data distribution (see Figure 2). The data are sorted into a series of equally sized intervals, known as bins, and plotted along the x-axis, and the number of observations that occur within each bin are plotted along the y-axis. Like the box plot, the histogram provides a method to determine the shape, spread, symmetry, and skewness of the data. Histograms can also be used to determine whether one or more populations exist in a data set or to compare the distributions between two or more data sets, for example, site versus background concentrations.

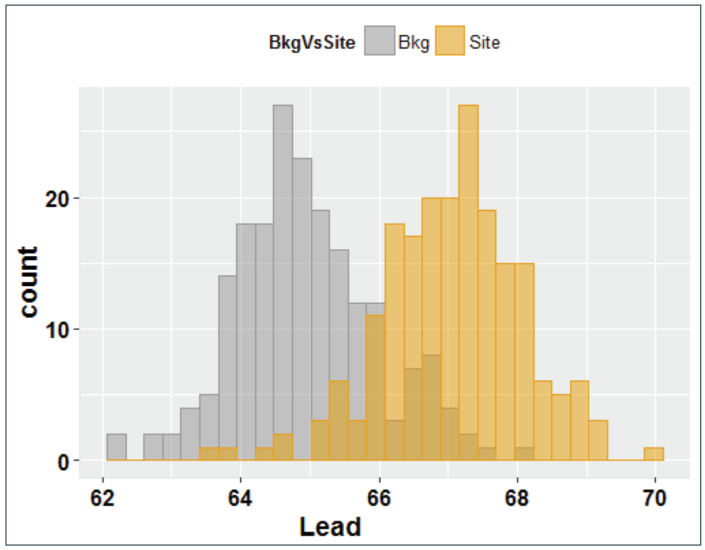

Figure 2. Side-by-Side Histograms ***(Source: Geosyntec)***

Quantile-Quantile (QQ) Plots. A QQ plot compares the percentiles of a data set to the percentiles of a theoretical probability distribution, e.g., normal distribution (see Figure 3). For example, a normal QQ plot has the normal percentiles on the x-axis and the concentration data percentiles on the y-axis. If the points follow a strong linear pattern, then that suggests the data follow a normal distribution. The observations that are separated from the bulk of the data may represent potential outliers needing further investigation. Also, significant and obvious jumps and breaks in a normal QQ plot can be indications of the presence of more than one population (e.g., background and contaminated populations) and/or data gaps due to lack of sufficient data (data sets of smaller sample sizes).

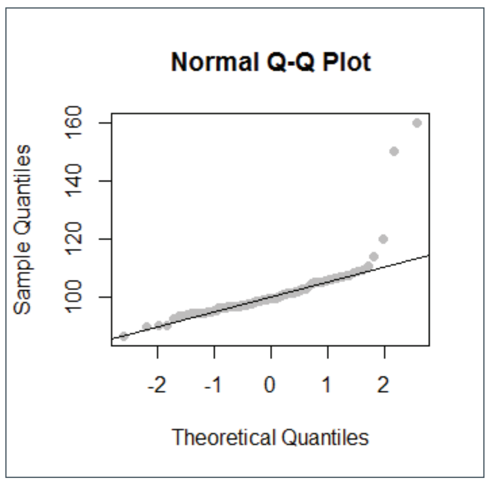

Figure 3. Standard Normal QQ Plot ***(Source: Geosyntec)***

Regression Analysis

Regression analysis is a parametric method that attempts to determine the strength of the relationship between two or more variables. The most commonly used method is simple linear regression (SLR), which determines whether there is a linear relationship between two variables, for example whether there is an increasing or decreasing trend in contaminant concentrations over time. As a test for trend, concentration data are plotted against time and a line is fit to the data. If the slope of the line is statistically different than zero, then there is sufficient evidence to suggest that a trend exists. If the slope is negative, there is a decreasing trend, and if the slope is positive, there is an increasing trend.

The slope also quantifies the magnitude of the trend, i.e., how the mean concentration changes over time. In the special case where the log concentration is plotted against time, also known as a log-linear regression or first-order decay rate model, then the magnitude of the slope is an estimate of the bulk attenuation rate, which can be used to estimate, cleanup timeframes (see USEPA, 2002 for more information on estimating bulk attenuation rates). Other applications of SLR include evaluating behavior in concentrations with increasing distance from a source and assessing the correlation between two contaminants (i.e., two metals in determining background). Other regression models are generally necessary to detect other non-linear relationships, such as cyclical trends.

Regression methods require the assignment of a value (e.g., RL, half the RL) for NDs; therefore, there should be few if any NDs when calculating a linear regression. Significant increasing or decreasing trends should be based on data sets with mostly detected measurements (>85%), otherwise a trend may be an artifact induced by changes in RLs over time. For data sets with <85% detected concentrations, it is recommended to use a nonparametric trend test, such as the Mann-Kendall test. Regression methods are also sensitive to violations of model assumptions (i.e., independent and normally distributed residuals with constant variance) and to the presence of outliers.

Mann-Kendall and Theil-Sen Line Tests

The Mann-Kendall (MK) and Theil-Sen (TS) tests are the nonparametric alternatives to SLR for detecting and quantifying trends over time. MK is a rank-based test used to detect monotonic, i.e., consistently increasing or decreasing over time, trends in contaminant concentrations over time. MK does not require data to conform to any specific probability distribution and concentrations do not need to be measured at equaled-spaced time intervals.

MK only detects the direction of a trend; therefore, TS must be used to estimate the magnitude of a trend. Generalizations of the MK tests are also available to accommodate seasonality in time series and to evaluate regional trends, i.e., simultaneous evaluation across several wells in a region (Gilbert, 1987).

The TS line is an alternative to SLR that can be used to estimate the slope of the trend line. However, the TS line models how the median concentration changes linearly with time, instead of the mean concentration. A test is conducted on the estimate of the TS line to determine if it is statistically different than zero.

Unlike SLR, results from the MK and TS tests are not impacted by the magnitude of extreme values. Both methods can be used with data sets containing NDs (i.e., <50%), but the test is difficult to apply to data sets containing multiple RLs; therefore, NDs should be substituted with a value smaller than the lowest detected concentration, e.g., half of the minimum RL.

Hypothesis Testing

Statistical hypothesis testing is a method of statistical inference, where two complementary hypotheses, the null and alternative hypothesis, are used to answer a study question about a population parameter. The null hypothesis is a baseline condition presumed to be true in the absence of strong evidence to the contrary, and the alternative hypothesis is the opposite condition that bears the burden of proof. Both parametric and non-parametric versions of these tests are available. The following are the three main types of hypothesis testing approaches:

One-Sample Hypotheses Tests. One-sample tests are used to compare the data to a fixed criterion, for example, the mean concentration of a contaminant in a compliance well to the MCL. These tests include: Student's t-test, Sign test, Wilcoxon Signed Rank test, and the Proportion test concern.

Two-Sample Hypotheses Tests. Two-sample tests are used to compare the means/medians of two populations, such as site versus background, surface versus subsurface soils, and upgradient versus downgradient wells. The two-sample hypotheses tests include: Student's t-test, Wilcoxon Signed Rank test, Gehan test, and Tarone-Ware test.

Multiple Populations. Other hypothesis tests are used to compare the means/medians of three or more populations, such as the spatial variability among multiple upgradient background wells. Tests to compare multiple populations include the one-way ANOVA and the Kruskal-Wallis test.

Statistical Limits

Statistical limits, also known as statistical intervals, are used to quantify the uncertainty in the estimates of population parameters, e.g., mean concentration. These limits represent the upper and lower boundaries of the estimated parameter. There are three main types of statistical limits: confidence limits, tolerance limits, and prediction limits. Both parametric and non-parametric methods are available to compute all of them. The presence of outliers will result in inflated upper limits; therefore, outliers should be excluded before the computation of these limits in order to avoid incorrect decisions. A confidence level must be pre-specified when constructing statistical limits. This level of confidence may be determined by federal or state regulatory requirements or guidance, or by project-specific needs, with the most common level set to 95%. These limits can be one-sided (i.e., just one of the upper or the lower) or two-sided (i.e., both upper and lower), but in most applications, interest lies in the upper limit, e.g., upper confidence limit.

Confidence Limits. Confidence limits are typically constructed on a mean or median and are used to determine if concentrations are above or below a criterion, to estimate exposure point concentrations, and to calculate uncertainty associated with a slope (i.e., TS line or SLR line). It should be noted that only mean concentrations can be compared to these limits, i.e., not individual measurements.

Tolerance Limits. Tolerance limits are typically constructed on an upper percentile, e.g., 95th percentile, of a background data set to estimate a background threshold value. They can also be used to establish an alternate compliance limit. Individual site measurements can be compared to this limit.

Prediction Limits. Prediction limits are constructed on a mean, or the ranking of measurements in a background data set and are also used to estimate a background threshold value. The difference between a prediction limit and the other limits is that it explicitly accounts for the comparison to a pre-specified number of future observations. Therefore, it is wider than a confidence interval, but individual site measurements can be compared to this limit.

Geospatial Methods

Geospatial methods are often used in combination with traditional statistical methods to address data that may be biased, clustered, or spatially correlated. More specifically, they are used to examine the spatial relationships between sample locations and provide a means of estimating (or interpolating) values at unsampled locations. For contaminated soil and groundwater sites, these methods are used for a variety of purposes: interpolation to create contour maps, hot spot detection, estimation of quantities (e.g., average site or background concentrations, mass/volumes, etc.), defining treatment zone footprint and depths, plume attenuation over time, monitoring program optimization, remedy performance, and verifying attainment of cleanup goals. Though these methods have a variety of applications, it is important that they be checked to ensure that the results are appropriate for, and consistent with, the conceptual site model.

Geospatial methods range from simple methods, such as inverse distance weighting (IDW) and Thiessen polygons, to more advanced geostatistical methods, such as kriging. The simple methods use a deterministic model, i.e., does not include a statistical error component, and does not require any assumptions about the data other than spatial correlation exists. In contrast, kriging is a stochastic model, i.e., requires estimation of a statistical error component, and requires that a set of statistical assumptions are met. Regardless of the method, the quality of the model fit to the data should be evaluated using cross-validation or validation techniques and other goodness-of-fit metrics.

A brief overview of a few of these methods is provided below. For more information, see ITRC's web-based guidance document Geospatial Analysis for Optimization at Environmental Sites (ITRC, 2016), which provides a series of fact sheets on geospatial methods, a detailed overview of these methods, and various case studies demonstrating the applications of these methods.

Inverse Distance Weighted. IDW predicts a value at an unsampled location by calculating the distance-weighted average of neighboring data points within a specified window. IDW can be used for interpolation to create contour maps and grids used in surface and volume calculations. This method performs best with larger data sets collected at a high spatial density. IDW can produce contours with bullseye's or mounds in the surface, so it is not well suited to contour groundwater elevations.

Thiessen Polygons. Thiessen polygons define the area of influence around each sample point. The area of each polygon and its corresponding data point is used to calculate an area-weighted average value. Thiessen polygons are often used to calculate exposure point concentrations for exposure areas that have clustered and unevenly spaced sampling points. However, sample points in sparsely sampled regions will have a larger area of influence, which could lead to misleading estimates if that is not a practical assumption.

Kriging. Kriging is a stochastic geostatistical method similar to IDW. It differs in that it assumes the spatial relation between interpolated values are dependent on observed values, which is modeled using a semivariogram. It also provides an estimate of the uncertainty in the interpolated values. Hence, both the interpolated values and their associated uncertainties can be mapped and used to evaluate the spatial distribution of contaminants. For example, these two maps can be used to delineate areas of contamination and to decide if further sampling is needed, for example, in areas of low sampling density. These maps can also be used to determine whether site-wide concentrations are above or below a comparison value, e.g., regulatory limit or background values.

What are Some Examples of Emerging Statistical Methods?

Monte Carlo

Monte Carlo simulation is a statistical tool for analyzing the variability and uncertainty associated with estimates from a deterministic model, such as exposure risks and dense non-aqueous phase liquid (DNAPL) mass estimates (see Figure 4). Monte Carlo methods allow for key input parameters to vary according to a known probability distribution, e.g., normal, lognormal, etc., as opposed to a single fixed value. A large number of realizations (hundreds or thousands) are generated from these probability distributions and the model is solved for each realization. The end result is a probability distribution of possible model outputs, as opposed to a single point estimate, that can be graphed and used to assess the mean and spread in model outputs and the probability of their occurrence.

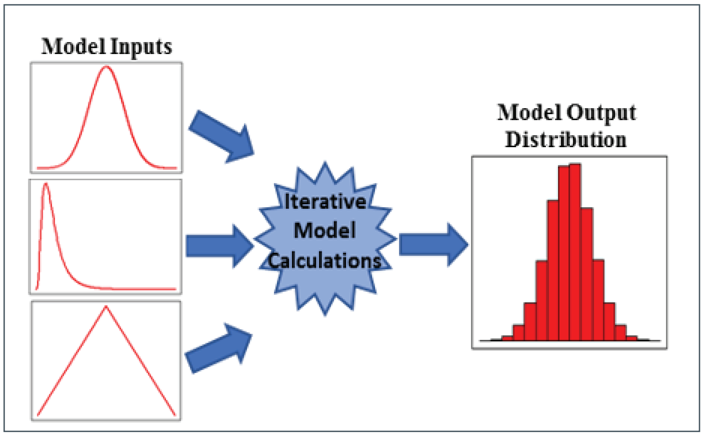

Figure 4. Schematic of Monte Carlo Simulation *(Source: Geosyntec)*

Multivariate Methods

Groundwater and soil concentration data are inherently correlated; therefore, multivariate methods allow for the analysis of all variables simultaneously. By using these methods, these complex data sets can be explored for patterns and relationships that could be missed using the traditional univariate methods, i.e., variable by variable analysis, presented in the previous section. Two of the most common multivariate methods are described below.

Hierarchical Cluster Analysis (HCA). HCA can identify common groups (i.e., clusters) within a large data set consisting of multiple contaminants across many samples. Samples are successively linked together in a dendrogram (hierarchal tree diagram, see Figure 5) based on increasing dissimilarity in contaminant distributions. This hierarchical method does not require the number of clusters to be specified beforehand, unlike other clustering methods, such as k-means. It is particularly useful when there is uncertainty whether there are patterns in the data that suggest a grouping structure. For example, clustering can be used to distinguish background data from site data based on the similarity of their chemical compositional patterns (i.e., patterns associated with background versus site-related releases).

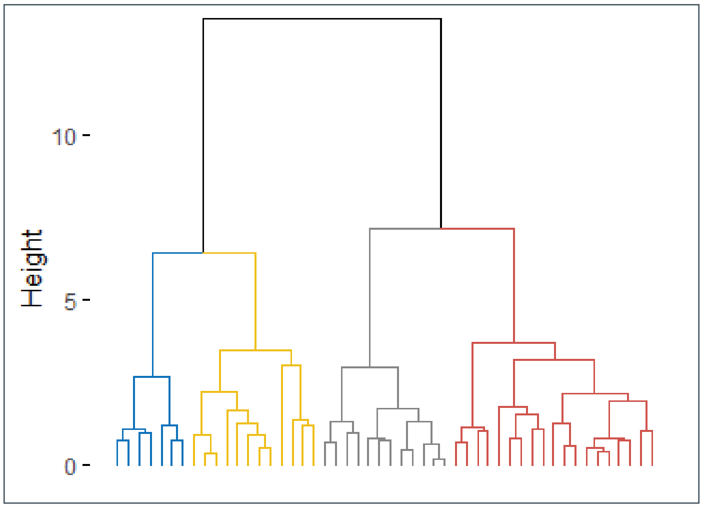

Figure 5. Example HCA Dendrogram *(Source: Geosyntec)*

Principal Component Analysis (PCA). PCA is used to summarize variation in a data set that consists of multiple correlated variables. PCA is also known as a data reduction technique because it reduces the number of variables by transforming them into smaller sets of uncorrelated variables without losing the most important information. These new variables correspond to a linear combination of the original variables and are called the principal components. The principal components can then be interpreted to better understand the processes or mechanisms driving the chemical compositions. For example, PCA can be applied to groundwater data sets to inform natural attenuation potential, degradation processes, and remediation strategies.

Case Study 1: Groundwater Statistics

The effectiveness of monitored natural attenuation (MNA) integrated with a large-scale pump-and-treat program as a remedial strategy was evaluated for multiple constituents of concern (COCs) at a site in Australia. The effectiveness of MNA was evaluated using statistical techniques to assess the natural attenuation behavior and quantify the dissolved phase degradation rate using data collected over a 12-year period between 2005 and 2017. Statistical analyses were conducted to evaluate the temporal and spatial trends (both lateral and vertical) of COC concentrations at the individual wells, as well as the dissolved phase mass in the plume by hydrostratigraphic layer (four in total). Temporal trends were evaluated using linear regression (assuming a first-order decay rate model), while spatial trends in COC concentrations and dissolved phase mass were evaluated using geostatistical methods (kriging).

For each well and COC, time series plots were examined to identify increasing and decreasing trends, anomalous and outlying data, non-detect frequencies, changes in detection limits over time, and appropriate regression model (e.g., zero- versus first-order decay rate). During this process it was noted that in one area of the plume, a high-concentration slug of groundwater was migrating downgradient from the source area over time, which was reflected in the time trends as an initial increase followed by a decrease once the center of the slug passed each location. For these data sets, the assessment of the natural attenuation behavior was truncated to focus only on the timeframe during which attenuation could be clearly observed and was not over-shadowed by plume advection.

For data sets that had statistically significant decreasing trends and met the model assumptions, attenuation rates and the half-lives were estimated using a first-order decay rate model. The estimated half-lives were then plotted on a map to evaluate the spatial distribution for each vertical layer of the model. Figure 6 shows the half-lives for ethylene dichloride (EDC). Plotting the spatial distribution of EDC half-lives provides insight into areas of the site where natural attenuation is effective at reducing mass (i.e., areas with a predominance of black [non-detect], blue or green symbols on Figure 6) as well as identifying areas where natural attenuation is less effective at reducing mass (i.e., areas with a predominance of red [no bulk attenuation or NBA], orange or purple symbols). Significant differences in attenuation were identified not only in different areas of the plumes (see red and black circled areas on Figure 6), but also vertically. Additionally, natural attenuation in the shallower intervals is progressing much faster than some

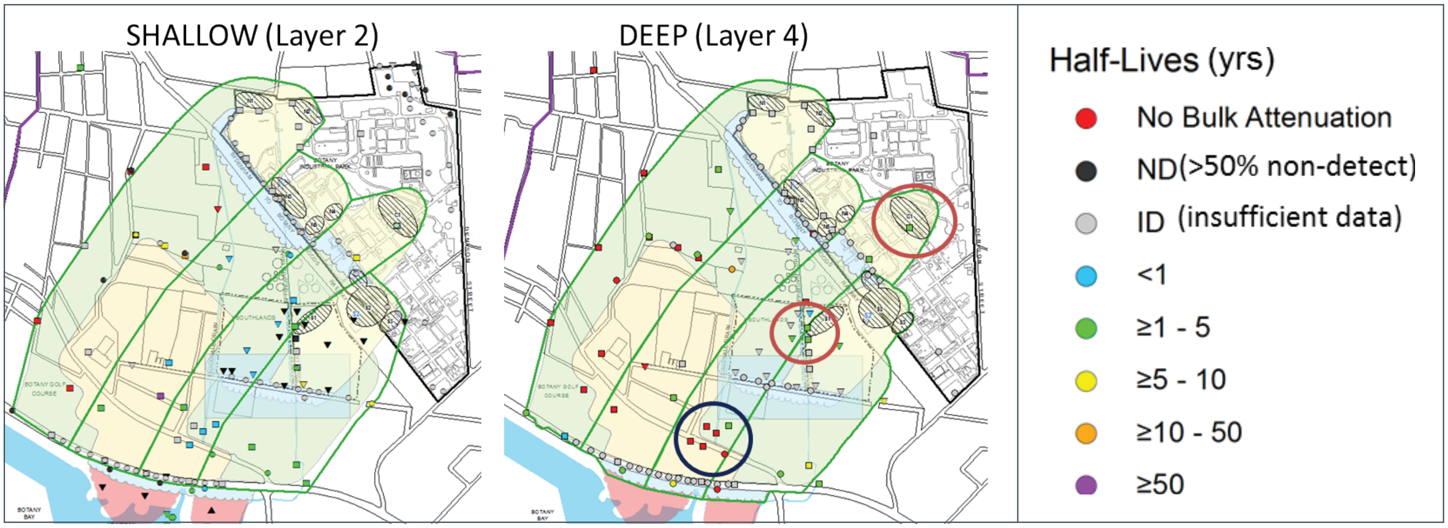

Figure 6. Spatial Distribution of EDC Half-Lives by Layer ***(Source: Geosyntec)***

Note: Areas of the plumes with a predominance of orange, purple or red (no bulk attenuation; NBA) symbols indicate areas/depths where persistence of high concentrations and slow attenuation is observed under natural conditions, even with ongoing pump-and-treat. Areas with a predominance of black (non-detect), blue or green symbols indicate areas/depths where natural attenuation combined with pump-and-treat is contributing significantly to reductions in plume concentrations/mass.

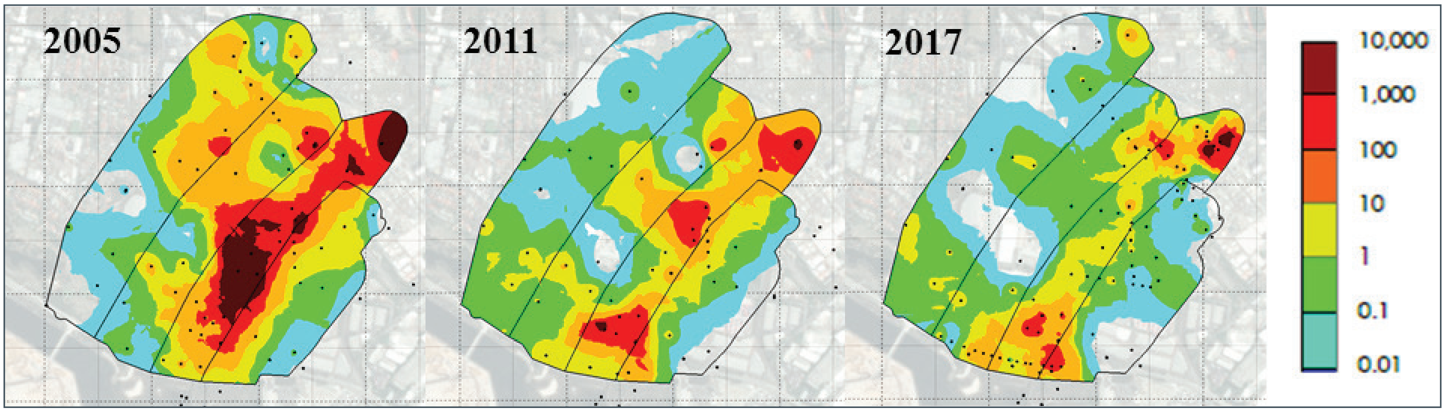

Figure 7. Temporal Changes in Maximum EDC Concentrations ***(Source: Geosyntec)***

areas of the deeper layers, and the deeper interval of the toe of the plume has had little attenuation whereas attenuation has been observed even in upgradient DNAPL source areas (indicated as hatched areas on Figure 6).

COC concentrations across the site were interpolated using a three-dimensional (3D) kriging model. Maximum interpolated concentrations across all depths were projected into plan view maps for every two-year time point between 2005 and 2017 to evaluate the overall changes in the spatial distribution of COC concentrations over time (i.e., before and during remedy operation). Figure 7 shows the EDC trend maps at three time points. These maps illustrate the disappearance of EDC from the majority of the plume over time, and also highlight where concentrations are persisting in the source area and the toe of the plume. The 3D models were also used to estimate the dissolved phase mass in the plume over time. First-order decay rate models were used to calculate plume attenuation rates and quantify source decay behavior, which provided insight into overall plume lifespans and potential MNA timeframes.

In summary, this case study highlights how multiple statistical methods can be incorporated into remedial strategy development and decision making. Understanding the spatial and temporal changes in attenuation behavior provides quantitative data to input into a cost-benefit analysis tool, where different remedial strategies can be evaluated, including continuing MNA, incorporating active remediation into areas of slower attenuation behavior to reduce plume lifespans, or active treatment over the entire plume. Each of these options requires remedial and monitoring durations to be estimated for generation of costs, which could be quantified through this statistical evaluation.

Case Study 2: Soil Statistics

The mass estimates of a primary COC in soils were evaluated for a confidential site. The potential uncertainty of mass estimates was evaluated using statistical methods to better understand the pre-remediation conditions and to provide a reference point that could be used to track remedial progress. The controlling factors for estimating COC mass in soils included the quantity of DNAPL mass and its spatial distribution. The distribution and range of parameters that influenced the COC mass estimates were evaluated using Monte Carlo methods, while the spatial distribution of soils representing DNAPL mass was estimated using spatial IDW.

The presence of DNAPL was indicated by soil concentrations exceeding the DNAPL partitioning threshold, which was calculated based on chemical properties of the COC and the soil properties at the site, which are inherently uncertain and variable. Two parameters were selected to evaluate the uncertainty in the DNAPL partitioning threshold, which were assumed to follow a triangular distribution and have ranges based on either site-specific values or published ranges for geological units observed at the site. A total of 1,000 realizations were generated from each distribution, and each one was used to calculate the DNAPL partitioning threshold concentration. The resulting distribution of 1,000 estimates of the DNAPL partitioning threshold concentration was then used to evaluate the uncertainty in the spatial distribution of the DNAPL mass. COC concentrations across the study area were interpolated using a 3D IDW model. Several model options were evaluated by adjusting the key interpolation input parameters. These model options were evaluated using cross-validation methods to evaluate model goodness-of-fit and prediction performance.

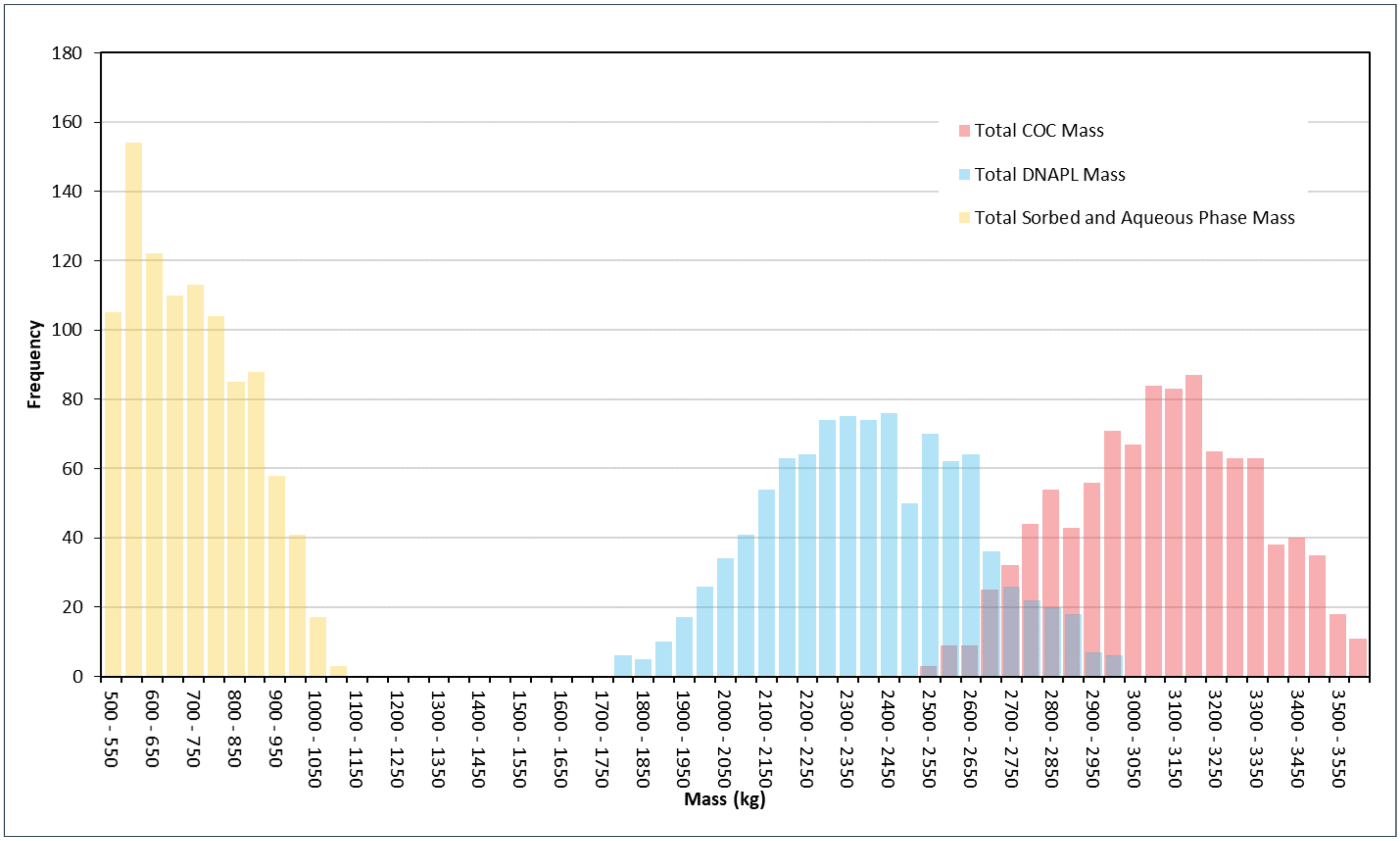

***Figure 8. Distributions of Total DNAPL, Total Sorbed and Aqueous Phase, and Total COC Mass** (Source: Geosyntec)*

The final step of the analysis was to use the 3D interpolations and the distribution of DNAPL partitioning threshold concentrations to evaluate the uncertainty of the total COC mass in the soil. For each of the 1,000 realizations, the total interpolated soil volumes greater than and less than or equal to the DNAPL partitioning threshold concentration value were calculated. The soil volume greater than the DNAPL partition threshold concentration was used to calculate the total DNAPL mass, while the soil volume less than or equal to the DNAPL partition threshold concentration was used to calculate the total sorbed and aqueous phase mass. Figure 8 shows the resulting distributions of the DNAPL mass, sorbed and aqueous phase mass, and total COC mass. These distributions were then used to understand the variability and uncertainty in the mass estimates and to define probable ranges in the mass estimates (mean and upper and lower bounds).

In summary, this case study highlights how statistical methods (Monte Carlo) and geospatial methods (IDW) can be used to quantify and evaluate the uncertainty of the contaminant mass in soils. Understanding the uncertainty in the mass estimates allows decision makers to effectively delineate the extent of contamination while reducing the cost for additional sampling. Future mass estimates will be derived using the same statistical approach during the remediation and post-remediation stages to track remedial progress towards meeting the remedial goal.

Resources

Gilbert, R.O. 1987. Statistical Methods for Environmental Pollution Monitoring. New York: John Wiley & Sons, Inc.

Interstate Technology and Regulatory Council (ITRC). 2013. "Groundwater Statistics and Monitoring Compliance: Statistical Tools for the Project Life Cycle." Washington, DC: ITRC Groundwater Statistics and Monitoring Compliance Team.

ITRC. 2016. Geospatial Analysis for Optimization at Environmental Sites Website (GRO-1). https://gro-1.itrcweb.org/

U.S. Army Corps of Engineers. 2008. "Environmental Statistics". No. 200-1-16. Washington, DC.

United States Environmental Protection Agency (USEPA). 1989. "Methods for Evaluating the Attainment of Cleanup Standards, Vol. 1 Soils and Solid Media." EPA 230/02-89-042 United States Environmental Protection Agency.

USEPA. 2005. "Uniform Federal Policy for Quality Assurance Project Plans." EPA-505-B-04-900A/ DTIC ADA 427785. Washington D.C.: United States Environmental Protection Agency. https://www.epa.gov/sites/production/files/documents/ufp_qapp_v1_0305.pdf.

USEPA. 2006a. "Guidance on Systematic Planning Using the Data Quality Objectives Process." EPA QA/G-4. Washington D.C.: United States Environmental Protection Agency. http://www2.epa.gov/quality/guidance-systematic-planning-using-data-quality-objectives-process-epa-qag-4.

USEPA. 2006b. "Data Quality Assessment: A Reviewer's Guide." EPA QA/G-9R. Washington D.C.: United States Environmental Protection Agency.

USEPA. 2009. "Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities." Unified Guidance EPA 530/R-09-007. Washington DC: United States Environmental Protection Agency.

USEPA. 2015. "ProUCL Version 5.1.00 Technical Guide." EPA/600/R-07/041. Washington, DC: United States Environmental Protection Agency. https://www.epa.gov/land-research/proucl-version-5100-documentation-downloads

Naval Facilities Engineering Command (NAVFAC). 2002. *Guidance for Environmental Background Analysis, Volume I: Soil*. UG-2049-ENV.

NAVFAC. 2004. *Guidance for Environmental Background Analysis, Volume III: Groundwater.* UG-2059-ENV.

NAVFAC. 2010. DON *Guidance for Planning and Optimizing Monitoring Strategies.* UG-2081-ENV.

NAVFAC. 2012. *Guidance for Optimizing Remedial Action Operation (RA-O).* UG-NAVFAC EXWC-EV-1301.

ESTS N39430-16-D-1802