NAVAL FACILITIES ENGINEERING COMMAND
Washington, DC 20374-5065

NFESC
User's Guide
UG-2049-ENV

GUIDANCE FOR ENVIRONMENTAL BACKGROUND ANALYSIS

Volume I: Soil

Prepared by

Battelle Memorial Institute
Earth Tech, Inc
NewFields, Inc.

April 2002

PREFACE

This guidance document supports and implements the September 2000 Navy background policy document, *Navy Interim Final Policy on the Use of Background Chemicals*. It is part of a series devoted to background analysis that provides instructions for characterizing background conditions at sites where past uses of the property have resulted in actual or suspected chemical releases. Each volume in this series is devoted to a particular medium. This volume focuses on analytical methods and procedures that can be used to identify background chemicals in the soil medium (whether from anthropogenic or natural sources), and estimate the background chemical concentration ranges that represent soil background conditions at the site.

For the sake of completeness, each volume in the series opens with a summary of the September 2000 Navy background policy, and discussions of both state and federal regulatory requirements and guidance. Each volume then describes data review and assessment procedures, explains the Geochemical and Comparative Methods of background analysis, and presents medium-specific case studies that illustrate application of the background analysis methods. Because each volume in this series is intended to serve as a stand-alone document, some identical or similar discussions occur across the volumes. The preface of each volume identifies the sections that are unique to that volume and the medium of concern.

The soil background analysis techniques presented in this volume focus primarily on naturally occurring metals. Methods and procedures designed specifically for background analysis of naturally occurring metals in soil are presented in Section 3 (Geochemical Method), and Section 5 (Case Studies). Appendix A (Geochemistry for Background Analysis) provides a summary of the geochemical principles that form the basis of the Geochemical Method.

This guidance document expands on the procedures established in the following documents:

- Procedural Guidance for Statistically Analyzing Environmental Background Data, San Diego: DON SWDIV and EFA West, 1998.
- Handbook for Statistical Analysis of Environmental Background Data, San Diego: DON SWDIV and EFA West, 1999.
- Site Management Plan Update for the Pearl Harbor Naval Complex, Pearl Harbor, Hawaii: DON PACDIV, 2001.

EXECUTIVE SUMMARY

This guidance document provides instructions for characterizing background conditions at sites where past uses of the property have resulted in actual or suspected chemical releases. Background analysis is necessary to identify background chemicals—those derived from natural or anthropogenic sources not related to activities conducted at the site—and to estimate the chemical concentration ranges that represent background conditions at the site.

According to the September 2000 Navy background policy document *Navy Interim Final Policy on the Use of Background Chemicals*, cleanup efforts at Navy sites should address only those risks associated with chemical concentrations that are elevated as a result of a site-related release. Cleanup efforts therefore must address only chemicals that have been released at the site—not background chemicals. At some sites, unacceptable risks may be associated with chemical concentrations within the background range. These risks are outside the scope of the Navy's Environmental Restoration Program; however, Navy policy requires restoration program personnel to convey information regarding all identified risks to stakeholders. The Navy background policy was developed to ensure compliance with federal and state laws and regulations, and is consistent with U.S. Environmental Protection Agency (U.S. EPA) technical guidance.

The background analysis techniques presented in this document focus on naturally occurring metals, and are based on well-established statistical methods and geochemical relationships. The data analysis and statistical testing methods closely follow U.S. EPA's *Guidance for the Data Quality Objectives Process* and *Guidance for Data Quality Assessment: Practical Methods for Data Analysis.*

Background analysis is an integral component of most environmental investigations. Existing site data that may be relevant to background should be reviewed and assessed during the initial phase of an investigation. The operational history, physical characteristics, and chemical characteristics of the site should be evaluated in order to (1) determine the conditions that may affect the distribution of both contaminants and background chemicals, and (2) develop a list of suspected COPCs (target chemicals). If additional or supplementary data are required, a sampling and analysis plan should be developed. The sampling and analysis program should be carefully designed and implemented to ensure that all data needed to evaluate background conditions are collected. After the additional data have been collected and incorporated into the dataset, the detected chemical concentrations should be screened against appropriate risk-based criteria to determine which chemicals should be carried forward for background analysis.

Step-by-step instructions for background analysis techniques or methods are presented. For many sites and target chemicals, background concentration ranges can be estimated by the spatial analysis and probability plotting techniques presented in Exploratory Data Analysis (Section 2.2). If this initial analysis does not produce a technically defensible and reliable estimate of the background concentration range for a particular target chemical, the project team should identify appropriate methods for further background analysis (Section 2.3). Detailed instructions for analysis using the Geochemical Method (Section 3) and the Comparative Method (Section 4) are then presented. Case studies (Section 5) illustrate the application of each method.

The Geochemical Method uses techniques based on geochemical principles to estimate background metal concentration ranges. Chemical concentrations in soil are controlled by the chemical composition of the parent rocks and the geochemical processes that occur during soil formation; therefore, in natural soils, certain metals tend to occur together, and metal/metal concentration ratios often are restricted to relatively narrow ranges. These natural geochemical relationships can be characterized and used to estimate the chemical concentration ranges that represent background conditions. The Geochemical Method requires only data from the suspected contaminated site to determine background concentration ranges. Therefore, the method is particularly useful when it is not possible to identify and collect data from a background (or reference) area (as required for the Comparative Method).

The Comparative Method is based on comparison of chemical concentration data from the investigation site to data from a reference area. Soil at the reference area (usually but not always an off-site area) must be physically and geochemically similar to native soil at the investigation site. Background concentration ranges are estimated by applying various statistical tests to both datasets to determine whether concentrations of the target chemicals at the investigation site are statistically similar to reference area concentrations.

CONTENTS

FIGURES

TABLES

BOXES

ACRONYMS AND ABBREVIATIONS

AF	absolute frequency
AOC	area of concern
APHA	American Public Health Association
ARAR	applicable or relevant and appropriate requirement
ASCE	American Society of Civil Engineers
ASTM	American Society for Testing and Materials
ATSDR	Agency for Toxic Substances and Disease Registry
bgs	below ground surface
BRAC	Base Realignment and Closure (Act)
Cal/EPA	California Environmental Protection Agency
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
CF	cumulative frequency
CNO	Chief of Naval Operations
COPC	chemical of potential concern
CV	coefficient of variation
DDT	dichlorodiphenyltrichloroethane
DL	detection limit
DoD	United States Department of Defense
DON	Department of the Navy
DQA	data quality assessment
DQO	data quality objective
DTSC	Department of Toxic Substances Control (of California)
EFA	Engineering Field Activity
EFD	Engineering Field Division
FDEP	Florida Department of Environmental Protection
FDER	Florida Department of Environmental Regulation
GIS	Geographic Information System
H_a	alternative hypothesis
H_o	null hypothesis
IR	Installation Restoration
J	estimated concentration
MDEQ	Michigan Department of Environmental Quality
MERA	Michigan Environmental Response Act
mg/kg	milligram(s) per kilogram
MSSL	media-specific screening level

NCP	National Oil and Hazardous Substances Pollution Contingency Plan
NFA	no further action
NPL	National Priorities List
NQ	not qualified
NRC	Nuclear Regulatory Commission
OERR	Office of Emergency and Remedial Response (of U.S. EPA)
OSWER	Office of Solid Waste and Emergency Response (of U.S. EPA)
OU	Operable Unit
NRCS	Natural Resources Conservation Service
PACDIV	Pacific Division Naval Facilities Engineering Command
PAH	polycyclic aromatic hydrocarbon
PCB	polychlorinated biphenyl
POI	point of interest
ppb	parts per billion
ppm	parts per million
PRG	preliminary remediation goal
PZC	point of zero charge
QA	quality assurance
QAPP	quality assurance project plan
QC	quality control
RAGS	Risk Assessment Guidance for Superfund
RCRA	Resource Conservation and Recovery Act
RI	Remedial Investigation
SAP	Sampling and Analysis Plan
SARA	Superfund Amendments and Reauthorization Act
SVOC	semivolatile organic compound
SWDIV	Southwest Division Naval Facilities Engineering Command
SWMU	solid waste management unit
TOC	total organic carbon
TPH	total petroleum hydrocarbons
U	nondetect
UCL	upper confidence limit
UJ	nondetect estimated
USC	United States Code
USCS	Unified Soil Classification System
USDA	United States Department of Agriculture
U.S. EPA	United States Environmental Protection Agency
USGS	United States Geological Survey
UST	underground storage tank
VSP	Visual Sample Plan
VOC	volatile organic compound
WRS	Wilcoxon Rank Sum test

GLOSSARY

α	Alpha is the tolerated probability of a Type I error in a hypothesis test.
β	Beta is the tolerated probability of a Type II error in a hypothesis test.
ε	Epsilon is the proportion of a site at which chemicals are present at concentrations greater than background levels.
Alternative Hypothesis (H_a)	The hypothesis that is accepted if the null hypothesis is rejected.
Anthropogenic Background	Chemicals present in the environment due to human activities that are not related to specific point sources or site releases.
Background Chemicals	Chemicals present in the environment due to naturally occurring geochemical processes and sources, or to human activities not related to specific point sources or site releases.
Basalt	A dark, fine-grained, extrusive (volcanic) igneous rock with a low silica content (40 to 50%), but rich in iron, magnesium, and calcium. Generally occurs in lava flows, but also as dikes. Basalt makes up most of the ocean floor and is the most abundant volcanic rock in the Earth's crust.
Box and Whisker Plot	A graphic way of summarizing a set of data measured on an interval scale. Often used in exploratory data analysis, a box and whisker plot is a type of graph that shows the shape of the distribution, its central value, and variability. The picture produced consists of the most extreme values in the dataset (maximum and minimum values), the lower and upper quartiles, the median, and the mean.
Censored Dataset	A dataset that contains one or more nondetects.
Clay Minerals	Finely crystalline, hydrous silicates formed from weathering of such silicate minerals as feldspar, pyroxene, and amphibole. Most common clay minerals belong to kaolinite, montmorillonite, and illite groups.
Comparative Method	Application of statistical two-sample tests for identifying COPCs. In this method, results collected from a nearby uncontaminated or "background" area are statistically compared to results from samples collected at a site suspected to be contaminated.
Confidence Interval	The interval within which an unknown population parameter, such as the mean or the expected value of a predicted value, falls with a given probability. The estimated interval is calculated from a given set of sample data.
Confidence Limits	The lower and upper boundaries of a confidence interval for a given probability (i.e., the values that define the range of a confidence interval). For example, the upper and lower bounds of a 95% confidence interval are the 95% confidence limits.

Correlation Coefficient (*r*)	A number between –1 and 1 that measures the degree to which two variables are linearly related. If there is a perfect linear relationship with positive slope between the two variables, the correlation coefficient is 1; if there is positive correlation, whenever one variable has a high (low) value, so does the other. If there is a perfect linear relationship with negative slope between the two variables, the correlation coefficient is –1; if there is negative correlation, whenever one variable has a high (low) value, the other has a low (high) value. A correlation coefficient of zero means that there is no linear relationship between the variables.
Correlation Matrix	A matrix that represents the correlation coefficient between paired measurements of different variables.
Data Quality Assessment (DQA) Process	The scientific and statistical evaluation of data to determine if data obtained from environmental data operations are of the right type, quality, and quantity to support their intended use.
Data Quality Objective (DQO) Process	A series of planning steps based on the scientific method which are designed to ensure that the type, quantity, and quality of environmental data used in decision-making are appropriate for the intended application.
Degree of Freedom (*f*)	Describes the number of values in the final calculation of a statistic that are free to vary.
Distribution	The frequency (either relative or absolute) with which measurements in a dataset fall within specified classes. A graphic display of a distribution is referred to as a *histogram*.
Enrichment Factor Analysis	A study of metal enrichments and potential metal-contaminated concentrations in soils. The metals are compared to concentrations of naturally occurring metals found in source rocks such as igneous, sedimentary, and metamorphic formations. The enrichment factor (E_i) is defined as the ratio of the concentration of element *i* (contaminated) to a naturally occurring metal in the soil sample divided by this ratio in source rocks.
Exploratory Data Analysis	A statistical and graphic procedure for examining data in order to describe the data's main distributional features.
Factor Analysis	A multivariate statistical method of computing some underlying factors representing variability of measured data. Factor analysis reduces a large number of correlated variables to a smaller, more manageable number of factors. A factor is a linear combination of related variables that can replace those variables in future analysis. The numerical characteristics of computed factors are attributed to the underlying correlations among investigated variables.
Geochemical Method	Application of geochemical and statistical methods to extract background data from site data.
Geostatistics	A class of techniques for the analysis of spatially correlated data. In these analyses, variograms or related techniques are used to quantify and model the spatial correlation structure. Also includes various estimation techniques, such as kriging, that utilize spatial correlation models.

Granite	A coarse-grained intrusive (plutonic) igneous rock with high (≥65%) silica (SiO_2) content typical of continental regions. Quartz, plagioclase feldspar, and potassium feldspar make up most of the rock and give it a fairly light color. Granite has more potassium feldspar than plagioclase feldspar. Usually contains biotite, and also may contain hornblende.
Histogram	A method of displaying graphically the characteristics of a distribution of items in a given population or sample. In a histogram, each measure usually is represented by a single block that is placed over the midpoint of the class interval into which the measure falls.
Hypothesis	An assumption about a property or characteristic of a set of data under study. The goal of statistical inference is to decide which of two complementary hypotheses is likely to be true. The null hypothesis (H_o) describes what is assumed to be the true state of nature; the alternative hypothesis (H_a) describes the complementary situation.
Igneous Rock	Rock formed from the cooling and solidification of magma, and that has not been changed appreciably by weathering since its formation.
Inflection Point	A point on a curve where the direction of the curvature changes. For probability plots in background analysis, an inflection point that marks a distinct increase in slope typically represents the upper bound of the background concentration range. (As noted in Section 2, probability plot inflection point analysis alone is not sufficient to conclusively define the background range.)
Inter-Quartile Range (IQR)	A measure of the spread of or dispersion within a dataset. The IQR is difference between the 25th and 75th percentiles of the measured values of the sample. IQR is not affected by outliers.
Kriging	In geostatistics, a weighted-moving-average interpolation method in which the set of weights assigned to samples minimizes the estimation variance.
Lognormal Distribution	A family of positive-valued, skewed distributions commonly used in environmental work.
Mean	A measure of the central tendency of a distribution, which is computed by averaging (totaling the various individual results and dividing by the number of results involved) a dataset.
Median	A measure of the central tendency of a distribution, which is obtained by ranking the individual results in a dataset from smallest to largest and selecting the middle value. For an even number of results, the median is computed as the arithmetic average of two middle values.
Metamorphic Rock	Rock derived from preexisting rocks that have been altered physically, chemically, and/or mineralogically as a result of natural geological processes, principally heat and pressure, originating within the earth. The preexisting rocks may have been igneous, sedimentary, or another form of metamorphic rock.
Mode	A measure of the central tendency of a distribution, which is defined as the value in the population that occurs most frequently.

Naturally Occurring Background	Ambient concentrations of chemicals present in the environment that have not been influenced by human activity.
Nondetects	Measurements that are reported by the analytical laboratory to be below some quantitative upper limit, such as the detection limit or the quantitation limit.
Nonparametric Test	A statistical test that does not require any specific assumptions about the exact form of the underlying probability distributions of the investigated measures. Consequently, nonparametric tests are generally valid for a fairly broad class of distributions.
Normal (Gaussian) Distribution	A family of bell-shaped distributions described by the mean and variance.
Null Hypothesis (H_o)	The hypothesis that represents a theory that has been put forward, either because it is believed to be true or because it is to be used as a basis for argument, but has not been proved. The null hypothesis is assumed to be true, unless data and other evidence demonstrate otherwise with sufficient confidence.
Outlier	A measurement that is unusually large or small relative to others in the same dataset, and which therefore is suspected of misrepresenting the population from which it was collected.
Parametric Test	A test that requires specific assumptions about the exact form of the underlying probability distributions of the investigated measures. Parametric tests are not valid unless the underlying assumptions are met.
Partial Correlation	The correlation between two continuous variables that remains after the influence of one or more variables has been controlled or eliminated
Population	The entire collection of items that constitute the variable of interest.
Power	The probability that a test will reject the null hypothesis, when the alternative hypothesis is true. Power is defined to be $1 - \beta$.
Prediction Interval	The interval within which a new observation of the predicted variable falls with a given probability. Often plotted on scatter plots, this interval displays the limit around the fitted line in which a single new observation can be expected to fall with a given probability.
Probability Distribution	The probability distribution of a discrete random variable is a list of probabilities associated with each of its possible values. It sometimes is referred to as the probability function or the probability mass function.
Pth Percentile	The specific value of a distribution that divides the set of measurements in such a way that the *P* percent of the measurements fall below (or are equal to) this value, and 100 – P percent of the measurements exceed this value.
Pth Quantile	The specific value of a distribution that divides the set of measurements in such a way that the proportion, *P*, of the measurements fall below (or are equal to) this value, and the proportion 1 – P of the measurements exceed this value.

Random Sample	A set of items that have been drawn from a population in such a way that each time an item was selected, every item in the population had an equal opportunity to appear in the sample. In environmental field investigations, random samples imply data that are collected in an unbiased, uncorrelated, and nonclustered manner.
Range	In descriptive statistics, the difference between the highest and lowest measured value. In geostatistics, the separation distance between any pair of measured values beyond which the pair are uncorrelated.
Regression	A set of techniques to characterize the manner in which one of the measures changes as the other measure changes.
Scatter Plot	A plot of a set of bivariate (two variables) data. A scatter plot gives a visual picture of the relationship between the two variables and aids the interpretation of the correlation coefficient or regression model.
Sedimentary Rock	A rock formed from materials deposited from suspension or precipitated from solution and usually being more or less consolidated. The principal sedimentary rocks are sandstones, shales, limestones, and conglomerates.
Shale	Sedimentary rock derived from mud. Commonly finely laminated (bedded). Particles in shale are commonly clay minerals mixed with tiny grains of quartz eroded from pre-existing rocks.
Significance Level	In statistical hypothesis tests, the significance level is a fixed probability tolerated of wrongly rejecting the null hypothesis (H_o). It is the probability of a Type I error and is set by the investigator in relation to the consequences of such an error. Usually, the significance level is chosen to be 0.01, 0.05, or 0.10 (i.e., 1%, 5%, or 10%).
Skewness	A measure of asymmetry of the distribution of the sample data values.
Soil Horizon	A layer of soil or soil material approximately parallel to the land surface and differing from adjacent genetically related layers in physical, chemical, and biological properties or characteristics such as color, structure, texture, consistency, kinds and number of organisms present, and degree of acidity or alkalinity.
Standard Deviation (*s*)	A measure of dispersion of the distribution of the sample data values.
Standard Error	A measure of the variability (or precision) of a sample estimate, such as the computed arithmetic mean. Standard errors are needed to construct confidence intervals for the computed statistics.
Statistic	A measure of a statistical property of a population, computed based on sample results. An example of a statistic is the mean (i.e., average) of the measures in the sample.
Target Population	For the purposes of this document, the set of environmental space/time units within spatial and time boundaries for which a decision is needed on whether a chemical of interest is a COPC.
Transformation to Linearity	A transformation of a response variable, or independent variable, or both, that produces an approximate linear relationship between the variables.

Type I Error	Falsely rejecting the null hypothesis when it is in fact true, and accepting the alternative hypothesis.
Type II Error	Falsely accepting the null hypothesis as being true.
Variance	A measure of dispersion of the distribution of a set of data values. The variance is the square of the standard deviation.
Variogram	A plot of the variance (one-half the mean squared difference) of paired sample measurements as a function of the distance (and optionally of the direction) between samples. Typically, all possible sample pairs are examined. Variograms provide a means of quantifying the commonly observed relationship that samples close together tend to have more similar values than samples far apart.

1. INTRODUCTION

This guidance document supports and implements Chief of Naval Operations (CNO) background policy by providing detailed instructions for evaluating background chemicals in soil. Background chemicals are derived from natural and anthropogenic sources (i.e., sources not related to site-specific activities or operations), and are not associated with site-related chemical releases. Background analyses are essential for distinguishing between soils that have been impacted by a site-related chemical release and soils that have not.

1.1 Navy Policy and Guidance

The Navy has issued policy and guidance documents to address the role of background data in the Environmental Restoration Program, and present techniques for background analysis. Navy risk assessment policies also specify requirements for the use of background data in human health and ecological risk assessments.

1.1.1 Navy Background Policy

The *Navy Interim Final Policy on the Use of Background Chemical Levels* was released by the CNO in September 2000 (DON, 2000a). The policy stresses the importance of eliminating background chemicals from the list of chemicals of potential concern (COPCs) carried through a risk assessment, and setting cleanup levels above the background range.

The policy specifically requires the following:

- ❑ Chemicals that may have been released at the site must be clearly identified to ensure that the Navy is focusing on remediating COPCs associated with the release.
- ❑ Chemicals detected at concentrations below the upper bound of the background range must be excluded from the full baseline risk assessment. All chemicals screened out as a result of background considerations should be documented in the baseline risk assessment report.
- ❑ Cleanup levels must not be below the upper bound of the background range.

Conducting a background analysis early in the site investigation process helps to ensure that the Navy's cleanup responsibilities are clearly defined. The background analysis must be supported by adequate chemical and ambient geochemical information. Limited data (i.e., a few "background" samples) may not be sufficient to develop a defensible background analysis. The background analysis also considers both naturally occurring and anthropogenic sources, as shown on Figure 1-1. The COPC selection process (which includes elimination of chemicals on the basis of the background analysis) should be discussed with regulators and conveyed to the community as early as possible. The methods used for background analysis must be scientifically based, technically defensible, and cost-effective.

As noted in the Navy background policy, in some cases unacceptable risks may be associated with chemical concentrations within the background range. Although this risk is outside the scope of the Navy's Environmental Restoration Program, restoration program personnel should convey the information to stakeholders: chemicals screened out due to background considerations should be evaluated against the appropriate risk-based screening criteria, and the results should be documented in the risk characterization sections of the baseline risk assessment report.

Cleanup levels should be risk-based, but must not be within the background range. Therefore, cleanup efforts must be limited only to risks associated with contaminant concentrations that are elevated as a result of a site-related release.

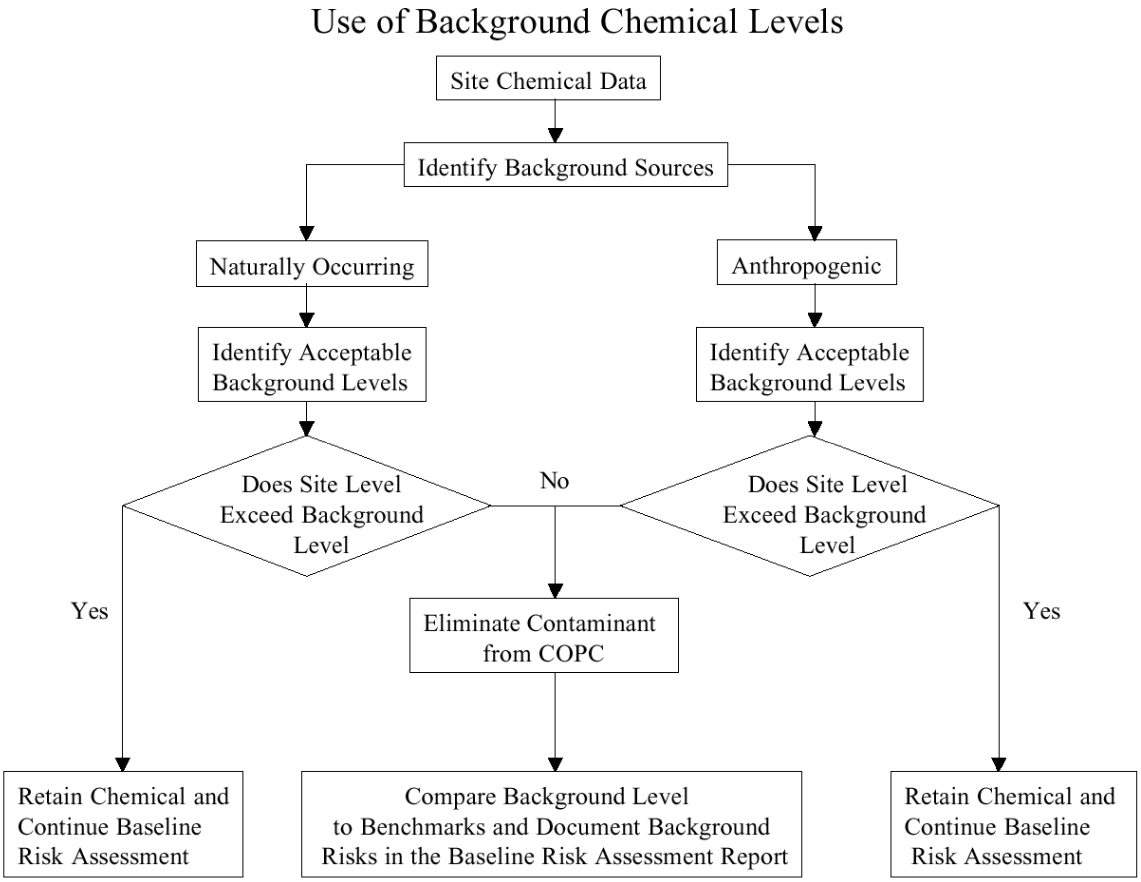


FIGURE 1-1. Navy policy on use of background chemical levels

1.1.2 Navy Background Guidance

Navy Engineering Field Divisions/Activities (EFDs/EFAs) have generated technical guidance documents for background data analysis, including the following:

- ***SWDIV/EFA West:*** *Procedural Guidance for Statistically Analyzing Environmental Background Data* (DON, 1998) and *Handbook for Statistical Analysis of Environmental Background Data* (DON, 1999a). These guidance documents provide detailed step-by-step instructions for graphic and statistical background analysis using the Comparative Method (i.e., statistical comparison of site data to reference area background data).
- ***PACDIV:*** *Protocol for Background Evaluation*. Presented in Appendix C of the *Site Management Plan Update for the Pearl Harbor Naval Complex* (DON, 2001b), this protocol serves as the basis for all background analyses conducted in support of environmental investigations and cleanup at the Pearl Harbor Naval Complex Superfund site as well as all other PACDIV Installation Restoration (IR) and Base Realignment and Closure Act (BRAC) projects. The protocol provides guidance on extracting background chemical concentrations from on-site datasets using a series of geochemically and statistically based tools. It was developed in cooperation with the United States Environmental Protection Agency (U.S. EPA) Region 9.

The guidance and procedures presented in this document are based on principles established in the above SWDIV/EFA West and PACDIV guidance documents. Specifically, the Geochemical Method (Section 3) is based primarily on the PACDIV background analysis protocol (DON, 2001b); and the Comparative Method (Section 4) is based on principles presented in the SWDIV/ EFA West Handbook (DON, 1999a).

1.1.3 Navy Risk Assessment Policy

The Navy has issued the following policies to provide guidance on the use of background chemical concentrations in human health and ecological risk assessments:

- *Navy Policy for Conducting Human Health Risk Assessments Related to the Installation Restoration Program* (DON, 2001a)
- *Navy Policy for Conducting Ecological Risk Assessments* (DON, 1999b).

According to the policies listed above, background chemicals should be screened out as early in the risk assessment process as possible. In human health risk assessments, background chemicals should be considered during the Tier 1A portion of the evaluation. In ecological risk assessments, background chemicals should not be evaluated until Step 3a of the Tier 2 baseline ecological risk assessment. The focus of subsequent risk calculations should be COPCs (i.e., chemicals detected at concentrations above the upper level of the background range).

1.2 Scope of Soil Background Analysis Guidance Document

This guidance document presents techniques for evaluating chemical data and soil characteristics to distinguish between soils that have been impacted by a site-related chemical release and those that have not. Both natural processes (e.g., erosion of naturally occurring metallic mineral deposits) and anthropogenic processes (e.g., deposition of chemicals from internal combustion engine exhaust) may result in elevated concentrations of various chemicals—including hazardous substances—in otherwise nonimpacted soils. These background chemicals are derived from natural or anthropogenic sources, and are not associated with site-related chemical releases.

To evaluate the nature and extent of potential contamination at a site or facility, soil samples typically are analyzed for chemicals suspected to have been released as a result of site activities. However, detection of elevated chemical concentrations in soil samples does not necessarily indicate that a release has occurred—the elevated concentrations may represent background conditions. Therefore, to ensure that only soils that have been impacted by a chemical release are targeted for cleanup, *background analysis should be conducted early in the site investigation process*. Failure to distinguish between concentrations associated with a site-related chemical release and background concentrations may lead to establishing cleanup levels within the background range, resulting in unnecessary and costly remediation, and potentially delaying property transfer and re-use. Furthermore, as discussed in Section 1.4, cleanup of chemicals present at concentrations within the background range is not consistent with established environmental regulations, policies, and guidelines.

The procedures presented in this guidance document will allow Navy environmental restoration personnel to technically and defensibly differentiate between site-related releases and background conditions. The techniques are designed to evaluate background levels of organic and inorganic chemicals. Background chemicals, particularly metals in soil, occur naturally in all soils and are often present at concentrations high enough to represent unacceptable risks to human and ecological receptors and exceed promulgated standards. Therefore, for most sites, naturally occurring background metal concentrations must be characterized in order to accurately evaluate the nature and extent of site-related metal contamination and assess the associated risks.

Techniques for statistical analysis of chemical data are fundamental components of the background analysis process. However, statistical analysis alone is not sufficient to fully understand and define the background conditions that exist at a particular site. In many instances, an understanding of the geological and geochemical processes involved in formation of naturally occurring chemicals also is essential. Therefore, this guidance document also provides guidance for evaluating the geology of the site and the geochemical characteristics of site soils as they relate to background analyses.

This guidance document presents detailed instructions for two commonly accepted methods of background analysis:

- ***Geochemical Background Analysis*** (i.e., the Geochemical Method) (Section 3) relies on the geochemical association among naturally occurring background chemicals. The Geochemical Method can distinguish between background concentrations and concentrations that represent a release by evaluating the geochemical characteristics of site soils and using various statistical techniques to identify anomalies.
- ***Comparative Statistical Analysis*** (i.e., the Comparative Method) (Section 4) compares chemical concentrations detected in site soils to ranges of chemical concentrations that exist at a nonimpacted "reference area"—i.e., an area that has not been impacted by an on-site chemical release. For this method to be successful, soils at the reference area must have physical, geochemical, biological, and anthropological characteristics similar to native soils at the contaminated area.

1.3 Roles of Background Analysis in the Environmental Restoration Program

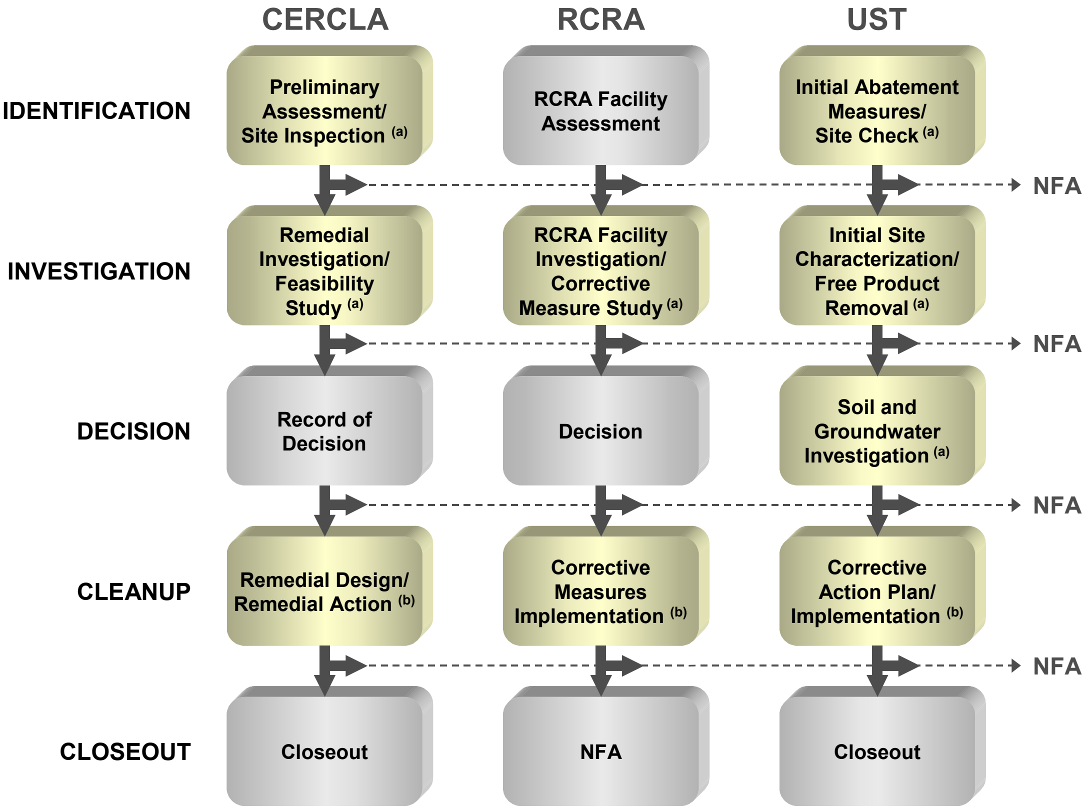
As noted in Section 1.4, federal law requires the Navy to protect human health and the environment and comply with applicable or relevant and appropriate requirements (ARARs) at all Navy chemical release sites. The Navy's responsibilities for site cleanup or other response actions necessary to comply with environmental laws and regulations cannot be defined until the nature and extent of contamination associated with a chemical release have been characterized. To evaluate the nature and extent of contamination, the Navy must distinguish between contamination associated with a chemical release and naturally occurring or anthropogenic background conditions; therefore, background analysis is an integral part of the environmental assessment, decision-making, and cleanup process.

Figure 1-2 illustrates the roles of background analysis in environmental investigation and restoration activities conducted under the three primary regulatory frameworks (CERCLA, RCRA, and the UST Program). Background analysis should be incorporated into the site identification, investigation, decision, and cleanup phases as necessary to achieve the following objectives:

- Identify chemicals that are present as a result of site-related releases (i.e., COPCs) for further consideration during the risk assessment process.
- Eliminate naturally occurring and anthropogenic chemicals that occur at concentrations within the background range from the list of suspected COPCs.
- Ensure documentation and discussion of chemicals eliminated from the list of suspected COPCs.
- Ensure adequate delineation of COPC-impacted soil at concentrations above background levels.
- Ensure that the cleanup level established for each COPC is not below the upper bound of the site-specific background range.
- Obtain stakeholder concurrence with site characterization and remedial conclusions.

It should be noted that retention of a chemical as a COPC does not necessarily imply that it will require cleanup. Decisions regarding cleanup should be made only after the baseline risk assessment and ARAR review are complete.

Background conditions should be accurately quantified to eliminate uncertainties introduced by qualitative or semiquantitative background investigations. If background conditions are not accurately known, it may be impossible to determine whether a chemical release has occurred, to adequately define the nature and extent of contamination, or to assess the residual contribution to risk made by the release. Background conditions must be characterized before the risk assessment process is completed so that appropriate cleanup or other response actions can be recommended and implemented. Background analysis is essential for



Background Analysis Applications:

(a)	**Site Characterization:**	Delineation of impacted media above background levels
	Risk Assessment:	Determination of COPCs/computation of residual risks
(b)	**Cleanup:**	Ensuring cleanup goals ≥ background levels

CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
RCRA	Resource Conservation and Recovery Act
UST	Underground Storage Tank Program
NFA	no further action

FIGURE 1-2. Roles of background analysis within regulatory cleanup frameworks

setting reasonable and attainable cleanup goals if cleanup is required. Project teams should never define cleanup goals that are within the background range established for the site.

The statistical and geochemical analyses presented in this guidance document will reduce the probability of decision error, and therefore can help to reassure stakeholders and regulators that the Navy has made the correct decisions regarding response actions for a particular site. In addition, by minimizing the probability of decision error, the Navy can avoid the unnecessary expense associated with cleaning up sites that do not require action to protect human health and the environment and thereby help to ensure that funding is appropriately focused on contaminated sites. Finally, background data are essential to provide scientifically defensible evidence to support a decision of no further action for a site; and, if cleanup is necessary, background data can aid in the selection of technically feasible and cost-effective remedial alternatives.

1.4 Statutory Requirements, Regulations, and Guidance

Numerous federal and state laws, regulations, and guidance are pertinent to background analysis. Federal law requires the Navy to protect human health and the environment and comply with ARARs at all Navy chemical release sites. The Navy policy and guidance documents summarized above were developed to ensure compliance with laws and regulations that address background analysis and its role in the site assessment and cleanup process. The *Navy/Marine Corps Installation Restoration Manual* (DON, 2000b) provides an overall synopsis of the environmental laws and regulations that define and affect the Navy IR Program, and describes the procedures the Navy has developed to ensure compliance with these laws and regulations. According to the *Installation Restoration Manual*, all actions at IR sites shall comply with the following:

- ❑ Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA)
- ❑ Superfund Amendments and Reauthorization Act (SARA)
- ❑ National Oil and Hazardous Substances Pollution Contingency Plan (NCP).

Although compliance with U.S. EPA policy and guidance documents is not mandatory, the *Installation Restoration Manual* also specifies that IR project teams shall reasonably interpret and apply U.S. EPA policy and guidance to make cleanup decisions and plan response actions.

1.4.1 Federal Laws and Regulations

Federal laws, including the CERCLA/NCP, require protection of human health and the environment at sites where CERCLA hazardous substances have been released. To comply with the CERCLA/NCP, the Navy must determine whether CERCLA hazardous substances have been released to the environment, assess the nature and extent of contamination, and evaluate associated risks to human health and the environment. To achieve these objectives, background analysis must be included as an integral component in the site assessment and cleanup process.

U.S. EPA requirements acknowledge the importance of background analysis. In 42 USC §9604(a)(3)(A), CERCLA recognizes that remediation may be impractical or impossible if naturally occurring background levels are higher than regulatory criteria. According to this section:

> "The President shall not provide for a removal or remedial action under this section in response to a release or threat of a release of a naturally occurring substance in its unaltered form, or altered solely through naturally occurring processes or phenomena, from a location where it is naturally found."

This provision clearly indicates that cleanup of background chemicals is not practical, even when their concentrations exceed federal, state, or local regulatory criteria.

RCRA requirements also recognize the importance of background analysis. RCRA gives the U.S. EPA authority to require cleanup of releases that impact environmental media within designated solid waste management units (SWMUs) (RCRA 3004[u]).

1.4.2 U.S. EPA Guidance

Several U.S. EPA technical guidance documents address issues related to background, including criteria for determining when a release has occurred, identifying cleanup requirements, and evaluating background concentration ranges. Two examples are summarized in Box 1-1.

Regional offices of the U.S. EPA also have issued technical guidance on background analysis. Examples are summarized in Box 1-2.

1.4.3 State Requirements, Guidance, and Technical Publications

Several states have developed regulations and guidance specifically addressing background data evaluation. Representative examples are summarized in Box 1-3.

1.5 Key Definitions

Some of the keywords and concepts used in this guidance document are defined below; a more comprehensive list is provided in the glossary section.

BOX 1-1. Examples of U.S. EPA background-related guidance

- ***U.S. EPA Office of Emergency and Remedial Response (OERR):*** OERR has published a number of guidance documents that address how background levels should be considered under the CERCLA and RCRA programs. The OERR guidance manual *Risk Assessment Guidance for Superfund* (RAGS) stresses the importance of background analysis. Volume I, the *Human Health Evaluation Manual (Part A)* (U.S. EPA, 1989b), notes that "Background sampling is conducted to distinguish site-related contamination from naturally occurring or other nonsite-related levels of chemicals." According to U.S. EPA's *Guidance for Data Usability in Risk Assessment* (U.S. EPA, 1992b), one of the primary objectives of sampling and analysis programs for sites of suspected environmental contamination should be to determine "whether site concentrations are sufficiently different from background." Similarly, U.S. EPA's *Guidance for Conducting Remedial Investigations and Feasibility Studies Under CERCLA* (U.S. EPA, 1988) states that background sampling should be conducted to differentiate between chemical releases resulting from site operations and background conditions.
- ***U.S. EPA Office of Solid Waste:*** According to *RCRA Facility Investigation Guidance* (U.S. EPA, 1989a): "High variability in the chemical composition of soils makes determination of background levels for the constituents of concern essential. This is particularly important for quantification of toxic metals, because such metals commonly occur naturally in soil."
- ***U.S. EPA Office of Solid Waste and Emergency Response (OSWER):*** OSWER has published an Engineering Forum Issue paper, *Determination of Background Concentrations of Inorganics in Soils and Sediments at Hazardous Waste Sites* (U.S. EPA, 1995), for educational use by its project managers. This publication provides a highly informative summary of the technical issues that should be considered to determine whether the concentrations of inorganic chemicals detected at a site are elevated relative to local background conditions. The first portion of the paper presents definitions and important factors influencing background concentrations. Issues addressed include selection of background sampling locations, considerations for selecting sampling procedures, and statistical analyses for determining if chemical concentrations at a background area and a waste site are significantly different. The paper notes in particular that it is not feasible to establish a single bright-line concentration value to define background for a particular chemical— instead, background should be expressed as a *concentration range* determined by statistical analysis of the chemical data. The second portion of the paper is divided into two parts. Part A presents procedures for determining whether hazardous waste site-related activities have resulted in increased inorganic chemical concentrations in soils compared to background concentrations. These procedures are based on the approach employed by the State of Michigan (MDEQ, 1991, 1994). Part B presents approaches for determining background levels of inorganic chemicals at CERCLA sites and is a modification of a U.S. EPA issue paper addressing background (U.S. EPA, 1992a). Both sets of procedures are based on the comparative statistical approach to establishing background, which requires sampling of an off-site reference area.

BOX 1-2. Examples of U.S. EPA regional background guidance

- ***U.S. EPA Region 1:*** *Risk Updates (No. 5): COPC Selection Process Update* (U.S. EPA, 1999). In this bulletin, U.S. EPA Region 1 clarifies their intent to ensure that background chemicals are carried through the risk evaluation process. The document suggests that the relevance of background concentrations should be discussed in the risk characterization or uncertainty sections of the risk assessment.
- ***U.S. EPA Region 4:*** *Statistical Tests for Background Comparison at Hazardous Waste Sites* (Interim Draft Supplemental Guidance to RAGS) (U.S. EPA, 1998). This regional guidance document provides details of the statistical approaches that the Region 4 Office of Technical Services considers appropriate for comparing site chemical concentrations to background levels when selecting COPCs (see Section 4).
- ***U.S. EPA Region 8:*** *Evaluating and Identifying Contaminants of Concern for Human Health* (U.S. EPA, 1994). This regional guidance document is intended to clarify the evaluation process for selecting COPCs for the human health baseline risk assessment process. In this bulletin, U.S. EPA Region 8 recommends the use of distributional tests (statistical tests used to determine if the central tendencies of two datasets are similar) to compare measured on-site datasets to background datasets (see Section 4).

BOX 1-3. Examples of state background guidance

- ***Florida Department of Environmental Protection (FDEP; formerly Department of Environmental Regulation):*** FDER published a technical document entitled *A Guide to the Interpretation of Metal Concentrations in Estuarine Sediments* (FDER, 1988). This publication, although specifically developed to address sediments, describes a very useful general approach to distinguishing between chemical concentrations associated with pollution and chemical concentrations that represent natural (background) conditions. The method is based on natural geochemical associations of metals with one another, and serves as the basis for one of the primary background analysis tools used in the Geochemical Method (see Section 3 of this document). In 1995, the FDEP released a special notice related to this publication describing the importance of using appropriate analytical methods for metals analyses (FDEP, 1995).
- ***Michigan Department of Environmental Quality (MDEQ):*** The Michigan Environmental Response Act or MERA (307 Protection Act, 1982) specifies that background conditions must be identified to determine whether chemical concentrations in site soils are elevated due to a chemical release. *Operational Memorandum No. 15* (1993) to MERA established default background concentrations based on a comprehensive background soil survey completed in 1991. Simple alternative statistical methods for calculating background concentrations are documented in the *Verification of Soil Remediation Guidance Document* (MDEQ, 1994).
- ***Pennsylvania Department of Environmental Protection:*** Title 25, §250.202 of the Pennsylvania Code sets requirements for establishing background chemical concentrations.
- ***California Department of Toxic Substances Control (DTSC):*** DTSC provided a Final Policy entitled *Selecting Inorganic Constituents as Chemicals of Potential Concern at Risk Assessments at Hazardous Waste Sites and Permitted Facilities* (Cal/EPA, 1997). This policy promotes the use of graphic techniques and describes various statistical tests used to distinguish background chemicals from site-related COPCs.

1.5.1 Site and Background Areas

The following terminology is used throughout this document:

❑ *Site* describes a zone designated for investigation due to actual, suspected, or potential chemical releases. A site usually consists of both impacted and nonimpacted areas. Site-specific field data are used to evaluate the extent of each area.

❑ *Background area* or *reference area* refers to an area where detected chemicals are attributed to natural or anthropogenic (i.e., nonsite-related) sources only. These nonimpacted background or reference areas may be located off site, or within the designated site boundaries.

❑ *Installation* or *facility* describes the extent of a Navy property at which one or more activities have been or are being conducted. An installation may contain a number of sites, as well as both impacted and nonimpacted (background) areas.

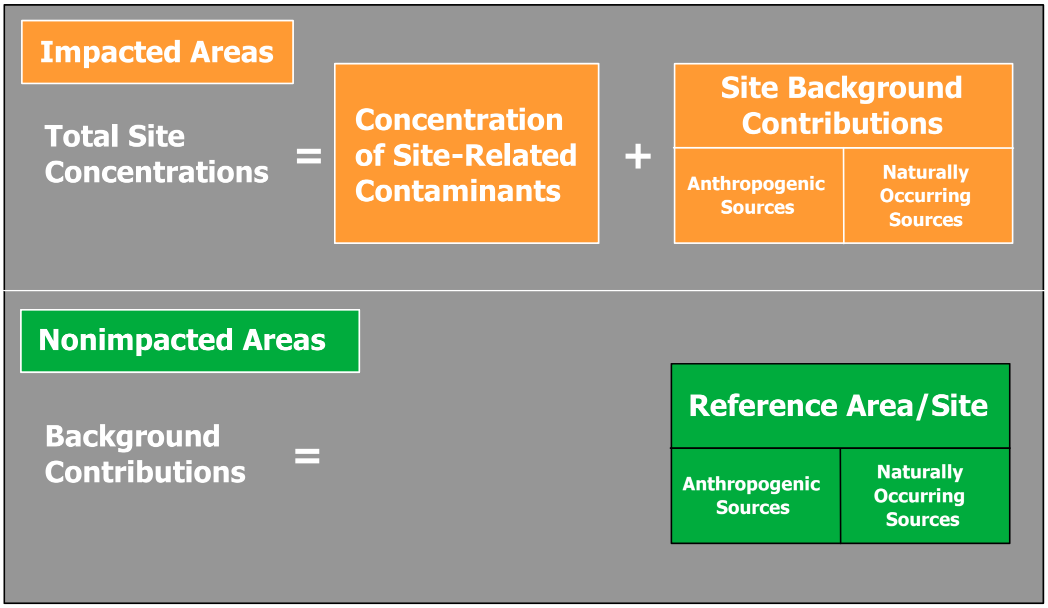
The concept of impacted versus nonimpacted areas is depicted on Figure 1-3, which illustrates the contributions to total concentrations made by site-related and background chemicals.

1.5.2 Background and Site-Related Chemicals

Background Chemicals

Chemicals derived from natural or anthropogenic sources not associated with site-related chemical releases (i.e., sources not related to activities or operations conducted at the site) are referred to as *background chemicals*. According to U.S. EPA (1989b), background chemicals fall into two categories:

❑ ***Naturally Occurring or Nonanthropogenic Chemicals:*** Chemicals present as a result of geochemical processes that have not been influenced by human activity. Naturally occurring organic and inorganic background chemicals in soil and water are attributable to the natural geological or hydrogeological characteristics of the area. These chemicals



Source: Adapted from U.S. EPA (1995)

FIGURE 1-3. Concept of impacted versus nonimpacted areas

have not been altered by human activity. Some examples include organic compounds derived from natural oil seeps, metals from bedrock, polycyclic aromatic hydrocarbons (PAHs) generated by forest fires, and naturally occurring asbestos.

- ***Anthropogenic Chemicals:*** Synthetic or natural substances that have been released to the environment as a result of human activities, but are not related to specific activities conducted at the site. These chemicals are usually ubiquitous in the environment, and may impact very large areas. Anthropogenic background chemicals are generated by human activities, but are unrelated to specific point sources or site releases. U.S. EPA (1989b) cites the following sources of anthropogenic background chemicals: agricultural runoff, urban runoff, septic systems, air pollution, irrigation; agricultural and residential application of pesticides (e.g., arsenicals, DDT); industrial discharges, landfills, municipal sludge land application; and urban pollution (e.g., lead and PAHs from automobiles and combustion processes, salts used for road de-icing). Anthropogenic background chemicals typically are: (a) widely distributed in the environment due to human activities; (b) not related to site sources or releases; (c) not related to other point sources or releases; and (d) attributable to past or present legal applications or sources.

Site-Related Chemicals

Site-related chemicals are chemicals released during past or current operations at an installation. In this document, such chemicals are referred to as COPCs. At some sites, elevated chemical concentrations may be the combined result of natural/anthropogenic (background) sources and a site-related chemical release.

Defining Background Conditions

Chemicals associated with background conditions and site-related releases, as well as chemicals that represent only background conditions, are routinely detected during sampling and analysis. Project teams may unnecessarily remediate chemicals and areas where no releases have occurred, unless background conditions are accounted for in one of the following two ways:

- ***Distinguishing COPCs from Background Chemicals.*** In this case, the project team must screen out background chemicals from the list of detected chemicals. Failure to make these distinctions could confound the investigation and remedial decisions. Chemicals detected at concentrations that do not exceed the upper bound of the background concentration range should be eliminated from consideration at the appropriate point in the risk assessment process (see Section 1.1). The remaining chemicals then are carried forward as COPCs for further evaluation during the risk assessment.

- ***Determining Background Levels of COPCs.*** If an individual chemical detected at a site is present due to both site-related and background sources, then the project team will need to quantify the concentration range that represents background conditions. The upper bound of the background concentration range must be identified to: (a) delineate the extent of a site-related chemical release; (b) calculate residual risks caused by a site-related release; and (c) determine the scope of required cleanup, should remediation become necessary. As shown in Table 1-1, some chemicals may pose unacceptable risks even at background levels. However, the cleanup goal should never be set at a point below the upper bound of the background concentration range.

1.6 Background Analysis Components

The flowchart shown in Figure 1-4 was developed to display the typical sequence of steps that should be followed in order to quantify background concentration ranges for chemicals in soil at Navy installations. As indicated on the flowchart, the first series of steps involves data review and assessment for screening and exploratory data analysis purposes. The data review and assessment process may allow investigators to estimate background ranges for some or all of the suspected COPC metals. No further background analysis is necessary if all the estimated background ranges are technically defensible and acceptable to stakeholders. If further background analysis is necessary, then an appropriate method (or methods) should be selected. Either or both of the methods described below may be selected for this purpose.

The Figure 1-4 flowchart illustrates the full background analysis sequence for a typical site. In many instances, certain steps can be combined or bypassed depending on project-specific requirements and conditions.

TABLE 1-1. Examples of naturally occurring and anthropogenic background concentrations with corresponding risk levels

Analyte	Background Concentration Range in Soils (mg/kg)	Background Risk (Cancer Risk Range)[c]	Residential PRGs (mg/kg)[e]
Naturally Occurring Background			
Arsenic	0.1–97[a]	3E–07 to 2E–04	0.39
Chromium (hexavalent form)	1.0–2,000[a]	3E–08 to 7E–05 5E–06 to 1E–02[d]	30.0 0.2[f]
Anthropogenic Background			
PAHs (Benzo[a]pyrene)	0.4–1.3 (rural soils)[b] 0.4–650 (urban soils)[b]	7E–06 to 2E–05 7E–06 to 1E–02	0.062

(a) Shacklette et al. (1971).
(b) ATSDR (2000).
(c) Background risks correspond to background concentrations and are based on residential exposure assumptions used to derive U.S. EPA Region 9 preliminary remediation goals (PRGs).
(d) Background risks correspond to background concentrations and are based on residential exposure assumptions used to derive California-modified PRGs.
(e) U.S. EPA (2000).
(f) California-modified PRG (California Environmental Protection Agency [Cal/EPA], 1994).

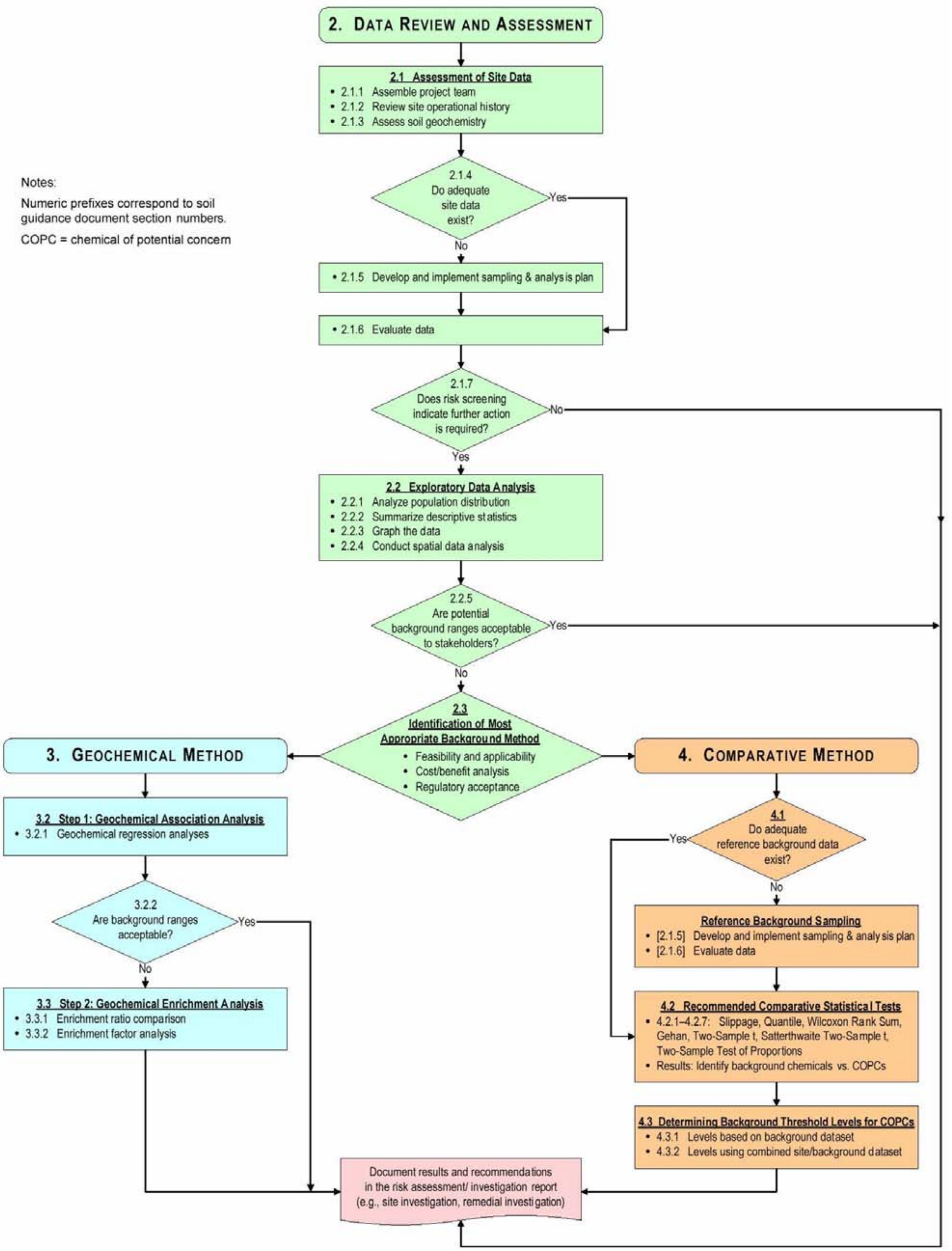


FIGURE 1-4. Typical background data evaluation process for inorganic constituents in soil

1.6.1 Data Review and Assessment

During the data review and assessment phase, the data should be evaluated to identify data gaps, determine the scope of supplementary sampling and analysis (if required), and compare suspected COPC concentrations to risk-based screening criteria. Supplementary sampling and analysis may be necessary to complement the existing dataset and ensure adequate site coverage. After the dataset is complete, representative exposure concentrations (such as upper confidence limits [UCLs] of mean concentrations) are computed and compared to their corresponding risk-based criteria. Chemicals that do not pose unacceptable risks will require no further action. The remaining chemicals should be subjected to exploratory statistical and spatial data analysis to estimate background ranges. If exploratory statistical and spatial data analysis do not yield technically defensible and acceptable background range estimates for all suspected COPC metals, then investigators should select a method (or methods) for further background analysis as described below.

1.6.2 Method Selection

If exploratory statistical and spatial data analysis do not yield satisfactory background range estimates, then the data should be analyzed further by the Geochemical Method and/or the Comparative Method (both methods may be implemented in some situations). Investigators should select the method most likely to yield technically defensible and acceptable background range estimates for each suspected COPC. Criteria that should be considered during the method selection process include: feasibility and applicability, regulatory acceptance, and cost/benefit ratios. Ideally, the identified method(s) will be applicable to the investigated chemicals, require minimal supplementary data, be acceptable to the regulatory community, and provide maximum cost savings. Also, communication channels between stakeholders should be established early in the process to ensure regulatory acceptance of the selected method.

1.6.3 Geochemical Method

The Geochemical Method is particularly useful for evaluating background metal concentrations in soil, and can be used when it is not possible to identify a suitable nonimpacted reference area. The method requires data only from the site—no off-site or reference area data are needed. The Geochemical Method utilizes various bivariate statistical tools to identify background ranges by evaluating concentration distributions and geochemical interrelationships.

Stakeholder acceptance of the validity of the geochemical interrelationships that form the basis of the Geochemical Method is critical to the success of the method. This acceptance can be attained through effective communication of geochemical processes and elemental relationships, supported by regional and site-specific field data.

1.6.4 Comparative Method

The Comparative Method consists of a series of statistical tests for comparing the site dataset to a reference area dataset, in order to determine whether the site and reference area concentrations are statistically similar. For this purpose, adequate reference-area sampling data are required in addition to the on-site data. Comparative tests are categorized as parametric or nonparametric. Parametric tests are based on specific distributional assumptions, such as normality of mean concentrations, whereas nonparametric tests are free of such assumptions. Each selected test is designed to assess specific aspects of the investigated data. For example, certain tests are designed to evaluate the similarity of the extreme site and background concentrations, whereas others are aimed at assessing central tendencies (median or mean) of observed concentrations.

To ensure the success of the Comparative Method, stakeholder acceptance of the background (nonimpacted) areas and their corresponding measured concentrations is most critical. This acceptance can be attained by effectively communicating the geochemical/anthropogenic characteristics of both the site and the reference area to all stakeholders.

1.6.5 Background Analysis Documentation

Success of background analysis is highly dependent on the adequate and appropriate presentation of results and findings. Documentation of the background analysis should contain at least the following components:

- Descriptions of site history and setting
- A summary of the major soil associations, soil families, and soil series that occur at the site.
- Technically defensible rationale for selection of chemicals that occur at the site as a result of natural and/or anthropogenic processes. This rationale should be based on regional and site-specific anthropogenic, geochemical, soil science, soil chemistry, and vegetation information.
- Descriptions of the site and reference area datasets
- Rationale to demonstrate that the datasets are adequate for background analysis
- Description of the method selection process
- Complete results of the analysis in adequate details in order to allow a comprehensive review of computations.

Consistent with the Navy background policy (DON, 2000a), the background analysis report must address the following issues:

- Chemicals that may have been released at the site must be clearly identified to ensure that the Navy is focusing on remediating COPCs associated with the release.
- Chemicals detected at concentrations below the upper bound of the background range must not be included in the full baseline risk assessment. All chemicals screened out as a result of background considerations must be discussed and documented in the risk characterization section of the baseline risk assessment report.
- Cleanup levels must not be below the upper bound of the background range.
- The methods presented in the background analysis report must be scientifically based, technically defensible, and cost effective.

1.7 References

Agency for Toxic Substances and Disease Registry (ATSDR). 2000. Toxicological Profile for Benzo(a)pyrene. PB/90/258245/AS. Atlanta.

California Environmental Protection Agency (Cal/EPA). 1994. *Preliminary Endangerment Assessment Guidance Manual (PEA).* Sacramento: Department of Toxic Substances Control.

California Environmental Protection Agency (Cal/EPA). 1997. *Selecting Inorganic Constituents as Chemicals of Potential Concern at Risk Assessments at Hazardous Waste Sites and Permitted Facilities. Final Policy.* Sacramento: Department of Toxic Substances Control (DTSC), Human and Ecological Risk Division. February.

Department of the Navy (DON). 1998. *Procedural Guidance for Statistically Analyzing Environmental Background Data.* San Diego: DON SWDIV and EFA West. September.

Department of the Navy (DON). 1999a. *Handbook for Statistical Analysis of Environmental Background Data.* San Diego: DON SWDIV and EFA West. April.

Department of the Navy (DON). 1999b. *Navy Policy for Conducting Ecological Risk Assessments.* Chief of Naval Operations Letter Ser. N453E/9U595355. Washington, DC. April 5.

Department of the Navy (DON). 2000a. *Navy Interim Final Policy on the Use of Background Chemical Levels.* Memo from Chief of Naval Operations (N45) (Rear Admiral L.C. Baucom, Director, Environmental Protection, Safety, and Occupational Health Division), to: Commander, Naval Facilities Engineering Command. 5090, Ser. N453E/OU595690. September 18.

Department of the Navy (DON). 2000b. *Navy/Marine Corps Installation Restoration Manual.* Alexandria, VA: Naval Facilities Engineering Command. August.

Department of the Navy (DON). 2001a. *Navy Policy for Conducting Human Health Risk Assessments Related to the Installation Restoration Program.* Chief of Naval Operations Letter Ser. N453E/1U595168. Washington, DC. February 12.

Department of the Navy (DON). 2001b. *Site Management Plan Update for the Pearl Harbor Naval Complex, Pearl Harbor, Hawaii.* Pearl Harbor, Hawaii: PACNAVFACENGCOM. October.

Florida Department of Environmental Protection (FDEP). 1995. *Special Notice from the FDEP Contaminated Sediment Management Group.* Tallahassee, FL: FDEP Technical Review Section. April 5.

Florida Department of Environmental Regulation (FDER). 1988. *A Guide to the Interpretation of Metal Concentrations in Estuarine Sediments.* Eds. S.J. Schropp and H.L. Windom. Tallahassee, FL. April.

Michigan Department of Environmental Quality (MDEQ). 1991. *Michigan Background Soil Survey.* Formerly Michigan Department of Natural Resources. Prepared by Waste Management Division, Geotechnical Sampling and Support Unit. April.

Michigan Department of Environmental Quality (MDEQ). 1993. *MERA Operational Memorandum No. 15: Default Type A Cleanup Criteria.* From Alan J. Howard, Chief, DEQ Environmental Response Division, to Environmental Response Division Staff. Lansing, MI. September 30.

Michigan Department of Environmental Quality (MDEQ). 1994. *Verification of Soil Remediation Guidance Document.* Revision 1. Lansing, Mich.: DEQ Environmental Response and Waste Management Divisions. April.

Shacklette, H.T., J.C. Hamilton, J.G. Boerngen, and J.M. Bowles. 1971. *Elemental Composition of Surficial Materials in the Conterminous United States.* USGS Professional Paper 574-D. Washington, DC: U.S. Government Printing Office.

United States Environmental Protection Agency (U.S. EPA). 1988. *Guidance for Conducting Remedial Investigations and Feasibility Studies Under CERCLA.* Interim Final. EPA/540/G-89/004. Washington, DC: Office of Emergency and Remedial Response. October.

United States Environmental Protection Agency (U.S. EPA). 1989a. *Interim Final RCRA Facility Investigation (RFI) Guidance.* Vols. 1–4. EPA/530/SW-89/031. Washington, DC: Office of Solid Waste, Waste Management Division.

United States Environmental Protection Agency (U.S. EPA). 1989b. *Risk Assessment Guidance for Superfund, Volume I, Human Health Evaluation Manual (Part A).* Interim Final. EPA/540/1-89/002. Washington, DC: Office of Emergency and Remedial Response. December.

United States Environmental Protection Agency (U.S. EPA). 1992a. "Options for Addressing High Background Levels of Hazardous Substances at CERCLA Sites." Draft Final Issue Paper. Office of Emergency and Remedial Response. January.

United States Environmental Protection Agency (U.S. EPA). 1992b. *Guidance for Data Useability in Risk Assessment (Part A).* Final. Publication 9285.7-09A. PB92-963356. Washington, DC: Office of Emergency and Remedial Response. April.

United States Environmental Protection Agency (U.S. EPA). 1994. *Evaluating and Identifying Contaminants of Concern for Human Health.* U.S. EPA Region 8 technical guidance. Prepared by U.S. EPA Region 8, Denver, CO.

United States Environmental Protection Agency (U.S. EPA). 1995. *Determination of Background Concentrations of Inorganics in Soils and Sediments at Hazardous Waste Sites.* U.S. EPA Engineering Forum Issue paper. EPA/540/S-96/500. Washington, DC: Office of Solid Waste and Emergency Response. December.

United States Environmental Protection Agency (U.S. EPA). 1998. *Statistical Tests for Background Comparison at Hazardous Waste Sites.* Interim Draft Supplemental Guidance to RAGS. Prepared by U.S. EPA Region 4, Atlanta, GA.

United States Environmental Protection Agency (U.S. EPA). 1999. *Risk Updates (No. 5): COPC Selection Process Update*. Prepared by U.S. EPA Region 1, Boston, MA.

United States Environmental Protection Agency (U.S. EPA). 2000. Region 9 Preliminary Remediation Goals (PRGs) Table. Prepared by U.S. EPA Region 9, San Francisco, CA. URL: <http://www.epa.gov/region09/waste/sfund/prg/index.html>. November 22.

2. DATA REVIEW AND ASSESSMENT

A background analysis is a component of a site environmental investigation that must be carefully planned and implemented to differentiate between background and site-related occurrences. Because some chemicals may pose risks to human health or the environment even at natural background levels, identification of these differences can facilitate decision-making and avoid potentially unnecessary remedial action. For example, some characteristics of naturally occurring metals (e.g., arsenic) can complicate the remedy decision process:

- ❑ They have low risk-based criteria (i.e., their background levels may exceed risk-based criteria).
- ❑ They are often used in industrial, military, or commercial activities.
- ❑ They can occur naturally over a wide concentration range.

Thorough data review is key to effectively evaluating these factors and ensuring that background analyses are scientifically accurate and technically defensible for appropriate decision-making. The procedures for data review and assessment described in this section should be implemented as part of any well-planned environmental site investigation, and are not intended to duplicate the site investigation process. It is important to involve all stakeholders throughout the planning and execution phases of a background analysis. As noted in Section 1.3, evaluation of background conditions can greatly reduce the probability of decision error, and therefore can help to reassure stakeholders and regulators that the Navy has made the correct decisions, and has minimized the probability of decision errors regarding response actions.

Hypothesis Development. The project team should develop hypotheses based on the nature of the suspected chemical release, the types of background chemicals that may occur at the site, and the potential extent of the impacted area. The following hypotheses should be considered:

- ❑ ***Hypothesis 1*:** The targeted chemicals are potential background chemicals at the site, and if these chemicals were released in the past, they are likely to have impacted the entire site.

  If the entire site is impacted by a chemical release, the concentration distribution will be the result of an overall shift to higher values than the original (background) distribution. (It should be noted, however, that a release that covers the entire areal extent of a site, and affects both surface and subsurface soils, is extremely uncommon.) In this case, the on-site dataset alone may not be enough to distinguish between background levels and concentrations that represent a chemical release. If evidence suggests that a release has impacted the entire site and the contamination extends beyond the site boundaries, the collection of off-site data will be necessary to delineate the extent of the release. These additional data then can be used to identify background concentration ranges.

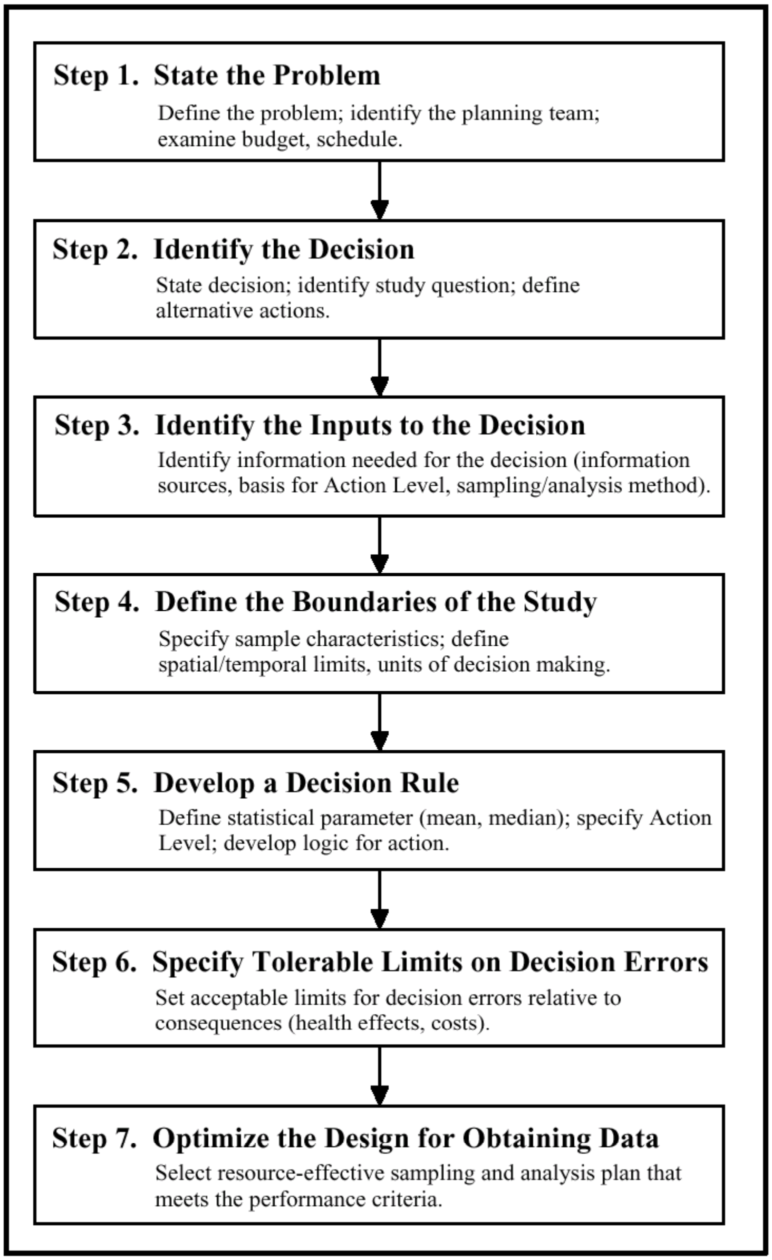
- ❑ ***Hypothesis 2*:** The targeted chemicals are potential background chemicals at the site, and if these chemicals were released in the past, they are likely to have impacted only portions of the site.

  If portions of the site are impacted, chemical concentrations in certain areas of the site will be elevated relative to the underlying background concentrations. Consequently, the dataset will contain two distinct populations (background concentrations and above-background concentrations) that can be distinguished by the background analysis methods presented in this document.

- ❑ ***Hypothesis 3*:** The targeted chemicals that exist at the site are present due to natural or anthropogenic processes only (i.e., no chemical releases are likely to have occurred at the site).

> If no chemical releases have occurred at the site, all concentrations represent background. In this case, the background analysis methods presented below will show that the dataset represents only one population.

Upon development of a defensible hypothesis, it should be presented to the stakeholders. Upon concurrence of all the stakeholders on the appropriateness of the chosen hypotheses, appropriate methods for conducting the subsequent background analyses should be selected.



Step 1. State the Problem
Define the problem; identify the planning team; examine budget, schedule.

Step 2. Identify the Decision
State decision; identify study question; define alternative actions.

Step 3. Identify the Inputs to the Decision
Identify information needed for the decision (information sources, basis for Action Level, sampling/analysis method).

Step 4. Define the Boundaries of the Study
Specify sample characteristics; define spatial/temporal limits, units of decision making.

Step 5. Develop a Decision Rule
Define statistical parameter (mean, median); specify Action Level; develop logic for action.

Step 6. Specify Tolerable Limits on Decision Errors
Set acceptable limits for decision errors relative to consequences (health effects, costs).

Step 7. Optimize the Design for Obtaining Data
Select resource-effective sampling and analysis plan that meets the performance criteria.

FIGURE 2-1. Seven-step DQO planning process (Source: U.S. EPA, 2000c)

2.1 Assessment of Site Data

Background analysis should be based on a well-defined data quality objectives (DQOs) decision-making framework in order to answer the following key questions:

- ❑ Is additional sampling and analysis required to determine site-specific background levels?
- ❑ If additional sampling and analysis is necessary, what quality and quantity of samples and analyses are needed?

Available site data must be reviewed and assessed to avoid the unnecessary effort and expense associated with collecting data that are not needed for the analysis. The data assessment should be consistent with the most recent U.S. EPA *Guidance for the Data Quality Objectives Process* (U.S. EPA, 2000c). Specifically, the seven-step DQO planning process (Figure 2-1) should be used to determine the type, quantity, and quality of environmental data needed to support the decision-making process. Proper use of the DQO process will provide the scientific foundation for defensible decision-making by helping to assure that representative field samples are collected at appropriate locations and times, that appropriate techniques are used for graphic and statistical analysis of the resulting data, and that the graphic and statistical test results are properly interpreted. When the DQO planning process is complete, appropriate supplementary site field and/ or background samples, if necessary, are collected at locations and times according to the specified sampling design.

After the samples have been processed and analyzed for the specified chemicals and parameters, the measurements must be evaluated to ensure that they are of the type, quantity, and quality specified during the DQO process. For example, if statistical tests are pursued, this evaluation should be conducted according to the data quality assessment (DQA) process (U.S. EPA, 2000b), which consists of the following steps:

1. Review of DQOs (output of each step of the DQO process) and sampling design
2. Conduct preliminary data review
3. Select the statistical test
4. Verify the assumptions
5. Draw conclusions from the data.

In many instances, the selected background analysis method may require information in addition to soil concentrations of the target chemicals. For example, if the Geochemical Method is selected for analysis of naturally occurring metals, then the concentrations of other analytes, such as aluminum, iron, magnesium and silicon (for clay soils), and/or calcium (for carbonate soils) should be measured. These supplementary data then should be combined with the existing site data.

2.1.1 Assemble Project Team

The first step in the data review and assessment phase of a background analysis is to assemble an appropriate project team. This team should include personnel with a broad range of expertise, including:

- ☐ ***Statisticians.*** Statisticians provide expertise in designing background-specific sampling plans, selecting appropriate statistical tests and procedures, and interpreting the analytical results.
- ☐ ***Geochemists.*** Geochemists provide expertise in characterizing site-specific geochemical conditions, identifying parameters that must be investigated, developing appropriate sampling plans, and determining site-specific elemental relationships.
- ☐ ***Geologists/Soil Scientists.*** Geologists and soil scientists provide expertise in characterizing regional/site geology, identifying site soil types, determining the adequacy of data coverage, and developing site background conceptual models.
- ☐ ***Toxicologists/Risk Assessors.*** Toxicologists/risk assessors provide expertise in characterizing site-specific risks posed by background chemicals and COPCs, determining appropriate risk-based criteria, and documenting the results of the background analysis.

If it is not feasible to obtain input from all of the disciplines listed above, at a minimum, personnel experienced in statistics and geochemistry are essential to plan a technically defensible background analysis strategy.

2.1.2 Review Site Operational History and Identify Site Conditions

A well-planned site investigation requires information on the operational history of the site, as well as on the site's physical and chemical characteristics. This information is helpful to identify potential site-specific COPCs, to evaluate the need for further sampling, and, if necessary, to select appropriate sampling locations and analyses.

Operational History

Site investigations usually consider the operational history of a site by focusing on:

- ☐ Storage of hazardous substances or chemicals at the site
- ☐ Transport, handling, or use of hazardous substances or chemicals at the site
- ☐ Potential or known releases of hazardous substances or chemicals at the site.

The primary source(s) of information on the operational history are the owners or operators of the site. Previous environmental or regulatory investigations that have compiled information directly applicable to the evaluation of background conditions at the site will be of particular interest. The following information also should be acquired, if possible:

- ❑ History of the industrial, commercial, or residential uses of the site
- ❑ Types of materials that were stored, handled, manufactured, or disposed of at the site
- ❑ Locations of site activities involving potential COPCs and potential releases
- ❑ Site-specific maps and "as-built" diagrams of current and former site structures related to storage, handling, manufacture, or disposal of site COPCs
- ❑ Locations of visible signs of potential COPC release (e.g., soil staining, discoloration, odor, and/or stressed vegetation)
- ❑ Contaminant release incident report databases from federal, state, or local regulatory agencies
- ❑ Title search agencies
- ❑ Master plans for larger facilities
- ❑ Tax maps, United States Geological Survey (USGS) maps, other maps of the site and surrounding area; and aerial photographs of the site and surrounding area
- ❑ Other potential sources of site history information, particularly former site employees and area residents.

Several guidance documents present procedures for assessment of the operational history of an investigation site. The American Society for Testing and Materials (ASTM) presents protocols for Phase I (ASTM, 2001f) and Phase II (ASTM, 2001g) site assessments.

Physical Setting

The natural background chemicals that occur at a site, and the fate and transport of COPCs, depend on physical characteristics such as topography, geology, and the types of soil that occur at the site and in the surrounding area. Therefore, these physical characteristics should be evaluated carefully during the site data assessment phase.

Topographic and Geographic Information

Accurate maps of the investigation site and the surrounding region are essential for background analysis. Two types of maps are required: (1) a small scale regional map for placing the site in a regional context, and (2) a detailed large scale site map to plot key site features, contaminant sources, impacted areas, and nonimpacted areas.

Maps should provide topographic and geographic representations of several types of information needed for the site investigation and background analysis:

- ❑ Terrain conditions and directions of site and regional surface drainage to identify potential off-site sources of chemicals
- ❑ Locations of existing sampling points to evaluate the spatial distribution of detected chemicals
- ❑ Known or potential on-site and off-site disposal or release locations
- ❑ Other pertinent information, including property boundaries, right of ways, and utility corridors.

USGS quadrangle maps, or "quads," are common regional-scale maps that show regional topography, water bodies, landforms, streets, and general land use. Quads are available for most regions of the United States at different scales, most typically ranging from 1 inch = 24,000 feet, to 1 inch = 250,000 feet. Other maps at larger scales are available for limited areas. A list of available maps for the United States can be located at the USGS "Online Map Lists" Web page (USGS, 2001). The availability of large-scale maps will vary for different sites. Depending on the location of the site, the following large scale maps and aerial photographs may be available:

- ❑ ***Tax maps.*** Tax maps showing land ownership are available from all states or local municipalities. Knowledge of landowners may provide key information concerning land use or activities at a site.

- ❑ ***Photogrammetric aerial survey (i.e., aerial photographs).*** Photogrammetry may show features not illustrated on USGS quads, such as stressed vegetation.
- ❑ ***Regional stereoscopic photo maps.*** Some photogrammetric surveys take photographs in pairs from slightly offset angles. When viewed stereoscopically, the photograph pairs produce three-dimensional images that clearly show topography and other site features.
- ❑ ***Historical/archaeological maps.*** Historical and archaeological maps can identify areas that have been affected by natural processes or human activities.

Site-specific information may be difficult to obtain for some Naval facilities. As with many United States Department of Defense (DoD) facilities, aerial photography may not be allowed in areas or regions that are considered sensitive, and maps for sensitive areas may be classified. However, information and maps may be available for neighboring sites, or non-DoD areas.

A Geographic Information System (GIS) can be very useful for compiling and integrating site data, particularly for sites with the following characteristics:

- ❑ Large areal extent
- ❑ Multiple potential contaminant types and sources
- ❑ Complex geologic/hydrogeologic structure.

GIS is a highly effective tool for interpreting and co-analyzing different datasets that can be effectively analyzed only in a spatial format. For example, changes in the extent of contamination can be detected by overlaying locations of contamination detected during different site investigations. In addition, previously unidentified contaminant sources can be revealed by overlaying contaminant concentration data on historical aerial photographs. The older photographs may reveal relationships between clusters of detected contaminants and potential contaminant sources that no longer show up on USGS quads.

Geologic, Soil, and Vegetation Information

The distribution of background chemicals at a site depends upon the types of soil and rock that occur at the site; therefore, the physical, geochemical, and biological characteristics of the soil and its parent rock should be evaluated carefully. Accurate field geological observations and descriptions of site soils and rocks are essential to plan a technically defensible background analysis strategy (see Section 2.1.3).

Concentrations of naturally occurring metals usually are related closely to soil type. Therefore, the absence of site-specific information on soil types greatly diminishes the reliability of the analysis. For example, spatial data analysis (Section 2.2.4) and the Comparative Method (Section 4) both require accurate characterization of soil conditions. In this document, the term “soil type” is used to differentiate soils based on characteristics including texture (e.g., sand, silt, clay), structure, and composition (i.e., mineralogy, organic content, moisture content).

Vegetation also can affect the distribution of background chemicals in soil. For example, some metals (e.g., silver, arsenic, cadmium, copper, mercury, lead, antimony, and zinc) can be significantly concentrated in surface soils due to uptake by plants and adsorption by organic matter. Therefore, information regarding the plant communities that occur onsite and differences in vegetation density can be useful to identify geochemical relationships for background analysis, and provide evidence to support conclusions of the analysis. Technical information regarding the effects of vegetation on the distribution of heavy metals in soil is available in Kabata-Pendias (2001) and Adriano et al. (1997).

Sources of geologic, soil, and vegetation information include:

- ❑ ***Geologic survey maps.*** Geologic maps indicate the rock and soil types that occur in different areas. Geologic survey maps of certain quads can be obtained from the USGS.

- ❑ ***Soil surveys.*** Soil surveys and maps indicate major soil associations, soil families, and soil series, and can be obtained from the U.S. Department of Agriculture (USDA)/Natural Resources Conservation Service (NRCS)—National Cooperative Soil Survey, and/or state soils offices located in the corresponding county seat, or state capital.
- ❑ ***Vegetation maps.*** Vegetation maps of certain areas can be obtained from the USDA Forest Service. However, a survey by a qualified biologist may be necessary if vegetation communities vary significantly over the site area. Guidance for classifying and mapping vegetation is available in FGDC (1997), Grossman and Faber-Langendoen (1998), and O'Neil and Hill (2000).

A qualified soil scientist, geologist, or geotechnical engineer should evaluate soil samples in the field to determine soil type, grain-size distribution, and classification according to Soil Science Society of America (SSSA), USDA/NRCS, or ASTM standards. Sand, silt, and clay percentages should be estimated for each sample. In addition, to accurately quantify soil characteristics, geotechnical testing and general chemistry analyses are strongly recommended. These tests and analyses should be conducted according to ASTM geotechnical testing standards and U.S. EPA analytical methods. Measurement of the following geotechnical parameters is strongly recommended:

- ❑ Moisture content (ASTM, 2001a)
- ❑ Visual identification and classification of soils (ASTM, 2001b)
- ❑ Soil density (ASTM, 2001c)
- ❑ Particle size distribution (ASTM, 2001d)
- ❑ Atterberg limits (ASTM, 2001e).

The following general chemistry analyses are also strongly recommended:

- ❑ Oxidation-reduction (redox) conditions (APHA [1998] Standard Method 2580A)
- ❑ Soil pH (U.S. EPA [2001] SW-846 Method 9045)
- ❑ Cation exchange capacity of soils (U.S. EPA [2001] SW-846 Method 9081)
- ❑ Total organic carbon (TOC) content (U.S. EPA [2001] SW-846 Method 9060).

Target Chemicals

After evaluating the operational history of the site and its geochemical and anthropogenic characteristics, a list of target chemicals should be developed. The target chemical list should include chemicals used at the site during its history, potential chemical breakdown products, potential site-specific background chemicals, COPCs that also are present at background levels, and parameters required for background analysis (see Box 2-1).

Nonimpacted Background Areas (Reference Areas)

Initial assessment of background conditions may involve identification of potentially nonimpacted portions of the site or the surrounding area. Areas that are apparently free of site-related impacts

BOX 2-1. Identifying target analytes

COPCs – Select based on site history and previous investigations.

Examples:

- Antimony, copper, and lead (firing ranges)
- Arsenic (wood preservative, pesticide equipment rinsing)
- Chromium (metal plating, alloys)
- Lead (leaded gasoline, battery acid disposal)

Additional metals required for background data analysis – Select based on geology and geochemistry.

Examples:

- Aluminum, iron, magnesium, and silicon (terrestrial – clays)
- Calcium (carbonate formations – limestone)

should be identified for this purpose. Such background (or reference) areas must be geochemically and anthropogenically similar to the impacted portions of the site. Existing regional background data for soils with geochemical and anthropogenic characteristics similar to those of site soils may be useful for screening purposes (e.g., during site inspections), and to support background analysis conclusions. The operational history and chronological aerial photographs of the site, as well as soil and geological maps, can provide valuable information for defining potential background areas. These areas may be targeted for sampling if needed for the background analysis (U.S. EPA, 1992a).

The process for selecting potential background areas should include the following steps:

❑ ***Identify pertinent operations and historical uses of the property.*** Current operations and past uses or activities at the potential background area should be examined to evaluate known or potential contaminant sources, locations of past releases, or locations of contaminant disposal. The examination should include identifying known or potential upgradient and off-site sources of contamination.

❑ ***Compare physical characteristics.*** The topographic, geographic, geologic, and hydrogeologic characteristics of the potential background area should be compared to the corresponding characteristics of the suspected chemical release site. Any surface drainage pathways or groundwater recharge locations should be related to fill areas, rock outcrops, or different soil types in order to evaluate probable fate and transport pathways.

❑ ***Formulate a conceptual site model.*** Information on known or potential contaminant sources should be integrated with the potential fate and transport pathways.

2.1.3 Assess Soil Geochemistry

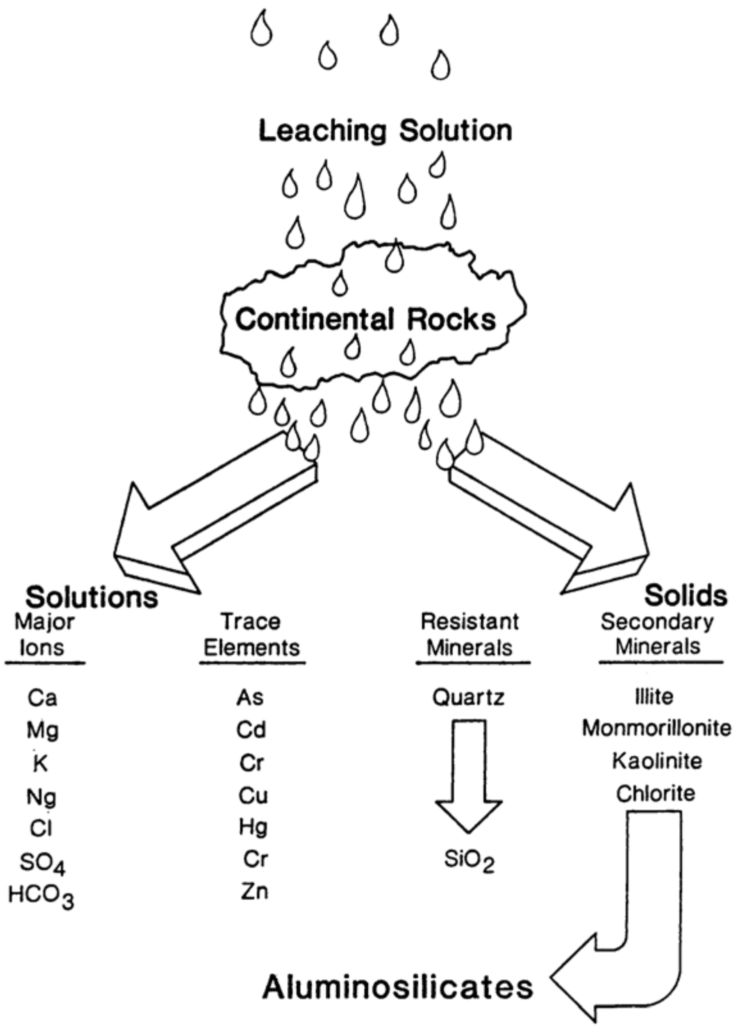
Geochemical processes transport and redistribute naturally occurring chemicals, particularly metals, in the geologic environment. Geochemical knowledge is the key to understanding the underlying causes of natural concentration anomalies in soil. It also is very useful when differentiating between background metals and metals that may be associated with a chemical release. The following overview of the important geological and geochemical principles used in background analysis is supplemented with further detail provided in Appendix A.

The metals that occur naturally in soils originate in the minerals that form the rocks of the earth's crust. Soil is formed when rocks and their component minerals are broken down by physical and chemical processes known as weathering. Physical weathering refers to the mechanical disintegration of rock by natural forces. Chemical weathering, the most important process in soil formation, mobilizes the chemical elements that make up rock-forming minerals. This process, shown in Figure 2-2, results in the physical disintegration of the rock, formation of new (secondary) minerals, and changes in the overall chemical composition. The physical and chemical breakdown of rocks and minerals that occurs during weathering creates soils with distinct geochemical characteristics.

Rock-forming minerals are natural chemical elements or compounds with distinct chemical compositions and (in most cases) characteristic crystalline structures. The distribution of metals and other elements in soil depends primarily on the chemical composition of the minerals that make up the source (or parent) rock from which the soil is derived. Therefore, the natural geochemical relationships that exist in the parent rock usually are reflected in the soil. For example, in natural soils, certain metals tend to occur together, and metal/metal concentration ratios are often restricted to relatively narrow ranges. These geochemical relationships can be characterized and used to evaluate soil background conditions. The electronic structure and chemical properties of the elements (as reflected in the periodic table of the elements [Figure A-2]) can be very useful to predict geochemical relationships for different soil types and to identify relationships that may be useful for background analysis.

Metal Concentrations in Soil

Background metal concentrations within a particular soil reflect the chemical composition of the parent rock from which the soil is formed, the



Source: FDER (1988, Figure 1).

FIGURE 2-2. Conceptual model of the chemical weathering process

geochemical processes that occur both during and after soil formation, and any anthropogenic processes that may be occurring. Heavy metals occur in natural soils over very broad concentration ranges. Weathering, biologically mediated chemical reactions, and other natural geochemical processes can cause certain metals to become enriched or depleted in soil. Table 2-1 lists the ranges of heavy metal concentrations encountered in agricultural soils worldwide. A detailed list of generic background metal concentration ranges for a variety of soil types is presented in Appendix A, Table A-9. Metal concentrations not exceeding the upper bounds of the ranges shown on Table 2-1 or Table A-9 may be attributable to background sources.

The type of parent rock is one of the factors that affect metal concentrations in soils. Weathering and soil formation processes also can significantly enrich or deplete the concentrations of certain metals. For example, metallic ions tend to adsorb strongly to clay and organic material; therefore, metal concentrations tend to be highest in fine-grained soils that contain high percentages of clay and organic particles. The generic background concentration table presented in Appendix A illustrates the high levels of enrichment that can be encountered in fine-grained soils.

The strong affinity of certain metals for organic carbon and clay particles suggests that, in some cases, it may be advantageous to normalize metal concentrations to the organic carbon (i.e., TOC) or clay content of each soil sample before implementing the background analysis methods described in this document. However, because normalization is not necessary to evaluate background at most sites, it is not addressed in the procedural guidance presented in subsequent sections of this document. Adsorption to organic carbon and

TABLE 2-1. Ranges and means of total concentrations of heavy metals in surface soils (mg/kg)

Element	Typical Range	Mean	Element	Typical Range	Mean
Arsenic	<0.1–66.5	8.6	Mercury	0.008–1.11	0.11
Barium	10–1,500	413	Molybdenum	0.1–7.35	1.82
Cadmium	0.01–2.7	0.53	Nickel	0.2–450	22
Chromium	1–1100	54	Selenium	0.005–1.9	0.33
Cobalt	0.1–70	7.9	Tin	80–17,000	3,300
Copper	1–140	19.8	Vanadium	6.3–500	74.4
Lead	1.5–176	28.6	Zinc	3.5–770	64
Manganese	7–9,200	437			

Source: Kabata-Pendias (2001, Table 9).

clay is discussed further in Appendix A.

Elemental Correlation

Soil background metals usually occur in groups. Nickel and chromium are examples of metals that commonly occur together in rocks and soils.

In rocks of the same type, certain groups of elements are often highly correlated (i.e., their concentration ratios tend to be relatively constant). In addition, although their concentrations may vary considerably among different rock types, certain elements still tend to be correlated. For example, as shown on Figure 2-3, nickel/chromium concentration ratios tend to be relatively constant among rocks classified as granite, shale, clay, and basalt. Elemental correlation such as that observed for nickel and chromium is common for many pairs and groups of metals, and therefore provides a very powerful tool for background analysis. If correlation exists, background ranges of the metals can be defined based on the relatively constant ratios of one metal to another. If a metal is found at an elevated concentration that does not fit the background ratio, a chemical release should be suspected. Elemental correlation is central to background analysis by geochemical association analysis and is discussed in more detail in Section 3.2 and Appendix A.

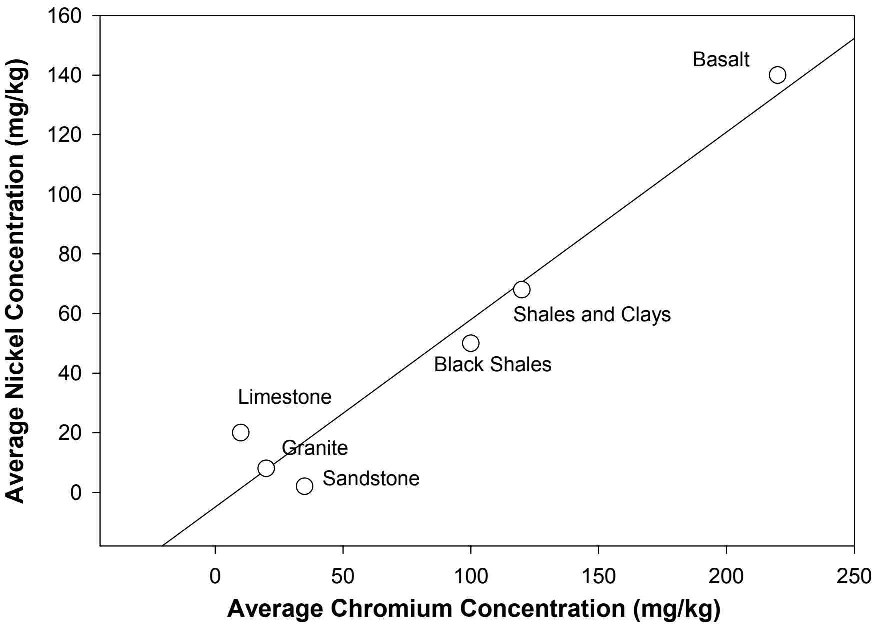


FIGURE 2-3. Correlation plot of chromium vs. nickel concentrations in various rocks

2.1.4 Determine Whether Adequate Site Data Exist

Upon compilation and review of site quantitative and qualitative information, the project team must determine whether existing data are adequate. If additional and supplementary data are needed, a Sampling and Analysis Plan (SAP) should be developed within the framework of the DQO process and as described below.

2.1.5 Develop and Implement Sampling and Analysis Plan

The sampling and analysis program should be carefully designed and implemented to ensure that all data needed to evaluate background conditions are collected. The SAP should specify the DQOs and the procedures that will be used to collect the data, and should address factors that must be considered when new data are combined with a previously existing dataset (see Section 2.1.6). As additional data needs become evident (e.g., if it is necessary to expand the sampling area or collect data from a reference site), the SAP should be amended or a SAP addendum should be prepared.

When using the Comparative Method (Section 4) for background analysis, the extent of the reference or background area must be clearly defined and justified in the SAP. Soils in the reference area should be physically, geochemically, biologically and anthropogenically similar to site soils. The Comparative Method is likely to be successful only if all stakeholders agree that the selected reference area is appropriate for comparison to the site. The absence of such a consensus could lead to prolonged and unproductive disagreements among various stakeholders.

Sampling Locations

Sampling locations and the total number of additional samples necessary to meet the DQOs should be identified. In many instances, the background analyses rely on existing site data, which typically are collected by judgmental sampling (e.g., sampling of locations where contamination is believed likely to occur based on information regarding the site history). In general, under such conditions, additional sampling will focus on areas of the site where data have not been previously collected. The datasets that result from this sampling sequence often are biased, clustered, and correlated. A thorough analysis of correlated datasets can be conducted through the use of geostatistical analyses. A detailed description of these techniques is beyond the scope of this guidance document; further information is available in Matheron (1971), Journel and Huijbregts (1978), Isaaks and Srivastava (1989), and ASCE (1990a,b).

Ideally, the background analysis should be supported by representative datasets. The use of a probability-based sampling design strategy would assure the representativeness of the selected sampling locations. Three such designs are systematic sampling, simple random sampling, and stratified random sampling (U.S. EPA, 2000a). If systematic sampling is used, sample locations could be positioned at the nodes of a square or triangular grid system that is placed at a random starting spot in the area to be sampled. The use of a triangular grid sampling design is suitable if the grid nodes (where samples are collected) are spaced far enough apart for the measurements to be independent. It also is necessary that the grid pattern not coincide with a pattern of contamination in soil in such a way such that the estimated average concentration determined from the measurements is biased high or low.

The use of a simple random sampling design, where all locations are equally likely to be chosen, also would be an acceptable design. However, simple random sampling may lead to large unsampled portions of the investigated area. A stratified random sampling system can eliminate such spatial gaps; under this system, the investigated area is divided into zones, where an equal number of sample locations are selected randomly within each zone. This system yields random samples while avoiding large unsampled areas. These three sampling designs and others are discussed in U.S. EPA (2000a).

Target Analytes and Parameters

Chemicals that may be associated with a suspected chemical release (e.g., suspected COPC metals), as well as metals that may be important for background analysis, should be targeted for chemical analysis. When using the Geochemical Method (Section 3) for background analysis, analytical data may be necessary to characterize the on-site concentration distribution of metals that are expected to represent background conditions. These potential reference metals can be identified on the basis of site-specific soil characteristics and geochemistry.

The SAP should specify laboratory reporting limits for all analytes to ensure that reporting limits are low enough to meet project-specific DQOs. Reporting limits must be low enough to allow comparison to risk-based screening criteria, and, to ensure that data quality is adequate for background analysis, should be set at levels well below published background ranges for typical soils. In addition, the SAP should focus on the collection of only those analytical data necessary to evaluate contamination, estimate background ranges, or achieve other site-specific DQOs—unnecessary data should not be collected. The SAP also must establish quality assurance (QA) and quality control (QC) procedures for the field sampling and laboratory analytical programs (for example, the SAP should specify requirements for field quality control sampling).

Finally, because soil types play a very important role in the distribution of background chemicals, the SAP should specify procedures for field logging of soil samples and for geotechnical sampling and testing procedures (e.g., grain-size distribution, density, etc.). In addition to soil analysis and testing, whole rock analysis may be necessary if background is to be determined by geochemical enrichment analysis (Section 3.3).

The project team must seek concurrence of stakeholders (e.g., U.S. EPA and state regulators). A draft version of the SAP should be prepared and submitted for review. The SAP should be implemented only after concurrence by all stakeholders. Finally, the project team must ensure compliance with all requirements and procedures established by the SAP during the field sampling and laboratory analytical programs.

2.1.6 Evaluate Data

After the analytical data are received from the laboratory, the data should be evaluated to ensure that they are of the right type, quality, and quantity for the intended background analysis method. The U.S. EPA guidance manual *Guidance for Data Quality Assessment* (2000b) provides general guidance for data quality assessment, and describes how it fits into the DQO process. Data validation is a key component in the data evaluation process.

Validate Data

Data should be assessed for quality according to the U.S. EPA functional guidelines and data validation procedures (1994, 2000b). The data validation procedures assign qualifiers to the data that give the end users a qualitative measure of data usability. Data may be assigned the following qualifiers:

J	estimated concentration
N	presumptive evidence of the identification of an analyte
R	rejected data (unusable)
U	not detected at the reporting limit.

These qualification codes will be assigned to each qualified data point. Combinations of qualifiers, such as UJ and NJ, also are possible. Guidelines for using qualified data in a background analysis are presented in Section 2.2.4.

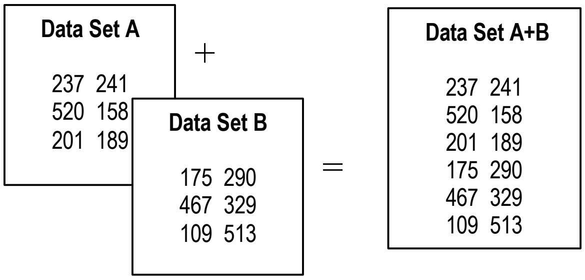
Combine Datasets

In many instances, background analysis involves combining different datasets, such as existing and additional field measurements. Combining two or more datasets to form a larger dataset may improve the ability of the analysis to differentiate between background and COPCs. For example, soil samples may have been collected and analyzed at several different times from the same site. Pooling the data will increase the number of data points available for background analysis, which improves the reliability of the results. However, an inappropriate combination of datasets can have the opposite effect.

Before datasets are combined, it is necessary to carefully define the spatial and temporal boundaries within which a chemical is evaluated. From a statistical point of view, such areas are considered target populations. If multiple datasets are to be combined, all datasets must be representative of the same target population. Datasets from geochemically, anthropogenically, biologically, and/or physically different soil types may not be suitable for combination as a single target population. For example, the placement of fill soils can drastically change the chemical composition of surface soil. Combining surface soil datasets from before and after fill placement is not likely to be appropriate.

Ideally, the datasets being considered for pooling should be obtained using the same sampling design and analytical methods. For example, it may not be appropriate to combine spatially clustered, biased data with data from random sampling locations, because chemical concentrations could be much higher in the clustered, biased area. In such instances, the clustered dataset may have to be declustered for subsequent analyses using one of several available declustering alternatives (see Section 4.1.2). Similarly, combining judgmental/purposeful (often clustered) sampling data with grid sampling data could lead to an unrepresentative site dataset.

If good evidence indicates that site chemical concentrations are generally uniform, sampling locations will not be a critical issue of concern. However, this assumption should not be made without



Should Datasets be Combined?

substantial supporting evidence. It also is important to verify that measurements in all the datasets being considered for pooling have similar quality characteristics. For example, the detection limits, quantitation limits, and measurement biases should all be sufficiently low, and an adequate number of blank and duplicate samples should be taken to check for the magnitude of bias and variability. Furthermore, to ensure consistency, the same sample collecting, compositing, handling, and measuring methods should be used for all the datasets that are pooled. Graphic and statistical methods may be used to assure that the pooled datasets are representative of the same target population (Box 2-2).

2.1.7 Conduct Risk Screening

After the data have been validated and combined, the maximum detected chemical concentrations should be compared to appropriate risk-based screening criteria. Typical examples of risk-based screening criteria are the U.S. EPA Region 9 PRGs and U.S. EPA Region 6 Media-Specific Screening Levels (MSSLs). If the maximum concentrations of a chemical are equal to or less than its corresponding risk-based screening level, then no background analysis is necessary for that particular chemical.

If the maximum concentrations of a chemical exceed its risk-based screening level, a representative exposure concentration should be calculated and compared to the screening level in order to determine whether background analysis is necessary. This step typically is done as part of the human health and ecological risk assessment process (see Section 1.1.3). As indicated in U.S. EPA (1992b, 1997b, 2000b), representative exposure concentrations are average concentrations over an exposure area. For small datasets (with less than 20 measurements), the maximum detected value is conservatively used as the representative exposure concentration. For larger datasets, this concentration is computed as the UCL of the mean concentration.

The choice of the appropriate method for computing the UCL depends on the distribution of measured values. For example, if the dataset does not exhibit a lognormal distribution, then the normality of mean concentration, supported by the central limit theory, can be applied in order to use Student's t distribution for UCL computation (U.S. EPA, 1992b). Otherwise, lognormal H-statistics (U.S. EPA, 1992b) can be used to compute UCLs. This latter procedure, however, is demonstrated to be nonrobust when the number of samples is small (Singh et al., 1997).

2.2 Exploratory Data Analysis

Upon compilation of the combined site dataset, the measured concentrations of target chemicals should be evaluated using a series of statistical techniques and tests (Box 2-3). These tools, described below, are used to:

1. Determine the probability distribution of the investigated data
2. Compute descriptive summary statistics of measured values

BOX 2-2. Statistical methods for comparing datasets

Graphic Methods: Histograms, boxplots, and probability plots (Section 2.2.3) of individual datasets are useful when assessing the similarity between datasets.

Two Datasets: The Comparative Method (Section 4) can be used to evaluate the difference between mean or median concentrations. Differences in the variance of measurements for the two datasets that have a normal distribution (with possibly different means) could be tested using the F test described in U.S. EPA (2000b, p. 4-33) and Conover (1998). The Squared Ranks Test of variances (Conover, 1998, p. 300) may be used to test for equality of variances. This test may be used regardless of the shape of the data distributions.

Multiple Datasets: The Kruskal-Wallis test (Gilbert, 1987, p. 250; Conover, 1998, p. 288) may be used to assess differences among their median concentrations. Equality of variances of more than two datasets can be tested as described by Conover (1998, p. 303). Both of these tests may be applied regardless of the shape of the underlying distribution.

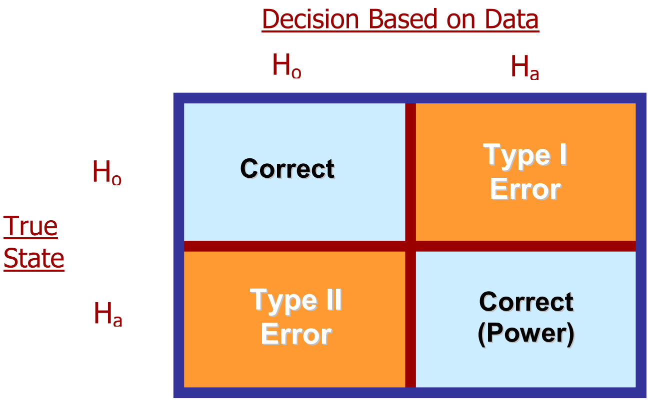
BOX 2-3. What is a statistical test?

A statistical test is a comparison of some data-based quantity (test statistic) with a critical value that usually is obtained from a special table. The test is conducted to determine if a statistically significant result has occurred.

Statistical testing determines whether the data are convincing beyond a reasonable doubt that a specified null hypothesis, H_o, is false and should be rejected in favor of a specified alternative hypothesis, H_a, that is true and should be accepted.

The choice of a null hypothesis is not arbitrary and is commonly governed by the available site knowledge. Statistical testing can be viewed as a means to quantitatively assess the validity of a hypothesis. For example when testing for normality, the following H_o and H_a are used:

H_o: The dataset is derived from an underlying normal distribution.
H_a: The dataset is not derived from an underlying normal distribution.



		Decision Based on Data	
		H_o	H_a
True State	H_o	Correct	Type I Error
	H_a	Type II Error	Correct (Power)

If the test rejects H_o in favor of H_a, then it can be concluded the dataset does not support the null hypothesis, and therefore the data should be examined closely to see if the data are derived from another distribution. This decision could involve two types of errors, as follows:

Type I error, or false positive decision error, is the error made by rejecting H_o when H_o is true. Type II error, or false negative decision error, is the error made by not rejecting H_o when H_o is false. The quantity α is the probability of making a Type I error. $100(1 - \alpha)\%$ is defined as the confidence level of the test. If the test does not reject H_o, either the data confirm the null hypothesis, or the information in the dataset is simply not sufficient for the test to reject H_o at the required confidence level. The quantity β is the probability of making a Type II error. $100(1 - \beta)\%$ is referred to as the Power of the test (i.e., the probability of correctly rejecting the null hypothesis). Tolerable limits on the probability of making a decision error should be specified, and compared to the decision error probability associated with the statistical test and the available data (see *Guidance for the Data Quality Objectives Process* [U.S. EPA, 2000c]).

Statistical tests are grouped as either parametric or nonparametric. Parametric tests assume specific distributions for the investigated variable. For example, the two-sample t test assumes that the mean concentrations of the investigated datasets are normally distributed. In contrast, nonparametric tests can be applied to any dataset regardless of their distributions.

3. Compute representative exposure concentrations for risk screening

4. Identify potential outliers

5. Determine background ranges.

The statistical tests and procedures discussed in this document assume that the measured values are independent and representative of the target population. Representative data usually can be acquired by a suitable random sampling, or by a systematic square or triangular grid sampling design, as discussed in Section 2.1.5.

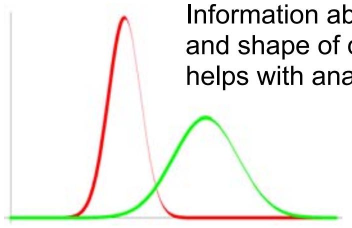
The concentration data associated with each sampling location should be posted on a suitable base map of the site to enable investigators to look for anomalies in the spatial distribution of each suspected COPC. Visual inspection of such spatial plots can reveal potentially impacted areas of the site, variations in metal concentrations with soil type, and other spatial features of the dataset.

Samples from clustered and biased locations may not provide data that are representative of the site. The data also may exhibit spatial correlation, and therefore cannot be viewed as independent datasets. In many instances, the presence of potential spatial correlation can be assessed through visual inspection of concentration maps of the chemicals of concern. (Spatially uncorrelated data usually are characterized by random occurrences of elevated measurements surrounded by lower concentrations.) To identify the source of spatial correlation, check the procedure that was used to determine the sampling locations and produce the corresponding concentration maps.

Clustered and biased datasets can be declustered prior to statistical testing, as noted in Section 2.1.6. For example, the investigated area can be divided with a grid into equally spaced squares. Each grid square then can be represented by the average concentration of measured values within the square, or a predefined number of samples can be selected randomly from each square.

2.2.1 Analyze Population Distribution

A number of statistical techniques and tests require normally distributed data. Appendix B.1 presents several test methods (including the Shapiro-Wilk W and D'Agostino Tests) that can be used to determine whether the data are normally distributed. These tests also can be used to test the lognormality of a dataset if each datum is transformed to its natural logarithm before the test is applied. In such cases, if the test indicates the transformed data are normally distributed, the original (untransformed) data are likely to be lognormally distributed.



Information about the location and shape of data distributions helps with analysis of the data.

2.2.2 Summarize Descriptive Statistics

This section discusses the summary statistics used to describe site and/or reference area datasets. These descriptive measures, along with graphic tools (Section 2.2.3), are used to develop an understanding of the range, variability, and shape of the underlying distribution of the measurements, as well as the number of nondetects and possible outliers that are present. This information is needed to decide which statistical test(s) or method should be pursued.

Datasets without Nondetects

Several descriptive summary statistics should be computed for the site and/or background datasets; these statistics are listed and defined in Box 2-4. The number of measurements in a dataset is denoted by *n*. The *n* measurements are denoted by $x_1, x_2, \ldots, x_n$. Examples that show how to calculate the descriptive summary statistics are provided in Box 2-5.

Datasets with Nondetects

Nondetects are measurements that the analytical laboratory reports are below some quantitative upper limits such as the detection limit or the limit of quantitation. Datasets that contain nondetects are said to be censored datasets.

The methods used to compute descriptive statistics when nondetects are present should be selected based on the number of nondetects and the total number of measurements, *n* (detects plus nondetects). If *n* is large (i.e., $n > 25$) and less than 15% of the dataset are nondetects, the general guidance in DON (1998) and U.S. EPA (2000b) is to replace the nondetects with a surrogate value (e.g., the detection limit [DL], one-half the DL, or zero). The descriptive summary statistics in Box 2-4 then may be computed using the now full dataset; however, some of the resulting statistics will be biased to some degree. (The median, pth sample percentile, and the interquartile range may not be biased if the number of nondetects is sufficiently small.) The biases may be large when *n* is small (i.e., $n<25$).

If 15% to 50% of the dataset are nondetects, the guidance offered in DON (1998) and U.S. EPA (2000b) is to forgo replacing nondetects with a surrogate value. Instead, the mean and standard deviation should be computed using the Cohen method or computing a trimmed mean or a Winsorized mean and standard deviation. These methods are described in detail in Appendix B.

BOX 2-4. Descriptive summary statistics for datasets with no nondetects

Descriptive Statistics	Definitions and Computation
Arithmetic Mean ($\bar{x}$)	$\bar{x} = (x_1 + x_2 + \ldots + x_n) / n$
Median (when *n* is an odd integer)	The middle value of the *n* measurements after they are arranged in order of magnitude from smallest to largest.
Median (when *n* is an even integer)	The arithmetic average of the middle two of the ordered measurements.
pth Sample Percentile	The value (not necessarily an observed measurement) that is greater than or equal to p% of the values in the dataset and less than or equal to (1 – p)% of the data values, where $0 < p < 1$. Compute $k = p(n + 1)$, where *n* is the number of measurements. If k is an integer, the pth percentile is the kth largest measurement in the ordered dataset. If k is not an integer, the pth percentile is obtained by linear interpolation between the two measurements in the ordered dataset that are closest to k.
Range	The difference between the maximum and minimum measurements.
Interquartile Range	The 75th sample percentile minus the 25th sample percentile.
Sample Standard Deviation (*s*)	A measure of dispersion (spread or variation) of the *n* measurements in a dataset that is computed as follows: $s = \{[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2] / (n - 1)\}^{1/2}$
Sample Variance	The sample variance is the square of the sample *s*, that is, sample variance = s^2.
Coefficient of Variation (CV)	The CV is a measure of relative standard deviation that is computed as follows: $CV = s / \bar{x}$.

BOX 2-5. Examples of descriptive summary statistics for datasets with no nondetects

Descriptive Statistics	Example Calculations
Arithmetic Mean ($\bar{x}$)	Suppose there are five data, say 50, 34, 52, 62, 60. Then the arithmetic mean is: $\bar{x} = (50 + 34 + 52 + 62 + 60) / 5 = 51.6$
Median (when *n* is an odd integer)	For the five data (after being ordered from smallest to largest) 34, 50, 52, 60, 62, the median is 52.
Median (when *n* is an even integer)	Suppose there are 6 data, which when ordered from smallest to largest are 0.1, 0.89, 2.0, 3.01, 3.02, 4.0. Then the median is (2.0 + 3.01) / 2 = 2.50.
pth Sample Percentile	Suppose the dataset (after being ordered) is 34, 50, 52, 60, 62, and we want to estimate the 60th percentile, that is, p = 0.6. Now, k = 0.6 (5 + 1) = 3.6. Because k is not an integer, we linearly interpolate between the 3rd and 4th largest measurements, that is, the 0.60 sample percentile is 52 + 0.6 (60 – 52) = 56.8.
Range	For the dataset 50, 34, 52, 62, 60, the range is 62 – 34 = 28.
Interquartile Range	The 75th sample percentile of the (ordered) dataset 34, 50, 52, 60, 62 is 60 + 0.5(62 – 60) = 61. The 25th sample percentile is 34 + 0.5(50 – 34) = 42. Therefore, the interquartile range is 61 – 42 = 19.
Sample Standard Deviation (*s*)	The sample *s* of the dataset 50, 34, 52, 62, 60 is: $s = \{[(50 - 51.6)^2 + (34 - 51.6)^2 + (52 - 51.6)^2 + (62 - 51.6)^2 + (60 - 51.6)^2] / 4\}^{1/2} = 11.08$
Sample Variance	The sample variance of the dataset 50, 34, 52, 62, 60 is the square of the sample *s*, that is, variance = $(11.08)^2 = 122.77$.
Coefficient of Variation (CV)	The CV for the dataset 50, 34, 52, 62, 60 is CV = 11.08 / 51.6 = 0.21.

2.2.3 Graph Data

The measured concentrations of chemicals with representative exposure concentrations that exceed risk-based screening levels should be evaluated graphically. If data from a reference area are to be used in the background analysis, these data also should be graphically evaluated. Graphic evaluation of the site and/or reference area datasets is used to:

- ❑ Conduct exploratory data analyses to develop hypotheses about possible differences in the means, variances, and shapes for the site and reference area distributions
- ❑ Visually depict and communicate differences in the distribution parameters (means, variances, and shapes) for the site and/or reference area data distributions
- ❑ Determine whether the site and/or background data are distributed normally, lognormally, or according to some other distribution
- ❑ Identify potential background ranges for the target chemicals.

Graphic techniques include histograms, boxplots, and probability plots. Much of this discussion is drawn from *Guidance for Data Quality Assessment* (U.S. EPA, 2000b), which offers a more thorough survey of graphic methods, including plots for two or more variables and for data collected over time and space. The methods included in this document, summarized in Table 2-2, were selected because they are easy to run and well suited for background analysis. The methods in Table 2-2 can be performed easily using the DataQUEST (U.S. EPA, 1997a) statistical software.

Histograms

The histogram is a bar chart used to provide an overview of the data distribution. Range intervals of the measured parameter are plotted along the x-axis, and the percentage of observations that occur within each range interval are plotted along the y-axis. All interval widths in a histogram should be the same size, as shown in Figures 2-4 and 2-5.

The general shape of the histogram is used to assess whether a large portion of the data is tightly clustered around a central value (the mean or

TABLE 2-2. Summary of selected graphic methods and their features

Method	Description	Method Features	
Histogram	A graph constructed using bars that describes the approximate shape of the data distribution.	• Easy to construct, understand, and explain. • Shows the shape, spread (range), and central tendency (location) of the data distribution.	• The choice of interval width for the histogram bars can affect the perception of the shape of the distribution.
Boxplot	A box graph with extended lines (*whiskers*) that depicts the central tendency and shape of the distribution.	• Easy to construct, understand, and explain. • Shows the 25th, 50th, and 75th percentiles as well as the mean, spread of the data, and extreme values. • Good for comparing multiple datasets on a common scale on the same page of report (for example, site and background sets).	• Provides less detailed information about the shape of the data distribution than is conveyed by the histogram.
Probability Plot	A plot of the estimated quantiles of a dataset versus the quantiles of a hypothesized distribution for the dataset.	• A graphic method for testing whether a dataset may be well fit by a hypothesized distribution (e.g., lognormal or normal). • Provides guidance about whether the dataset might be composed of two or more distinct populations (for example, background and site contamination populations), and can help identify outliers.	• A separate plot is required for each hypothesized distribution. • Subjective judgment is used to decide if the plot indicates the dataset may have the same distribution as the hypothesized distribution. • The plot should be used in conjunction with other graphic and statistical methods, such as those described in Section 2.2.4.

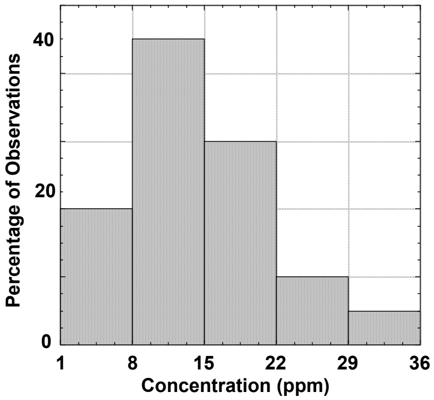


FIGURE 2-4. Example histogram with larger interval widths

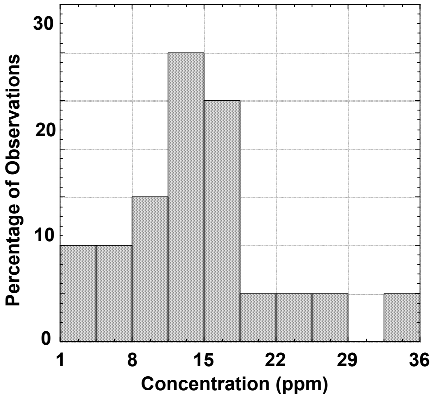


FIGURE 2-5. Example histogram with smaller interval widths

median) or spread out over a larger range of measured values. A symmetrical histogram suggests that the underlying population might be normally distributed, whereas an asymmetric histogram with a long tail of high measurement values suggests a lognormal or some other skewed distribution. These hypotheses can be evaluated using probability plots.

A histogram typically is constructed by first dividing the range of measured values into equal intervals. The number of measurements within each interval is counted, and the count is divided by the total number of measurements in the dataset to obtain a percentage. The length of the bar for that interval is the magnitude of the computed percentage. The sum of the bar percentages is 100%. Directions for constructing a histogram are provided in Box 2-6. An example is provided in Box 2-7.

The visual impression conveyed by a histogram is quite sensitive to the choice of the range interval (width of the bar). The histogram in Figure 2-4 graphs 22 measurements using a concentration range of 7 ppm. The histogram in Figure 2-5 is based on the same data as that used for Figure 2-4, but it uses an interval (bar width) of 3.5 ppm rather than 7 ppm. Note that Figure 2-5 gives the impression the data distribution is more skewed to the right (toward larger values) than does Figure 2-4. That impression is only due to the use of a smaller interval: only three data values are greater than 22 ppm, so the amount of information available to define the shape and extent of the right tail of the distribution is very limited. To guard against misinterpretation of histograms, the number of data points used to construct the histogram must always be reported. The bar widths should not be too narrow if the dataset is small.

BOX 2-6. Directions for constructing a histogram (after U.S. EPA, 2000b)

STEP 1: Let x_1, x_2, …, x_n represent the *n* measurements. Select the number of intervals (bar widths), each of equal width[(a)]. A rule of thumb is to have between 7 and 11 intervals that cover the range of the data. Specify a rule for plotting values that equal interval endpoints (i.e., plot in the higher interval or in the lower interval).

STEP 2: Count the number of measurements within each interval.

STEP 3: Divide the number of measurements within each interval by *n* (the total number of measurements in the dataset) to compute the percentage of measurements in each interval.

STEP 4: For each interval, construct a box which has a length that corresponds to the percentage value computed in Step 3.

(a) U.S. EPA (2000b) considers the case where the bar widths are not of equal size.

BOX 2-7. Example: Constructing a histogram (from U.S. EPA, 2000b)

STEP 1: Suppose the following n = 22 measurements (in ppm) of a chemical in soil have been obtained:

17.7, 17.4, 22.8, 35.5, 28.6, 17.2 19.1, <4, 7.2, <4, 15.2, 14.7, 14.9, 10.9, 12.4, 12.4, 11.6, 14.7, 10.2, 5.2, 16.5, and 8.9.

These data range from <4 to 35.5 ppm. Suppose equal sized interval widths of 5 ppm are used, that is, 0 to 5, 5 to 10, 10 to 15, etc. Also, suppose we adopt the rule that a measurement that falls on an interval endpoint will be assigned to the higher interval. For example, a measurement of 5 ppm will be placed in the 5 to 10 ppm interval instead of the 0 to 5 ppm interval. For this particular dataset, no measurements happen to fall on 5, 10, 15, 20, 25, 30, or 35. Therefore, the rule is not needed for this dataset.

STEP 2: The table at right shows the number of observations within each interval defined in Step 1.

STEP 3: The table contains n = 22 measurements, so the number of observations in each interval will be divided by 22. The resulting percentages are shown in column 3 of the table.

STEP 4: For the first interval (0 to 5 ppm), the vertical height of the bar is 9.10. For the second interval (5 to 10 ppm), the height of the bar is 13.6, and so forth for the other intervals.

Interval	Number of Data in Interval	Percent of Data in Interval
0 to 5 ppm	2	9.10
5 to 10 ppm	3	13.60
10 to 15 ppm	8	36.36
15 to 20 ppm	6	27.27
20 to 25 ppm	1	4.55
25 to 30 ppm	1	4.55
30 to 35 ppm	0	0.00
35 to 40 ppm	1	4.55

Boxplots

The boxplot, sometimes called a box-and-whisker plot, simultaneously displays the full range of the data, as well as key summary statistics. Figure 2-6 is an example a boxplot of the data listed in Step 1 of Box 2-7. (In this plot, the two <4 values were set equal to 4.)

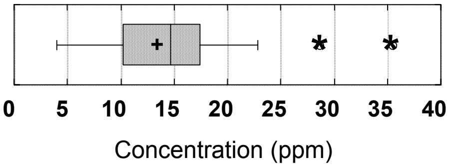


FIGURE 2-6. Example boxplot (box-and-whisker plot)

A boxplot provides a visual picture of the symmetry or asymmetry of the dataset. The boxplot is composed of a central box divided by a vertical line placed at the median value of the dataset, and two lines extending out from the box (called the whiskers). If the data distribution is symmetric, the central box will be divided into two equal halves by the median, the mean will be approximately equal to the median, the whiskers will be approximately the same length, and approximately the same number of extreme data points (if any exist) will occur at either end of the plot.

The arithmetic mean of the dataset is displayed using a + sign. The length of the central box (the interquartile range; see Box 2-4 for definition) indicates the spread of the central 50% of the data, whereas the lengths of the whiskers show the extent that measurements are spread out below and above the central 50% box. The upper end of the whisker that extends to higher concentrations corresponds to the largest data value that is less than the 75th percentile plus 1.5 times the length of the 50% box. Similarly, the lower end of the whisker that extends to lower concentrations corresponds to the smallest data value that is greater than the 25th percentile minus 1.5 times the length of the 50% box. Any data values that fall outside the range of the whiskers are plotted as asterisks. Horizontal boxplots (such as the plot shown in Figure 2-4) may be rotated counterclockwise 90°, so that the box and whiskers are oriented vertically. U.S. EPA (2000b) also illustrates how to construct a boxplot.

Probability Plots

A probability plot is a graph of data versus the quantiles of a user-specified distribution (quantiles are defined in the Glossary). In background analysis, probability plots are used for three purposes: (1) to determine how well data fit a hypothesized distribution (e.g., lognormal or normal), (2) to identify outliers, and (3) to identify separate populations within the dataset and thus determine background concentration ranges. These uses are discussed further below.

Probability plots can be constructed by plotting the expected quantiles of the hypothesized distribution on standard graph paper (this method is described in Box 2-8). However, as described in the example shown in Box 2-9, a special type of graph paper called probability plotting paper can be used in order to avoid the need to determine the expected quantiles of the hypothesized distribution. Probability plots also can be constructed with the aid of a statistical software program, such as U.S. EPA DataQUEST software (U.S. EPA, 1997a), thus saving the effort of determining quantiles from special tables, or plotting points manually on probability plotting paper.

Use of a Probability Plot to Evaluate a Population Distribution

Probability plots often are used to visually evaluate the null hypothesis that the data are well fit (modeled) by a specified distribution. Frequently, the null hypothesis is that the dataset has either a normal or lognormal distribution; however, other distributions such as the Weibull and Gamma distributions (Gilbert, 1987, p. 157) sometimes are used. If the graph of plotted points in a probability plot appears linear with little scatter or deviation about the line, the results indicate that the data appear to be well fit by the hypothesized distribution that was being tested. If the hypothesized distribution is the normal distribution, the data values are not transformed before they are plotted. If the hypothesized distribution is the lognormal distribution, the procedures are the same, except the logarithms of the data are plotted (or a log-scale is used).

Figure 2-7 is a probability plot constructed to test the null hypothesis that the data have a normal distribution. Note that the x-axis for Figure 2-7 represents cumulative percentages for the standard normal distribution.

If a probability plot does not exhibit a linear pattern for the hypothesized distribution, the

BOX 2-8. Directions for constructing a normal probability plot (from U.S. EPA, 2000b)

Let $x_1, x_2, \ldots, x_n$ represent the *n* data points. To determine whether the data are normally distributed, construct a normal probability plot.

STEP 1: Order all the *n* data from smallest to largest and denote the ordered *distinct* (different) data values by $x_{(1)}, x_{(2)}, \ldots, x_{(n')}$, where n' may be less than n. For each distinct data value, compute the absolute frequency, AF_i. The absolute frequency is the number of times each distinct value occurs. If a data value occurs only once, the absolute frequency for that value is 1. If a data value occurs more than once, count the number of times the distinct value occurs. For example, consider the dataset 1, 2, 3, 3, for which $n = 4$ and $n' = 3$. The absolute frequency of value 1 is 1, that is, $AF_1 = 1$. The absolute frequency of value 2 is 1, that is, $AF_2 = 1$. But the absolute frequency of value 3 is 2, that is, $AF_3 = 2$, as 3 appears two times in the dataset.

STEP 2: Compute the cumulative frequency, CF_i, for each of the n' distinct data values. The CF_i is the number of data points that are less than or equal to $x_{(i)}$, that is, $CF_i = \sum_{j=1}^{i} AF_j$. Using the data given in Step 1, the CF for value 1 is 1, the CF for value 2 is 2 (that is, 1+1), and the CF for value 3 is 4 (that is, 1+1+2).

STEP 3: Compute $Y_i = \frac{CF_i}{(n+1)}$ for each distinct data value

STEP 4: Determine from the standard normal distribution (Table C-1) the quantile associated with each value of Y_i. Denote the quantile of the ith distinct data value by Z_i.

STEP 5: Plot the pairs (x_i, Z_i). If the plot of these points is well fit by a straight line, the data most likely fit a normal distribution. Otherwise, the data may be better fit by another distribution.

BOX 2-9. Example: Constructing a probability plot by graphing cumulative percentages on probability plotting paper

Consider the following n = 14 data points that have been ordered from smallest to largest: 5, 6, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, 12, 13. To test the hypothesis that the data are normally distributed, construct a normal probability plot.

STEP 1: Because there are no duplicate values in the dataset, the AF of each value is 1.

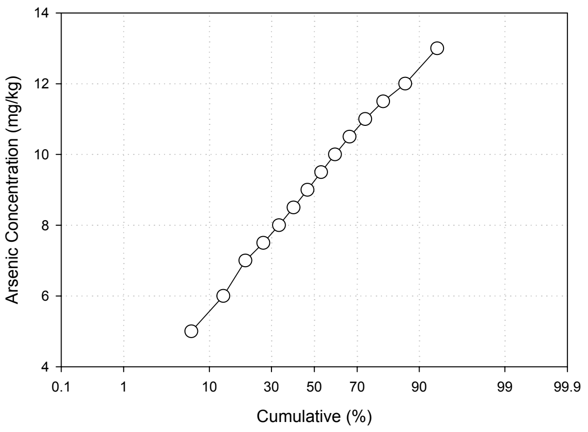
STEP 2: Compute the cumulative frequency for each data value as shown in the table below.

STEP 3: The cumulative percentages $Y_i = 100[CF_i / (n+1)]$ for each of the 14 distinct data values are shown in the last column of the table. The cumulative percentage associated with an individual data value is the probability (expressed as a percentage) that a randomly selected value from the dataset will be less than or equal to that individual data value.

STEP 4: Plot the n = 14 pairs of X_i, Y_i on probability plotting paper.

STEP 5: It appears the plot is approximately linear; therefore, the data can be assumed to be normally distributed.

i	Individual X_i	Absolute Frequency AF_i	Cumulative Frequency CF_i	Cumulative Percentage Y_i
1	5	1	1	6.7
2	6	1	2	13.3
3	7	1	3	20.0
4	7.5	1	4	26.7
5	8	1	5	33.3
6	8.5	1	6	40.0
7	9	1	7	46.7
8	9.5	1	8	53.3
9	10	1	9	60.0
10	10.5	1	10	66.7
11	11	1	11	73.3
12	11.5	1	12	80.0
13	12	1	13	86.7
14	13	1	14	93.3



characteristics of the curve may indicate that the data fit another type of distribution. Three typical distribution characteristics that will cause probability plots to deviate from a straight line are asymmetry (skewness), outliers, and heavy tails of the distribution. (Helsel and Hirsch [1992, pp. 30–33] describe these three conditions in detail.) If a probability plot is constructed to test the null hypothesis that the data are normally distributed, but the dataset is actually skewed to the right, the normal probability plot will be concave when viewed from above. If the dataset is skewed to the left, the graph will be convex when viewed from above. The plotted points in Figure 2-7 form a concave curve, indicating that the dataset is skewed to the right. Because lognormal distributions are right-skewed, it is logical to test the hypothesis that the dataset is well fit by a lognormal distribution. Figure 2-8 shows a probability plot of the logarithms of the data. The plotted line is well fit by a straight line; therefore, it may be tentatively accepted that the data are lognormally distributed. However, this result should be checked by conducting the Shapiro-Wilk W test (discussed in Appendix B).

Boxes 2-8 and 2-9 provide examples of the procedures used to construct a probability plot when the null hypothesis is that the data are normally distributed. The investigated data can include both detected and nondetected values (Box 2-10). The same procedures can be used to test the null hypothesis that data are lognormally distributed, by using logarithms of the data instead of the untransformed data.

Use of Probability Plots to Identify Outliers

Probability plots are useful for identifying potential outliers. Data points that are near a straight line and form a continuous distribution are likely to represent natural conditions, whereas data points that are not near the line or do not fit a continuous distribution (outliers) may represent contamination (or the presence of more than one natural population). A data value (or a few data values) much larger or much smaller than the rest will cause the other data values to be compressed into the middle of the graph.

Log-scale or log-transformation is appropriate for most situations encountered during background analysis. When log-scale or log-transformed data are plotted on a probability graph, data points from a lognormally distributed population lie near a single straight line, and the distribution is continuous (i.e., there will be no large gaps). Log-scale plotting is often appropriate even though some of the datasets encountered in background analysis may not be lognormally

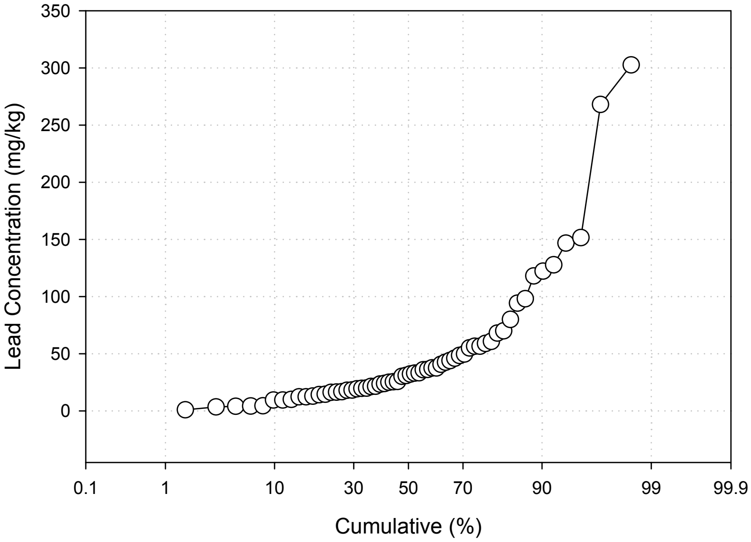


FIGURE 2-7. Example of a probability plot (linear y-axis 100 × cumulative probability on the x-axis)

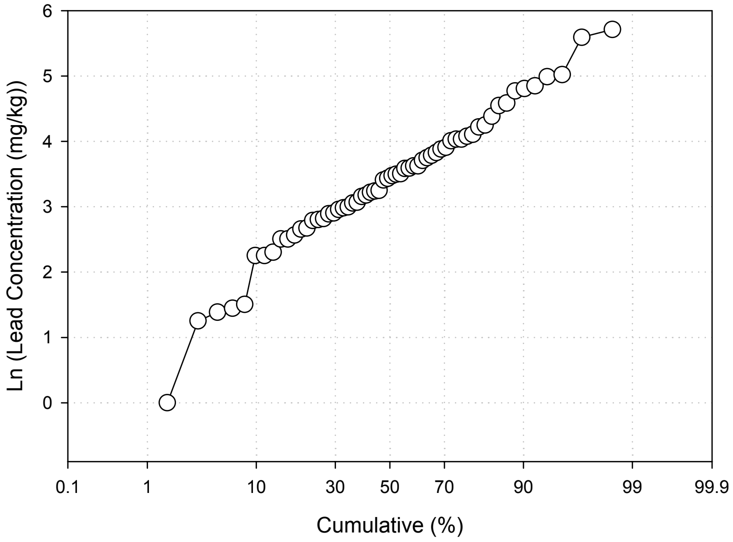


FIGURE 2-8. Example of a log-transformed probability plot

BOX 2-10. Use of multiple nondetects in probability plots

If the investigated datasets contain multiple nondetects, two possible ways to construct a probability plot are:

- Replace nondetects by one-half of the detection limit for each nondetect, or
- Assign all nondetects a dummy value at or below the lowest detected value.

Alternatively, Akritas et al. (1994, p. 227) and Michael and Schucany (1986, p. 476, Equation 11.8) have developed statistical procedures for constructing probability plots when multiple nondetects are present. However, these methods are somewhat complicated and their use for constructing probability plots for background analysis has not been evaluated.

distributed. For example, if data from a normal population distribution are log-transformed and graphed on a probability plot, it will still be possible to identify outliers.

Use of Probability Plots to Identify Background Ranges

Site datasets can contain both impacted and non-impacted measurements. The presence of these multiple populations in a dataset results in a segmented probability plot. Therefore, probability plots can be used to assess whether the measurements should be separated into different populations. A change in the slope or existence of an inflection point in a probability plot could indicate the threshold values separating different populations in the investigated dataset. Singh et al. (1994) present a procedure for identifying inflection points as a means to determine background ranges.

The probability plotting method for establishing background typically includes the following elements:

- ❑ Inflection points in the probability plot suggest multiple populations, including possible outliers. A straight-line plot with no inflection points indicates a single population.
- ❑ Usually ambient (local background) conditions are conservatively defined as the range of concentrations associated with the low-concentration segment of the population. The different populations may be distinguished by visual inspection.

Probability plots offer a simple way of graphically describing data and determining background ranges. However, to avoid any misleading conclusions, interpretation of these plots should always be supported by other accompanying analyses, as discussed in Section 2.2.4. The *Handbook for Statistical Analysis of Environmental Background Data* (DON, 1999, Section 2.5.6) provides further discussion of the potential limitations of probability plots.

At some sites "co-contamination" may occur if a chemical release contained both metals and organic compounds. As noted in Box 2-11, such co-contamination has no effect on metal background concentration ranges, and does not interfere with probability plot analysis, or any of the other background analysis techniques presented in this document.

2.2.4 Conduct Spatial Data Analysis

In spatial data analysis, univariate and probability plots are constructed for each suspected COPC metal to visually distinguish between background concentrations and elevated concentrations (i.e., outliers that may represent contamination). Inspection of concentration data posted on a site map also can be useful for evaluating the spatial characteristics of a dataset, and for distinguishing concentrations that may be associated with a chemical release from concentrations that could represent background conditions.

Univariate Plots

Univariate plotting is used to evaluate the spatial distribution of suspected COPC metals in soil. Information on spatial distribution is useful for evaluating preliminary conclusions based on the probability plot, and for assessing the significance of outliers. Univariate plots show the distribution of metal concentrations in a series of plots with a single numeric axis, the y-axis. The first step in univariate plotting is to segregate the metals data into four categories: data qualifier, sampling depth, soil type, and sampling location. A separate univariate plot then is constructed for each of the four categories by plotting the metal concentrations that correspond to each variable within the category. The concentration values are usually log-transformed or plotted against a log-scale; however, it may be useful to plot the data against a linear scale (particularly if the population is not lognormally distributed). The categories and examples of the variables within each category are shown in Table 2-3. By plotting metal concentrations for each variable, the spatial distribution of metals in site soils can be visualized. Investigators then can begin to make decisions regarding potential background concentration ranges on the basis of physical, geochemical, and statistical principles. Univariate plotting is described in more detail in the following subsections.

BOX 2-11. Co-contamination: Organic compounds and metals

At some sites, the presence of organic contamination at a particular location may indicate that metal contamination also is likely to be present at that location. This "co-contamination" may occur if a chemical release contained both metals and organic compounds (e.g., a release of both organic and inorganic pesticides, or a waste oil release). However, for most sites, there will be little or no correlation between the metal and organic compound distributions. Chemical releases that contain both types of contaminants are relatively uncommon and, more importantly, organic compounds and metals have very different fate and transport properties. It also is important to note that the presence of organic co-contaminants has no effect on metal concentration background ranges, or the results of any of the background analysis techniques presented in this document.

Organic compounds generally are less stable and more mobile in soil than metals; therefore, any spatial correlation between metals and organic contaminants observed at a site can be misleading. Because most organic compounds are far more volatile than metals, and are subject to attenuation due to biodegradation and other natural processes, they tend to have shorter residence times in soil. Organic compounds also tend to migrate through the soil to a greater extent than metals (e.g., leaching and transport with groundwater). As a result of their mobility, organic compounds tend to move away from a chemical release area, whereas the relatively immobile metals tend to remain within the release area.

If the metal and organic contaminants are correlated, the correlation may or may not be helpful to evaluate the spatial distribution of contamination at a site. However, it is not necessary to evaluate this relationship in order to establish background metal concentration ranges. The background distribution of metals in soil is not affected by an overprint of organic contamination (or any other type of contamination). The Geochemical Method is based on the identification of outliers that do not fit the background concentration population distribution. All of the techniques used in the Geochemical Method are capable of distinguishing between a population that represents background levels and a population that represents contamination regardless of the presence of organic contamination—even if the organic contaminant source also is a metals source. If metal contamination exists at a site, then the distinction between the population representing background concentrations and the population representing contamination can be detected by inspecting a probability plot and associated univariate plots. Similarly, geochemical association and enrichment of natural metals in soil are not affected by organic contamination. Background levels therefore can be distinguished from contamination by geochemical association analysis or geochemical enrichment analysis (see Section 3).

A rank plot can be used to evaluate the spatial distribution of metals relative to the distribution of organic chemicals at a site. As shown on the figure below, data points that represent samples with elevated concentrations of organic compounds are marked on the rank plot. Risk-based screening criteria such as U.S. EPA PRGs can be used to define elevated concentrations of organic compounds. The rank plot then can be used to compare data points that correspond to elevated levels of organic compounds with the rest of the population.

The rank plot shown here indicates that samples with elevated levels of organic compounds occur throughout the distribution. Elevated organic concentrations are associated with metal concentrations that represent background (the relatively flat left side of the rank plot), and with metal concentrations that represent contamination (the far right side of the rank plot). Therefore, the organic data shown on the example plot do not provide any additional information that can be used to make conclusions regarding background.

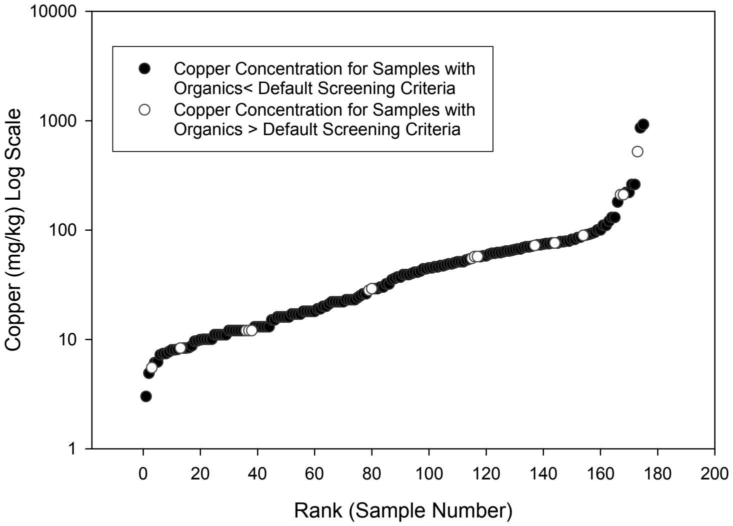


TABLE 2-3. Univariate plot categories and variables

Univariate Plot Category	Variables
Data qualifier	NQ, J, U, UJ
Sampling depth	Surface soil, subsurface soil
Soil type	Clay, sand, silt, etc.
Sampling location	AOC-1, POI-1, OU-1, etc.

AOC = area of concern.
J = estimated concentration.
NQ = not qualified.
OU = operable unit.
POI = point of interest.
U = nondetect.
UJ = nondetect estimated.

Data Qualifier Univariate Plot

Analytical data should be evaluated for quality first, in order to eliminate inferior data from subsequent analysis, and thus ensure that all decisions made during the analysis are based on a solid foundation. A data qualifier univariate plot provides a profile of the overall quality of the dataset, and classifies concentration values according to their reliability and usability. Data qualifiers are parameters used to indicate the quality of the data with respect to the established QC acceptance criteria. Data qualifiers are assigned during the data validation process in accordance with U.S. EPA (1994) *National Functional Guidelines for Inorganic Data Review*. Data qualifiers that should be considered during background metals analysis are listed in Table 2-4. Figure 2-9 shows an example data quality distribution plot for copper.

NQ and J values are preferred for background analysis; however, U and UJ values can be used if the guidelines listed below are observed. If concentration data are rejected as a result of serious QC deficiencies, an R qualifier is attached. No rejected data should be included in any phase of background analysis. The following guidelines should be used to select data for subsequent analysis:

1. U and UJ values for the metal can be plotted initially along with NQ and J values in the univariate plot; however, any U or UJ values that exceed the maximum NQ value should be eliminated from the univariate plot and subsequent evaluations.

2. Only NQ and J values should be used to make the final estimate of the background range for the metal. If U or UJ values fall above the level initially identified as the upper bound of the background range, they should be eliminated, the affected plots should be recreated without the eliminated values, and the background range should be reevaluated.

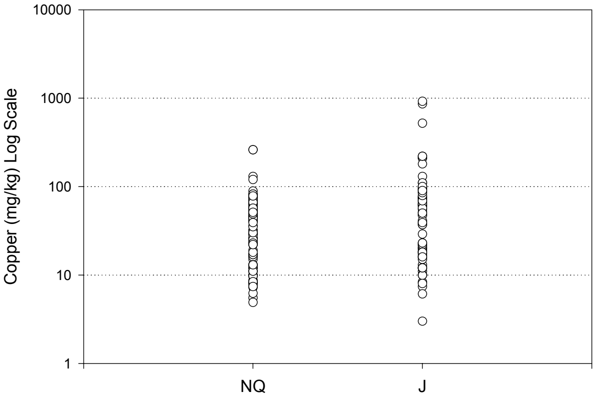


FIGURE 2-9. Univariate plot of copper concentrations vs. data qualifier (noncensored data)

TABLE 2-4. Data qualifiers used in background metals analysis

Qualifier	Definition	Explanation
NQ	Not qualified	All QC criteria associated with the analytical result were within acceptance criteria and the metal was quantified at a concentration above the laboratory reporting limit.
J	Estimated concentration	The associated concentration value is an estimated quantity.
U	Nondetect	The sample was analyzed for the metal, but the metal was not detected at a concentration above the associated value (either the sample quantitation limit or the sample detection limit).
UJ	Nondetect estimated	The sample was analyzed for the metal, but the metal was not detected. The associated value is an estimate and may be inaccurate or imprecise.

3. If most of the data are qualified UJ or U (as illustrated in Figure 2-10), the data should not be used to establish the background concentration range for the suspected COPC. In this case, it may be necessary to reanalyze samples using a laboratory analytical method that will yield lower reporting limits.

Procedures for plotting the data qualifier univariate distribution are presented in Box 2-12.

Sampling Depth Univariate Plot

A sampling depth univariate plot shows the metal concentration ranges that correspond to surface and subsurface soil (Figure 2-11). By inspecting the sampling depth univariate plot, investigators can reach the following preliminary conclusions:

- ❑ If concentrations in surface soil tend to be higher than in subsurface soil, then contamination should be suspected. Although differences in soil horizons and natural chemical and biological processes can cause metal concentrations to be higher in surface soils, contamination should be suspected if concentrations in surface soil are higher than in subsurface soil because chemical releases often occur above ground. For example, Figure 2-11 indicates that copper concentrations are higher in surface soil than in subsurface soil, suggesting that surface soil may be contaminated with copper. However, uptake of copper by surface vegetation and adsorption to organic particles in surface soil should be considered as a possible cause of the surface enrichment.

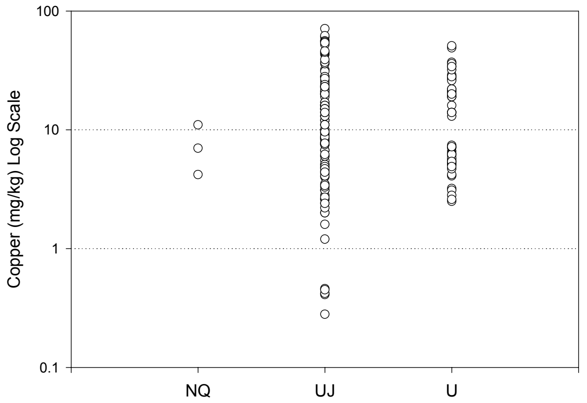


FIGURE 2-10. Univariate plot of copper concentrations vs. data qualifier (censored data)

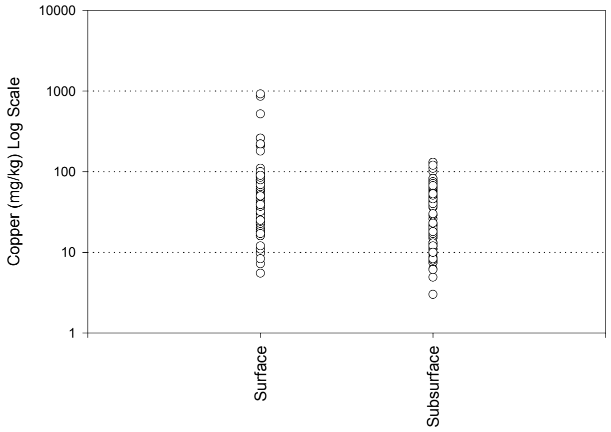


FIGURE 2-11. Univariate plot of copper concentrations vs. sampling depth

- ❑ If concentrations in subsurface soil tend to be higher than in surface soil, the metal is likely to be naturally occurring (unless an underground release/subsurface source is suspected, or soil conditions and mobility of the target metal allow penetration to deeper soils).

BOX 2-12. Procedures for plotting the data qualifier univariate distribution

1. Sort the data according to data qualifier.
2. Enter the data in columns, as shown in the table at right.
3. Create a "multiple Y data" scatter plot graph with data qualifiers along the x-axis and concentration along the y-axis.
4. Convert the y-axis scale to a common log scale (or perform log-transformation).

	NQ	J	UJ
Analytical Concentrations	100	10	5
	10	12	9
	95	95	10
	80	108	20
	20	98	
		96	
		20	

- If there is no significant difference between surface and subsurface soil concentrations, the metal is likely to be naturally occurring. This conclusion can be confirmed by inspecting the other univariate plots (geochemical association analysis and/or geochemical enrichment analysis also can be used for confirmation).

The "resolution" of the sampling method must be considered when segregating data according to sampling depth. Equipment used to collect boring soil samples may not be capable of collecting samples over discrete intervals less than 2 feet long. In addition, the boring action may mix soil from near the surface with deeper soils. Therefore, surface and subsurface soil samples typically are defined according to the following criteria:

- Soil samples collected between the surface and 0.5 foot below ground surface (bgs) with hand tools ("grab samples") are considered surface soil samples. Grab samples collected at depths greater than 0.5 feet bgs are considered subsurface soil samples.

- Soil samples collected from borings between the surface and 2 feet bgs are considered surface soil samples. Boring soil samples collected at depths greater than 2 feet bgs are considered subsurface soil samples.

Procedures for plotting the sampling depth univariate distribution are presented in Box 2-13.

In some circumstances, it may be useful to construct sampling depth univariate plots that segregate the data according to soil horizon or depth interval (e.g., 0-2, 2-4, 4-6, and 6-8 feet bgs). This approach is particularly useful at sites where the soil profile varies significantly with depth, or where soil is potentially impacted over a discrete subsurface depth interval. However, this type of univariate plot can be constructed only if sufficient data are available.

Soil Type Univariate Plot

A soil type univariate plot illustrates the metal concentration ranges that correspond to the different soil types that exist at the site. If the concentration ranges for a given metal vary significantly according to soil type, the metal distribution may reflect natural geochemical processes rather than metal contamination. As discussed in Section 2.1.3, if enrichment is due to natural geochemical processes, metals concentrations tend to be high in fine-grained soils such as clays and silts, and low in coarse-grained soils such as sands and gravels. The soil type univariate plot shown in Figure 2-12 indicates that the highest copper concentrations occur in the clay soil; therefore, the elevated copper concentrations are likely to represent natural background conditions.

A chemical release can impact any type of soil; therefore, if metal concentration ranges show relatively little variation in the different soils at a site, contamination should be suspected. For example, Figure 2-13 indicates that relatively high copper concentrations occur in the sandy gravel, silt, and gravelly sand; therefore, these elevated copper concentrations may be the result of a chemical release. Because concentrations of naturally occurring metals tend to be highest in soils with the smallest grain size, the maximum copper concentration detected in the clay can be used to

BOX 2-13. Procedures for plotting the sampling depth univariate distribution

1. Sort the data according to sampling depth (i.e., surface, subsurface).
2. Enter the data in columns as shown in the table at right.
3. Create a "multiple Y data" scatter plot graph, with sampling depth along the x-axis, and concentration along the y-axis.
4. Convert the y-axis scale to a common log scale (or perform log-transformation).

Analytical Concentrations

Surface	Subsurface
100	108
10	98
95	96
80	20
20	5
10	9
12	10
95	20

estimate the upper bound of the copper background concentration range.

Procedures for plotting the soil type univariate distribution are presented in Box 2-14.

When evaluating soil types, investigators should assess the geology of the site and the geochemical characteristics of the soil to determine whether soils in different areas of the site may be derived from different parent rock types. If this is the case, it may be appropriate to identify separate background ranges for each area (see Box 2-15).

Sampling Location Univariate Plot

A sampling location univariate plot illustrates the metal concentration ranges that correspond to different areas of a site. Concentration anomalies that correspond to individual areas may represent contamination.

The following assumptions are used to evaluate sampling location univariate plots:

- Each distinct area of the site has a different operational history; therefore, significant differences in metal concentrations among the areas may be related to the different activities conducted at each site. For example, in Figure 2-14, elevated copper concentrations occur at three areas (POI-05, -06, and -07). Because these areas were historically used as hazardous substance storage areas, the elevated copper concentrations are likely to represent contamination.

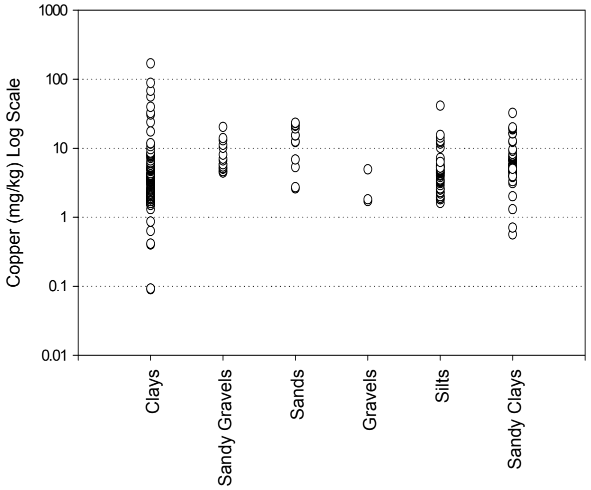


FIGURE 2-12. Univariate plot of copper concentrations vs. soil type (nonimpacted site)

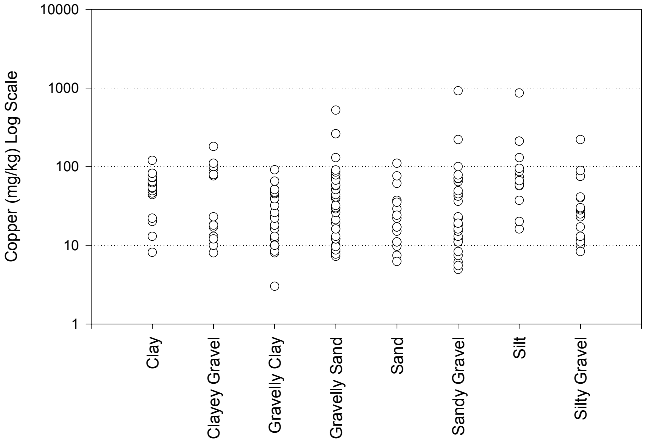


FIGURE 2-13. Univariate plot of copper concentrations vs. soil type (impacted site)

BOX 2-14. Procedures for plotting the soil type univariate distribution

1. Sort the data according to soil type (e.g., clay, clayey gravel, gravelly clay).
2. Enter the data in columns, as shown in the table at right.
3. Create a "multiple Y data" scatter plot graph, with soil type along the x-axis, and concentration along the y-axis.
4. Convert the y-axis scale to a common log scale (or perform log-transformation).

	Clay	Clayey Gravel	Gravelly Clay
Analytical Concentrations	100	10	5
	10	12	9
	95	95	10
	80	108	20
	20		30
	85		
	80		

BOX 2-15. Soils derived from different parent rocks: Are multiple background ranges required?

Background metal concentration ranges in soils derived from different parent rocks can differ widely, even after geochemical redistribution and enrichment. Therefore, investigators should be aware that certain sites might require special consideration: sites where soils derived from different types of parent rock occur in different areas of the site, and sites where the areas can be clearly differentiated in the field.

In such cases, it may be necessary to evaluate background for each area—just as, under the Comparative Method (Section 4), separate off-site background samples representative of each soil type would be collected and analyzed. Such sites are rare, however; even if site soils are derived from more than one type of parent rock, separate background estimates are usually not required. Unless the site is very large, soils derived from different parent rocks usually are not segregated into discrete areas large enough to necessitate separate background analyses.

An example of a site that may require separate analyses is one where soil in one area of the site is derived from weathered volcanic rocks, but the rest of the site contains soils derived from a coral reef formation. At such a site, background metal concentration ranges for the two areas can be very different. If the two soils occur in discrete areas, each with a significant areal extent relative to the overall area of the site, then the site should be divided into two "subsites," and a separate background range should be established for each area.

In some cases, it may be possible to identify separate data populations that represent background concentrations for distinct areas of the site based on analysis of a comprehensive (i.e., sitewide) dataset. For example, the probability plot may reveal the presence of several different populations. If each population can be correlated with a geographically and geologically distinct area of the site, it may be possible to estimate background ranges for each area. However, if the comprehensive analysis does not clearly show the separate populations, it may be necessary to segregate the data, and perform a separate analysis for each area.

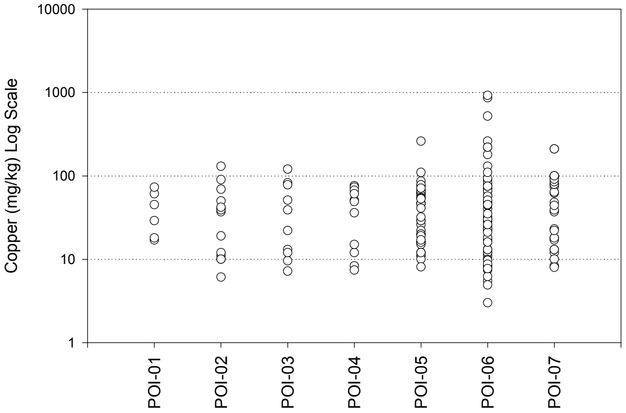


FIGURE 2-14. Univariate plot of copper concentrations vs. sampling location

- If a metal shows only small concentration differences among areas, the concentrations are likely to represent naturally occurring (background) levels. For example, in Figure 2-14, the maximum detected copper concentration at POI-01 differs little from concentrations at POI-02, -03, and -04.

Procedures for plotting the sampling location univariate distribution are presented in Box 2-16.

Probability Plots

Although probability plots can be constructed by plotting the data versus quantiles of the hypothesized distribution on standard graph paper (as

BOX 2-16. Procedures for plotting the sampling location univariate distribution

1. Sort the data according to sampling location (e.g., POI-01, POI-02, POI-03).
2. Enter the data in columns, as shown in the table at right.
3. Create a "multiple Y data" scatter plot graph with sampling location along the x-axis and concentration along the y-axis.
4. Convert the y-axis scale to a common log scale (or perform log-transformation).

	POI-01	POI-02	POI-03
Analytical Concentrations	100	10	5
	10	12	9
	95	95	10
	80	108	20
	20		30
			40
			50

shown in Box 2-8), or by manually plotting cumulative percentages on special probability paper (as shown in Box 2-9), probability plots for background analysis usually are plotted with the aid of a statistical software program. Concentrations of the suspected COPC metal are plotted with respect to the y-axis, and the cumulative percentages associated with the concentration values are plotted with respect to the x-axis. To facilitate comparison with the univariate plots, the scale used for the y-axis should match the scales used for the univariate plots. An example log-scale probability plot is shown in Figure 2-15.

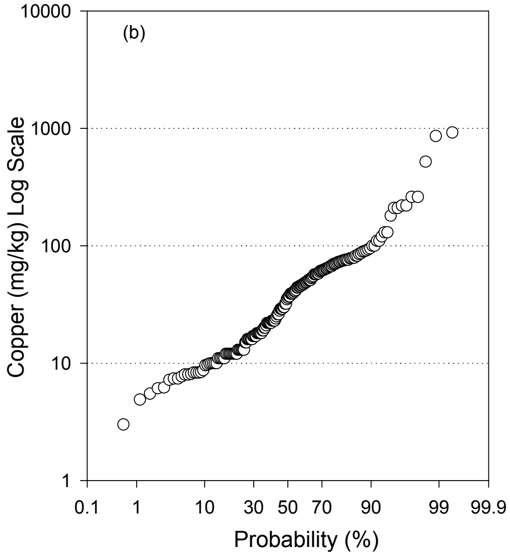


FIGURE 2-15. Log-scale probability plot of lead concentrations

A continuous straight-line plot with no large gaps indicates that the data fit the hypothesized distribution, and represent a single population—most likely a naturally occurring population. Significant deviations or data gaps indicate that more than one population exists at the site, suggesting that contamination may potentially be present. If, as shown in Figure 2-16, the data approximate a straight line for concentrations in the lower range, but concentrations in the upper range depart from the line (i.e., an increase in slope occurs), then the upper range of the concentration distribution represents a separate population. In this case, the lower-range concentrations are likely to represent natural (background) conditions, whereas the upper-range concentrations are likely to represent contamination. If data clustering or gaps do not confound the analysis, the point at which the slope changes (inflection point) is likely to represent the upper bound of the background concentration range (see Figure 2-16). This hypothesis should be evaluated by inspecting combined plots (described in the following subsection).

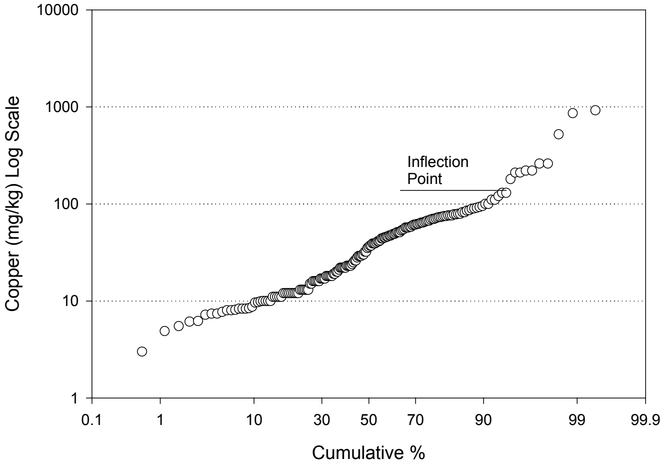


FIGURE 2-16. Log-scale probability plot of copper concentrations

Combined Plots

To combine univariate and probability plots, place them next to each other on the same page, with equivalent y-axis scales (Figure 2-17). The combined plots provide a comprehensive view of the characteristics of the dataset, and allow investigators to combine several lines of evidence to make decisions about the possible background range for the suspected COPC metal. Procedures for constructing the combined plots figure are presented in Box 2-17. Decision-making guidelines for identifying the range of concentrations that represent background are presented in Table 2-5.

Begin the combined plots analysis by inspecting the probability plot to identify an inflection point that may mark the upper bound of the background range. In the example shown on Figure 2-17, an inflection point is observed on the probability plot at approximately 130 mg/kg. At this point, it should be determined whether this initial estimate

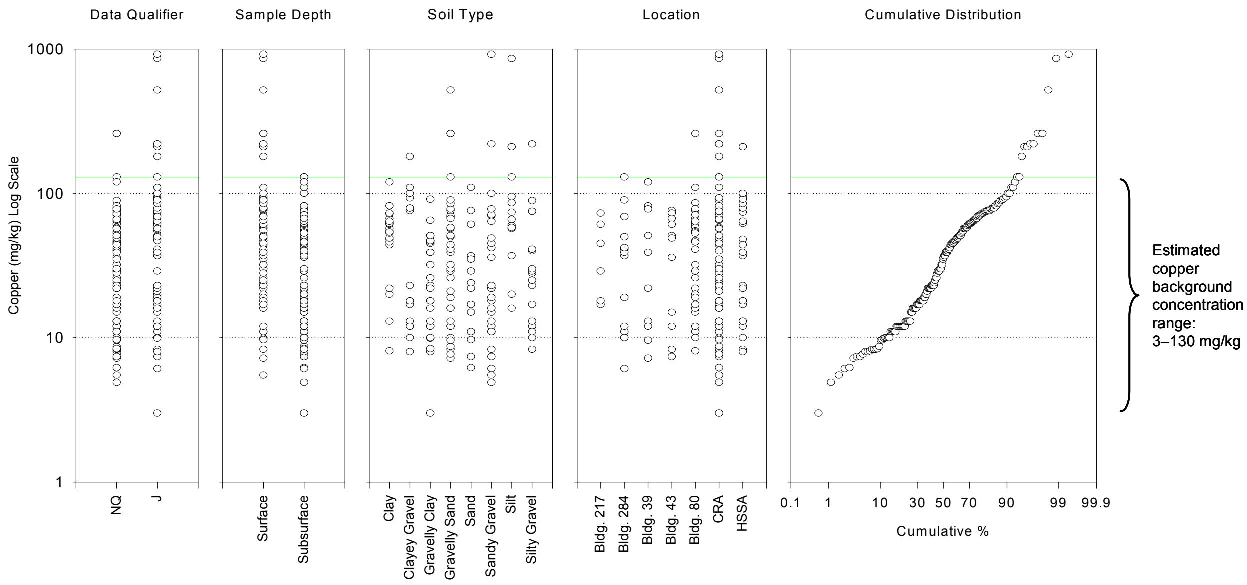


FIGURE 2-17. Combined plots (univariate and cumulative probability)

of the upper bound of the background concentration range is likely to represent the true upper bound of the background range at the site. The decision questions listed in Table 2-5 should be used to make this determination.

Potential problems encountered during the univariate and probability plot analysis include the following:

- Metals in natural soils may occur over wide concentration ranges (as noted in Section 2.1.3). If a wide natural concentration range exists at a site, detected concentrations that represent high background levels may be erroneously identified as outliers (i.e., the apparent upper bound of the background range would be too low).
- The dataset may be too small to produce a reliable estimate of the true range of background concentrations by univariate and probability plotting.
- At certain sites, metal contamination may be distributed relatively uniformly among sampling locations, and between surface and subsurface soils. However, this situation is not commonly encountered.

If the background range for a target metal can be estimated by analysis of the combined plots, the estimated range should be compared to the ranges reported for similar types of soil in the literature (see Section 2.1.3 and Appendix A). This comparison will help to ensure that a reasonable conclusion has been reached. However, if regional background

BOX 2-17. Procedures for creating the combined plots figure

1. After creating the univariate plots (data qualifier, sampling depth, soil type, and sampling location) and the probability plot, copy and paste all of the graphs onto one page. Typically, the page orientation is landscape.
2. Ensure that the y-axis scales are equivalent and align them horizontally across the page.
3. After all figures have been placed on the same page and the background concentration range has been estimated, mark the upper bound of the background range with a horizontal line that intersects all plots.

TABLE 2-5. Decision questions

Decision Question	Conclusion
1. Does the estimate of the upper bound of the background concentration range depend on any U or UJ values?	If the answer is yes, then U and UJ values in the vicinity of the upper bound of the estimated background concentration range should be eliminated, the plots should be recreated, and the analysis will have to be repeated. If the answer is no, then the data qualifier univariate plot supports the conclusion that the probability plot inflection point represents the upper bound of the background range. In the Figure 2-17 example, all copper concentrations are either NQ or J values and thus were retained for use in the data qualifier univariate plot; the background range conclusion is based on reliable and high quality data.
2. Does the sampling depth univariate plot indicate that concentrations vary according to sampling depth?	If the answer is yes (as in Figure 2-17), then the sampling depth univariate plot provides evidence to support the conclusion that the inflection point at 130 mg/kg represents the upper bound of the background range. As noted above, unless an underground release is suspected, surface soil is more likely to be contaminated. In the Figure 2-17 example, the maximum concentration detected in subsurface soil corresponds to the inflection point on the probability plot (130 mg/kg), and is likely to represent the upper bound of the background concentration range. Concentrations above 130 mg/kg are likely to represent contamination. If the answer is no (i.e., there are no significant differences between surface and subsurface soil concentrations), then the metal is likely to be naturally occurring, and the maximum concentration detected at the site is likely to represent the upper bound of the background range.
3. Does the soil type univariate plot indicate that relatively high concentrations tend to occur only in certain types of soil at the site?	If the answer is yes (i.e., relatively high metals concentrations tend to occur only in certain types of soil at the site), then the maximum concentration detected in the fine-grained soil can be used to estimate the upper bound of the background range. If the answer is no, then the soil type univariate plot supports the conclusion that the probability plot inflection point represents the upper bound of the background range. The soil type univariate plot shown in Figure 2-17 indicates that copper concentrations above 130 mg/kg occur in five of the eight different soil types (i.e., the elevated concentrations are not related to soil type). Because a chemical release can impact any of the different types of soil that occur at a site, the concentrations above 130 mg/kg are likely to be the result of a chemical release. In addition, the highest copper concentration was detected in coarse soil material (sandy gravel). However, as discussed in Section 2.1.3, the highest concentrations of naturally occurring metals tend to occur in fine-grained soils (e.g., clays). This provides further evidence that concentrations above 130 mg/kg are above the background range.
4. Does the sampling location univariate plot indicate that relatively high concentrations occur only in certain areas of the site?	If the answer is yes, then the sampling location univariate plot supports the conclusion that the probability plot inflection point represents the upper bound of the background range. In the Figure 2-17 example, copper concentrations above the estimated background range occur at only three of seven locations. The three POIs were suspected areas of copper contamination. If the answer is no (i.e., relatively high metals concentrations do not tend to occur only in certain areas of the site), then the soil is not likely to be contaminated. The maximum concentration detected at the site is therefore likely to represent the upper bound of the background range.

ranges are used for comparison, characteristics of the site soils (e.g., grain size and geochemistry) must be similar to the regional soil used for comparison.

2.2.5 Determine Acceptability of Background Ranges

After completing the analysis outlined above, investigators must decide whether the analysis has produced technically defensible and reliable estimates of the background concentration ranges for the suspected COPC metals. The following criteria should be considered:

- ❑ The identified background ranges must be derived according to technically defensible procedures, and must be supported by adequate data.
- ❑ The identified background ranges must be consistent with the known physical characteristics of the site (e.g., geology and geochemistry) and must be deemed appropriate from a remedy decision perspective.
- ❑ The identified background ranges must be acceptable to all stakeholders as representative of natural or anthropogenic ambient conditions.

If the above conditions are met, then the background analysis is completed for the target metal: the background concentration range has been established and should be documented for use in subsequent site-specific human health and ecological risk evaluations.

If the above conditions are not met, further background analysis is warranted. Section 2.3 discusses the methods that may be appropriate for further background analysis, and the criteria that should be used to identify an appropriate method (or methods).

2.3 Identification of Appropriate Background Methods

If the data analysis procedures described above have not produced technically defensible and reliable estimates of the background concentration ranges for the suspected COPC metals, the project team should identify appropriate methods for further analysis. As discussed in Section 1.2, background analysis methods can be divided into two main classes.

- ***Geochemical Method:*** This method consists of a set of tools for analyzing site data using geochemical principles. These tools are used to evaluate the relationships between naturally occurring background chemicals that tend to occur together as a result of geochemical processes, such as weathering and natural enrichment. Prime examples include naturally occurring inorganic chemicals that often are detected in groups. The co-presence of such background chemicals is governed by the chemical composition of the underlying parent rocks, and the geochemical processes that occur during and after soil formation.
- ***Comparative Method:*** This method is specifically designed to compare site data to data from a reference (or background) area. The statistical tests used in the Comparative Method are aimed primarily at evaluating whether on-site concentrations of the target chemicals are statistically similar to reference area concentrations. Adequate background sampling data are required for this purpose.

In many instances, depending on the nature of the investigated chemicals, both methods may be appropriate. The project team may choose to implement the Geochemical Method alone, or in conjunction with Comparative Method. Some of the key characteristics of these two classes of background analysis are listed in Table 2-6.

The Geochemical Method is particularly likely to be successful if evidence suggests that geochemical processes such as weathering, precipitation, or sorption to fine-grained soil particles control the concentrations of the target metals at the site. The method also can be applied to any set of correlated naturally occurring or anthropogenic chemicals that can be attributed to background sources. The Geochemical Method uses various bivariate statistical techniques to characterize the features and interrelationships of the targeted soil chemicals, and can identify background concentration ranges by evaluating on-site data only.

The Comparative Method is able to distinguish between COPCs and background chemicals by comparing the on-site dataset to a reference area

TABLE 2-6. Features of background analysis methods

Feature	Method	
	Geochemical Method	**Comparative Method**
Scientific/technical basis	Geochemical correlations	Statistical two-sample tests
Target chemicals	Correlated background chemicals (e.g., naturally occurring inorganics)	Naturally occurring and anthropogenic chemicals
Data needs	On-site dataset	On-site and reference area datasets
Challenges	Stakeholder acceptance of geochemical relationships	Identifying and delineating suitable reference areas; and demonstrating the representativeness of the reference area dataset

dataset. The method is capable of processing both naturally occurring and anthropogenic chemicals. Upon identification of COPCs, other statistical procedures or measures (such as probability plots or pth percentiles) may be used to identify background ranges for the targeted chemicals.

The feasibility and appropriateness of the Geochemical and Comparative Methods depend upon a number of factors, which are discussed in the following section. Table 2-7 provides some practical rules for identifying the most appropriate background analysis method(s).

2.3.1 Feasibility and Applicability

At many sites, either method will be successful; however, due to the difficulties associated with identifying and collecting data from a suitable reference area, the Comparative Method will not be feasible for some sites. The Geochemical Method, therefore, is particularly useful when reference area data are not available.

The Geochemical Method can identify background concentration ranges by evaluating correlated background chemicals using on-site data only (i.e., a separate [reference area] dataset is not required). If portions of a site have been impacted, the Geochemical Method can identify background ranges by distinguishing between the two populations (background concentrations and above-background concentrations). If the site has not been impacted by a chemical release, the Geochemical Method can show that only one population exists at the site, and that the range of this population is the background range at the site. Finally, geochemical association analysis (part of the Geochemical Method) is capable of identifying background ranges even if the entire site is impacted by a chemical release.

The Comparative Method also can be successful for sites that are nonimpacted, partially impacted, or entirely impacted. However, the Comparative Method is applicable to a broader range of chemicals, including both naturally occurring and anthropogenic background chemicals that may or may not be correlated. The Comparative Method does not assume any correlation among the targeted chemicals. The method requires data from a reference area that is geochemically and anthropogenically similar to the site. In some instances, due to physical, geographical, or cost constraints, it will not be possible to identify or adequately sample a suitable reference area. Furthermore, the Comparative Method requires that both the site and background area are adequately characterized, and that the areal extent of the background area, as well as the representativeness of the datasets, are acceptable to all stakeholders. Without such stakeholder concurrence, the Comparative Method is not likely to be successful.

Table 2-8 provides practical rules for evaluating the applicability of each method to analysis of organic and inorganic background chemicals.

TABLE 2-7. Practical rules for identification of appropriate background analysis methods

Conditions Favorable for the Geochemical Method	Conditions Favorable for the Comparative Method
• The targeted chemicals are likely to be correlated with naturally occurring chemicals (e.g., metals). • The site includes both potentially impacted and nonimpacted areas. • Adequate reference area data are not available due to physical, geographical, or budgetary constraints. • Regulatory acceptance of the method is expected due to precedent, guidance, or pre-approval.	• The targeted background chemicals are either naturally occurring or anthropogenic. • The site can be either completely or partially impacted. • A reference area that is geochemically and anthropogenically similar to the site can be identified. • The spatial extent of the reference area has been defined, and all stakeholders agree that the reference area is appropriate for the comparison. • There are no time or budgetary constraints that would prevent additional site and background investigations. • Regulatory acceptance of the method is expected due to precedent, guidance, or pre-approval.

TABLE 2-8. Practical rules for applicability of background analysis methods

Class of Chemicals	Method Applicability	
	Geochemical Method	**Comparative Method**
Organic	Applicable when the co-presence of organics is attributable to background sources or processes.	Applicable when adequate data from a nonimpacted reference area are available.
Inorganic	Applicable when the co-presence of inorganics is attributable to geochemical processes.	Applicable when adequate data from a nonimpacted reference area are available.

2.3.2 Cost/Benefit Analysis

Both methods require significant effort, cost, and time to implement. For example, the Geochemical Method will require additional computational effort and regulatory interaction. The Comparative Method involves effort to identify and characterize a suitable reference area, as well as additional computational effort and regulatory interaction.

Decision-makers should assess whether the potential benefits of the selected method justify its associated costs. In general, the benefits of background analysis are most likely to justify the associated costs when a site is nonimpacted or only marginally impacted.

2.3.3 Regulatory Acceptance

The selected method must be acceptable to all stakeholders. The most favorable conditions are: (a) the stakeholders promote the use of the selected method; (b) the stakeholders have accepted the selected method for similar sites; (c) the stakeholders do not insist on the use of alternative, overly conservative procedures; and (d) the stakeholders are willing to accept the decision process prior to sampling and analysis. Examples of U.S. EPA and state technical guidance on background analysis procedures are presented in Section 1.

2.4 References

Adriano, D.C., Z.-S. Chen, and S.-S. Yang, eds. 1997. *Biogeochemistry of Trace Metals.* Hanover, NH: I.K. Iskander, Cold Regions Research and Engineering Laboratory. September.

Akritas, M.G., T.F. Ruscitti, and G.P. Patil. 1994. Statistical Analysis of Censored Environmental Data. Pp. 221–242 in *Handbook of Statistics, Volume 12, Environmental Statistics*, eds. G.P. Patil and C.R. Rao. New York: North-Holland.

American Public Health Association (APHA). 1998. *Standard Methods for the Examination of Water and Wastewater.* 20th ed. Washington, DC.

American Society of Civil Engineers (ASCE). 1990a. Review of Geostatistics in Geohydrology, I. Basic Concepts. Prepared by ASCE Task Committee on Geostatistical Techniques in Geohydrology, S. Rouhani, Chairman and Principal Author. *ASCE Journal of Hydraulic Engineering* 116(5): 612–632.

American Society of Civil Engineers (ASCE). 1990b. Review of Geostatistics in Geohydrology, II. Applications. Prepared by ASCE Task Committee on Geostatistical Techniques in Geohydrology, S. Rouhani, Chairman and Principal Author. *ASCE Journal of Hydraulic Engineering* 116(5): 633–658.

American Society for Testing and Materials (ASTM). 2001a. *D2216-98 Standard Test Method for Laboratory Determination of Water (Moisture) Content of Soil and Rock by Mass*. West Conshohocken, PA.

American Society for Testing and Materials (ASTM). 2001b. *D2488-00 Standard Practice for Description and Identification of Soils (Visual-Manual Procedure)*. West Conshohocken, PA.

American Society for Testing and Materials (ASTM). 2001c. *D2937-00 Standard Test Method for Density of Soil in Place by the Drive-Cylinder Method*. West Conshohocken, PA.

American Society for Testing and Materials (ASTM). 2001d. *D422-63(1998) Standard Test Method for Particle-Size Analysis of Soils*. West Conshohocken, PA.

American Society for Testing and Materials (ASTM). 2001e. *D4318-00 Standard Test Methods for Liquid Limit, Plastic Limit, and Plasticity Index of Soils*. West Conshohocken, PA.

American Society for Testing and Materials (ASTM). 2001f. *E1527-00 Standard Practice for Environmental Site Assessments: Phase I Environmental Site Assessment Process*. West Conshohocken, PA.

American Society for Testing and Materials (ASTM). 2001g. *E1903-97 Standard Guide for Environmental Site Assessments: Phase II Environmental Site Assessment Process*. West Conshohocken, PA.

Conover, W.J. 1998. *Practical Nonparametric Statistics*. 3rd ed. Wiley Series in Probability and Statistics, Applied Probability and Statistics Section. New York: Wiley & Sons.

Department of the Navy (DON). 1998. *Procedural Guidance for Statistically Analyzing Environmental Background Data.* San Diego: DON SWDIV and EFA West. September.

Department of the Navy (DON). 1999. *Handbook for Statistical Analysis of Environmental Background Data.* San Diego: DON SWDIV and EFA West. July.

Federal Geographic Data Committee (FGDC). 1997. *Vegetation Classification Standard.* FGDC-STD-005. Reston, VA: U.S. Geological Survey Federal Geographic Data Committee, Vegetation Subcommittee.

Florida Department of Environmental Regulation (FDER). 1988. *A Guide to the Interpretation of Metal Concentrations in Estuarine Sediments.* Eds. S.J. Schropp and H.L. Windom. Tallahassee, FL. April.

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring.* New York: Van Nostrand Reinhold.

Grossman, D.H., and D. Faber-Langendoen. 1998. *International Classification of Ecological Communities: Terrestrial Vegetation of the United States; Vol. I, The National Vegetation Classification System: Development, Status, and Applications.* Arlington, VA: The Nature Conservancy.

Helsel, D.R., and R.M. Hirsch. 1992. *Statistical Methods in Water Resources*. New York: Elsevier.

Isaaks, E.H., and R.M. Srivastava. 1989. *Applied Geostatistics*. Cambridge, England: Oxford University Press.

Journel, A.G., and C.J. Huijbregts. 1978. *Mining Geostatistics*. London: Academic Press.

Kabata-Pendias, A. 2001. *Trace Elements in Soils and Plants.* 3rd ed. Boca Raton, FL: CRC Press, LLC.

Matheron, G. 1971. *The Theory of Regionalized Variables and Its Applications*. Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, No. 5.

Michael, J.R., and W.R. Schucany. 1986. Analysis of Data from Censored Samples. Pp. 461–496 in *Goodness-of-Fit Techniques*, eds. R.B. D'Agostino and M.A. Stephens. New York: Marcel Dekker.

O'Neil, L.J., and A. Hill, eds. 2000. *Guidelines for Mapping Vegetation on Army Installations.* ERDC TR-00-8, U.S. Army Engineer Research and Development Center, Vicksburg, MS, and SFIM-AEC-EC-CR-99006, Army Environmental Center, Aberdeen Proving Ground, MD.

Singh, A., A.K. Singh, and G. Flatman. 1994. Estimation of Background Levels of Contaminants. *Mathematical Geology* 26(3).

Singh, A.K., A. Singh, and M. Englehardt. 1997. *The Lognormal Distribution in Environmental Applications.* EPA Technology Support Center Issue. EPA/600/R-97/006. U.S. EPA, Office of Research and Development and Office of Solid Waste and Emergency Response. December.

United States Environmental Protection Agency (U.S. EPA). 1992a. Options for Addressing High Background Levels of Hazardous Substances at CERCLA Sites. Draft Final Issue Paper. Office of Emergency and Remedial Response. January.

United States Environmental Protection Agency (U.S. EPA). 1992b. *Supplemental Guidance to RAGS: Calculating the Concentration Term*. EPA Publication 9285.7-081. May.

United States Environmental Protection Agency (U.S. EPA). 1994. *USEPA Contract Laboratory Program National Functional Guidelines for Inorganic Data Review.* EPA/540/R-94/013 (PB94-963502). Office of Emergency and Remedial Response. February.

United States Environmental Protection Agency (U.S. EPA). 1997a. *Data Quality Evaluation Statistical Toolbox (DataQUEST), User's Guide, EPA QA/G-9D, QA96 Version.* EPA/600/R-96/085. Washington, DC: Office of Research and Development. December.

United States Environmental Protection Agency (U.S. EPA). 1997b. *Ecological Risk Assessment Guidance for Superfund: Process for Designing and Conducting Ecological Risk Assessment.* Interim final. EPA/540/R-97/006. Office of Solid Waste and Emergency Response. June.

United States Environmental Protection Agency (U.S. EPA). 2000a. *Guidance for Choosing a Sampling Design for Environmental Data Collection: Use in the Development of a Quality Assurance Project Plan*. EPA QA/G-5S. Peer Review Draft. Washington, DC: Quality Staff, Office of Environmental Information.

United States Environmental Protection Agency (U.S. EPA). 2000b. *Guidance for Data Quality Assessment: Practical Methods for Data Analysis. EPA QA/G-9, QA 00 Update.* EPA/600/R-96/084. Washington, DC: Office of Environmental Information. July.

United States Environmental Protection Agency (U.S. EPA). 2000c. *Guidance for the Data Quality Objectives Process. EPA QA/G-4.* EPA/600/R-96/055. Washington, DC: Office of Environmental Information. August.

United States Environmental Protection Agency (U.S. EPA). 2001. *SW-846 On-Line, Test Methods for Evaluating Solid Waste, Physical/Chemical Methods*. URL: <http://www.epa.gov/epaoswer/hazwaste/test/sw846.htm>. Office of Solid Waste.

United States Geological Survey (USGS). 2001. *Online Map Lists.* URL: <http://mapping.usgs.gov/mac/maplists.html>.

3. GEOCHEMICAL METHOD

The Geochemical Method uses statistical techniques based on geochemical principles to analyze on-site data and identify background metal concentration ranges. The techniques graphically and numerically distinguish between metal concentrations that reflect natural background conditions and concentrations that may represent a chemical release. Physical, chemical, geological, geochemical, and biological processes that affect metal concentrations in soils are carefully considered during the analysis to ensure that the calculated background ranges accurately reflect natural background conditions.

The Geochemical Method usually requires only data from the suspected contaminated site, which typically represent a combination of “impacted” and “nonimpacted” areas. Chemical concentrations detected in nonimpacted areas of the site represent background conditions. In most cases, no reference area or additional off-site sampling is necessary.

Also, a chemical release normally will impact only certain areas of a site; however, on rare occasions, the entire site may be impacted. In this event, additional (off-site) sampling will be required to define the extent of the site-related chemical release. The additional data then can be used to identify background concentration ranges. It should be noted that geochemical association analysis (Section 3.2), in conjunction with the data review and assessment procedures presented in Section 2, can identify background ranges when a chemical release affects the entire site and no off-site data are available.

The Geochemical Method is based primarily on the following fundamental geological and geochemical observations:

- ❑ Aluminum, iron, calcium, and magnesium are major components of the minerals that form the rocks of the earth’s crust, and therefore are major constituents of nearly all soils. Unless a release is suspected at a site, concentrations of these metals will most likely be within the background range. This can be confirmed by the background analysis methods presented in this section, or by comparison to typical concentration ranges presented in the literature (see Section 2.1.3 and Appendix A).
- ❑ Based on a wide range of observations, metal concentrations in natural rocks and soils generally tend to be lognormally distributed. However, it should be noted that the Geochemical Method can be used even if the data are not lognormally distributed. As long as outliers that do not fit the overall population distribution can be recognized, the background range can be identified.
- ❑ Based on geological and geochemical principles and observation, certain groups of metals tend to occur together in natural rocks and soils (i.e., they exhibit elemental association) (see Section 2.1.3).

The Geochemical Method is particularly useful when it is not possible to identify and collect background data from a reference area. The method has been successfully used at Naval installations in Hawaii and Guam, and in other locations where it is difficult or impossible to find suitable reference areas.

3.1 Overview of Geochemical Method

The Geochemical Method includes two general techniques: (1) geochemical association analysis and (2) geochemical enrichment analysis. This sequence of techniques is appropriate for most sites; however, it should not constrain the project team. For example, if investigators have reason to believe that geochemical enrichment analysis is likely to be the most successful approach, it can be done first. Additionally, it may be appropriate to use both techniques and combine the results to provide more than one line of evidence to stakeholders.

1. ***Geochemical association analysis.*** Geochemical association analysis is usually the

first step of the Geochemical Method. This technique is used to identify the background range by evaluating the association relationship between two metals. The scatter plots constructed during this step generally plot suspected COPC metal concentrations against concentrations of non-COPC metals. If the plots show a high degree of correlation between the two metals, it can be used to evaluate background. High metal concentrations that do not fit an observed strong relationship are likely to represent contamination.

2. ***Geochemical enrichment analysis.*** If the results of geochemical association analysis are inconclusive or unacceptable to stakeholders, geochemical enrichment analysis is usually the next step. Natural enrichment (enrichment associated with natural geochemical processes such as weathering, adsorption, and precipitation) can be distinguished from "unnatural enrichment," (enrichment associated with a chemical release) by calculating geochemical enrichment factors and constructing a probability plot.

After completing the analysis, the investigators and stakeholders must decide whether a technically defensible and reliable estimate of the background concentration range has been identified for each suspected COPC metal. The estimated range must be acceptable to all stakeholders. After an acceptable background range has been established, it should be documented for use during the human health and ecological risk assessments.

Use of a robust statistical software program (e.g., SigmaPlot or Statview; see Appendix D.3) is strongly recommended for the Geochemical Method. Although standard spreadsheet software offers some statistical functions, such programs typically lack many of the features that facilitate production of the required graphs.

All graphs should be constructed by plotting metal concentrations or enrichment factors with respect to a log scale (for the y-axis). (If it is necessary to plot using a linear scale, then log transformation should be applied to the concentrations or enrichment factors before they are plotted.) Metal concentrations in natural soils are often lognormally distributed (Gilbert, 1987); therefore, the log scale tends to make outliers (i.e., data points that do not fit the background population distribution) more visible. In addition, use of a log scale yields more compact and readable graphs than can be produced using a linear scale.

3.2 Geochemical Association Analysis

Geochemical association analysis is usually based on the association of metals identified as suspected COPCs (e.g., lead, arsenic, cadmium) with non-COPC metals (e.g., aluminum, iron, calcium). However, the relationship between two suspected COPC metals also can be used. (Metals that tend to occur together in natural rocks and soils are discussed in Section 2.1.3 and Appendix A.)

Construction of a scatter plot (see Figure 3-1) is the first step in the evaluation. The method can be used with as few as three data points (i.e., three concentration values for each of the two metals).

Although the results of this method are not as dependent upon sample size as other statistical methods, large datasets are more likely to allow investigators to reach clear and technically defensible conclusions than small datasets. Figure 3-1

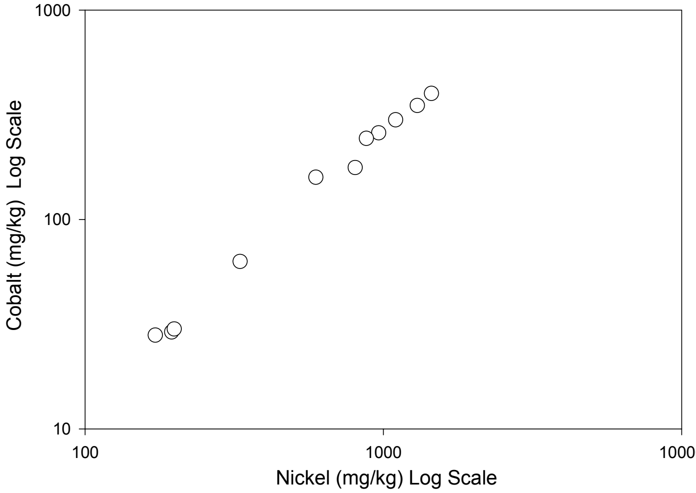


FIGURE 3-1. Scatter plot of nickel vs. cobalt concentrations (log scale) showing strong association

shows a log-log nickel vs. cobalt scatter plot. Strong correlation exists if the data tend to occur along or near a straight line. Linear regression (least-square regression) analysis can be used to evaluate the geochemical relationship, identify outliers that may represent contamination, and estimate the background concentration range.

As discussed in Section 2.1.3, the distribution of metals in natural soils depends primarily on the source rock, weathering processes, and surface adsorption phenomena. The most important factors that cause metal association in natural (nonimpacted) soils are summarized as follows:

- Certain groups of metals are closely associated due to their atomic structures (i.e., electron valence states) and chemical properties in nature. For example, the chemical properties of cobalt and nickel are very similar, and they are highly associated with each other in certain types of rocks, such as basalt.

- Distribution of trace metals is controlled by major chemical constituents of soil, including aluminum, iron, and calcium. Soil is composed chiefly of particles of chemically resistant minerals such as quartz and clay minerals—the secondary minerals formed by weathering of the parent rock. Adsorption of metals to fine grained soil particles—primarily clay and colloidal organic material—can result in elevated metal concentrations in natural soils. Because the highly sorptive clays contain high concentrations of aluminum, natural metal concentrations often show a high degree of correlation with aluminum (see Figure 3-2). Weathering of rocks with high iron concentrations (e.g., basalts) results in the formation of stable clay minerals and iron oxides that also tend to adsorb other metals; therefore, in many soils, natural metal concentrations also show a high degree of correlation with iron. The natural relationships between non-COPC reference metals (particularly aluminum and iron) and metals identified as suspected COPC metals for a site should be evaluated first. Follow-on evaluation of the relationships between suspected COPC metals then can be conducted.

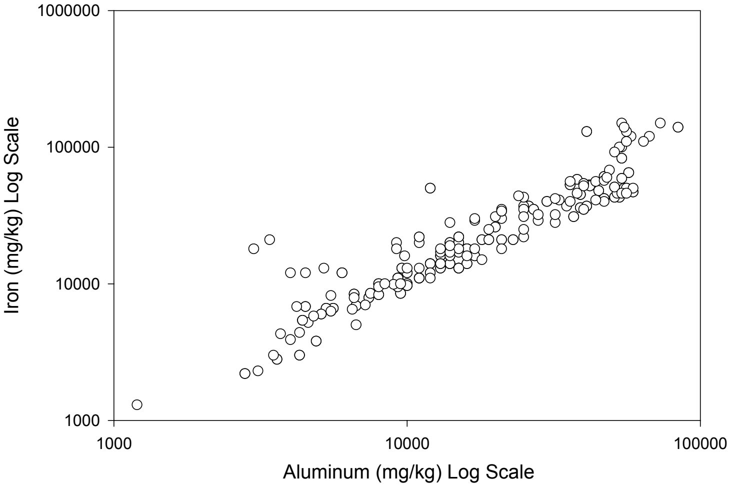


FIGURE 3-2. Scatter plot of aluminum vs. iron concentrations (log scale) showing strong association

3.2.1 Geochemical Regression

Geochemical regression involves the following elements:

1. A log-log scatter plot is constructed (see Box 3-1). An individual data point on a scatter plot

BOX 3-1. Procedures for constructing a scatter plot

1. Select the response variable (i.e., the suspected COPC metal) for which the background range is to be evaluated.
2. Select the reference variable (usually a non-COPC metal). Information regarding site soil types and geochemistry should be used to select the reference variable.
3. Mark corresponding values of the reference variable on the x-axis and values of the response variable on the y-axis.
4. Plot the data point corresponding to each sample on the graph.

represents the detected concentrations of two metals in a single sample. Concentrations of the reference variable (usually a non-COPC metal) are plotted with respect to the x-axis, and concentrations of the response variable (the suspected COPC metal) are plotted with respect to the y-axis.

2. Linear regression is used to draw the best-fit straight line through the data points. The slope and y-intercept of the best-fit line define the expected relationship between the two metals. By defining and plotting the expected relationship (i.e., the best-fit line), outliers can be identified, and the background range can be estimated. The strength of the association relationship is evaluated to ensure that conclusions are based on a strong geochemical association. The strength of the relationship is evaluated by examining the amount of data scatter around the best-fit line, and is quantified by the correlation coefficient.

It should be noted that statistical associations are overall tendencies, not ironclad rules. Geochemical association relationships between metals must be identified on the basis of the geochemical characteristics of the site. Outliers associated with elevated concentrations of the suspected COPC metal are likely to represent contamination.

The reference variable used in geochemical regression analysis can be thought of as a controlling variable. The response variable (the suspected COPC metal concentration) is controlled by (i.e., is a function of) the reference variable. For example, a sample with a high aluminum (reference metal) concentration is likely to have a high concentration of the suspected COPC metal because the degree of sorption onto clay particles (aluminosilicates) controls the concentration of the suspected COPC metal. Generally, it is assumed that all suspected COPC metals are response variables and non-COPCs (e.g., aluminum, iron, calcium) are reference variables. If a suspected COPC metal is strongly associated with a non-COPC, the metal also should be considered a non-COPC; therefore, the relationship between suspected COPC metals and non-COPCs should be evaluated first. If two suspected COPC metals are associated, the geochemical factors that may result in their association in the natural environment should be evaluated.

Scatter Plot Characteristics

The scatter plot provides a graphic representation of the characteristics and strength of the relationship between two metals. Potential site-specific geochemical relationships should be considered when selecting reference metals for scatter plotting. In order to establish a reliable and defensible estimate of the background concentration range, a strong geochemical relationship between two metals should be identified. It may be necessary to construct two or more scatter plots using different reference metals in order to identify a strong relationship that can be used to estimate the background range for the suspected COPC metal. Knowledge of natural geochemical conditions and relationships in the soil is required to select suitable reference metals (see Section 2.1.3 and Appendix A). Visual inspection of the example scatter plot shown in Figure 3-3 reveals a strong relationship between the response variable (chromium) and the reference variable (aluminum).

To examine and interpret the scatter plot, look for the *overall pattern* of the relationship and for any striking *deviations* (i.e., outliers) from the pattern.

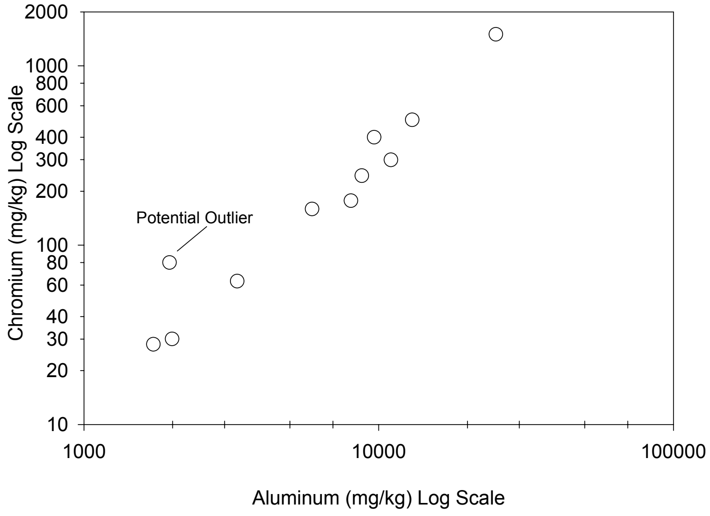


FIGURE 3-3. Scatter plot of aluminum vs. chromium concentrations (log scale) showing strong association

The overall pattern can be described in terms of *form* (e.g., a linear relationship), *direction* (e.g., the slope of a the best-fit line through the data), and *strength* (e.g., the degree of correlation). An example of an outlier is shown on Figure 3-3. The graph shows a relatively consistent relationship between chromium and aluminum for all but one data point. The form of the chromium-aluminum relationship shown in Figure 3-3 is linear, and the direction is positive (i.e., chromium increases as aluminum increases), so the strength of the relationship is relatively high.

Figure 3-4 is an example of a scatter plot with eight outliers. The data points that fit the overall pattern are likely to represent natural concentrations of the suspected COPC metal, whereas the outliers are part of a separate population and are likely to represent contamination. As shown in Figure 3-5, when the outliers are removed, the geochemical association relationship is very clear. The highest concentration that fits the linear relationship represents the estimated upper bound of the background concentration range.

When a scatter plot displays a linear relationship, the overall pattern can be described by drawing a straight line through the data points by linear (least-squares) regression (see Figure 3-6 and Box 3-2). The slope of the regression line is a function of the correlation coefficient, *r*. For background analysis, the correlation coefficient is used primarily to quantify the strength of the relationship between the reference and response variables. However, the scatter plot also is necessary to evaluate the strength of the relationship: A linear relationship is strong if most of the data points lie close to the regression line, whereas the relationship is weak if they are widely scattered about the line.

Figure 3-6 is an aluminum vs. chromium scatter plot showing the least-squares regression line representing the relationship between the two variables. The correlation coefficient in this example, 0.942, is relatively high, indicating a strong relationship (i.e., most of the data points are close to the regression line). The significance of the correlation coefficient is summarized in Box 3-3. Although the plot itself does not contain any information that would indicate which variable controls the other, an understanding of the underlying geochemical principles suggests that aluminum is the reference variable (i.e., high concentrations of chromium exist in certain soil samples because of their high aluminum concentrations). This strong association between aluminum (a non-COPC metal) and chromium (the suspected COPC metal) provides strong evidence that the high chromium concentrations observed in this dataset represent natural background conditions. Therefore, the maximum concentrations likely represent the upper bound of the background concentration range.

Figure 3-7 shows a dataset with a high correlation coefficient (0.9393). The regression coefficient associated with the regression line is the square of the correlation coefficient; in this case, $r^2 = 0.8823$. Least-square regression minimizes the distances of the regression line from the data points in the *y* (target metal) direction. The r^2 value represents the fraction of the variation in the target metal concentration values that is explained by the least-square regression.

Figure 3-7 also shows two intervals centered on the regression line: a 95% confidence interval, and a 95% prediction interval. A statistical software program (e.g., SigmaPlot; see Appendix D.3) can be used to calculate confidence and prediction intervals.

The confidence interval also is known as the confidence interval for the regression line. The true mean value of *y* (e.g., target metal concentration) that corresponds to a given value of *x* (e.g., reference metal concentration) has a 95% probability of being within the 95% confidence interval (Draper and Smith, 1981).

The prediction interval also is called the confidence interval for the population. A 95% prediction interval is the range within which the *y* value (e.g., target metal concentration) that corresponds to a given *x* value (e.g., reference metal concentration) is predicted to fall 95% of the time (Hahn and Meeker, 1991). For example, the prediction interval shown on Figure 3-7 indicates that, if a sample contains 10,000 mg/kg of aluminum, there is a 95% probability that the chromium concentration is between 15 and 50 mg/kg.

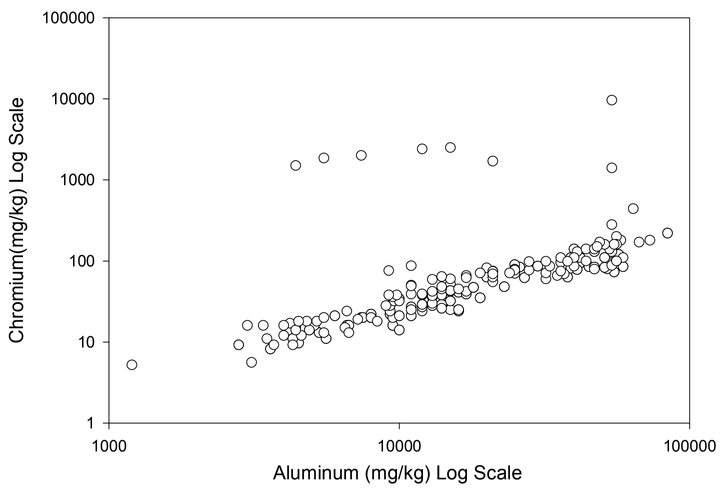


FIGURE 3-4. **Scatter plot of aluminum vs. chromium concentrations (log scale) showing unclear association**

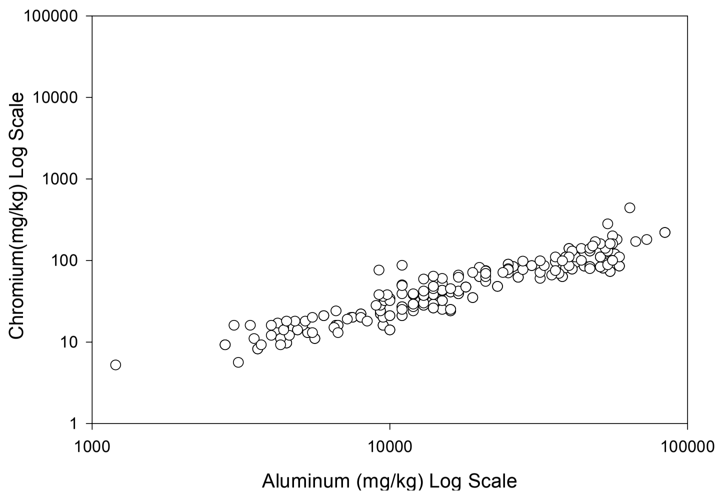


FIGURE 3-5. **Scatter plot of aluminum vs. chromium concentrations (log scale) with outliers removed**

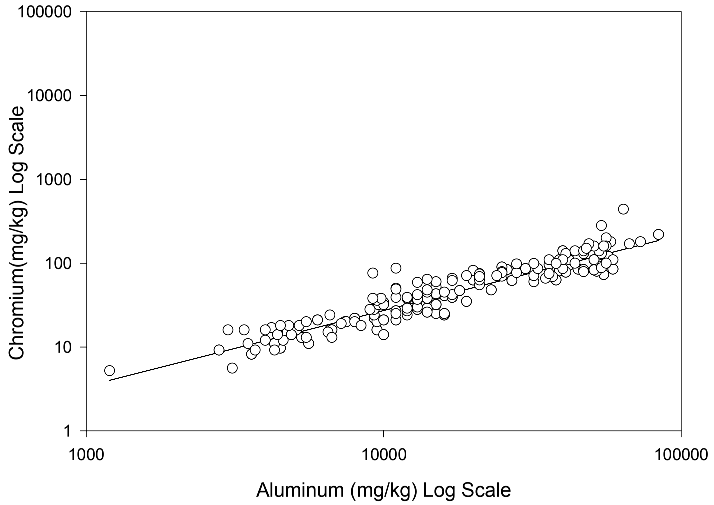


FIGURE 3-6. **Scatter plot of aluminum vs. chromium concentrations showing least-squares regression line**

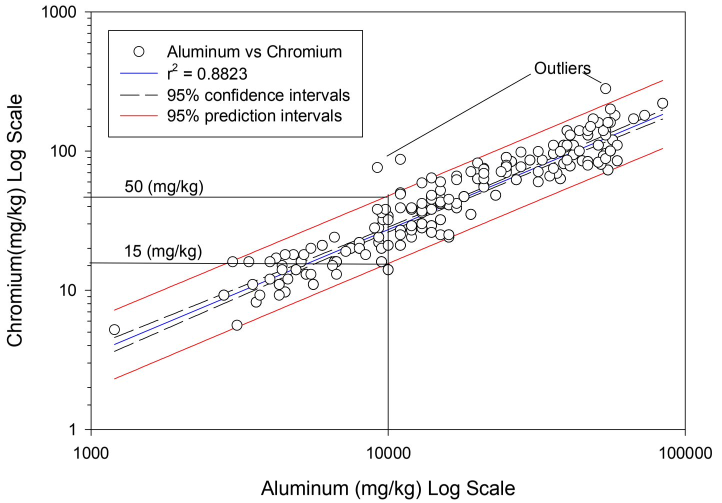


FIGURE 3-7. **Geochemical regression: aluminum vs. chromium concentrations (log scale)**

BOX 3-2. Linear (least-squares) regression

When a scatter plot displays a linear relationship, the overall pattern can be described by drawing a straight line through the data points. Of course, no single straight line will pass exactly through all data points. Fitting a line to the data means drawing a line that comes as close as possible to the points. The straight line that most closely fits the data is plotted by least-squares regression. The best-fit line provides a quantitative description of the relationship between two metals that can be used to estimate the background range of the suspected COPC metal.

Suppose that y is the concentration of a response metal (plotted on the y-axis) and x is the concentration of a reference metal (plotted on the x-axis). A straight line relating log (y) to log (x) has an equation of the form:

$$\log (y) = a + b \log (x)$$

In this equation, b is the slope, i.e., the change in log (y) per unit change in log (x). The y-intercept is a, i.e., the value of log (y) when log (x) = 0. A straight line describing the relationship between aluminum and chromium concentrations has the form:

$$\log (Cr) = a + b \log (Al)$$

where: Cr = the chromium concentration
Al = the aluminum concentration.

The equation of the regression line can be used to predict the value of the response variable, y, that corresponds to each value of the reference variable, x. The accuracy of predictions depends on how close the data lie to the regression line. A regression line that makes these prediction errors as small as possible is preferred. The least-squares method is the most common method used to fit a line to the data. The least-squares method minimizes the sum of the squares of the vertical distances between the data points and the line.

The equation of a least-squares regression line also can be expressed in terms of the mean and standard deviation of each dataset, and the correlation coefficient.

The slope can be expressed as: $b= r(s_y/s_x)$
The y-intercept can be expressed as: $a = \mu_y - b\mu_x$

where: μ_x and s_x are the mean and standard deviation of the log (x) distribution
μ_y and s_y are mean and standard deviation of the log (y) distribution
r is the correlation coefficient, a measure of the strength of the relationship between log (x) and log (y).

A computer can be used to quickly find the equation of the least-squares regression line, and plot the line on the scatter plot. A calculator with a regression function also can be used to find the equation of the least-squares regression line.

After calculating the correlation coefficients for each pair of metals evaluated by the geochemical regression analysis method, the coefficients can be tabulated to determine which pairs of metals show the highest correlation. An example of the resulting table, or correlation matrix, is shown in Table 3-1. If there is strong correlation between two metals, the relationship is likely to be very useful for geochemical association background analysis. In addition, strong correlation between a suspected COPC metal and a non-COPC reference metal is a strong indication that the suspected COPC metal is present at background levels.

3.2.2 Decision Point

After completing the analysis, investigators must decide whether the analysis has produced a technically defensible and reliable estimate of the background concentration range for the suspected COPC metal. The estimated range must be acceptable to all stakeholders.

❑ If the answer is yes, then the background analysis is completed for the metal: the background concentration range has been

BOX 3-3. Correlation coefficient

The correlation coefficient, *r,* is a dimensionless quantity that provides a measure of the strength of the linear relationship between two quantitative variables (e.g., the concentration of a reference metal, and the concentration of a response metal). Correlation does not depend upon the distinction between reference and response variables.

For a dataset that contains *n* observations of each of two concentration variables *v* and *w*:

$$(v_1,w_1), (v_2, w_2),\ldots, (v_n,w_n)$$

Log-transformation yields the following dataset:

$$(x_1, y_1), (x_2, y_2),\ldots, (x_n, y_n)$$

where: $x_i = \log (v_i)$, and $y_i = \log (w_i)$

The equation of the correlation coefficient associated with this dataset is:

$$r = \Sigma\ [(x_i - \mu_x)(y_i - \mu_y)]/(n - 1)\ s_x s_y$$

where: μ_x is the mean of the x data distribution, μ_y is the mean of the y data distribution, s_x is the standard deviation of the x data distribution, and s_y is the standard deviation of the y data distribution

Although this equation can be used to manually calculate the value of the correlation coefficient, the calculation is usually performed automatically with the aid of a computer (or calculator). The significance of the different possible values of the correlation coefficient is summarized below:

- The value of *r* is restricted to the range between –1 and +1. If *r* is positive, the response variable tends to increase as the reference variable increases. In this case, the slope of the regression line is positive. If *r* is negative, the response variable tends to decrease as the reference variable increases. In this case, the slope of the regression line is negative. Relationships that show negative correlation should not be used in the Geochemical Method to establish background concentrations.
- Extreme values ($r = -1$ or $r = +1$) occur only when a dataset exhibits perfect linear correlation (i.e., when every data point lies exactly on the regression line). Perfect correlation between two different metals will never be encountered in background analysis. High *r* values indicate very strong correlation; therefore, the associated relationship can be used to estimate a technically defensible background concentration range.

established and should be documented for use in subsequent site-specific human health and ecological risk evaluations.

If the answer is no, investigators may decide to proceed with geochemical enrichment analysis (Section 3.3), or the Comparative Method (Section 4).

3.3 Geochemical Enrichment Analysis

The techniques used for geochemical enrichment analysis are based on the assumption that natural processes tend to enrich or deplete metal concentrations in soil relative to its parent rock according to a consistent pattern. Concentrations of a suspected COPC metal that do not fit the pattern are likely to represent contamination. Geochemical enrichment analysis may be successful when spatial and geochemical association analyses yield inconclusive or unacceptable results.

The enrichment factor (*E*) corresponding to each soil sample is defined as the ratio of the suspected COPC metal concentration in the sample to the normalizing metal (usually a non-COPC metal) concentration in the sample, divided by the same ratio in the parent rock. Metal ratios for the parent rock usually are determined by referring to published metal concentration data. However, if necessary, site-specific metal concentrations and ratios can be established by whole rock analysis.

For example, to evaluate background levels of chromium in soils derived from basalt, the enrichment factor is expressed as:

$$E = (C_{Cr}/C_{Al})_{sample}/(C_{Cr}/C_{Al})_{basalt}$$

TABLE 3-1. Example correlation coefficient matrix

	Aluminum	Antimony	Arsenic	Barium	Beryllium	Cadmium	Calcium	Chromium	Cobalt	Copper	Iron	Lead	Magnesium	Manganese	Mercury	Nickel	Potassium	Selenium	Silver	Sodium	Thallium	Vanadium	Zinc
Aluminum	1.000	−0.298	0.306	−0.121	0.030	0.059	−0.713	0.781	−0.036	−0.042	0.718	−0.126	0.593	−0.058	−0.265	0.562	0.487	0.400	0.199	0.093	0.034	0.677	−0.214
Antimony		1.000	−0.006	0.264	0.320	0.186	0.317	−0.295	−0.042	0.203	−0.184	0.198	−0.303	0.000	0.220	−0.160	−0.187	0.021	0.376	0.332	0.395	−0.236	0.626
Arsenic			1.000	−0.209	−0.015	0.069	−0.199	0.238	−0.096	−0.114	0.269	−0.041	0.138	−0.005	−0.225	0.151	0.014	0.186	0.169	−0.032	0.110	0.113	0.047
Barium				1.000	0.019	−0.170	0.024	−0.348	0.788	0.371	0.191	0.068	−0.150	0.787	0.253	0.436	0.315	−0.123	−0.045	0.036	−0.015	0.138	0.364
Beryllium					1.000	0.626	0.187	0.060	−0.149	−0.012	0.064	−0.017	−0.219	−0.089	−0.246	−0.154	−0.036	0.333	0.641	0.715	0.720	0.121	0.046
Cadmium						1.000	−0.018	0.167	−0.188	−0.168	0.048	−0.028	−0.102	−0.126	−0.146	−0.227	−0.231	0.295	0.618	0.710	0.519	0.079	−0.057
Calcium							1.000	−0.594	−0.148	0.071	−0.724	0.013	−0.423	−0.104	0.157	−0.464	−0.342	−0.127	−0.068	0.017	0.156	−0.651	0.162
Chromium								1.000	−0.290	−0.174	0.597	−0.105	0.369	−0.303	−0.250	0.232	0.155	0.339	0.112	0.050	0.023	0.620	−0.281
Cobalt									1.000	0.226	0.253	0.068	−0.009	0.899	0.215	0.590	0.255	−0.182	−0.105	−0.054	−0.082	0.194	0.060
Copper										1.000	0.105	0.060	−0.062	0.215	0.110	0.177	0.181	−0.021	−0.032	0.023	−0.011	0.083	0.241
Iron											1.000	0.013	0.185	0.249	−0.200	0.515	0.511	0.201	0.125	0.100	0.018	0.897	−0.097
Lead												1.000	−0.148	0.045	0.034	−0.037	−0.068	−0.021	−0.042	−0.008	−0.007	0.020	0.096
Magnesium													1.000	−0.114	−0.013	0.495	0.320	0.046	−0.031	−0.105	−0.221	0.127	−0.230
Manganese														1.000	0.116	0.502	0.199	−0.090	−0.030	−0.047	0.001	0.173	0.114
Mercury															1.000	−0.025	−0.103	−0.327	−0.253	−0.129	−0.264	−0.220	0.271
Nickel																1.000	0.526	0.011	0.012	−0.042	−0.050	0.462	−0.005
Potassium																	1.000	0.240	−0.025	−0.021	−0.066	0.555	−0.059
Selenium																		1.000	0.395	0.259	0.418	0.325	−0.057
Silver																			1.000	0.726	0.791	0.100	0.054
Sodium																				1.000	0.638	0.112	0.080
Thallium																					1.000	0.058	0.109
Vanadium																						1.000	−0.179
Zinc																							1.000

Note: Correlation coefficients ≥0.5 (in shaded boxes above) warrant further evaluation; however, correlation coefficients <0.5 may require further evaluation if the corresponding elemental associations are expected due to local or regional geochemical characteristics. Because the strength of an association relationship cannot be determined from the correlation coefficient alone, correlation matrices should be used only in conjunction with scatter plots, univariate plots, and probability plots.

where: $(C_{Cr}/C_{Al})_{sample}$ = the ratio of the chromium concentration detected in a sample to the aluminum (normalizing metal) concentration detected in the same sample.

$(C_{Cr}/C_{Al})_{basalt}$ = the ratio of the average chromium concentration in the basalt parent rock to the average aluminum concentration in the basalt parent rock.

Therefore, the enrichment factor corresponding to a particular sample is a measure of the number of times a suspected COPC metal is enriched in the soil relative to basalt. The enrichment factor will equal 1 if no enrichment has occurred during soil formation. If natural enrichment has occurred through processes such as chemical precipitation or surface adsorption, enrichment factors will be >1. Metal contamination also will result in enrichment factors >1. Depletion may result in enrichment factors <1 for the suspected COPC metal.

By plotting enrichment factors on a probability graph, investigators can evaluate the pattern of enrichment, and identify outliers than represent “unnatural” enrichment (i.e., contamination).

3.3.1 Enrichment Ratio Comparison

The first step in the analysis is to investigate the geology of the site and the soil characteristics to identify the parent rock from which the soil is derived. After identifying the parent rock, average literature values for the concentration of the suspected COPC metal and the normalizing metal should be identified to determine the parent rock metal ratio. (If actual metal concentration data for the parent rock beneath or near the site are available, they should be used in place of literature values.) The suspected COPC metal and normalizing metal ratios in site soil samples are calculated, and the data are used to construct a probability plot.

Procedures for enrichment factor calculation and plotting are presented in Box 3-4.

The procedures presented in Box 3-4 were used to create the data listed in Table 3-2 and the probability plot shown in Figure 3-8. Chromium is the suspected COPC metal, and aluminum is the normalizing metal. Basalt was identified as the source rock. According to published data, the average chromium concentration for basalt is 114 mg/kg, and the average aluminum concentration is 79,400 mg/kg (Krauskopf and Bird, 1995). Therefore, the average $[C_{Cr}/C_{Al}]_{basalt}$ ratio is 0.0014. Although chromium concentrations may not be higher in the soil than in the basalt parent rock, the enrichment factors ($[C_{Cr}/C_{Al}]_{sample}/[C_{Cr}/C_{Al}]_{basalt}$) indicate that the soil is enriched in chromium relative to the parent rock.

BOX 3-4. Procedures for enrichment factor calculation and plotting

1. Select the suspected COPC metal for which the background range is to be evaluated.
2. Select the appropriate parent rock type (e.g., granite, basalt, limestone).
3. Select the non-COPC normalizing metal (e.g., aluminum).
4. Determine the average concentrations of the suspected COPC metal and the normalizing metal in the selected parent rock, and calculate the metal ratio for the parent rock (e.g., $[C_{Cr}/C_{Al}]_{basalt}$).
5. Calculate the metal ratios for the site soil samples (e.g., $[C_{Cr}/C_{Al}]_{sample1,}$ $[C_{Cr}/C_{Al}]_{sample\ 2,}$....$[C_{Cr}/C_{Al}]_{sample\ n}$)
6. Calculate the enrichment ratio for each sample (e.g., $[C_{Cr}/C_{Al}]_{sample}/[C_{Cr}/C_{Al}]_{basalt}$).
7. Arrange the data in order of increasing enrichment factor ratios.
8. Calculate the cumulative percentage that corresponds to each data point, and construct a probability plot using the paired enrichment factor–cumulative percentage data (see Section 2.2.3).
9. Evaluate the pattern shown on the probability plot: identify outliers that may represent contamination, and estimate the background concentration range.

TABLE 3-2. Enrichment factor probability plot data table

Chromium (mg/kg)	Aluminum (mg/kg)	$[Cr/Al]_{sample}$	$[Cr/Al]_{basalt}$	Enrichment Factor[(a)]	Cumulative Percentage
30.00	1,990.00	0.02	0.0014	10.77	5.56
28.00	1,720.00	0.02	0.0014	11.63	11.11
63.00	3,310.00	0.02	0.0014	13.60	16.67
177.00	8,050.00	0.02	0.0014	15.71	22.22
248.00	9,500.00	0.03	0.0014	18.65	27.78
159.00	5,940.00	0.03	0.0014	19.12	33.33
299.00	11,000.00	0.03	0.0014	19.42	38.89
244.00	8,770.00	0.03	0.0014	19.87	44.44
456.00	14,000.00	0.03	0.0014	23.27	50.00
500.00	13,000.00	0.04	0.0014	27.47	55.56
80.00	1,950.00	0.04	0.0014	29.30	61.11
400.00	9,650.00	0.04	0.0014	29.61	66.67
1,500.00	25,000.00	0.06	0.0014	42.86	72.22
976.00	15,000.00	0.07	0.0014	46.48	77.78
3,200.00	27,000.00	0.12	0.0014	84.66	83.33
4,500.00	22,000.00	0.20	0.0014	146.10	88.89
9,600.00	24,000.00	0.40	0.0014	285.71	94.44

(a) $[Cr/Al]_{sample}/[Cr/Al]_{basalt}$.

3.3.2 Enrichment Factor Analysis

High enrichment factors that deviate from the overall pattern depicted on the probability plot are likely to represent contamination. For example, results of the enrichment factor analysis shown in Table 3-2 and Figure 3-8 indicate that chromium has been enriched in site soils by factors of up to approximately 286. The inflection point (increase in slope) on the probability plot marks the approximate upper bound of the background concentration range. Therefore, the upper bound of the estimated background concentration range is 1,500 mg/kg. Three chromium concentration values (3,200, 4,500, and 9,600 mg/kg) are associated with enrichment factors that do not fit the overall population distribution; therefore, these elevated concentrations are likely to represent contamination.

3.4 Determination of Acceptable Background Ranges

After completing the analysis, investigators must determine whether the estimated background concentration ranges for each suspected COPC metal are technically defensible, reliable, and acceptable to all stakeholders. If the answer is yes, then the background analysis is completed, and the ranges should be documented for use in the site-specific human health and ecological risk evaluations.

If background cannot be determined for a suspected COPC metal after completing spatial analysis as outlined in Section 2.2.4 and both steps

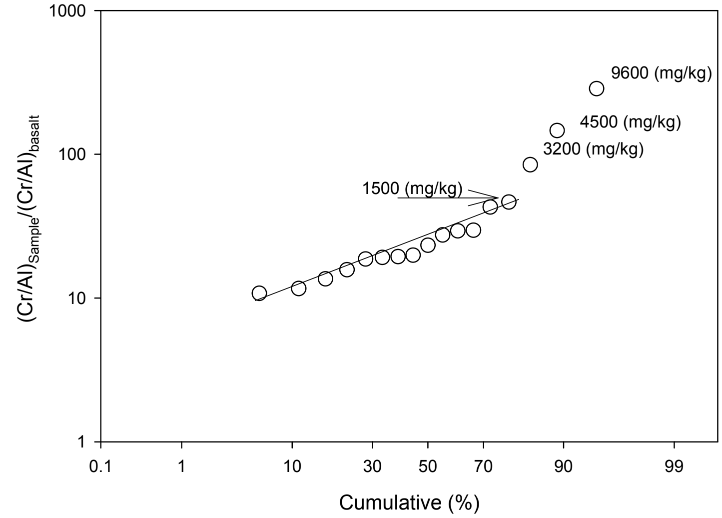


FIGURE 3-8. Probability plot of geochemical enrichment factors

outlined in this section, and the information is critical for completion of the risk assessment, it may be necessary to collect additional data. Additional samples may be collected from the site and/or surrounding area to provide data for further spatial or geochemical analysis, or reference area data may be collected and analyzed by the Comparative Method. If the inconclusive results are associated a high percentage of U or UJ data points in the dataset, it may be necessary to reanalyze samples by a method that will yield lower reporting limits. In some cases, a comprehensive evaluation of additional parameters, such as vegetation patterns, soil and plant biology, soil weathering patterns, and other soil science characteristics may provide additional evidence needed to justify elimination of a chemical from the list of COPCs, refine the estimated background range, or confirm conclusions of the analysis.

3.5 References

Draper, N.R., and H. Smith. 1981. *Applied Regression Analysis*. 2nd ed. New York: John Wiley & Sons.

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring.* New York: Van Nostrand Reinhold.

Hahn, G.J., and W.Q. Meeker. 1991. *Statistical Intervals, A Guide for Practitioners*. New York: John Wiley & Sons.

Krauskopf, K.B., and D.K. Bird. 1995. *Introduction to Geochemistry.* 3rd ed. New York: McGraw-Hill.

4. COMPARATIVE METHOD

4.1 Overview of Comparative Method

The Comparative Method involves the use of statistical hypothesis tests to compare various properties of site datasets and background datasets. The bases of these tests are hypotheses concerning the presence of naturally occurring or anthropogenic chemicals at the site. These hypotheses must be supported by available site-specific information on the nature of potential releases at the site. (Examples of such hypotheses are provided in Section 2.) The use of the Comparative Method requires datasets that represent background or reference (i.e., nonimpacted) areas. Ideally, both site and background datasets should be unbiased and representative of geochemically and anthropogenically similar domains. Furthermore, the two datasets should be nearly of the same size. However, in practice, site datasets usually are large and biased or clustered toward areas of concern, whereas background datasets are small and randomly collected. In such cases, site datasets may have to be declustered prior to statistical testing (see Section 2.2.).

If background datasets do not exist, then appropriate reference areas must be identified for soil sampling. This sampling process must be conducted in accordance with DQO principles. The scope of the background chemical analyses also must be adequately comprehensive in order to demonstrate the geochemical and anthropogenic similarity of the site and background areas. For this purpose, additional data parameters may be measured (see Section 2.1). Finally, concurrence of various stakeholders on the appropriateness and representativeness of the background dataset is necessary for a successful application of the Comparative Method.

4.1.1 Definition and Purpose of Comparative Statistical Tests

This section provides detailed instructions for computing descriptive statistics and conducting graphic and statistical analyses to determine if concentrations of chemicals in soil at a site are significantly elevated relative to concentrations in ambient (local) nonimpacted background areas. If chemicals are present at elevated concentrations, the chemicals are declared to be COPCs and are carried forward to subsequent risk calculations. Uncertainty in these decisions resulting from a limited number of soil samples (due to inevitable resource constraints) is accounted for by using statistical tests of hypotheses. The key questions that are addressed in this section are:

- ❑ What statistical procedures or tests should be used to determine if a chemical is a COPC?
- ❑ What two testing approaches should be avoided in order to not falsely conclude that a chemical is a COPC?
- ❑ How is the number of samples to be taken for the selected statistical test determined?
- ❑ How is the selected statistical test performed?

4.1.2 Common Comparative Statistical Tests

An initial, tentative selection of the most appropriate statistical test(s) should be made during the DQO planning process. This selection should be based on the number of samples required for the various tests to achieve the specified performance goals (DQOs), the particular distribution (normal or lognormal) expected of the data to be collected, and information in published statistical papers that demonstrate the performance of the candidate tests for various data distributions and contamination scenarios. However, after the new data have been collected and the preliminary graphic and distribution data analyses have been conducted (as discussed in Section 2), a final selection of the statistical test(s) can be made.

The assumptions and advantages and disadvantages of each of the tests discussed in this chapter are provided in Table 4-1 as a guide for users of this document to select the most appropriate

TABLE 4-1. Assumptions and advantages/disadvantages of statistical tests to detect when site concentrations tend to be larger than background concentrations

Test Statistic	Objectives/Assumptions	Advantages	Disadvantages
Slippage Test	• Objective is to test for differences in the right tail (largest values) of the site and background concentration distributions. • More less-than values are allowed than for other tests considered here. • At least one detected (quantified) background measurement is present and it is larger than the largest less-than value. • No assumptions are required with regard to the shape of site and background data concentration distributions.	• Very simple to conduct the test. • No distribution assumptions are necessary. • Many less-than values are permitted. • Can be used in conjunction (in tandem) with tests that focus on the detecting differences in the mean or median.	• May require large number of measurements to have adequate power to detect differences in site and background concentrations.
Quantile Test	• Objective is to test for differences in the right tail (largest values) of the site and background concentration distributions. • Less-than values are not among the largest *r* data values in the pooled set of site and background data. • No assumptions are required with regard to the shape of the site and background data concentration distributions.	• Relatively simple to conduct the test. • No distribution assumptions are necessary. • Can have more power to detect differences in the right tail of site and background distributions than tests like the WRS, Gehan, or two-sample t tests that focus on the mean or median. • Can be used in conjunction (in tandem) with tests that focus on detecting differences in the mean or median.	• May require large number of measurements to have adequate power to detect differences in site and background concentrations. • Test may be inconclusive if less-than values are present among the largest *r* data values.
Wilcoxon Rank Sum Test	• Objective is to test for differences in the medians of the site and background populations. • Only one detection limit (all less-than values have the same value), which is less than the smallest detected datum. • No more than 40% of both the site and background datasets are less-than values. • The site and background data concentration distributions have the same shape (variance).	• No distribution assumptions necessary. • In general, the test has more power to detect shift in site median than the two-sample t tests when the site and background data distributions are asymmetric (skewed to the right, to high concentrations). • Can be used in conjunction (in tandem) with Slippage and Quantile tests so that differences in the right tails of the site and background distributions, as well as differences in medians, can be detected.	• Relatively more complex to compute by hand. • Too many less-than values prevent use of the test.
Gehan Test	• Objective is to test for differences in the medians of the site and background populations. • Less-than values do not have the same value (multiple detection limits exist). • The censoring mechanism that generated the less-than values is the same for the site and background populations. • No assumptions are required with regard to the shape of the site and background data concentration distributions.	• Can be used when multiple detection limits are present. • Same advantages as for the WRS test.	• Relatively complicated to compute by hand. • The performance of the test is not known as well as that of the WRS test. • Must assume the same censoring mechanisms apply to the site and background data.

TABLE 4-1. Assumptions and advantages/disadvantages of statistical tests to detect when site concentrations tend to be larger than background concentrations (continued)

Test Statistic	Objectives/Assumptions	Advantages	Disadvantages
Two-Sample Test of Proportions	• Objective is to test for differences in the proportions of the site and background data above a given cutoff level. • Test may be used when more than 50% of the site or background datasets are less-than values. • No assumptions are required with regard to the shape of the site and background data concentration distributions.	• No distribution assumptions are necessary. • Relatively simple test to perform. • Can be used when many less-than values are present.	• A test based on proportions may not be what is really needed (e.g., it may be more appropriate to test for differences in means or medians).
Two-Sample t Test	• Objective is to test for differences in the means of the site and background populations. • Both site and background mean concentrations are normally distributed. • Less-than-values have no significant impact on computed means (e.g., less than 15% of measurements are below detection). • The site and background data distributions have the same shape (variance).	• Most powerful test for detecting a shift in the site mean from the background mean, if the site and background data are normally distributed.	• The test requires a statistical evaluation of the assumption of equal total variances for the site and background populations. • In general, the power will be less than that of the WRS test, if the data are not normally distributed. • Normal distribution assumption is often violated. • The results of the test can be affected by outliers. • Not well suited for datasets that contain less-than values.
Satterthwaite Two-Sample t Test	• Objective is to test for differences in the means of the site and background populations. • Both site and background data have a normal distribution. • No less-than values are present. • Site and background data distributions are expected or known to not have the same shape (variance).	• Test can be used when the site and background distributions have unequal variances.	• The test is relatively complicated to compute by hand. • Same disadvantages as for the two-sample t test.

statistical test(s). In this regard, note that the optimal selection of a test depends in part on whether:

- ❑ The entire distribution of the observed measurements from the site is expected to be simply shifted to higher values than the observed distribution of background measurements (i.e., the difference between the two datasets can be assessed by comparing their mean or median concentrations), or
- ❑ The true concentrations in relatively small areas at the site are expected to be elevated relative to the true background concentrations, in which case only a small portion of the distribution of site measurements would be expected to be shifted to higher concentrations than the distribution of background measurements (i.e., the difference between the two datasets can be assessed by comparing their highest concentrations).

For the case of a simple shift, the two-sample t test, the Satterthwaite two-sample t test, the Wilcoxon Rank Sum (WRS) test, and the Gehan test are preferred. However, the Slippage test, the Quantile test, and the two-sample test of proportions are better suited to identify chemicals that have elevated concentrations in only small areas at the site. Main features of these tests are discussed in Table 4-1 and later parts of this section.

Basic Assumptions of the Recommended Tests

All tests discussed in this section require that site and background measurements be independent (spatially or temporally uncorrelated) and representative of the underlying site and background populations. As discussed in Section 2.1.5, this assumption requires that (1) an appropriate probability-based sampling design strategy be used in order to determine the location of soil samples to be collected, and (2) the soil samples be far enough apart in space and time that spatial and temporal correlations among concentrations at different locations are not present.

In many instances, the background analyses must rely on existing site data, which are typically collected by judgmental sampling (e.g., sampling of locations where contamination is believed likely to occur based on information regarding the site history). Such datasets, however, are often biased, clustered, and correlated. Use of such measurements may entail overestimating soil chemical concentrations. Isaaks and Srivastava (1989, Chapter 10) provide a thorough introduction on the effects of the clustered data, as well as various declustering techniques.

In cases where excessive overestimation is anticipated, the existing clustered dataset should be declustered for subsequent analyses. A variety of declustering alternatives exist (Isaaks and Srivastava, 1989, Chapter 10). These techniques include cell declustering, in which a regular grid is used to divide the investigated area into equal cell areas. The data within each cell then are replaced using one of the following methods:

- ❑ Average data point: The data points within the cell are replaced by their average data point, which has a value equal to the average concentration of measured values within the cell. The location of this point is defined as the average easting and northing of the data points within the cell.
- ❑ Maximum data point: Only the data point with the highest measured value within each cell is retained.
- ❑ Randomly selected data points: Only a fixed number of randomly selected data points (e.g., 2 or 3 points) within each cell are retained.

Any of these declustering alternatives results in smaller datasets. To help guard against the tests having power that is too low to reliably detect a COPC, the number of samples (data values) in both the background and site datasets for all the statistical tests should be at least 10, and ideally will be more than 20.

4.1.3 Statistical Testing Approaches Not Recommended

This section describes two methods for comparing data that are not recommended for testing whether or not a chemical is a COPC. The methods are highlighted here as not acceptable because the probability that the tests will give erroneous answers is too great.

Comparing Maximum Site and Maximum Background Measurements

The first approach is to compare the maximum site measurement with the maximum background measurement, using the following decision rule:

- ❑ If the maximum site measurement exceeds the maximum background measurement, then declare the chemical is a COPC; otherwise, declare the chemical is not a COPC.

As discussed in O'Brien and Gilbert (1997), if the site and background have the same concentration distribution and if an equal number of samples are measured for the chemical for both the site and background, the probability is 50% that the maximum measurement occurs in the site dataset and 50% that it occurs in the background dataset. Thus, the chance is 50% that the chemical will be declared a COPC, when in fact the chemical is at background levels on the site. The probability of erroneously declaring a chemical to be a COPC increases if the site has more measurements that the background area. In fact, if the site dataset has n measurements and the background dataset has m measurements, the probability of an incorrect decision is $p = n/(n + m)$. For example, if $n = 20$ and $m = 10$, then $p = 20/30 = 2/3$. That is, the probability is 67% using the decision rule (of comparing the maximum site measurement with the maximum background measurement) will result in an incorrect identification of a chemical as a COPC.

Clearly, this decision rule is not acceptable because its performance in declaring whether or not a chemical is a COPC depends so critically on whether the site or background area has the most measurements.

Comparing the Maximum Site Measurement to a Background Threshold

The other approach is to use this following decision rule:

- ❑ If one or more site measurements exceed the 95th percentile of the background distribution, declare the chemical of interest to be a COPC.

Suppose the site and background distributions are identical and, thus, the chemical is not a COPC. Then, if the above decision rule is used, it can be shown the probability that one or more of *n* site measurements will exceed the 95th percentile is equal to $1 - (0.95)^n$, where 0.95 is the probability that any randomly drawn (representative) single site measurement is less than the 95th percentile of the background distribution. The expression $1 - (0.95)^n$ takes on the values shown in Table 4-2 for various values of *n*.

For example, if the background and site distributions are identical and if n = 21 site measurements of the chemical are obtained, the probability that one or more of the site measurements will exceed the 95th percentile of the background distribution is 0.67. In other words, there is 67% chance of falsely declaring a chemical as a COPC. If more extensive sampling is conducted at the site (for example, if n = 64), then the probability of falsely concluding the chemical is a COPC is 96%!

Threshold values, other than the 95th percentile, that might be used include the 90th or 99th percentiles. Also, the background mean, two times the background mean, or an upper confidence limit on the background mean might be suggested as appropriate threshold values. Regardless of which threshold value is selected, it will correspond to some percentile (perhaps unknown) of the background distribution. Therefore, no matter which threshold value is used, the basic problem of excessive decision errors remains if site measurements are individually compared to the threshold value. Only the specific probability of making an erroneous decision changes.

TABLE 4-2. Probabilities that one or more of *n* site measurements will exceed the 95th percentile of the background distribution if the site and background distributions are identical

n	$1 - (0.95)^n$
1	0.05
2	0.10
5	0.23
8	0.34
10	0.40
12	0.46
21	0.67
64	0.96

Overall, it is recommended that any background threshold comparison:

- ❑ Only be considered as a suggestion for additional investigation of whether a chemical is a COPC; and
- ❑ Never be the only criterion applied to determine if a chemical is a COPC.

4.2 Recommended Comparative Statistical Tests

Comparative statistical tests provide tools that do not have the problems of determining the background threshold values and elevated false decision error rates. In general, when using comparative statistical methods to determine which chemicals are COPCs:

- ❑ ***Datasets*: ** Use background datasets that are comparable to the site data. Ideally, both datasets should be unbiased and representative of geochemically and anthropogenically similar domains. Furthermore, the two datasets should be nearly of the same size. In practice, site datasets are usually large and biased or clustered toward areas of concern, and background datasets are small and randomly collected. Under such instances, the site datasets may have to be declustered prior to the test (Section 4.1.2).

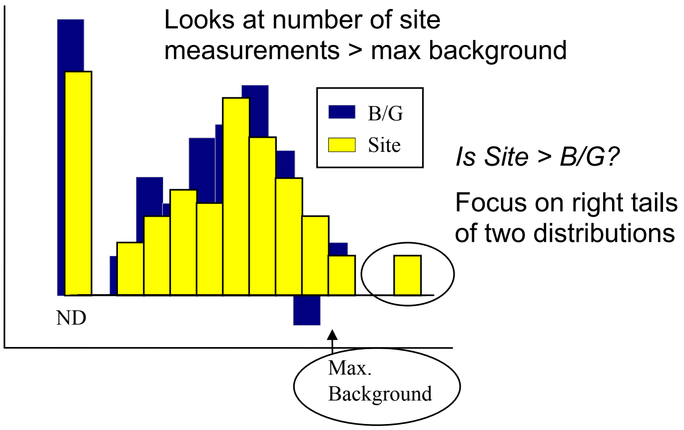
- ❑ ***Support Analyses*: Use a combination of** descriptive statistics, graphic methods and formal tests of hypotheses. Use graphic probability plotting methods as an initial step in determining COPCs.
- ❑ ***Assumption Verification*:** Always check the assumptions that underlie a formal statistical test of hypothesis for COPC. For example, some statistical tests require that the data be normally distributed or that the variances of the site and background data be equal.
- ❑ ***Preferred Tests*:**
  - ◆ Use the nonparametric Slippage test (Section 4.2.1) (comparing site measurements to the maximum background measurement) as a quick way to test for COPCs.
  - ◆ Use the Quantile test (Section 4.2.2) if an important criterion for deciding which chemicals are COPCs is whether the extreme site concentrations are higher than the extreme background concentrations.
  - ◆ Consider using the nonparametric WRS test (Section 4.2.3) if the assumptions that underlie the two-sample t test or the Satterthwaite two-sample t test are unreasonable for the site of interest.
  - ◆ Use the Gehan test (Section 4.2.4) instead of the WRS test if the background or site datasets contain multiple less-than values.
  - ◆ Use the two-sample t test (Section 4.2.5) if the mean concentrations of background and site data can be assumed to be normally distributed with about the same variance, and if very few or no below-detection (less-than) values are present.
  - ◆ Use the Satterthwaite two-sample t test (Section 4.2.6) if the mean concentrations of background and site datasets can be assumed to be normally distributed with different variances and if very few less-than values are present.
  - ◆ Use the two-sample test of proportions (Section 4.2.7) if more than 50% of the background or site measurements are less-than values.
- ❑ ***Most Common Tests*:** Expect to use nonparametric tests most of the time (Slippage, Quantile, WRS, Gehan, and the two-sample test of proportions) because they allow for the occurrence of more less-than values and do not require any specific distributions.

Finally, consult an experienced environmental statistician whenever questions arise regarding the most appropriate graphic or statistical testing methods to use. The application of statistics requires a thorough knowledge of statistical methods for environmental applications and the conditions for which they should be used.

4.2.1 Slippage Test

Site Contamination Scenario

Suppose certain factors indicate that operations at a site may have released small amounts of a chemical. Also, it is known that this particular chemical is present in soils in the natural environment in a defined background area located close to the facility. The decision to be made is whether the concentration levels of this chemical within the site exceed those in the natural background area. If so, the chemical will be declared to be a COPC. Knowledge of site operations suggests that if releases of the chemical did occur, the chemical may not be evenly spread across the site, although most parts of the site are expected to have relatively low concentrations.



Looks at number of site measurements > max background

Is Site > B/G?

Focus on right tails of two distributions

Role of the Data Quality Objectives Process

The DQO process is used to reach agreement with stakeholders and regulators regarding the methods that should be used to collect, handle, prepare, and measure the soil samples. For this scenario, assume that consensus was reached that less-than measurements may frequently occur. It also was agreed that the decision of whether the chemical is a COPC should be made (at least in part) on the basis of only the larger site and background measurements. The Slippage test was selected for this purpose because it uses only the largest few data values and does not require any assumptions about the underlying distribution of the site and background measurements. The assumptions behind using the Slippage test are summarized in Table 4-1.

The stakeholders and regulators also decided to use the WRS test in order to identify differences in the medians of the site and background distributions as a criterion for deciding if the chemical is a COPC.

Advantages and Disadvantages

- The Slippage test consists of counting the number of site measurements that exceed the largest background datum and then comparing that count with a critical value (see Box 4-1). Therefore, the Slippage test is extremely easy to conduct.

BOX 4-1. Procedure for conducting the Slippage test

1. Specify the tolerable probability, α, that the Slippage test will incorrectly declare that the site concentrations tend to be larger than the background concentrations. The probability α can only be selected to be 0.01 or 0.05 because critical values for conducting the test are only available for those two values of α (see Step 7 below). α is the probability the test will incorrectly declare the chemical is a COPC. Note: When both the Slippage test and the WRS test are conducted, the α level of the combined tests will be approximately the sum of the α level selected for each test.
2. Specify the values of ε and of the power $(1 - \beta)$ that the stakeholders and regulators have agreed are important for the Slippage test.
3. Determine the approximate minimum number of required measurements from Table 4-3.
4. Collect the same number of samples from each area (*n* samples from the site and *m* samples from the background area, where $n = m$), and measure the chemical of interest in each sample. Some of the measurements may be less-than values.
5. Determine the value of the largest *detected* background measurement. In making this determination, ignore all less-than values that may be present in the background dataset.
6. Count the number, *K*, of detected *site* measurements that are larger than the largest detected background measurement. In making this determination, ignore all less-than values in the site dataset.
7. If α was selected as approximately 0.01, determine the critical value K_c from Table C-2. If α was selected as approximately 0.05, determine K_c from Table C-3. Note that the value of K_c depends on *n* and *m* $(n = m)$.
8. If *K* is larger than the critical value K_c, declare the site concentrations for the chemical of interest tend to be larger than the background concentrations for that chemical, that is, the chemical is a COPC.

- The Slippage test cannot be applied if the largest background datum is a less-than value. However, the test can be conducted in a straightforward manner even if m − 1 of the *m* background data are less-than values, as long as the largest background less-than value is less than the largest background detected value.

- Because the Slippage test only uses the largest background measurement and the largest few of the site data measurements, it is important to verify that these values are not mistakes or errors made during sample collection, handling, measurement or data handling. A statistical test for outliers (Appendix B.3) can be used to help decide if the largest values are unusually large, relative to what is expected based on an assumed distribution for the other measurements in the dataset. If so, these outliers should be scrutinized to decide if they are mistakes or errors. To be safe, it is a good idea to scrutinize suspiciously large values, even if the outlier test does not indicate they are outliers.

- If the number of samples is sufficiently large, a high probability exists that the Slippage test will detect a COPC when the right tail of the site distribution is shifted to higher concentrations than the right tail of the background concentration distribution.

- In general, the Slippage test does not have high power to detect a shift in the mean or median of the site distribution relative to the mean or median of the background distribution. This is because the test looks at only the largest background measurement and the largest few site measurements.

- The Slippage test and the Quantile test are closely related. However, the Slippage test is so simple to perform that it takes essentially no additional effort to conduct. It can be viewed as a quick test to see almost at a glance whether it is likely that a given chemical is a COPC. However, if the Slippage test fails to declare that a chemical is a COPC, this result should not be used to make a final conclusion that the chemical is not a COPC. Additional statistical testing using the WRS test is needed.

- In general, the WRS test has better performance than the Slippage test to detect when the site concentrations are more or less uniformly greater across the entire site than background concentrations. The Slippage test performs better than the WRS test at detecting when only a portion of the site has concentrations much greater than the background area, assuming representative samples are collected from all regions of the site and background.

- Table 4-1 provides a summary of the advantages and disadvantages of the Slippage Test.

Guidance on Implementing the Slippage Test

The first step in implementing the Slippage test is to determine the number of site (n) and background measurements (m) required for the test to have adequate power to correctly declare the chemical of interest is a COPC. The required values of n and m depend not only on the required power, but also on the following design parameters:

- The proportion, ε, of the site that has concentrations greater than background.

- The magnitude of the difference between site and background concentrations.

- The tolerable probability, α, that the Slippage test will declare the chemical is a COPC when in fact it is not a COPC.

- The underlying distributions (for example, normal or lognormal) of the site and background concentration measurements.

Little information is present in the scientific literature concerning the best values of n and m for use in the Slippage test. However, Gilbert and Simpson (1990) provide enough information to construct a table for this guidance document (Table 4-3) that provides the approximate minimum number of measurements, n and m (for when $n = m$), that should be used in the Slippage test to achieve a power (probability) of approximately 0.80 and 0.90 for various values of ε. Their results are for the case where the tolerable value selected for α is between 0.025 and 0.05. Additional information on the power of the Slippage test is given in Gilbert and Simpson (1990, Figure 3).

It is important to note the following three points:

- The results in Table 4-3 are for the case where all site concentrations (in the ε region) are larger than any true background concentration. If it is suspected that some site concentrations in the ε region will be similar in value to background concentrations, but a few will be definitely larger than background measurements, the *n* and *m* in Table 4-3 will be too small to detect this small difference.
- If a value of α smaller than 0.025 is selected, the number of samples in Table 4-3 will have to be increased for the Slippage test to retain a power of 0.80 or 0.90. If a value of a larger than 0.05 is selected, the number of samples in Table 4-3 could be decreased somewhat and the Slippage test would still have a power of 0.80 or 0.90.
- If site and background measurements have already been collected and the budget does not allow for additional samples, the information in Table 4-3 can be used to approximately determine if a power of 0.80 and 0.90 can be achieved with the available number of measurements. If not, the data by themselves may not contain enough information for the Slippage test to make a confident decision about whether the chemical is a COPC. Other sources of reliable information, such as expert knowledge about installation operations at the site, should be used to the maximum extent in making COPC decisions.

Box 4-1 gives the procedure for conducting the Slippage test. Examples are provided in Boxes 4-2 and 4-3.

TABLE 4-3. Minimum number of measurements (*n* and *m*, n = m) required by the Slippage test to achieve a power of approximately 0.80 or 0.90 when a proportion, ε, of the site has concentrations substantially larger than background concentrations

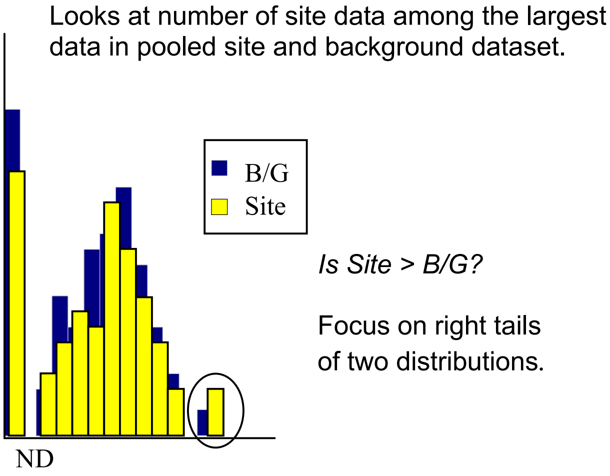
	Number of Required Measurements (*n* and *m*)	
Power	**0.80**	**0.90**
ε = 0.10	60	75
ε = 0.15	40	50
ε = 0.20	30	35
ε = 0.25	25	30
ε = 0.30	15	25
ε = 0.35	15	20
ε = 0.40	15	20
ε = 0.45	10	15
ε = 0.50	10	10
ε = 0.60	10	10

Source: Gilbert and Simpson (1990, Table 1 and Figure 3).

4.2.2 Quantile Test

Site Contamination Scenario

The site contamination scenario described for the Slippage test also applies to the Quantile test. A need exists to determine if concentrations of the target chemical within a site at the facility tend to be greater than those in a defined background area. If target chemical concentrations exceed background (and risk-based criteria), the chemical should be declared a COPC. Knowledge of site operations suggests that if releases of the chemical did occur, the contamination may not be evenly



Looks at number of site data among the largest data in pooled site and background dataset.

Is Site > B/G?

Focus on right tails of two distributions.

BOX 4-2. Example 1 of the Slippage test

1. Suppose $\alpha = 0.01$ is selected.
2. Suppose $\varepsilon = 0.50$ and a desired power of 0.80 are selected.
3. The approximate minimum number of measurements needed is $n = m = 10$ (from Table 4-3).
4. Suppose the following representative measurements of the chemical of interest are obtained (listed in order from smallest to largest):

    Background Data: 23, 36, 37, 37, 44, 57, 60, 61, 61, 79
    Navy Site Data: 15, 15, 20, 29, 30, 39, 60, 89, 90, 100

5. The value of the largest background measurement is 79.
6. $K = 3$ detected site measurements are larger than 79.
7. Using Table C-2 with $n = m = 10$, the critical value K_c is determined to be 6.
8. Therefore, the Slippage test declares that evidence is insufficient to declare the chemical is a COPC because $K = 3$ is not larger than $K_c = 6$.
9. However, do *not* conclude that the chemical is *not* a COPC. Instead, also conduct the WRS test (Section 4.2.3) on these data.

spread across the site, although most parts of the site are expected to have relatively low concentrations. This situation suggests that the Quantile test is appropriate, although the WRS test also should be performed.

Role of the Data Quality Objectives Process

Stakeholders and regulators use the DQO process to determine the methods that should be used to collect, handle, prepare and measure the soil samples. For this scenario, assume that consensus was reached that some less-than measurements would occur and the decision of whether the chemical of interest is a COPC should be made using the Quantile test in combination with the WRS test.

The Quantile test was selected because (1) it is a valid test regardless of the underlying distribution of the site and background data, (2) the test identifies differences in the right tails of the site and background concentration distributions, and (3) the

BOX 4-3. Example 2 of the Slippage test

1. Suppose $\alpha = 0.05$ is selected.
2. Suppose $\varepsilon = 0.30$ and a desired power of 0.80 are selected.
3. The approximate minimum number of measurements needed is $n = m = 15$ (from Table 4-3).
4. Suppose the following 30 representative measurements of the chemical of interest are obtained (listed in order from smallest to largest):

    Background Data: <3, <3, <4, <7, <7, <8, 8, 15, <16, <16, <17, <17, <22, <24, 25
    Navy Site Data: <5, <10, 11, 13, <22, 23, <24, <36, <40, 70, 89, <100, 115, 200, <300

5. The value of the largest detected background measurement is 25.
6. $K = 4$ detected site measurements are larger than 25.
7. Using Table C-3 with $n = m = 15$, the critical value K_c is determined to be 4.
8. Therefore, the Slippage test declares that the evidence is insufficient to declare that the chemical is a COPC because $K = 4$ is not larger than $K_c = 4$.
9. Normally, the WRS test would also be performed to confirm the results of the Slippage test. However, the datasets contain so many less-than values the WRS test cannot be computed (see Section 4.2.3). The Gehan test (Section 4.2.4) should be used in place of the WRS test.

test complements the WRS test, in that the WRS test is good at detecting shifts in the medians. The assumptions behind using the Quantile test are summarized in Table 4-1.

Advantages and Disadvantages

- The Quantile test is similar to the Slippage test. It consists of considering the largest *r* measurements in the pooled site and background datasets and counting the number of those measurements that are from the site. If *k* or more of the *r* measurements are site measurements, the chemical is considered to be a COPC. The Quantile test focuses on comparing the right tails of the site and background distributions rather than comparing the median or mean of the two distributions. For this reason, and because the WRS test focuses on detecting differences in the medians, the Quantile test should always be used in tandem with the WRS test.

- Any number of less-than values are permitted in the site and background datasets, as long as all less-than values are smaller than the smallest of the *r* largest detected measurements in the pooled dataset.

- In general, the WRS test has better performance than the Quantile test at detecting when the site concentrations are more or less uniformly greater across the entire site than background concentrations. The Quantile test performs better than the WRS test at detecting when only a portion of the site has concentrations greater than the background area (assuming a sufficient number of representative samples are collected from all regions of the site and background).

- Use of the Quantile test does not require knowledge of the underlying concentration distribution of the chemical of interest. For example, the measurements need not be normally or lognormally distributed.

- Table 4-1 provides a summary of the advantages and disadvantages of the Quantile test.

- The procedure for conducting the Quantile test is shown in Box 4-4. Boxes 4-5 and 4-6 provide two examples of its use.

BOX 4-4. Procedure for conducting the Quantile test

1. Select the tolerable probability, α, that the Quantile test will incorrectly declare that the site concentrations tend to be larger than background concentrations. The probability α may be selected to be 0.01, 0.025, 0.05, or 0.10. Note: When both the Quantile test and the WRS test are conducted, the α level of the combined tests will be approximately the sum of the α levels selected for each test.
2. Specify the values of ε and of the power ($1 - \beta = 0.80$ or 0.90) desired for the test.
3. Use the specified values of ε and power in Table 4-4 to determine the approximate number of site and background measurements needed. Table 4-5 may be used if it is important to detect site concentrations that are only slightly larger than background.
4. Collect the $n = m$ samples and measure the chemical of interest in each sample. Some of the measurements may be less-than values. If samples have already been collected and measured, verify their number is in agreement with Table 4-4 or Table 4-5. Collect additional samples, if necessary.
5. List from smallest to largest the pooled site and background measurements. The total number of pooled measurements is $n + m$.
6. Using the values of *n* and *m*, enter Table C-4, C-5, C-6, or C-7 (depending on whether α was selected to be 0.01, 0.025, 0.05, or 0.10, respectively) to find the values of *r* and *k* needed to conduct the Quantile test.
7. Determine from the ordered list of pooled site and background measurements if *k* or more of the largest *detected r* measurements are site measurements. (Note: ignore any less-than values when determining the largest detected *r* measurements). If so, the Quantile test indicates the chemical is a COPC. If not, the test indicates the measurements are insufficient for the Quantile test to conclude the chemical is a COPC. The WRS test should be computed.

BOX 4-5. Example 1 of the Quantile test

1. Suppose $\alpha = 0.05$ is selected.
2. Suppose $\varepsilon = 0.50$ is selected and a power of 0.80 is needed to detect when site concentrations are distinctly larger than background concentrations.
3. For these values of α, ε, and power, Table 4-4 indicates that a minimum of $n = m = 10$ measurements are required for the Quantile test.
4. Suppose the 20 measurements are as follows (the same data as was used to illustrate the Slippage test in Box 4-2):

   Background Data: 23, 36, 37, 37, 44, 57, 60, 61, 61, 79
   Navy Site Data: 15, 15, 20, 29, 30, 39, 60, 89, 90, 100

5. The 20 pooled and ordered background and site data are (S and B indicate Site and Background, respectively):

   S S S B S S B B B S B B S B B B B S S S
   15, 15, 20, 23, 29, 30, 36, 37, 37, 39, 44, 57, 60, 60, 61, 61, 79, 89, 90, 100

6. As $\alpha = 0.05$ was selected in Step 1, we find from Table C-6 for $n = m = 10$ that $r = k = 4$.
7. Among the largest $r = 4$ measurements in the pooled measurements (79, 89, 90, and 100), 3 are from the site. Therefore, because $3 < k$, that is, $3 < 4$, the Quantile test indicates the measurements are insufficient to conclude the chemical is a COPC. The WRS test should be performed.

BOX 4-6. Example 2 of the Quantile test

1. Suppose $\alpha = 0.01$ is selected.
2. Suppose $\varepsilon = 0.50$ and a power of 0.80 is needed to detect when site concentrations are distinctly larger than background concentrations.
3. For these values of α, ε, and power, Table 4-4 indicates that $n = m = 15$ measurements are required for the Quantile test.
4. Suppose the data are as follows:

   Background Data: <3, <3, <4, <7, <7, <8, 8, 15, <16, <16, <17, <17, 22, <24, <25
   Site Data: <5, <10, 11, 13, <22, 23, <24, <36, <40, 70, 89, 100, 115, 200, 300

5. The 30 pooled and ordered background and site data are:

   B B B S B B B B S S S B B B B B S B S B S B
   <3, <3, <4, <5, <7, <7, <8, 8, <10 11, 13, 15, <16, <16, <17, <17, <22, 22, 23, <24, <24, <25,

   S S S S S S S S
   <36, <40, 70, 89, 100, 115, 200, 300

6. As $\alpha = 0.01$ was selected in Step 1, we find from Table C-4 for $n = m = 15$ that $r = k = 6$.
7. Among the largest $r = 6$ detected measurements (70, 89, 100, 115, 200, 300), all 6 are from the site. Therefore, because k (that is, 6) of the largest 6 (that is, r) measurements are from the site, the Quantile test indicates the chemical is a COPC.

Guidance on Implementing the Quantile Test

The first step in implementing the Quantile test is to determine the number of site (n) and background (m) measurements required for the test to have adequate power to declare (when it is true) the chemical of interest is a COPC. Also, in common with the Slippage test, the required values of n and m depend on the following:

- The proportion, ε, of the site that has concentrations greater than background.
- The magnitude of the difference between site and background concentrations.
- The tolerable probability, α, that the Quantile test will declare, on the basis of measurements, the chemical is a COPC when in fact it is not a COPC.
- The underlying distribution (for example, normal or lognormal) of the site and background concentration measurements.

U.S. EPA (1994b, Tables A.2, A.3, A.4, and A.5) provides information on the values of *n* and *m* required for the Quantile test to achieve prescribed power to correctly declare a chemical is a COPC. A portion of those results is summarized in Tables 4-4 and 4-5. These tables show the approximate number of site and background measurements needed (n = m) for the Quantile test to have a power (probability) of approximately 0.80 and 0.90 to correctly declare that a chemical is a COPC. The *n* and *m* values presented in these tables correspond to α values of 0.01, 0.025, 0.05, or 0.10; where α is the tolerable probability that the Quantile Test will incorrectly declare that a chemical is a COPC (note that stakeholders and regulators should reach agreement on tolerable probability of error [α] in advance). The results in Tables 4-4 and 4-5 were obtained by assuming that the measurements are normally distributed. If it is suspected that measurements are skewed to the right and perhaps have a lognormal rather than a normal distribution, the number of samples should be increased somewhat to achieve the 0.80 and 0.90 power levels.

Table 4-4 lists the minimum number of measurements required to achieve a power of approximately 0.80 or 0.90 when approximately 85% of the actual (true) site concentrations (in the ε portion of the site) are larger than the vast majority of background concentrations. Table 4-5 lists the minimum number of measurements required to achieve a power of approximately 0.80 or 0.90 when many site and background concentrations in the ε region are similar in value, but about 5% of the site concentrations in the ε region are larger than the vast majority of background concentrations. The number of measurements is larger in Table 4-5 than in Table 4-4 because the results in Table 4-5 are for the case where site concentrations tend to be only slightly larger than background concentrations. Therefore, it takes more data (measurements) to achieve the same power to detect differences.

The Quantile test can be computed using the software EnvironmentalStats for S-Plus (Millard and Nagaraj, 2000).

TABLE 4-4. Minimum number of measurements (*n* and *m*, n = m) required by the Quantile test to achieve a power of approximately 0.80 or 0.90 when a proportion, ε, of the site has concentrations distinctly larger than background concentrations[(a)]

	Number of Required Measurements (*n* and *m*)							
Power	**0.80**	**0.90**	**0.80**	**0.90**	**0.80**	**0.90**	**0.80**	**0.90**
	α = 0.01		**α = 0.025**		**α = 0.05**		**α = 0.10**	
ε = 0.10	>100	>100	100	>100	80	100	55	70
ε = 0.20	55	60	40	40	35	40	25	35
ε = 0.30	25	30	20	25	20	20	15	15
ε = 0.40	20	25	15	20	15	15	10	15
ε = 0.50	15	20	15	15	10	10	10	10
ε = 0.60	10	15	10	10	10	10	10	10
ε = 0.70	10	10	10	10	10	10	10	10
ε = 0.80	10	10	10	10	10	10	10	10
ε = 0.90	10	10	10	10	10	10	10	10
ε = 1.0	10	10	10	10	10	10	10	10

(a) n = m were obtained for the case where the normal site concentration distribution is shifted to the right of the normal background concentration distribution by the amount Δ/s = 4 (U.S. EPA, 1994b, Tables A.2, A.3, A.4, and A.5). α is the tolerable probability (selected by stakeholders and regulators) that the Quantile test will incorrectly declare, on the basis of the measurements, that the chemical is a COPC.

TABLE 4-5. Minimum number of measurements (*n* and *m*, n = m) required by the Quantile test to achieve a power of approximately 0.80 or 0.90 when a proportion, ε, of the site has concentrations somewhat larger than background concentrations[(a)]

	Number of Required Measurements (*n* and *m*)							
	0.80	0.90	0.80	0.90	0.80	0.90	0.80	0.90
Power	α = 0.01		α = 0.025		α = 0.05		α = 0.10	
ε = 0.10	>100	>100	>100	>100	>100	>100	>100	>100
ε = 0.20	>100	>100	>100	>100	>100	>100	>100	>100
ε = 0.30	>100	>100	>100	>100	>100	>100	>100	>100
ε = 0.40	>100	>100	>100	>100	>100	>100	>100	>100
ε = 0.50	>100	>100	>100	>100	>100	>100	>100	>100
ε = 0.60	>100	>100	>100	>100	>100	>100	>100	>100
ε = 0.70	>100	>100	100	>100	75	>100	70	>100
ε = 0.80	>100	>100	75	>100	60	>100	50	>100
ε = 0.90	>100	>100	60	100	50	100	40	100
ε = 1.0	>100	>100	50	75	50	75	30	75

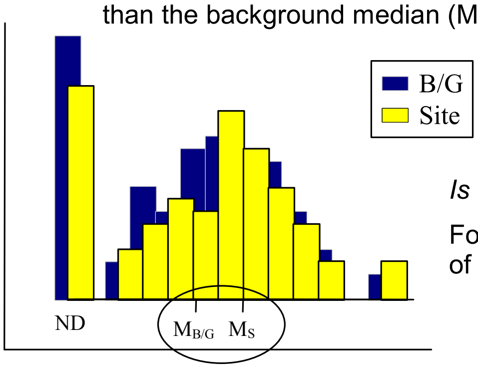
(a) n = m were obtained for the case where the normal site concentration distribution is shifted to the right of the normal background concentration distribution by the amount Δ/s = 1 (U.S. EPA, 1994b, Tables A.2, A.3, A.4, and A.5). α is the tolerable probability (selected by stakeholders and regulators) that the Quantile test will incorrectly declare, on the basis of the measurements, that the chemical is a COPC.

4.2.3 Wilcoxon Rank Sum Test

Site Contamination Scenario

For this scenario, assume that the stakeholders and regulators used Steps 1 and 2 of the DQO process with expert knowledge and all available past data to determine that, if contamination from the site operations has occurred, it probably is homogeneously distributed throughout the site rather than occurring in hot spots within that area.

WRS asks if the site median (M_S) is larger than the background median ($M_{B/G}$)



Is site > background?

Focus on medians of two distributions

Role of the Data Quality Objectives Process

Also, assume that stakeholders and regulators used the DQO planning process to agree on the following:

- The methods that should be used to collect, handle, prepare, and measure the soil samples
- It is unlikely that more than 40% of the measurements will be less-than values
- Both the WRS test and the Quantile test should be conducted
- The value of design parameters for determining the number of site and background measurements needed (see the subsection entitled “Guidance on Implementing the WRS Test” that follows).

The WRS test was selected for the following reasons:

- It is valid regardless of the underlying probability distribution of the site and of the background measurements
- The performance (power) of the test (in detecting when the median site concentration is shifted to the right of the median background

concentration) is known from theory and practice to be as high or higher than other statistical tests that test for shifts in averages.

The Quantile test was selected to be conducted with the WRS test because it has more power (better performance) than the WRS test to detect when only a portion of the site has concentrations greater than background. Therefore, using the Quantile test in conjunction with the WRS test will improve the probability of detecting either uniform or nonuniform contamination greater than background.

The underlying assumptions of the WRS test are:

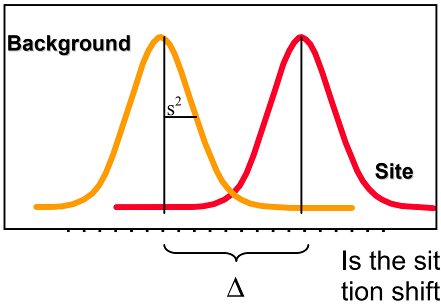
- The measurements obtained from the soil samples from the site and the background area are independent (not correlated). This assumption requires that (1) an appropriate probability-based sampling design strategy be used to determine soil sampling locations, and (2) the soil sampling locations are spaced far enough apart that a spatial correlation among concentrations at different locations is not present. In cases of clustered datasets, such measurements may have to be declustered prior to the test (Section 4.1.2).

- The underlying probability distribution of the site measurements has the same shape (variance) as the probability distribution for the background area. This assumption implies the two distributions are the same, except the distribution for the site may be shifted to higher concentrations than the distribution for the background area. This assumption of equal variances should be evaluated using descriptive statistics and graphic plots of the site and background data (see Box 2-2).

The assumptions behind using the WRS test are summarized in Table 4-1.

Advantages and Disadvantages

- If less-than values occur, all of them must have the same detection limit (the same less-than value), and that detection limit must be less than the smallest detected measurement.



Is the site data distribution shifted to the right of the background data distribution by an important amount Δ?

If multiple less-than values are scattered throughout the set of measurements, then the Gehan test should be used instead of the WRS test.

- The WRS test should not be used if more than 40% of the site or background datasets are less-than values.

- The WRS test does not place large importance (weight) on the larger site and background measurements. It uses and considers all measurements, rather than focusing on the largest measurements as is done by the Slippage test and the Quantile test.

- The WRS test should be used in conjunction with the Quantile Test so that either uniform contamination or nonuniform contamination can be detected with greater probability.

- The software EnvironmentalStat for S-Plus (Millard and Nagaraj, 2000) can be used to compute the WRS test and the Quantile test.

Guidance on Implementing the WRS Test

To implement the WRS test, determine the number of site and background measurements to collect, denoted by *n* and *m*, respectively. A formula for computing *n* and *m* is given in U.S. EPA (1994b, Equation 6.3). This sample-size formula requires the following inputs:

- The acceptable probability, α, that the WRS test will incorrectly declare that the chemical is a COPC. Often, α is set at a value in the range of 0.01 to 0.10.

- The power (probability) the WRS test should have to declare that the chemical is a COPC when that is in fact the case.

- The amount Δ/s (in units of standard deviation, *s*) by which the site median concentration exceeds the background median concentration that must be detected with the required power.

- The proportion of the total number of site and background soil samples that will be collected in the background area. If this proportion is 0.50, then n = m.

When n = m is desired (the ideal case), a formula for determining the number of site and background measurements is given in MARSSIM (NRC et al., 1997, Equation 5-1, p. 5-28).

However, rather than use the formulas in U.S. EPA (1994b) or NRC et al. (1997), it is simpler to select *n* and *m* from Table 4-6 if it is desired that n = m. The values of n = m in Table 4-6 of this document were obtained using Equation 5-1 in NRC et al. (1997) and then increasing that value by 20% to account for uncertainties and the likelihood that missing or unusable measurements will occur. The free computer software code Visual Sample Plan (VSP), developed by the U.S. Department of Energy's Pacific Northwest National Laboratory, also can be used to compute the number of samples for conducting a WRS test (see Appendix D.3).

Box 4-7 describes the steps to perform the WRS test when n<20 and m<20, and Box 4-8 provides an example of that procedure. Box 4-9 describes how to conduct the WRS test when n≥20 and m≥20, and Box 4-10 provides an example of that procedure.

4.2.4 Gehan Test

Site Contamination Scenario

The site contamination scenario is the same as for the WRS test in Section 4.2.3. That is, if contamination from site operations has occurred, it probably would be homogeneously distributed throughout the site.

Role of the Data Quality Objectives Process

Also, assume that stakeholders and regulators used the DQO planning process to agree that:

- The site and background datasets are likely to contain multiple less-than values of possibly different magnitudes, that is, all less-than values will not have the same detection limit.

- Both the Gehan and Quantile tests should be used.

The Gehan test was selected instead of the WRS test because less-than values with different detection limits were expected to occur. The assumptions behind using the Gehan test are summarized in Table 4-1.

Advantages and Disadvantages

- The Gehan test can be used when the background or site datasets contain multiple less-than values with different detection limits.

- The Gehan test is somewhat tedious to compute by hand.

- If the censoring mechanisms are different for the site and background datasets, then the test results may be an indication of this difference in censoring mechanisms rather than an indication that the chemical is a COPC.

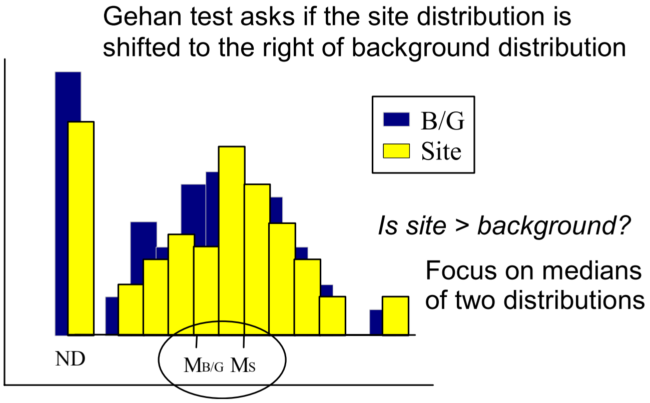


TABLE 4-6. Number of site and background measurements (*n* and *m*, n = m) required by the WRS test[(a)]

	α= 0.01					α= 0.025					α= 0.05					α= 0.10					α=0.25				
	Power					Power					Power					Power					Power				
Δ/σ	0.99	0.975	0.95	0.90	0.75	0.99	0.975	0.95	0.90	0.75	0.99	0.975	0.95	0.90	0.75	0.99	0.975	0.95	0.90	0.75	0.99	0.025	0.05	0.10	0.25
0.1	5,452	4,627	3,972	3,278	2,268	4,827	3,870	3,273	2,846	1,748	3,972	3,273	2,726	2,157	1,355	3,278	2,846	2,157	1,655	964	2,268	1,748	1,355	964	459
0.2	1,370	1,163	998	824	570	1,163	973	823	665	440	998	823	685	542	341	824	685	542	416	243	570	440	341	243	116
0.3	614	521	448	370	256	521	436	369	298	197	448	369	307	243	153	370	298	243	187	109	256	197	153	109	52
0.4	350	297	255	211	148	297	248	210	170	112	255	210	175	139	87	211	170	139	106	62	146	112	87	62	30
0.5	227	193	166	137	95	193	162	137	111	73	166	137	114	90	57	137	111	90	69	41	95	73	57	41	20
0.6	161	137	117	97	67	137	114	97	76	52	117	97	81	64	40	97	78	64	19	29	67	52	40	29	14
0.7	121	103	88	73	51	103	86	73	59	39	88	73	61	48	30	73	59	48	37	22	51	39	30	22	11
0.8	96	81	69	57	40	81	68	57	46	31	69	57	48	38	24	57	46	38	29	17	40	31	24	17	8
0.9	77	66	58	47	32	65	55	46	38	25	56	48	39	31	20	47	38	31	24	14	32	25	20	14	7
1.0	64	55	47	39	27	55	46	39	32	21	47	39	32	26	16	39	32	25	20	12	27	21	16	12	6
1.1	55	47	40	33	23	47	39	33	27	18	40	33	28	22	14	33	27	22	17	10	23	18	14	10	5
1.2	48	41	35	29	20	41	34	29	24	16	35	29	24	19	12	29	24	19	15	9	20	16	12	9	4
1.3	43	36	31	26	18	36	30	26	21	14	31	26	22	17	11	26	21	17	13	8	18	14	11	8	4
1.4	38	32	28	23	16	32	27	23	19	13	28	23	19	15	10	23	19	15	12	7	16	13	10	7	4
1.5	35	30	25	21	15	30	25	21	17	11	25	21	18	14	9	21	17	14	11	7	15	11	9	7	3
1.6	32	27	23	19	14	27	23	19	16	11	23	19	16	13	8	19	16	13	10	6	14	11	8	6	3
1.7	30	25	22	18	13	25	21	18	15	10	22	18	15	12	8	18	15	12	9	6	13	10	8	6	3
1.8	28	24	20	17	12	24	20	17	14	9	20	17	14	11	7	17	14	11	9	5	12	9	7	5	3
1.9	26	22	19	15	11	22	19	16	13	9	19	16	13	11	7	16	13	11	8	5	11	9	7	5	3
2.0	25	21	18	15	11	21	18	15	12	8	18	15	13	10	7	15	12	10	8	5	11	8	7	5	3
2.25	22	19	16	14	10	19	16	14	11	8	16	14	11	9	6	14	11	9	7	4	10	8	6	4	2
2.5	21	18	15	13	9	18	15	13	10	7	15	13	11	9	6	13	10	9	7	4	9	7	6	4	2
2.75	20	17	15	12	9	17	14	12	10	7	15	12	10	8	5	12	10	8	6	4	9	7	5	4	2
4.0	19	16	14	12	8	16	14	12	10	6	14	12	10	8	5	12	10	8	6	4	8	6	5	4	2
4.5	18	16	13	11	8	16	13	11	9	6	13	11	9	8	5	11	9	8	6	4	8	6	5	4	2
4.0	18	15	13	11	8	15	13	11	9	6	13	11	9	7	5	11	9	7	6	4	8	6	5	4	2

Source: NRC et al. (1997, Table 5.3).

(a) Power is the probability the WRS test correctly declares that the chemical is a COPC.

BOX 4-7. Procedure for conducting the WRS test when the number of site and background measurements is small (n <20 and m <20)

1. Specify the tolerable probability, α, that the WRS test will incorrectly declare that the site concentrations tend to be larger than the background concentrations (that is, of the test incorrectly declaring the chemical is a COPC). Note: When both the WRS and Quantile tests are conducted, the α level of the combined tests will be approximately the sum of the α levels selected for each test.
2. Specify the value of Δ/s and of power, where Δ/s is the magnitude of the difference in median site and background concentrations, that must be detected by the WRS test with the specified power. The notation Δ/s indicates that the shift is expressed in units of standard deviation (s) of the underlying background and site concentration distributions for the chemical of interest.
3. Use the specified values of α, Δ/s, and power in Table 4-6 to determine the number of site and background measurements needed when it is desired to have n equal to m. If having equal n and m is not desired, use U.S. EPA (1994b, Equation 6.3) and increase that value by 20% to guard against missing or unusable measurements.
4. Collect the same number of samples from each area *(n* samples from the site and *m* samples from the background area, where n = m), and measure them for the chemical of interest, some of which may be less-than values. If measurements are available from past sampling efforts, verify their number is at least as large as the number indicated in Table 4-6. Collect additional samples, if necessary, to achieve the required number of samples.
5. List and rank the pooled set of n + m site and background measurements from smallest to largest, keeping track of which measurements came from the site and which came from the background area. Assign the rank of 1 to the smallest value among the pooled data, the rank of 2 to the second smallest value among the pooled data, and so forth.

   If a few measurements are tied (identical in value) assign the average of the ranks that would otherwise be assigned to the tied observations. If several measurement values have ties, average the ranks separately for each of those measurement values.

   If a few less-than values occur (say, <10%), and if all such values are less than the smallest detected measurement in the pooled dataset, handle the less-than values as tied at an arbitrary value less than the smallest detected measurement. Assign the average of the ranks that would otherwise be assigned to these tied less-than values (the same procedure as for tied detected measurements).

   If between 10% and 40% of the pooled dataset are less-than values, and all are less than the smallest detected measurement, use the WRS test procedure in Box 4-9, even if n and m are less than 20. Note: The procedure in Box 4-9 is for the case where n and m are both of size 20 or larger. That procedure will provide only an approximate test if it is used when n and m are both smaller than 20. In that case, decisions of whether the chemical is a COPC should not be made until additional information is obtained by taking more samples and using a more sensitive measurement method.
6. Calculate the sum of the ranks of the *site* measurements. Denote this sum by R, then calculate W as follows:

   $$W = R - n(n + 1) / 2$$
7. Use the values of n and m and α to enter Table C-8 to find the critical value w_α, where α has been specified in Step 3 above. Table C-8 can be used only if α has been chosen to be 0.05 or 0.10.

   If $W > nm - w_\alpha$ the WRS test indicates the site concentration distribution is shifted to the right of the background concentration distribution, that is, that the chemical is a COPC.
8. If the WRS test declares the chemical is *not* a COPC, this conclusion may indicate either (1) the chemical is indeed not a COPC, (2) the assumptions that underlie the WRS test are not valid for the site and background measurements, or (3) an insufficient number of measurements (n and m) were obtained for the WRS test to detect the difference that actually exists in site and background concentration distributions.

BOX 4-7. (cont'd)

An evaluation should be made of the possibility the causes in cases 2 or 3 may have resulted in the WRS test declaring the chemical is not a COPC. Review the DQO planning process records to make sure the number of samples (n and m) collected agree with what was determined at that time to be necessary to detect a possible difference between site and background measurements that was considered important. For case 3, the shift in the concentration distribution may in fact be smaller than the shift selected by the stakeholders, in which case no additional measurements are needed.

Also, update the estimated number of site and background measurements needed by using a software package such as DataQUEST (U.S. EPA, 1997). Collect additional samples if needed.

BOX 4-8. Example of the WRS test when the number of site and background measurements is small (n <20 and m <20)

Suppose a need is present to determine if a chemical in surface soil in Region A on the Navy site is a COPC.

1. Suppose α was specified to be 0.05.
2. Suppose Δ/s and the power were specified to be 1.5 and 0.95, respectively. That is, the stakeholders and regulators specified that if the median of the site concentration distribution is shifted by the amount Δ/s greater than the median background distribution, then enough measurements should be obtained so that the WRS test has a power of 0.95 of detecting that fact.
3. Using these values of α, Δ/s, and power to enter Table 4-6, it is determined that $n = m = 18$ measurements are needed for the WRS test.
4. Then, 18 samples from both the site and the background area were collected using a suitable probability-based sampling design (for example, simple random sampling or sampling at the nodes of a square or triangular grid) and measurements made of the chemical of interest on each sample. Suppose the measurements were:

Background Data:	22,	32,	9,	12,	3,	7,	11,	2,	9,	11,	13,	16,	20,	25,	<1,	<1,	17,	21
Site Data:	24,	33,	5,	9,	36,	<1,	10,	50,	9,	19,	15,	10,	28,	9,	3,	15,	4,	19

5. Next, the data are pooled together and listed from smallest to largest. The ranks of the site data are determined (the site and background data and ranks are denoted by S and B, respectively):

	B	B	S	B	S	B	S	S	B	S	S	S	B	B	S	S	B	B
Data:	<1	<1	<1	2	3	3	4	5	7	9	9	9	9	9	10	10	11	11
Rank:	2	2	2	4	5.5	5.5	7	8	9	12	12	12	12	12	15.5	15.5	17.5	17.5

	B	B	S	S	B	B	S	S	B	B	B	S	B	S	B	S	S	S
Data:	12	13	15	15	16	17	19	19	20	21	22	24	25	28	32	33	36	50
Rank:	19	20	21.5	21.5	23	24	25.5	25.5	27	28	29	30	31	32	33	34	35	36

6. Sum the ranks of the site measurements to obtain $R = 2 + 5.5 + 7 + \ldots + 34 + 35 + 36 = 350.5$. Therefore,

$$W = R - n(n+1) / 2 = 350.5 - 18(19) / 2 = 179.5$$

7. Enter Table C-8 with $\alpha = 0.05$ and $n = m = 18$ to obtain $w_{0.05} = 110$.

   We compute $nm - w_\alpha = 18 \times 18 - 110 = 214$. Therefore, $W < nm - w_\alpha$, that is, $179.5 < 214$. The WRS has indicated the evidence in the data is insufficient to declare the chemical is a COPC.

   As the WRS did not declare that the chemical is a COPC, the DQO process notes are reviewed to make sure the number of measurements specified to meet the α, Δ/s, and power requirements were indeed obtained. Also, to update the estimated number of site and background measurements needed, use a software package such as DataQUEST (U.S. EPA, 1997). If the number of samples computed using that equation exceeds the number used in the WRS test, collect the indicated number of new site and background samples.

BOX 4-9. Procedure for conducting the WRS test when the number of site and background measurements is large ($n \geq 20$ and $m \geq 20$)

1. Specify the tolerable probability, α, that the WRS test will incorrectly declare that the site concentrations tend to be larger than the background concentrations, that is, of the test incorrectly declaring the chemical is a COPC. Note: When both the WRS test and Quantile test are conducted, the α level of the combined tests will be approximately the sum of the α levels selected for each test.
2. Specify the value of Δ/s and of power, where Δ/s is the magnitude of the difference in average site and background concentrations that must be detected by the WRS test with the specified power. The notation Δ/s indicates the shift is expressed in units of standard deviation (s) of the underlying background and site concentration distributions for the chemical of interest.
3. Use the specified values of α, Δ/σ, and power in Table 4-6 to determine the number of site and background measurements needed when it is desired to have n equal m. If no need is present to have equal n and m, use Equation 6.3 in U.S. EPA (1994b) and increase that value by 20% to guard against missing or unusable measurements.
4. Collect the same number of samples from each area *(n* samples from the site and *m* samples from the background area, where n = m), and measure them for the chemical of interest, some of which may be less-than values. If measurements are available from past sampling efforts, verify their number is at least as large as the number indicated in Table 4-6. Collect additional samples, if necessary to achieve the required number of samples.
5. List and rank the pooled set of n + m site and background measurements from smallest to largest, keeping track of which measurements came from the site and which came from the background area. Assign the rank of 1 to the smallest value among the pooled data, the rank of 2 to the second smallest value among the pooled data, and so forth.

   If <40% of the measurements in the pooled dataset are tied (identical in value) assign the average of the ranks that would otherwise be assigned to the tied observations. If several measurement values exist for which ties occur, average the ranks separately for each of those measurement values.

   If <40% of the pooled dataset are less-than values and if all such values are less than the smallest detected measurement in the pooled dataset, handle those less-than values as being tied at an arbitrary value less than the smallest detected measurement. Assign the average of the ranks that would otherwise be assigned to this group of tied values (the same procedure as for detected measurements that are tied). Note: The total number of tied detected measurements and tied less-than values should not exceed 40% of the total number of measurements.

   If more than 40% of the pooled data are less-than values, then do not use the WRS test. The Gehan test should be used instead (Section 4.2.4).
6. Calculate the sum of the ranks of the site measurements. Denote this sum by R.
7. Calculate

   $$w_{1-\alpha} = n(n+1) / 4 + z_{1-\alpha} [n (n + 1) (2n + 1)/24]^{1/2}$$

   where $z_{1-\alpha}$ is the 100(1−α) percentile of the standard normal distribution, which is tabulated in Table C-1. For example, if $\alpha = 0.05$, then $z_{1-\alpha} = z_{0.95} = 1.645$ from Table C-1.
8. The WRS test declares that the chemical is a COPC if $R > w_{1-\alpha}$.

Guidance on Implementing the Gehan Test

The Gehan test is one of several nonparametric tests that have been proposed to test for differences between two sites when the datasets have multiple censoring points. Among these tests, Palachek et al. (1994) indicate that they selected the Gehan test primarily because it was the easiest to explain, because the several methods generally behave comparably, and because the Gehan test reduces to the

BOX 4-10. Example of the WRS test when the number of site and background measurements is large ($n \geq 20$ and $m \geq 20$)

1. Suppose α is specified to be 0.01.
2. Suppose Δ/s and the power were specified to be 1.8 and 0.95, respectively. That is, the stakeholders and regulators specified that if the median of the site concentration distribution is $\Delta/s = 1.8$ (in units of standard deviation, *s*) units greater than the median background distribution, enough measurements should be obtained so the WRS test has a power of 0.95 of detecting that fact.
3. Using these values of α, Δ/s, and power to enter Table 4-6, we find that $n = m = 20$ measurements are needed for the WRS test, where *n* and *m* are the number of site and background measurements, respectively.
4. Then 20 samples from both the site and the background areas were collected using a suitable probability-based sampling strategy, for example, simple random sampling. Suppose the measurements were (listed in increasing magnitude):

Background Data: <10, <10, <10, <10, 12, 15, 15, 18, 22, 26, 27, 29, 29, 29, 55, 60, 77, 90, 101, 150
Site Data: <10, <10, <10, 25, 27, 27, 36, 36, 99, 101, 103, 140, 145, 150, 180, 190, 199, 200, 250, 300

5. Next, the data are pooled together and listed from smallest to largest. Then the ranks of the site data are determined (the site and background data and ranks are denoted by S and B, respectively).

	B	B	B	B	S	S	S	B	B	B	B	B	S	B	B	S	S	B	B	B	S
Data:	<10	<10	<10	<10	<10	<10	<10	12	15	15	18	22	25	26	27	27	27	29	29	29	36
Rank:	4	4	4	4	4	4	4	8	9.5	9.5	11	12	13	14	16	16	16	19	19	19	21

	S	B	B	B	B	S	B	S	S	S	S	B	S	S	S	S	S	S	S
Data:	36	55	60	77	90	99	101	101	103	140	145	150	150	180	190	199	200	250	300
Rank:	22	23	24	25	26	27	28.5	28.5	30	31	32	33	34	35	36	37	38	39	40

6. The sum of the ranks of the site data is $R = 4 + 4 + 4 + 13 + 16 + \ldots + 39 + 40 = 507.5$.
7. Also, $w_{0.99} = n(n+1) / 4 + z_{0.99} [n (n + 1) (2n + 1)/24]^{1/2}$

$= 20(21) / 4 + 2.33[20(21)(41)/24]^{1/2}$

$= 167.4$

where $z_{0.99} =$ 2.33 is the 99th percentile of the standard normal distribution, that is found in Table C-1.

8. Because $R > w_{0.99}$, that is, $507.5 > 167.4$, the WRS test determines the chemical to be a COPC.

WRS test, which is a relatively well-known test to environmental professionals. Palachek et al. (1994) used their computer code to conduct Gehan tests on data from the Rocky Flats Environmental Technology Site near Denver, CO. They recommend using the Gehan test rather than a more complicated procedure involving replacement of nondetects by a value such as one-half the detection limit, testing for distribution shape and variance, and then conducting appropriate t tests or the WRS test.

The number of samples (measurements) needed from the site and from background to conduct the Gehan test may be approximated using the method described for the WRS test in Section 4.2.3. The procedure for conducting the Gehan test, when $n \geq 10$ and $m \geq 10$, is given in Box 4-11. An example of the test is given in Box 4-12. If $n < 10$ or $m < 10$, the procedure in Box 4-13 may be used to conduct the Gehan test.

BOX 4-11. Gehan test procedure when n ≥10 and m ≥10

1. Specify the tolerable probability, α, that the Gehan test will incorrectly declare that the site median is larger than the background median, that is, of the test incorrectly declaring that the chemical is a COPC.
2. Specify the value of Δ/s and the power, where Δ/s is the magnitude of the difference in median site and background concentrations that must be detected by the Gehan test with the specified power. The notation Δ/s indicates the shift is expressed in units of standard deviation (*s*) of the underlying background and site concentration distributions for the chemical of interest. Recall that an underlying assumption is that the variances of the site and background data for the chemical are the same.
3. Use the specified values of α, Δ/s, and the power in Table 4-6 to determine the number of site and background measurements needed when it is desired to have *n* equal *m*. If no need exists to have equal *n* and *m*, use U.S. EPA (1994b, Equation 6.3) and increase that value by 20% to guard against missing or unusable measurements.
4. Collect the same number of samples from each area *(n* samples from the site and *m* samples from the background area, where n = m), and measure them for the chemical of interest, some of which are expected to be less-than values. If measurements are available from past sampling efforts, verify that their number is at least as large as the number indicated in Table 4-6. Collect additional samples if necessary to achieve the required number of samples.
5. List the combined *m* background and *n* site measurements, including the less-than values, from smallest to largest, where the total number of combined samples is N = m + n. The less-than symbol (<) is ignored when listing the *N* data from smallest to largest.
6. Determine the *N* ranks, R_1, R_2, …, R_N, for the *N* ordered data values using the method described in the example given in Box 4-12.
7. Compute the *N* scores, $a(R_1)$, $a(R_2)$,…,$a(R_N)$ using the formula $a(R_i) = 2R_i - N - 1$, where *i* is successively set equal to 1, 2, …, *N*.
8. Compute the Gehan statistic, *G*, as follows:

$$G = \frac{\sum_{i=1}^{N} h_i a(R_i)}{\left\{ mn \sum_{i=1}^{N} [a(R_i)]^2 / [N(N-1)] \right\}^{1/2}} \qquad (1)$$

where h_i = 1 if the *i*th datum is from the site population
= 0 if the *i*th datum is from the background population
N = n + m
$a(R_i)$ = $2R_i - N - 1$, as indicated above.

9. The Gehan test determines the chemical to be a COPC if $G \geq Z_{1-\alpha}$, where $Z_{1-\alpha}$ is the $100(1-\alpha)$th percentile of the standard normal distribution, which is obtained from Table C-1. Otherwise, the test indicates that the evidence is not strong enough to conclude that the chemical is a COPC.

BOX 4-12. Example of the Gehan test

1. Suppose α was specified to be 0.05.
2. Suppose Δ/s and the power were specified to be 2.0 and 0.90, respectively. That is, the stakeholders and regulators specified that, if the median of the site concentration distribution is greater than the median background distribution by the amount $\Delta/s = 2.0$ (in units of standard deviation, *s*), enough measurements should be obtained so the Gehan test has a power of 0.90 of detecting that fact.
3. Using the specified values of Δ/s and power in Table 4-6, it is determined that n = m = 10 measurements are needed to conduct the Gehan test.
4. The 10 samples from the site and the background area were collected using a suitable probability-based sampling design (for example, simple random sampling or sampling at the nodes of a square or triangular grid) and measurements were made of the chemical of interest on each sample. Suppose the measurements are:

Background:	1	<4	5	7	<12	15	18	<21	<25	27
Site:	2	<4	8	17	20	25	34	<35	40	43

5, 6 and 7. Use the following procedure to determine the N = 20 ranks $R_1, R_2, \ldots, R_{20}$ and the 20 scores $a(R_i)$. Refer to Table 1 below when going through the steps.

Table 1. Calculations to Determine the Ranks, R_i, and the Scores, $a(R_i)$

Data	h_i	Index_i	d_i	e_i	R_i	$a(R_i)$
1	0	0	1	0	4	–13
2	1	0	2	0	5	–11
<4	1	1	2	1	4.5	–12
<4	0	1	2	2	4.5	–12
5	0	0	3	2	9	–7
7	0	0	4	2	8	–5
8	1	0	5	2	9	–3
<12	0	1	5	3	6	–9
15	0	0	6	3	10.5	0
17	1	0	7	3	11.5	2
18	0	0	8	3	12.5	4
20	1	0	9	3	14.5	6
<21	0	1	9	4	8	–5
<25	0	1	9	5	8	–5
25	1	0	9	5	15.5	10
27	0	0	10	5	16.5	12
34	1	0	12	5	17.5	14
<35	1	1	12	6	9.5	–2
40	1	0	13	6	19	17
43	1	0	14	6	20	19

- List the combined *m* background and *n* site measurements, including the less-than values, from smallest to largest, as illustrated in column 1 of Table 1. Ignore the less-than symbol when listing the *N* data from smallest to largest.
- Place a 0 or 1 in the second column of Table 1 (the column with heading h_i) using the following rule:

  h_i = 0 if the *i*th measurement is from background
  = 1 if the *i*th measurement is from the site

- Place a 0 or 1 in the 3rd column of Table 1 (the column with heading Index_i) using the following rule:

  Index_i = 0 if the *i*th measurement is a detect
  = 1 if the *i*th measurement is a less-than value

- When moving down the data in column 1, determine the values of parameters d and e (columns 4 and 5 in Table 1) using the following rules:
  - If the first datum in column 1 is a detect, that is, if $\text{Index}_i = 0$, then set d = 1 and e = 0 in the first row of Table 1.
  - If the first datum in column 1 is a less-than value, that is, if $\text{Index}_i = 1$, then set d = 0 and e = 1 in the first row of Table 1.
  - For each successive row (rows 2 through n = 20) increase d by 1 whenever the datum in column 1 in that row is a detect, that is, whenever Index = 0
  - For each successive row increase e by 1 whenever the datum in column 1 in that row is a less-than value, that is, when Index = 1.

BOX 4-12. (continued)

- Let T denote the total number of less-than values in the pooled background and site datasets. For the previous data there are T = 6 less-than values. Compute the rank of the *i*th datum (i.e., of the datum in the *i*th row in the previous table) as follows:
  - $R_i = d_i + (T + e_i)/2$ if the datum in column 1 of the *i*th row is a detect, that is, if $h_i = 0$ for the *i*th row.
  - $R_i = (T + 1 + d_l)/2$ if the datum in column 1 of *i*th row is a less-than value, that is, if $h_i = 1$ for the *i*th row.
- Compute the n = 20 scores, $a(R_1)$, $a(R_2)$, …, $a(R_{20})$, using the formula

  $$a(R_i) = 2R_i - N - 1$$

  for successive values of i = 1, 2, …, 20.

8. Compute the Gehan statistic, G:

$$G = \frac{(-11)+(-12)+(-3)+2+6+10+14+(-2)+17+19}{\{10*10[(-13)^2+(-11)^2+(-12)^2+\ldots+(-2)^2+(17)^2+(19)^2]/20*19\}^{1/2}}$$

$$= 40 / [(100*1942) / (20*19)]^{1/2}$$
$$= 40 / 22.606$$
$$= 1.77$$

9. Step 1 specified that $\alpha = 0.05$. Using Table C-1 with $\alpha = 0.05$, it is determined that $Z_{1-\alpha} = Z_{0.95} = 1.645$. Because G > 1.645, that is, 1.77 > 1.645, the Gehan test determines the chemical to be a COPC.

BOX 4-13. Procedure for conducting the Gehan test when n <10 and m <10

1. Generate on a computer all possible orderings of the combined *n* site and background measurements. Denote the *number* of possible orderings by *M*.
2. Compute the *G* statistic (Box 4-11) for each of these orderings to generate an empirical distribution (histogram) of *M* values of *G*.
3. Determine the 100(1 – α)th percentile of the empirical distribution of *G* generated by Step 2 as follows (from Gilbert, 1987, p. 141) where α is the tolerable probability that the test procedure described in this box will incorrectly declare that the chemical is a COPC:
   - Order the *M* values of *G* from smallest to largest.
   - Compute $k = (1 - \alpha)(M + 1)$
   - If *k* is an integer, the (1 – α)th percentile is the kth largest value of the ordered *M* values of *G*.
   - If *k* is not an integer, determine the value of k′, where k′ is the largest integer less than *k*. Compute the (1 − α)th percentile by linear interpolation between the k′th and (k′ +1)th largest values of *G*.
4. If the value of *G* computed, *using the ordering actually observed for the collected background and site data,* equals or exceeds the 100(1–α)th percentile obtained above, conclude the chemical is a COPC.

4.2.5 Two-Sample t Test

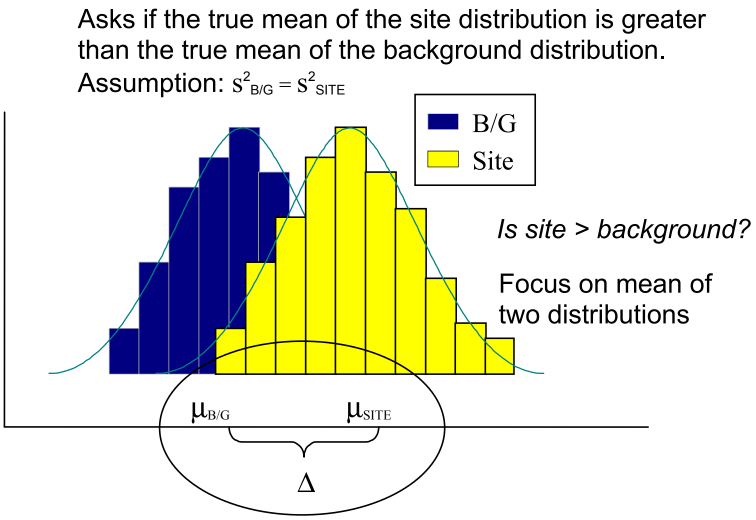
Site Contamination Scenario

The site contamination scenario for the two-sample t test is the same as that for the WRS test in Section 4.2.3. That is, if contamination from site operations has occurred, it probably would be homogeneously distributed throughout the site.

Role of the Data Quality Objectives Process

Also, for this scenario, assume that the stakeholders and regulators used the DQO planning process to agree:

- On the methods that will be used to collect, handle, prepare and measure the soil samples
- On the value of the design parameters for determining the number of site and background measurements needed (discussion following)



Asks if the true mean of the site distribution is greater than the true mean of the background distribution.

Assumption: $s^2_{B/G} = s^2_{SITE}$

Is site > background?

Focus on mean of two distributions

- That it is likely that very few less-than values will be reported by the laboratory
- That the mean concentrations are likely to be normally distributed (for large datasets, *n* and *m* greater than 30, this assumption can be viewed as valid based on the central limit theory; however, for small datasets, tests for normality of the measurements should be conducted to assure the validity of this assumption)
- That for small datasets (*n* and *m* less than 30), if the tests for normality indicate the measurements are not normally distributed so the estimated site and background means are not approximately normally distributed, the WRS test and the Quantile test should be used in place of the two-sample t test
- That the measurements from the site are expected to have approximately the same variance as the background measurements
- If a statistical test (an F test described in Conover [1998] and U.S. EPA [2000, p. 4-33]) indicate the site and background measurements may not have the same variance, but both mean concentrations appear to be normally distributed, then the Satterthwaite two-sample t test (Section 4.2.6) will be used to test for differences in the site and background means.

The two-sample t test was selected because the assumptions of normality of mean, equal variances for background and site data, and the absence of less-than values were expected to be valid. However, once the measurements are obtained, these assumptions will be evaluated by observation and by using statistical tests. If the site and background variances appear to be approximately equal but the mean concentrations are not expected to be normally distributed, the WRS test may be used in place of the two-sample t test. If the two datasets are not normally distributed and have unequal variances, the Quantile and Slippage tests may be used. The assumptions behind using the two-sample t test are summarized in Table 4-1.

Advantages and Disadvantages

- If less-than values occur and if those values are replaced by substitute values such as the detection limit or one-half the detection limit, then the two-sample t test could be computed. However, the test would give biased and perhaps misleading results.
- If there is only one detection limit (for example, if all less-than values are <10), and no more than about 40% of both the site and background data are less-than values, then the two-sample t test should be replaced with the WRS test. This recommendation is correct even though the data may be normally distributed. The Quantile test also may be used in conjunction with the WRS test. If the less-than values take on multiple values (<10, <15, etc.), the Gehan test should be used in place of the WRS test.

- Most statistical software packages compute the two-sample t test.

Guidance on Implementing the Two-Sample t Test

The number of site (*n*) and background (*m*) measurements required to conduct the two-sample t test should be approximated using the procedure outlined in Box 4-14. An example of the computation of Equation 1 in Box 4-14 is given in Box 4-15. After *n* and *m* have been determined, the samples collected, and measurements reported by the laboratory, summary statistics should be computed for both the site and background datasets. In particular, the computed sample variance of the site measurements should be compared with the computed sample variance of the background measurements to determine if they are approximately equal, a required assumption of the two-sample t test. A procedure (an F test) for testing if the two sample variances are equal is provided in Conover (1998) and U.S. EPA (2000, p. 4-33). This procedure is commonly found in statistical software packages.

If some measurements appear to be unusually large compared to the remainder of the measurements in the dataset, a test for outliers should be conducted (see Appendix B.3). Once any identified outliers have been investigated for being mistakes or errors and, if necessary, discarded, small site and background datasets (*n* and *m* less than 30) should be tested for normality using both probability plots and normality tests.

After the assumptions of equal variances (and normality for small datasets) have been shown to be reasonable, the two-sample t test can be conducted. The procedure for doing the test is given in Box 4-16, and an example of this procedure is given in Box 4-17.

BOX 4-14. Procedure for calculating the number of site and background measurements required to conduct the two-sample t test

The formula for calculating the number of site (*n*) and background (*m*) measurements required to conduct the two-sample t test is:

$$n = m \approx \frac{2s^2(Z_{1-\alpha} + Z_{1-\beta})^2}{(\mu_s - \mu_b)^2} + 0.5*(Z_{1-\alpha})^2$$

where: s^2 = expected variance of the measurements at both the site and background area (ideally, the value of s^2 used should be approximated using measurements previous obtained from the site and background or obtained in a special pilot study at the site and background).

α = the tolerable probability that the two-sample t test will incorrectly declare the chemical is a COPC (α is usually specified to be a small value such as 0.01, 0.025, 0.05 or 0.10).

$1 - \beta$ = the power (probability) required that the two-sample t test will declare the chemical is a COPC when that is indeed the case (β is usually specified to be ≥ 0.80).

$\mu_s - \mu_b$ = true site mean (μ_s) minus the true background mean (μ_b); the difference in the *true* (unknown) means of the site and background that the stakeholders and regulators have agreed needs to be detected by the two-sample t test with power (probability) equal to $1 - \beta$.

$Z_{1-\alpha}$ = the $100(1 - \alpha)$ percentile of the standard normal distribution, which is found in Table C-1 (for example, if $\alpha = 0.05$, Table C-5 indicates $Z_{1-0.05} = Z_{0.95} = 1.645$).

$Z_{1-\beta}$ = the $100(1 - \beta)$ percentile of the standard normal distribution, which is found in Table C-1 (for example, if $1 - \beta = 0.80$, then we find from Table C-1 that $Z_{0.80} = 0.84$).

The appropriate values of the parameters for the above equation should be determined by the stakeholders and regulators during the application of the DQO planning process.

BOX 4-15. Example of the procedure for calculating the number of site and background measurements required to conduct the two-sample t test

Suppose the values of the parameters in Equation 1 in Box 4-14 were specified by the stakeholders and regulators as follows:

$s^2 = 7.5$
$\alpha = 0.025$
$1 - \beta = 0.80$
$\mu_s - \mu_b = 4$

Table C-1 indicates that $Z_{1-\alpha} = Z_{0.975} = 1.96$ and $Z_{1-\beta} = Z_{0.80} = 0.84$

Therefore, Equation 1 is:

$$\begin{aligned} n &= m \approx 2{*}7.5{*}(1.96 + 0.84)^2 / 4^2 + 0.50{*}(1.96)^2 \\ &= 9.27 \text{ or } 10 \end{aligned}$$

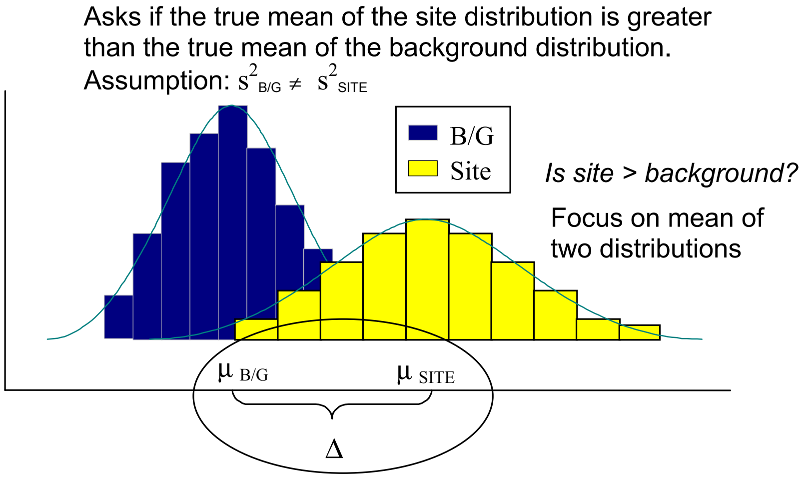
Therefore, 10 site and 10 background measurements are required for the two-sample t test to attain the performance specified (by the values of α and $1 - \beta$) to detect a difference in true means of size $\mu_s - \mu_b = 4$ when the variance of the data at the site and background areas is $s^2 = 7.5$.

The reader may want to try other values of s^2 and $\mu_s - \mu_b$ to see how $n = m$ change for the specific values of α and $1 - \beta$ given above.

4.2.6 Satterthwaite Two-Sample t Test

Site Contamination Scenario

The site contamination scenario for the Satterthwaite two-sample t test is the same as that for the WRS test in Section 4.2.3. That is, if contamination from site operations has occurred, it probably would be homogeneously distributed throughout the site.



Role of the Data Quality Objectives Process

The only difference between the DQOs for the usual two-sample t test and the Satterthwaite two-sample t test is that the stakeholders and regulators have concluded, based on prior data and statistical tests or on the basis of expert knowledge, that the measurements from the site are not expected to have approximately the same total variance (among measurements) as the background measurements. Recall that a procedure for testing statistically if two sample variances are equal is provided in Conover (1998) and U.S. EPA (2000, p. 4-33).

Advantages and Disadvantages

If less-than values should occur and if those values are replaced by substitute values, such as the detection limit or one-half the detection limit, then the Satterthwaite two-sample t test could be computed. However, the test would give biased and perhaps misleading results. The recommendation in this situation is to replace the Satterthwaite t test with the WRS and Quantile tests. If the less-than values take on multiple values, (for example, <10, <15, etc.,) the Gehan test should be used in place of the WRS test.

BOX 4-16. Procedure for conducting the two-sample t test

1. Use the DQO process to select values of s^2, α, $1-\beta$ and $\mu_s - \mu_b$. Then use the procedure in Box 4-14, as illustrated in Box 4-15, to determine the number of site (*n*) and background (*m*) measurements.
2. Collect the samples and obtain the *n* and *m* site and background measurements.
3. Suppose
   - the *n* site measurements are denoted by $x_1, x_2, \ldots, x_n$
   - the *m* background measurements are denoted by $y_1, y_2, \ldots, y_m$
4. Compute the two-sample t test statistic, denoted by T:

$$T = \frac{\bar{x} - \bar{y}}{\{(n+m)[n-1)s_x^2 + (m-1)s_y^2]/[nm(n+m-2)]\}^{1/2}}$$

where $\bar{x}$ = the arithmetic mean of the *n* site measurements
$\bar{y}$ = the arithmetic mean of the *m* background measurements
s_x^2 = the sample variance of the *n* site measurements (the formula for computing this variable is given in Box 2-4)
s_y^2 = the sample variance of the *m* background measurements (see Box 2-4)

5. The two-sample t test declares:
   - the chemical is a COPC if $T \geq t_{1-\alpha,\, n+m-2}$
   - insufficient evidence exists to conclude that the chemical is a COPC if $T < t_{1-a,\, n+m-2}$

   where $t_{1-\alpha,\, n+m-2}$ is the $100(1-\alpha)$ percentile of the t distribution that has $n + m - 2$ degrees of freedom (*f*). The value of $t_{1-\alpha,\, n+m-2}$ is determined from Table C-9 by using that table with the values of $1-\alpha$ and $n + m - 2$. Note the value of α was specified in Step 1, as part of the process for determining the number of site and background measurements required.

If the two-sample t test declares the chemical is not a COPC, it may indicate either (1) the chemical is indeed not a COPC, (2) the assumptions that underlie the t test are not valid for the site and background measurements, or (3) an insufficient number of measurements (*n* and *m*) were obtained for the two-sample t test to be able to detect the difference in site and background concentration distributions that actually exists.

An evaluation should be made of the possibility the causes in cases 2 or 3 may have resulted in the two-sample t test declaring the chemical is not a COPC.

- First, review the DQO planning process records to make sure the number of samples (*n* and *m*) collected agree with what was determined at that time to be necessary to detect a difference between site and background means that was considered important.
- Second, review the computations conducted to test for normality and equality before the t test was calculated. Verify that the tests were done correctly using the appropriate data. Redo the tests if necessary.
- Third, the shift in the site concentration distribution may in fact be smaller than the shift selected by the stakeholders as being important to detect, in which case no additional measurements are needed. However, as the true difference in means is unknown, update the estimated number of site and background measurements needed to detect the critical (important) shift in the site mean by calculating the variance of the site and background measurements (s_x^2 and s_y^2, respectively) and use the larger of these two variances estimated using the Box 4-14 equation. If this new value, denoted by n′, is larger than either the number of site or background measurements obtained and used in the t test, collect additional samples so n′ site and n′ background measurements are available. Then rerun the t test.

BOX 4-17. Example of the two-sample t test

1. Suppose the values of the parameters for calculating *n* and *m* (see Box 4-14) were specified by the stakeholders and regulators to be $s^2 = 7.5$, $\alpha = 0.025$, $1 - \beta = 0.80$, and $\mu_s - \mu_b = 4$. In Box 4-15 it was shown that n = m = 10 for these parameter values.
2. The n = m measurements were obtained.
3. Suppose the values were as follows:

   Site Measurements (*x*): 90, 77, 81, 210, 92, 130,110,120,140, 84
   Background Measurements (*y*): 23, 15, 78, 26, 90, 99, 87, 34, 17, 10

   There do not appear to be any potential outliers in either dataset. Therefore, tests for outliers do not appear to be needed. Each dataset should be used in a test for normality (Appendix B.1). The reader is encouraged to conduct these tests. Suppose the tests indicate the data can be assumed to be normally distributed.
4. The following calculations were conducted:

$$\bar{x} = 113.4$$
$$\bar{y} = 47.9$$
$$s_x^2 = 1623.82$$
$$s_y^2 = 1287.21$$

$$T = \frac{\bar{x} - \bar{y}}{\left\{(n+m)[(n-1)s_x^2 + (m-1)s_y^2]/[nm(n+m-2)]\right\}^{1/2}}$$

$$= \frac{113.4 - 47.9}{\{(10+10)\ [9*1623.82 + 9*1287.21]/[10*10\ (10+10-2)]\}^{1/2}}$$

$$= \frac{65.5}{17.06}$$

$$= 3.84$$

5. The value of $t_{1-\alpha,\, n+m-2}$, that is, of $t_{0.975,\, 18}$ is found from Table C-9 to be 2.101. Therefore, as T > 2.101, that is, 3.84 > 2.101, the two-sample t test declares that the chemical is a COPC.

Guidance on Implementing the Satterthwaite Two-Sample t Test

Ideally, the same number of measurements should be obtained for both the site and background areas. Let the number of such measurements be denoted by n. The number of site and background measurements should be approximated using the procedure in Box 4-14 that was used for the two-sample t test, where s^2 is now the larger of the site and background measurement variances.

When the *n* measurements have been obtained and the assumption of normality appears reasonable based on the use of statistical tests (Appendix B.1) and graphic methods (Section 2.2.3), the Satterthwaite two-sample t test can be conducted as described in Box 4-18. An example of the Satterthwaite two-sample t test is given in Box 4-19.

BOX 4-18. Procedure for conducting the Satterthwaite two-sample t test

1. Use the DQO process to select values of α, β, $\mu_s - \mu_b$ and the larger of the site and background variances (s^2). Then use the procedure in Box 4-14, as illustrated in Box 4-15, to determine the number of measurements (*n*) for both the site and the background area.
2. Collect the samples and obtain the *n* site and *n* background measurements
3. Suppose
   - the *n* site measurements are denoted by $x_1, x_2, \ldots, x_n$
   - the *n* background measurements are denoted by $y_1, y_2, \ldots, y_n$
4. Compute the Satterthwaite two-sample t test statistic, denoted by T_s:

$$T_s = \frac{\bar{x} - \bar{y}}{\left(s_x^2/n + s_y^2/n\right)^{1/2}}$$

where $\bar{x}$ = the arithmetic mean of the *n* site measurements
$\bar{y}$ = the arithmetic mean of the *n* background measurements
s_x^2 = the sample variance of the *n* site measurements (the formula for computing this variable is given in Box 2-4)
s_y^2 = the sample variance of the *n* background measurements (see Box 2-4).

5. Compute the approximate degrees of freedom, *f*, as follows:

$$f = \frac{\left(s_x^2/n + s_y^2/n\right)^2}{\left(s_x^2/n\right)^2/(n-1) + \left(s_y^2/n\right)^2/(n-1)}$$

Note: the Satterthwaite two-sample t test can be computed when the number of site and background measurements are not equal. In that case, *n* in these equations would be replaced by n_x and n_y, as appropriate.

6. The Satterthwaite two-sample t test declares that:
   - the chemical is a COPC if $T_s \geq t_{1-\alpha,\, f}$
   - insufficient evidence is offered to conclude that the chemical is a COPC if $T_s < t_{1-\alpha,\, f}$, where $t_{1-\alpha,\, f}$ is the $100(1-\alpha)$ percentile of the t distribution that has *f* degrees of freedom. The value of $t_{1-\alpha,\, f}$ is determined from Table C-9 by using that table with the values of $1-\alpha$ and *f*. Linear interpolation may be used to determine $t_{1-a,\, f}$ in Table C-9 if *f* is not an integer.

If the Satterthwaite two-sample t test declares the chemical is *not* a COPC, it may indicate either (1) the chemical is indeed not a COPC, (2) the assumptions that underlie the t test are not valid for the site and background measurements, or (3) an insufficient number of measurements (*n* and *m*) were obtained for the Satterthwaite two-sample t test to be able to detect the difference in site and background concentration distributions that actually exists.

An evaluation should be made of the possibility the causes in cases 2 or 3 may have resulted in the Satterthwaite two-sample t test declaring that the chemical is not a COPC.

1. First, review the DQO planning process records to make sure that the number of samples (*n* and *m*) collected agrees with what was determined at that time to be necessary to detect a possible difference between site and background means that was considered important.
2. Second, review the computations done for the tests for normality and equality of variance conducted on the measurements before the Satterthwaite t test was calculated. Verify the tests were done correctly using the appropriate data. Redo the Satterthwaite t tests if necessary.
3. Third, the shift in the concentration distribution may, in fact, be smaller than the shift selected by the stakeholders, in which case no additional measurements are needed. However, as the true difference in means is unknown, update the estimated number of site and background measurements needed by calculating the variance of the site and background measurements (s_x^2 and s_y^2, respectively) and use the larger of these two variances estimated using the Box 4-14 equation. If this new value, denoted by n', is larger than the number of site and background measurements obtained and used in the t test, then collect additional samples so that n' site and n' background measurements are collected. Then rerun the Satterthwaite t test.

BOX 4-19. Example of the procedure for conducting the Satterthwaite two-sample t test

1. Suppose a preliminary study was conducted to estimate the variance of the background and site measurements, and the variance of the site data was found to be significantly larger than the background data variance. Suppose the larger of the two estimated variances was 15. Therefore, that value was selected as the value for s^2. (If very few site and background measurements were obtained in the preliminary study, say less than 10 for each, the value for s^2 may be increased by 20% or so to guard against not taking enough measurements.) Also, suppose the values of the other parameters for calculating *n* and *m* (see Box 4-14) were specified by the stakeholders and regulators during the DQO process to be $\alpha = 0.10$, $1 - \beta = 0.90$ and $\mu_s - \mu_b = 4$. For these parameter values, n = 14.1, rounded down to n = 14.
2. Therefore, n = 14 site and n = 14 background samples were collected and measured using the methods specified during the DQO process and as documented in the QAPP.
3. Suppose the measurements are as follows:

   Site Measurements (*x*): 7.2, 4.3 10.9,11.5, 2.0, 6.4, 12.1, 2.2, 0.5, 0.9, 1.1, 2.0, 5.1, 10.5
   Background Measurements (*y*): 8.1, 14.2, 5.0, 2.5, 7.2, 4.9, 10.8, 1.1, 8.5, 11.3, 9.2, 2.7, 4.1, 9.1

4. No potential outliers appear to be present in either dataset. Therefore, tests for outliers do not appear to be needed. Each dataset should be evaluated graphically (Section 2.2.3) and using a formal statistical test (Appendix B.1) to evaluate if the data for each dataset can be reasonably assumed to be normally distributed. The reader may verify the assumption of normality appears to be a reasonable assumption for both datasets.
5. Next, the following calculations are conducted:

   $\bar{x} = 5.41$
   $\bar{y} = 6.84$
   $s_x^2 = 18.708$
   $s_y^2 = 14.316$

$$T_s = \frac{\bar{x} - \bar{y}}{\left(s_x^2 / n + s_y^2 / n\right)^{1/2}}$$

$$= \frac{5.41 - 6.84}{\left(18.708/14 + 14.316/14\right)^{1/2}}$$

$$= \frac{-1.43}{1.536}$$

$$= -0.931$$

$$f = \frac{\left(18.708/14 + 14.316/14\right)^2}{\left(18.709/14\right)^2 / 13 + \left(14.316/14\right)^2 / 13}$$

$$= \frac{5.564}{0.1374 + 0.08043}$$

$$= 25.54 \text{ degrees of freedom}$$

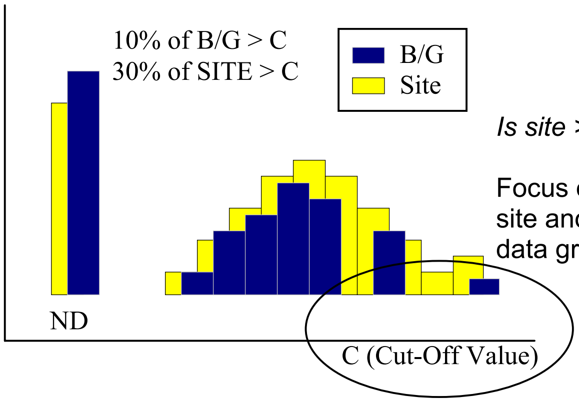
6. Using linear interpolation between $t_{0.90,\,25} = 1.316$ and $t_{0.90,\,26} = 1.315$ in Table C-9, we find that $t_{0.90,\,25.54} = 1.3155$. Therefore, as $T_s < 1.3155$, that is, as $-0.931 < 1.3155$, the Satterthwaite two-sample t test does *not* declare the chemical is a COPC. Indeed, the estimated mean of the site measurements is less than the estimated mean of the background measurements.

BOX 4-19. (cont'd)

7. As the test did not declare the chemical was a COPC, the DQO process records and QAPP should be reviewed to double check that all requirements for collecting the type, quantity, and quality of measurements were correctly followed. Next, evaluate whether the number of measurements used in the test (n = 14) was too small to achieve the allowable α and β decision error rates specified during the DQO process (see Step 1 in this example) and recorded in the QAPP. To do so, calculate *n* and *m* (see Box 4-14) using the larger of the estimated site and background variances, that is, using $s^2 = 18.7$, as computed in Step 5. We find that n = 16.1, which is rounded up to 17, when $s^2 = 18.7$, $\alpha = 0.10$, $1 - \beta = 0.90$ and $\mu_s - \mu_b = 4$. Therefore, three additional samples should be collected and measured in both the background area and at the site. Simple random sampling should be used to determine the locations in the field of the new samples. Also, the collection and measurement protocols specified in the QAPP for obtaining the new data should be exactly the same as for the original data. Then the Satterthwaite two-sample t test should be recomputed using the new background and site datasets, each of which consists of 14 old and three new measurements. Before conducting the Satterthwaite two-sample t test, the graphic methods and statistical test for normality should be conducted on the new datasets (n = 17) to reassess if the normality assumption is still reasonable.

4.2.7 Two-Sample Test of Proportions

Asks if a larger proportion of the site data than of the background data exceeds a concentration C.



Is site > background?

Focus on proportion of site and background data greater than C.

Site Contamination Scenario

For this scenario, assume that site operations may have released a chemical to surface soil at the site, but a distinct contamination pattern of high and low concentrations is not expected to be present. Therefore, a statistical test will be applied to the entire site to indicate whether to reject the null hypotheses indicating the chemical of interest is not a COPC and accept the null hypothesis that the chemical of interest is a COPC. If a distinct contamination pattern is expected, and if sufficient information on that pattern is available or could be obtained, then the site would be separated into separate strata (zones) that are relatively homogeneous. In that case, a separate statistical test and decision would be made for each zone.

Role of the Data Quality Objectives Process

Assume the DQO planning process was implemented. Suppose the DQO planning team, including regulators, believed it was highly likely that more than 50% of the background, and possibly site, measurements would be reported as less-than values. In this case, it is difficult to conduct a valid statistical test of whether the site average (mean or median) is shifted to the right (to higher concentrations) of the background average (mean or median). Therefore, the DQO planning team decided to conduct a statistical test to assess if a larger proportion of the site than of the background area had concentrations greater than a specified concentration C, where C is greater than the detection limit. The two-sample test of proportions is suitable for this situation.

The DQO planning team also agreed:

❑ That the null and alternative hypotheses that will be tested are

$$H_o: \quad P_s \le P_b$$
$$H_a: \quad P_s > P_b$$

where P_s and P_b are the true proportions of the site and background measurements, respectively, that exceed *C* (a predefined concentration level).

- ❑ On the methods that will be used to collect, handle, prepare, and measure the soil samples.
- ❑ That the value of the concentration *C* should be just slightly greater than the largest background less-than value and, therefore, *C* would need to be selected after the background data are obtained.
- ❑ On the parameters needed to compute the number of background and site measurements of the chemical of interest (discussed following).

The assumptions behind using the two-sample test of proportions are summarized in Table 4-1.

Advantages and Disadvantages

- ❑ The test may be conducted regardless of the underlying distribution of the measurements. That is, the test is a distribution-free (nonparametric) test.
- ❑ The test is rather easy to perform.
- ❑ The test requires that the measurements be independent (not spatially or temporally correlated) and that simple random sampling be used to determine the sampling locations in both the background and site areas. However, sampling on a grid pattern is acceptable if the grid pattern does not correspond (line up) with a pattern of changing concentrations for the chemical of interest in either the background or site areas.
- ❑ The test does not evaluate whether the site mean (median) exceeds the background mean (median).

The two-sample proportion test was selected in this case because it is not possible to avoid a large number of less-than values when the measurement method of choice is used.

Guidance on Implementing the Two-Sample Test of Proportions

The number of site (*n*) and background (*m*) measurements required to conduct the two-sample test of proportions should be approximated using the procedure outlined in Box 4-20. An example of the procedure is given in Box 4-21. After the data have been collected according to the specifications worked out during the DQO planning process, the datasets should be examined to look for outliers. A test for outliers should be conducted for any datum that appears to be unusually large, relative to the remaining data in the dataset (see Appendix B, Section B.3). Tests for normality or lognormality of the data need not be conducted.

After the data evaluation process (Section 2.1.6) has been completed (that is, once it has been determined that the data contain no errors, that they have been collected, handled, and measured according to the specifications developed during the DQO process), and the assumptions that underlie the use of the two-sample test of proportions have been shown to be reasonable, then the two-sample test of proportions may be conducted. The procedure for conducting the test is given in Box 4-22. An example is provided in Box 4-23.

4.3 Determination of Background Threshold Levels for COPCs

Once COPCs for a given site have been identified, then it may be necessary to determine background levels for each COPC. These background levels may be used for delineation of the impacted areas or for computing residual risks above ambient (background) conditions. For this purpose, either background datasets, or combined site and background datasets can be used, as discussed below.

4.3.1 COPC Background Levels Using Background Dataset

The following stepwise procedure is similar to those that have been proposed by some regulatory agencies to compute COPC background levels by solely relying on the background datasets.

BOX 4-20. Procedure for calculating the number of site and background measurements required to conduct the two-sample test of proportions

The formula for calculating the number of site (*n*) and background (*m*) measurements required to conduct the two-sample test of proportions is as follows (from U.S. EPA, 2000):

$$n = m = \frac{2(Z_{1-\alpha} + Z_{1-\beta})^2 \, \overline{P}\,(1 - \overline{P})}{D^2}$$

where: $\overline{P}$ = $(P_s + P_b) / 2$.

P_s = the proportion of the true site distribution of potential measurements that exceeds C.

P_b = the proportion of the true background distribution of potential measurements that exceeds C.

α = the tolerable probability that the two-sample test of proportions will incorrectly reject H_o, that is, will incorrectly declare the chemical is a COPC, (α is usually specified to be a small value such as 0.01, 0.025, 0.05 or 0.10).

$1 - \beta$ = the power (probability) required that the two-sample test of proportions will declare that the chemical is a COPC when that is indeed the case (β is usually specified to be ≥ 0.80).

D = the difference in the true (unknown) proportions of the site and background distributions of potential measurements that exceed the constant C *that must be detected with probability* $1 - \beta$. That is, the stakeholders and regulators have agreed that the difference D needs to be detected by the two-sample test of proportions with power (probability) equal to $1 - \beta$.

$Z_{1-\alpha}$ = the $100(1 - \alpha)$ percentile of the standard normal distribution, that is tabulated in Table C-1 (for example, if $\alpha = 0.05$, then Table C-1 indicates that $Z_{1-0.05} = Z_{0.95} = 1.645$).

$Z_{1-\beta}$ = the $100(1 - \beta)$ percentile of the standard normal distribution, that is tabulated in Table C-1 (for example, if $1 - \beta = 0.80$, Table C-1 indicates that $Z_{0.80} = 0.84$).

The appropriate values of the parameters in the above equation should be determined by the stakeholders and regulators during the application of the DQO planning process.

BOX 4-21. Example of the procedure for calculating the number of site and background measurements required to conduct the two-sample test of proportions

Suppose the values of the parameters for calculating *n* and *m* (see Box 4-20) were specified by the stakeholders and regulators as follows:

$D = 0.20$
$\alpha = 0.025$
$\beta = 0.20$
$Z_{1-\alpha} = Z_{0.975} = 1.96$ and $Z_{1-\beta} = Z_{0.80} = 0.84$ (from Table C-1).

Because P_s and P_b are true values and therefore are unknown, estimates of these true proportions must be supplied from a preliminary sampling study conducted at the background and site. This study must be conducted using the same sampling and analysis protocol that will be used in the main study. Suppose a preliminary study based on collecting 20 samples in the background area and 20 samples at the site yields estimates of P_s and P_b to be 0.30 and 0.15, respectively. In this case, $\overline{P} = (0.30 + 0.15) / 2 = 0.225$, and:

$$n = m = 2(1.96 + 0.84)^2 \, 0.225\,(1 - 0.225) / 0.20^2 = 68.35$$

and is rounded up to 69. Therefore, 69 samples are needed from the background area and 69 from the site. Because the 20 site and 20 background samples have already been collected, handled, and measured using the methods required for the full study, only 49 new site and 49 new background measurements need be collected.

BOX 4-22. Procedure for conducting the two-sample test of proportions (from U.S. EPA, 2000)

1. Stakeholders and regulators use the DQO process to select values of α, β, *D*, and *C* (recall that *C* is the concentration limit of interest).
2. Conduct a preliminary sampling and measurement study at the background area and for the region within the Navy site being examined (Region A) to obtain estimates of the true proportions P_s and P_b of the site and background populations that exceed *C*. Then use the procedure in Box 4-20 to determine *n* and *m*, the number of site and background measurements needed.
3. Collect, handle, and measure the *n* and *m* samples, as specified in the sampling and analysis plan and the QAPP.
4. Suppose

   - *n* site measurements are denoted by $x_1, x_2, \ldots, x_n$
   - *m* background measurements are denoted by $y_1, y_2, \ldots, y_m$

   Note: In this document it is recommended that n = m. However, the following formulas are for the more general case where the number of site measurements, *n*, and the number of background measurements, *m*, are not equal.
5. Let k_s and k_b be the number of site and background measurements, respectively, that exceed *C*.
6. Compute $p_s = k_s / n$, which is the estimated proportion of the true distribution of potential site measurements that exceed *C*.
7. Compute $p_b = k_b / m$, which is the estimated proportion of the true distribution of potential background measurements that exceed *C*.
8. Compute

$$p = (k_s + k_b) / (n + m)$$

9. Compute np_s, mp_b, $n(1 - p_s)$, $m(1 - p_b)$. If all of these quantities are greater than or equal to 5, continue with step 10. If not, consult a statistician for assistance, because the computations for the test become more complicated when these quantities are less than 5.
10. Compute the test statistic:

$$Z_p = (p_s - p_b) / [p(1- p)(1/n + 1/m)]^{1/2}$$

11. Use Table C-1 to find $Z_{1-\alpha}$
12. If $Z_p \geq Z_{1-\alpha}$ the test has declared that $P_s > P_b$, that is, that the true proportion of the potential site measurements greater than the concentration value C is greater than the true proportion of the potential background measurements greater than C.

    If $Z_p < Z_{1-\alpha}$, then not enough evidence is present from the data to conclude that $P_s > P_b$. In that case, go to Step 13.
13. Suppose the test declares not enough evidence is present from the data to conclude that $P_s > P_b$. This conclusion may indicate either (1) the chemical is not a COPC, (2) the assumptions that underlie the test are not valid for the site and background measurements, or (3) an insufficient number of measurements (*n* and *m*) were obtained for the test to be able to detect the difference D that actually exists. Evaluate if the causes in cases 2 or 3 may have resulted in the test declaring the chemical is not a COPC. Review the DQO planning process records to make sure the number of measurements (*n* and *m*) agree with what was determined at that time to be necessary to detect the specified difference D. For item 3, use the procedure in Box 4-20 to recompute the number of measurements required for the test. Those computations should be done using the estimates p_s and p_b in place of P_s and P_b, respectively. If the new value of *n* is greater than what was used to compute the test statistic, collect the additional samples needed and rerun the test.

BOX 4-23. Example of the two-sample test of proportions

1. Suppose the stakeholders and regulators specified that $\alpha = 0.025$, $\beta = 0.20$, D = 0.20, and C = 1 ppb for the chemical of interest.
2. Also suppose that a preliminary study was conducted at the site and background area to obtain estimates of the true proportions P_s and P_b. Suppose these estimates were 0.30 and 0.15, respectively. Then, as illustrated in Box 4-21, n = 69 measurements are needed from the site and 69 also from the background area.
3. A total of 138 measurements are obtained. Suppose $k_b = 19$ of the 69 background measurements were greater than C, that is, greater than 1 ppb. Furthermore, suppose that $k_s = 24$ of the site measurements were greater than C. Therefore,

   $$p_b = 19/69 = 0.275$$
   $$p_s = 24/69 = 0.347$$
   $$p = (k_s + k_b) / (n + m) = (19 + 24) / (69 + 69) = 0.3116$$

4. Also,

   $$mp_b = 69(0.275) = 19$$
   $$mp_s = 69(0.347) = 24$$
   $$m(1 - p_b) = 69(1 - 0.275) = 50$$
   $$n(1 - p_s) = 69(1 - 0.347) = 45$$

   all greater than 5. Therefore, we continue on with the test as described in Box 4-22.

5. The test statistic is computed as follows:

   $$\begin{aligned} Z_p &= (p_s - p_b) / [p(1 - p)(1/n + 1/m)]^{1/2} \\ &= (0.347 - 0.275) / [0.3116(1 - 0.3116)(1/69 + 1/69)]^{1/2} \\ &= 0.072 / [0.2145*(0.014493 + 0.014493)]^{1/2} \\ &= 0.072 / 0.0789 \\ &= 0.913 \end{aligned}$$

6. From Table C-1 we find that $Z_{1-\alpha} = Z_{0.975} = 1.96$
7. As $Z_p < 1.96$, that is, $0.913 < 1.96$, the data do not provide sufficient information for the test to reject H_o and declare the chemical is a COPC.
8. We recalculate *n* and *m* (see Box 4-20) to check if this lack of a statistically significant results (at the $1 - \alpha = 0.975$ confidence level) is due to collecting fewer measurements than required to achieve the power of $1 - \beta = 0.80$ when D = 0.20. We obtain:

   $$n = m = 2(1.96 + 0.84)^2 \, 0.3116(1 - 0.3116) / 0.2^2 = 84.09$$

   that indicates 85 site and 85 background measurements are needed. Therefore, too few measurements have been made.

In conclusion, the data indicate the true difference D is estimated to be 0.347 – 0.275 = 0.072. The two-sample test of proportions was not able to declare on the basis of the data this difference was large enough to conclude that $P_s > P_b$.

- ❑ Ensure the adequacy of the size of the background datasets, which usually implies more than 20 to 30 unbiased and representative measurements.
- ❑ Use a statistical test, such as the Shapiro-Wilk W test (Appendix B.1.1), to test the background dataset for normality and lognormality.
- ❑ Compute descriptive statistics for the background dataset (Section 2.2.2).
- ❑ Construct an appropriate normal or lognormal probability plot (Section 2.2.3) of the background data in order to ensure that background dataset represents a unique population (probability plots also can be used to assess outliers

or multiple populations within the background area).

- ❑ Select the background threshold value as the value that represents the upper range of ambient conditions. In the case of an adequate and comprehensive background dataset, the maximum concentration should be used as the background threshold value. If the number of background measurements is small, the threshold value should be selected either as the highest measured value, or the pth percentile (such as 95th or 99th percentile) of measured values, whichever is higher. In practice, a variety of measures (including twice the background mean concentration) have been used as the background threshold value (U.S. EPA, 1998).

4.3.2 COPC Background Levels Using Combined Site and Background Datasets

This approach relies on the incorporation of both site and background datasets. Such combined datasets usually include a large number of samples (greater than 30 samples). For this purpose, this procedure should be followed:

- ❑ Use a statistical test, such as the Shapiro-Wilk W test (Appendix B.1.1), to test the combined site/background dataset for normality and lognormality.
- ❑ Compute descriptive statistics (Section 2.2.2) for the combined site/background dataset.
- ❑ Construct the appropriate normal or lognormal probability plot (Section 2.2.3) of the combined data in order to determine the threshold levels of the ambient condition by determining the inflection or break points in the segmented probability plot.
- ❑ Select the background threshold value as the value that represents the upper range of the lowest concentration population.

This procedure is based on the assumption that in any segmented probability plot, the inflection or break points can be used to differentiate between background and site-impacted populations. Although inflection points may indeed indicate separate underlying populations, no assurance is given that separate populations do indeed exist. However, this potential problem can be addressed by the spatial data analysis techniques presented in Section 2.2.4.

4.4 References

Conover, W.J. 1998. *Practical Nonparametric Statistics*. 3rd ed. Wiley Series in Probability and Statistics, Applied Probability and Statistics Section. New York: Wiley & Sons.

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring.* New York: Van Nostrand Reinhold.

Gilbert, R.O., and J.C. Simpson. 1990. Statistical Sampling and Analysis Issues and Needs for Testing Attainment of Background-Based Cleanup Standards at Superfund Sites. Pp. 1–16 in *Proceedings of the Workshop on Superfund Hazardous Waste: Statistical Issues in Characterizing a Site: Protocols, Tools, and Research Needs,* eds. H. Lacayo, R.J. Nadeau, G.P. Patil, and L. Zaragoza. University Park, PA: Pennsylvania State University, Center for Statistical Ecology and Environmental Statistics.

Isaaks, E.H., and R.M. Srivastava. 1989. *An Introduction to Applied Geostatistics.* New York: Oxford University Press.

Millard, S.P., and K.N. Nagaraj. 2000. *Environmental Statistics with S-PLUS*. Boca Raton, FL.: CRC Press.

Nuclear Regulatory Commission, U.S. Department of Energy, U.S. Environmental Protection Agency, and Department of Defense (NRC et al.). 1997. *Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM).* Final report. NUREG-1575; EPA 402/R-97-016. December.

O'Brien, R.F., and R.O. Gilbert. 1997. Comparing Sites to Background, Part 1. *Environmental Testing & Analysis* (September/October): 12–14.

Palachek, A.D., D.R. Weier, T.R. Gatliffe, D.M. Splett, and D.K. Sullivan. 1994. *Statistical Methodology for Determining Contaminants of Concern by Comparison of Background and Site Data with Applications to Operable Unit 2.* Internal report. SA-94-010. Golden, CO: EG&G Rocky Flats Inc., Rocky Flats Plant, Statistical Applications.

United States Environmental Protection Agency (U.S. EPA). 1994b. *Statistical Methods for Evaluating* *the Attainment of Cleanup Standards, Vol. 3: Reference-Based Standards for Soils and Solid Media*. EPA 230-R-94-004. Washington, DC: Office of Policy, Planning, and Evaluation.

United States Environmental Protection Agency (U.S. EPA). 1997. *Data Quality Evaluation Statistical Toolbox (DataQUEST), User's Guide, EPA QA/G-9D, QA96 Version.* EPA/600/R-96/085. Washington, DC: Office of Research and Development. December.

United States Environmental Protection Agency (U.S. EPA). 1998. *Statistical Tests for Background Comparison at Hazardous Waste Sites.* Interim Draft Supplemental Guidance to RAGS. Atlanta, GA: U.S. EPA Region 4.

United States Environmental Protection Agency (U.S. EPA). 2000. *Guidance for Data Quality Assessment: Practical Methods for Data Analysis. EPA QA/G-9, QA 00 Update.* EPA/600/R-96/084. Office of Environmental Information. July.

5. CASE STUDIES

5.1 Geochemical Method Case Study

The Geochemical Method has been implemented successfully at a number of Navy sites. The case study presented in this section demonstrates the successful application of the Geochemical Method to a suspected chemical release site formerly owned by the Navy.

5.1.1 Site and Surrounding Area History

The site and surrounding area reportedly were used as a disposal area by several federal agencies. Between the early 1940s and the late 1970s, the central western portion of the site served as an encampment for construction workers, and was the location of a military hospital. A vehicle maintenance shop also was located in the vicinity of the site.

The discovery of stained soil and buried scrap metal during construction activities led to investigation of an area immediately adjacent to the site. Soil gas, surface soil, and subsurface soil samples were collected at the adjacent site. Elevated levels of volatile organic compounds (VOCs), semi-volatile organic compounds (SVOCs), PAHs, metals, explosive residues, and total petroleum hydrocarbons (TPH) were detected in a pile of scrap metal. The scrap pile was cleaned up in accordance with state and federal requirements. The site investigation report recommended that the Navy investigate the possibility that contamination extended onto their property.

A Remedial Investigation (RI) characterized the nature and extent of suspected contamination at the site. Objectives of the RI were as follows:

- ❑ Assess the nature and extent of potential contamination resulting from past disposal and burial practices.
- ❑ Evaluate potential human health and ecological risks associated with chemicals detected at the site.
- ❑ Assess the need for further action to address potential risks associated with the site.

5.1.2 Field Investigation

Soil samples were collected at the site during two separate sampling rounds.

During Round I, soil samples were collected to assess the horizontal and vertical extent of contamination (both solid debris and chemical contaminants), and the potential for migration of chemical contaminants. Soil samples were collected in topographically low areas, in areas of suspected contamination (near drums, debris, and tanks), and along the site boundaries. During Round I, 37 surface soil and 26 subsurface soil samples were collected at the locations shown on Figure 5-1. The Round I analytical data were compared to the regional U.S. EPA PRGs for industrial settings. During the initial data evaluation, VOCs, SVOCs, polychlorinated biphenyls (PCBs), TPH, and explosives were eliminated from the list of COPCs because they were either below analytical detection limits or below risk-based screening criteria. Three metals—arsenic, chromium, and lead—were detected at concentrations above their respective risk-based screening criteria. Metal concentrations above the screening criteria appeared to occur only at four localized hot spots. Metal concentrations detected in samples from all other areas of the site were below the screening criteria.

After preliminary evaluation of the Round I data, it was determined that a second round of sampling was necessary to better define the extent of arsenic, chromium, and lead contamination at each of the four areas shown in Figure 5-2. To provide the data required for on-site background analysis, soil samples collected during Round II were analyzed for three non-COPC metals (aluminum, calcium, and iron) in addition to the three target metals. A total of 42 surface soil and 39 subsurface soil samples were collected from the four areas.

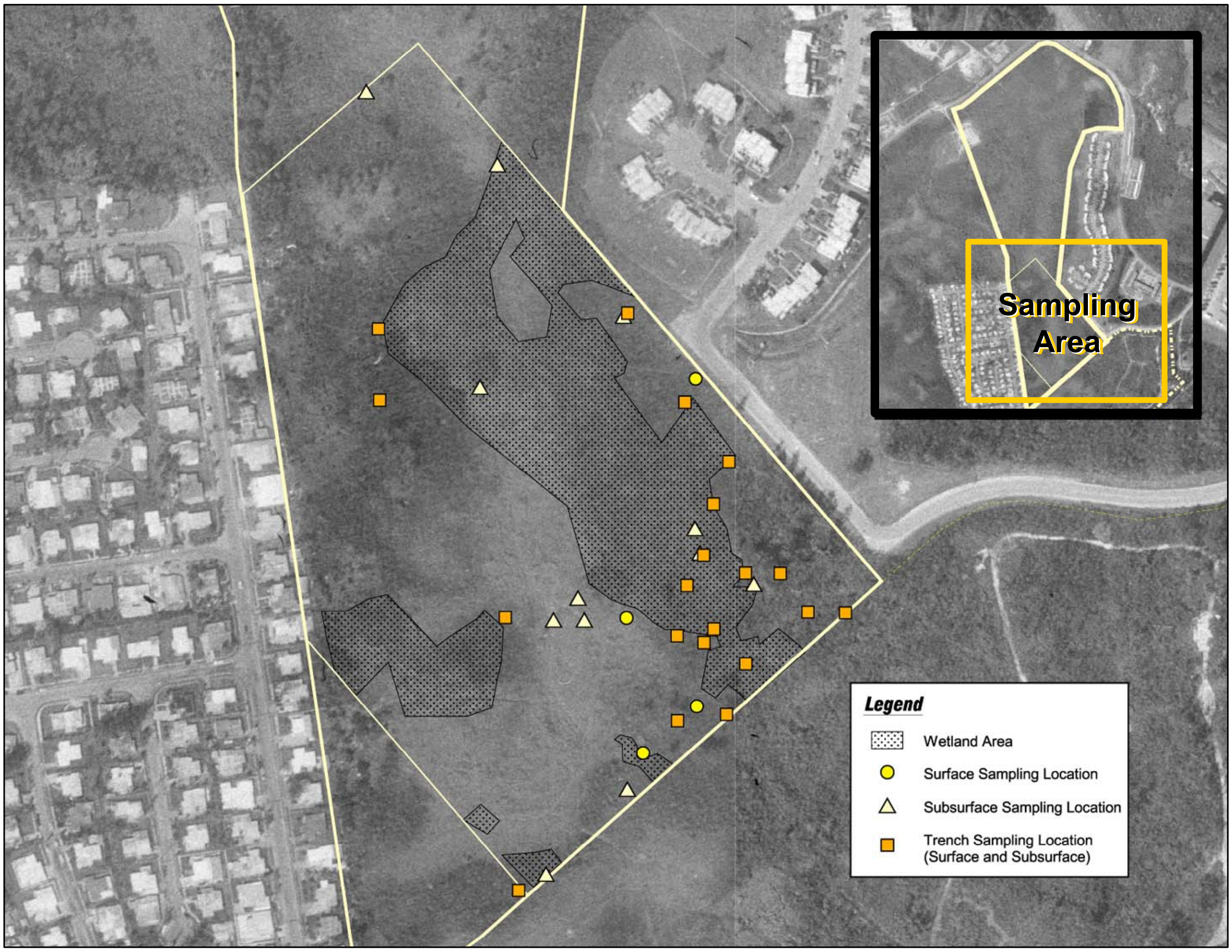


FIGURE 5-1. Round I sampling locations

Descriptive statistics of site data for aluminum, arsenic, chromium, and lead are provided in Table 5-1.

5.1.3 Background Evaluation

The following evaluation was based on metals data collected during both rounds of sampling. It was anticipated that the large dataset would provide an accurate representation of background metals concentrations. Before the Geochemical Method was initiated, the spatial distribution of metals, potential contaminant sources, and potential organic co-contamination were evaluated. As stated previously, concentrations of organic chemicals were either below analytical detection limits or below risk-based screening criteria; therefore, evaluation of organic contaminants was not required.

After evaluating the geochemical characteristics of the soil at the site, aluminum was identified as a non-COPC metal and was selected for use as a reference metal. The geological setting of the site is quite uniform, and only minor variations in soil type were observed among the soil samples; therefore, no soil type univariate plots were constructed.

Data Review and Assessment

Aluminum, arsenic, chromium, and lead concentrations were plotted on univariate plots as functions of data qualifier, sampling depth, and sample location. The univariate plots illustrate the spatial characteristics of the aluminum, arsenic, chromium, and lead concentration distributions. Concentrations of each metal were plotted in mg/kg on a $\log_{10}$ scale (as explained in Section 2.2.4). A

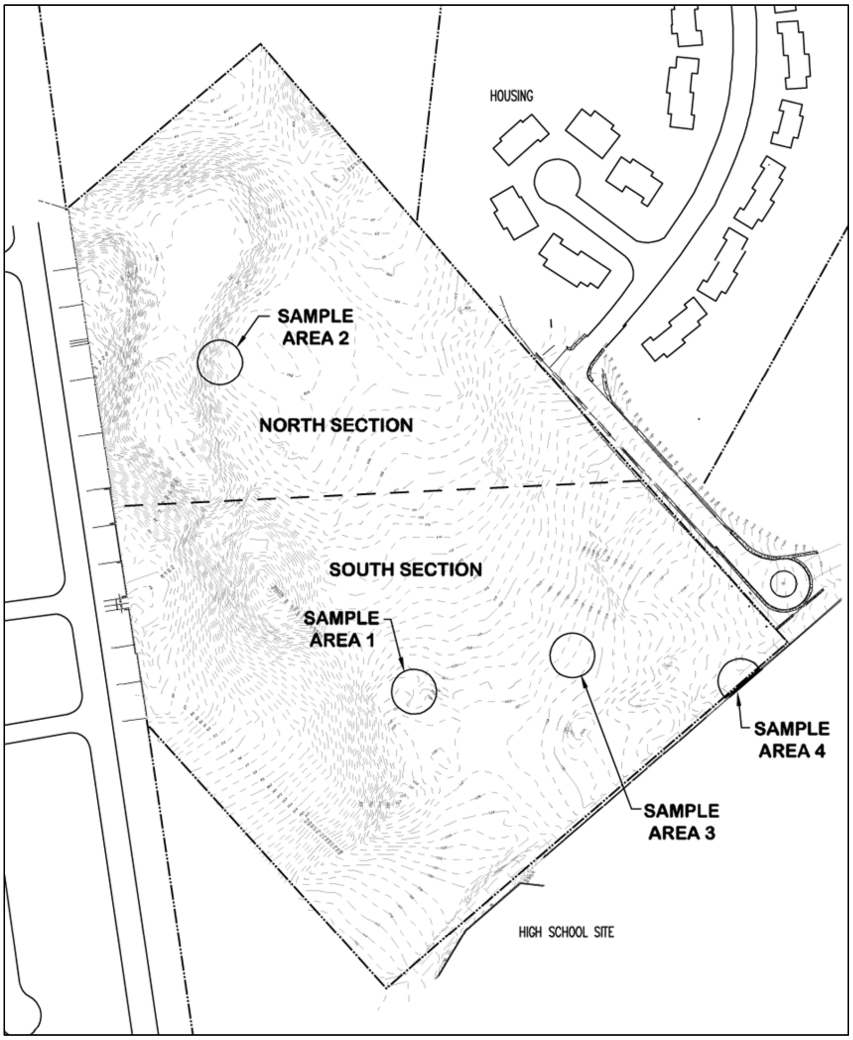


FIGURE 5-2. Round II sampling locations

TABLE 5-1. Descriptive statistics

	Aluminum (mg/kg)	Arsenic (mg/kg)	Chromium (mg/kg)	Lead (mg/kg)
Mean	39,980	5.82	250	69.6
Std.Dev	14,859	6.189	121	509
Count	144	144	144	144
Min.	1,710	0.6	10.9	0.37
Max.	71,000	55	533	6,010
Median	42,600	3.6	218	4.15

probability plot also was constructed for each of the four metals. The probability plots were prepared by plotting the metal concentration associated with each data point versus the cumulative percentage associated with the data point (on a probability scale). Each probability plot was reviewed to identify distinct increases in slope (i.e., inflection points) that may represent the upper bound of the background concentration range. The combined univariate and probability plots then were inspected to estimate the background concentration ranges. (Note that the site was divided into north and south areas for the first sampling round. Samples were collected from Areas 1 to 4 during the second sampling round.)

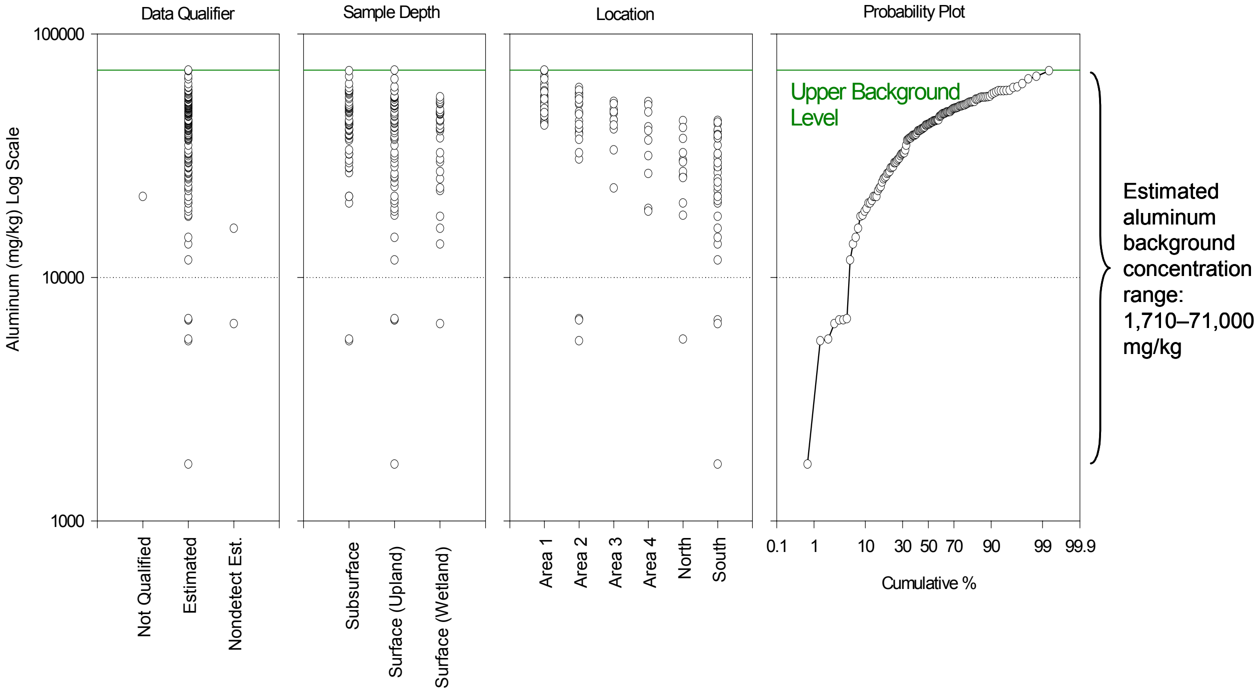
The following four decision questions (see Table 2-5) were addressed to evaluate the combined plots:

- Does the estimate of the upper bound of the background concentration range depend on any U or UJ values?

- ❑ Does the sampling depth univariate plot indicate that concentrations vary according to sampling depth?
- ❑ Does the soil type univariate plot indicate that relatively high concentrations tend to occur only in certain types of soil at the site?
- ❑ Does the sampling location univariate plot indicate that relatively high concentrations occur only in certain areas of the site?

Aluminum

Univariate plots and a probability plot were constructed to evaluate the spatial distribution of aluminum and confirm that it could be used as a non-COPC reference metal (see Figure 5-3). Because the probability plot shows no inflection point and no outliers in the upper concentration range, the initial hypothesis was that all observed aluminum concentrations are within the background concentration range. The combined plots then were evaluated to address the decision questions and confirm this hypothesis (see Table 5-2).



Estimated aluminum background concentration range: 1,710–71,000 mg/kg

FIGURE 5-3. Combined univariate and cumulative probability plots for aluminum

TABLE 5-2. Combined plot analysis for aluminum

Decision Question	Conclusion
• Does the estimate of the upper bound of the background concentration range depend on any U or UJ values?	No, only two aluminum concentration values were qualified as UJ, and none were qualified as U. Therefore, all conclusions are based on reliable and high quality data.
• Does the sampling depth univariate plot indicate that concentrations vary according to sampling depth?	No, aluminum concentrations in surface and subsurface soil samples were very similar. Therefore, the detected aluminum concentrations are likely to be within the background range.
• Does the soil type univariate plot indicate that relatively high concentrations tend to occur only in certain types of soil at the site?	*A soil type univariate plot was not prepared because only one basic soil type occurs at the site (no soil-type dependence exists).*
• Does the sampling location univariate plot indicate that relatively high concentrations occur only in certain areas of the site?	Yes, the maximum concentrations of aluminum occur in the vicinity of Area 1; however, based on the probability and sample depth plots, the high concentrations detected at Area 1 are within the background range.

Analysis of the combined plots indicated that aluminum is not a COPC at the site. The upper bound of the estimated background concentration range was determined to be the maximum detected concentration: 71,000 mg/kg.

Arsenic

The probability plot for arsenic (Figure 5-4) shows an inflection point at 17.2 mg/kg; therefore, the initial hypothesis was that 17.2 mg/kg represents the upper bound of the background concentration range. The combined plots were then evaluated to address the decision questions and confirm this hypothesis (Table 5-3).

Analysis of the combined plots confirmed the hypothesis that 17.2 mg/kg is the upper bound of the background range for arsenic. The three outliers that represent contamination are associated with two surface soil samples (22 and 23 mg/kg) and one subsurface soil sample (55 mg/kg) collected at Area 2 (Figure 5-2).

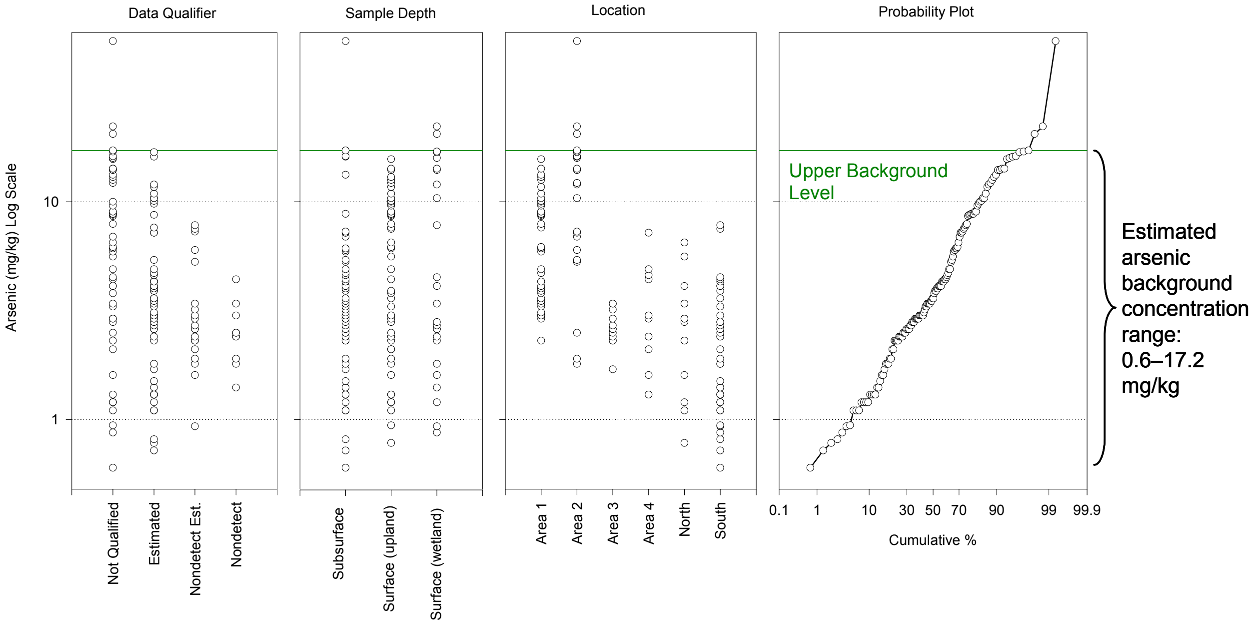


FIGURE 5-4. Combined univariate and cumulative probability plots for arsenic

TABLE 5-3. Combined plot analysis for arsenic

Decision Question	Conclusion
• Does the estimate of the upper bound of the background concentration range depend on any U or UJ values?	No, although a significant percentage of U and UJ data occur in the dataset, most of the data were qualified as NQ or J, and only NQ or J data were used to determine the inflection point.
• Does the sampling depth univariate plot indicate that concentrations vary according to sampling depth?	Yes, one very high concentration value was detected in subsurface soil. However, concentrations above the inflection point (17.2 mg/kg) also were detected in surface soil. Comparison of the sampling depth plot to the probability plot supports the hypothesis that the inflection point represents the upper bound of the background range.
• Does the soil type univariate plot indicate that relatively high concentrations tend to occur only in certain types of soil at the site?	A soil type univariate plot was not prepared because only one basic soil type occurs at the site (no soil-type dependence exists).
• Does the sampling location univariate plot indicate that relatively high concentrations occur only in certain areas of the site?	Yes, arsenic concentrations above the inflection point were detected only in samples from Area 2. This indicates that concentrations above 17.2 mg/kg are likely to represent arsenic contamination.

Chromium

The probability plot for chromium (Figure 5-5) shows no inflection point and no outliers in the upper concentration range; therefore, the initial hypothesis was that all observed chromium concentrations are within the background concentration range. The combined plots then were evaluated to address the decision questions and confirm this hypothesis (Table 5-4).

Analysis of the combined plots indicated that chromium is not a COPC at the site. Therefore, the upper bound of the estimated background concentration range was determined to be the maximum detected concentration: 533 mg/kg.

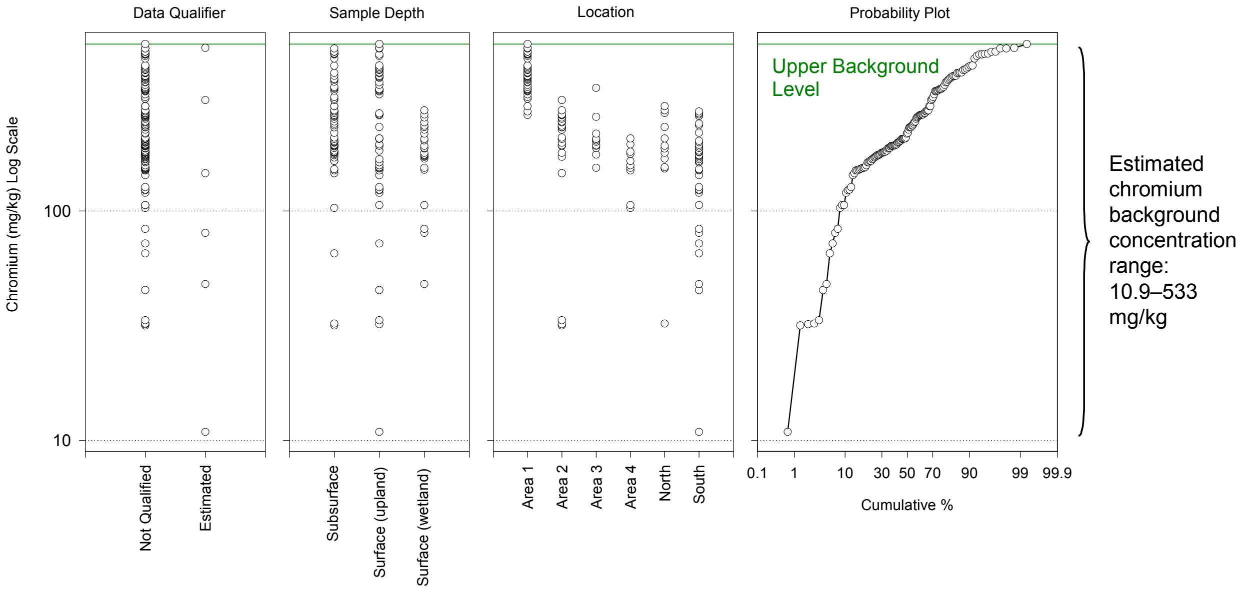


FIGURE 5-5. Combined univariate and cumulative probability plots for chromium

TABLE 5-4. Combined plot analysis for chromium

Decision Question	Conclusion
• Does the estimate of the upper bound of the background concentration range depend on any U or UJ values?	No, all chromium concentration data were qualified as either NQ or J. Therefore, all conclusions are based on reliable and high quality data.
• Does the sampling depth univariate plot indicate that concentrations vary according to sampling depth?	No, chromium concentrations in surface and subsurface soil samples were very similar. Therefore the detected chromium concentrations are likely to be within the background range.
• Does the soil type univariate plot indicate that relatively high concentrations tend to occur only in certain types of soil at the site?	A soil type univariate plot was not prepared because only one basic soil type occurs at the site (no soil-type dependence exists).
• Does the sampling location univariate plot indicate that relatively high concentrations occur only in certain areas of the site?	Yes, the highest chromium concentrations were detected only in samples from Area 1. However, the probability plot indicates that the high concentrations are within the background population distribution, and do not represent contamination. This conclusion also was confirmed by geochemical association analysis.

Lead

The probability plot for lead (Figure 5-6) shows an inflection point at 248 mg/kg; therefore, the initial hypothesis was that 248 mg/kg represents the upper bound of the background concentration range. The combined plots then were evaluated to address the decision questions and confirm the initial hypothesis (Table 5-5).

Analysis of the combined plots confirmed the hypothesis that 248 mg/kg is the upper bound of the background range for lead. The three outliers that represent contamination (394, 1,110, and 6,010 mg/kg) are associated with surface soil samples from Areas 3 and 4 (Figures 5-2 and 5-7).

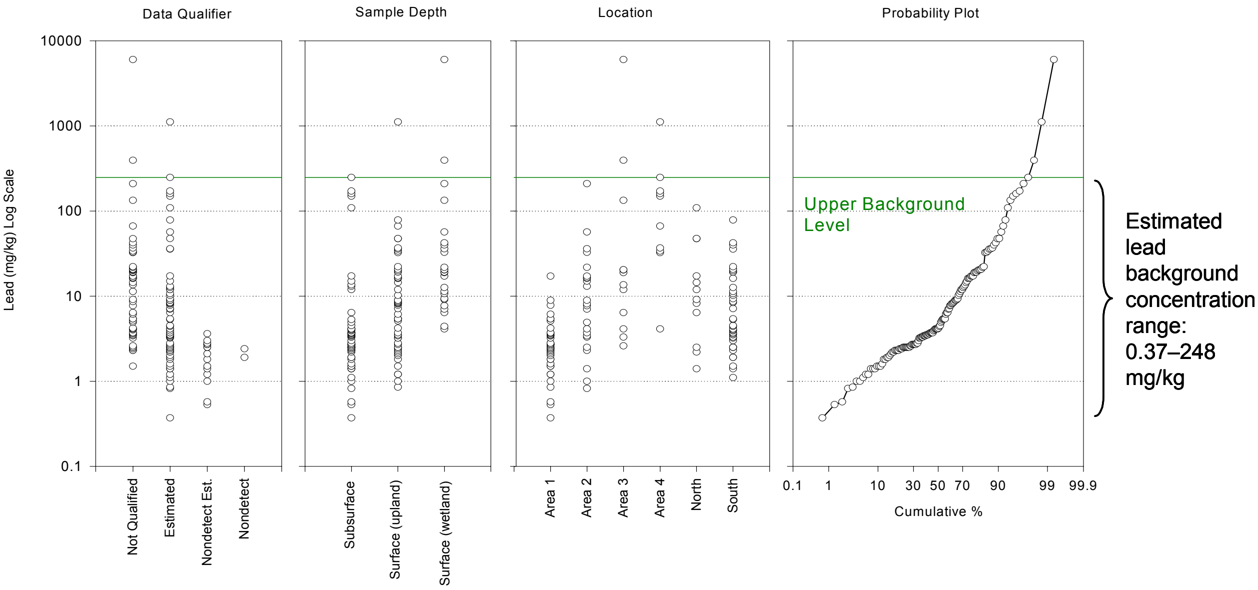


FIGURE 5-6. Combined univariate and cumulative probability plots for lead

TABLE 5-5. Combined plot analysis for lead

Decision Question	Conclusion
• Does the estimate of the upper bound of the background concentration range depend on any U or UJ values?	No, although the dataset includes a small percentage of U and UJ data, most of the data were qualified as NQ or J, and only NQ or J data were used to determine the inflection point.
• Does the sampling depth univariate plot indicate that concentrations vary according to sampling depth?	Yes, lead was detected in surface soil at concentrations well above the upper bound of the subsurface soil lead concentration range. The maximum lead concentration detected in subsurface soil coincides with the inflection point. This supports the hypothesis that 248 mg/kg represents the upper bound of the background range. The three outliers most likely represent lead contamination in surface soil.
• Does the soil type univariate plot indicate that relatively high concentrations tend to occur only in certain types of soil at the site?	A soil type univariate plot was not prepared because only one basic soil type occurs at the site (no soil-type dependence exists).
• Does the sampling location univariate plot indicate that relatively high concentrations occur only in certain areas of the site?	Yes, lead concentrations above the inflection point were detected only in samples from Areas 3 and 4. This indicates that concentrations above 248 mg/kg are likely to represent lead contamination.

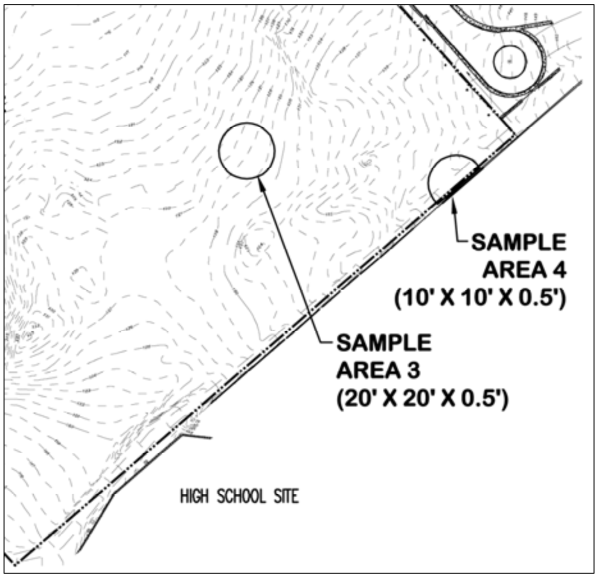


FIGURE 5-7. Surface soil removal areas

TABLE 5-6. Matrix of correlation coefficients

	Aluminum	Arsenic	Chromium	Lead
Aluminum	1.00	0.306	0.896	–0.126
Arsenic		1.00	0.238	–0.041
Chromium			1.00	–0.105
Lead				1.00

Geochemical Method

The combined plot analysis indicated that all detected chromium concentrations are within the background range. However, to illustrate the Geochemical Method and confirm the previous conclusion, chromium was evaluated further by using Step 1 of the Geochemical Method, geochemical association analysis (Section 3.2). Aluminum was used as a reference metal for the chromium geochemical association analysis because it is not a COPC at the site. Also, as shown in Table 5-6, the correlation coefficient (0.896) indicates a strong geochemical association relationship between arsenic and chromium.

A log-log scatter plot (Figure 5-8) was constructed to evaluate the background range. Linear regression was used to plot the best-fit straight line through the data, and the strength of the association between aluminum and each target metal was quantified by evaluating the correlation coefficient. The regression coefficient associated with the least-squares regression line (r^2) was 0.8025, indicating that the regression line shown in Figure 5-8 “explains” 80.25% of the variation in chromium concentrations with a 95% confidence level. All but one of the observed chromium concentrations lie within the 95% prediction interval; this is further evidence of the relatively high strength of the chromium-aluminum relationship. The one outlier represents an anomalously low chromium concentration, suggesting that chromium may have been depleted at the associated sampling location. Because the outlier lies near the middle of the concentration range, it does not affect the estimated background range.

The high degree of correlation between aluminum (a non-COPC) and chromium (the suspected COPC), and the lack of high-concentration outliers on the geochemical regression plot, indicate that all chromium concentrations are within the background range. The regression analysis confirms the conclusion that the relatively high chromium concentrations detected at Area 1 are related to high natural aluminum concentrations, and are therefore within the background range. Therefore, it can be concluded that the geochemical regression analysis supports the results of the combined plot analysis, and that the maximum detected chromium concentration (533 mg/kg) represents the

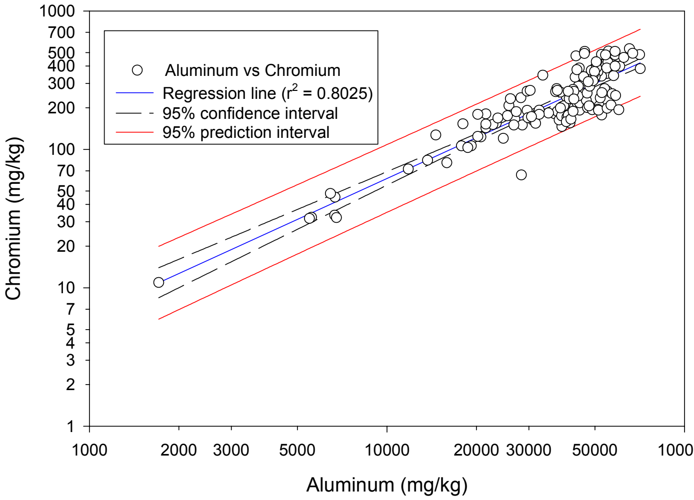


FIGURE 5-8. Geochemical regression: chromium vs. aluminum

upper bound of the estimated natural background range.

5.1.4 Summary of the Geochemical Method Case Study

Chromium was eliminated from the list of COPCs, and nearly all of the detected arsenic concentrations are within the background range; therefore, the site-specific background analysis and site-specific risk assessment conducted for this site allowed the Navy to eliminate two of the four areas from the list of cleanup sites. Cleanup was required only at Areas 3 and 4 (to address lead contamination in surface soil) (Figure 5-7). Although arsenic levels in three samples collected at Area 2 were above the upper bound of the background range, cleanup was not required because the site-specific risk assessment concluded that risks were within the acceptable range.

Table 5-7 lists the maximum concentration, the 95% UCL concentration, and the upper bound of the site-specific estimated background concentration range for each metal, and compares the concentration values to U.S. EPA regional PRGs. The maximum aluminum concentration is below the residential PRG. Although most of the detected arsenic concentrations are well above both the residential and industrial PRGs, arsenic concentrations above the background range were detected in only three samples. Chromium concentrations above the PRG values were detected at many locations; however, the background analysis indicates that all detected chromium concentrations are within the background range. Therefore, chromium was eliminated from the COPC list. Lead concentrations above the residential PRG were detected in two samples. The background analysis indicates that the lead concentrations detected in these samples are well above the upper bound of the background concentration range.

5.2 Comparative Method Case Study

The case study involves the environmental investigation at a former industrial facility (Figure 5-9). To streamline presentation of this case study, the steps leading to selection of the Comparative Method (as outlined in Section 4) are not detailed in this section; it is assumed that the method has already been selected.

5.2.1 Site and Surrounding Area History

This facility was in operation from 1945 to 1980. The operation included use and disposal of liquids containing inorganic constituents, such as arsenic. In 1985, the U.S. EPA conducted a preliminary

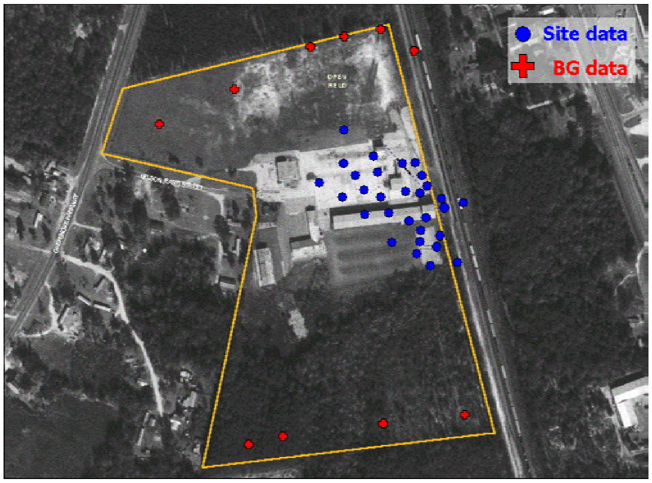


FIGURE 5-9. Site aerial photograph

TABLE 5-7. Final background ranges

Element	Maximum Detected Concentration at Site (mg/kg)	95% UCL of the Dataset (mg/kg)	Estimated Site-Specific Background Concentration (mg/kg)	U.S. EPA Screening Criteria (mg/kg)	
				Residential PRG	Industrial PRG
Aluminum	71,000	46,000	71,000	76,000	100,000
Arsenic	55	7	17.2	0.39	2.7
Chromium	533	289	533	210	450
Lead	6,010	34	248	400	1,000

site investigation and concluded that the site should not be placed on the National Priorities List (NPL), a published list of U.S. hazardous waste sites that are eligible for extensive, long-term cleanup under the Superfund program. In 1995, after reconsideration of the past data, the site was placed on the state hazardous sites inventory.

5.2.2 Field Investigation

After the site was listed as a state hazardous site, a site investigation was conducted. Most of the soil sampling effort was centered around the potentially impacted area in the vicinity of the former operation area (Figure 5-9). Upon concurrence of the regulatory agency, two zones, substantially down- and upgradient of the former operation area, were designated as background areas. These latter areas were sampled for inorganic constituents, including arsenic and copper. Summary results of site and background data are listed in Table 5-8. Descriptive statistics of these measured values are provided in Table 5-9.

TABLE 5-8. Site and background data

Site Data			Background Data		
Sample	Arsenic (ppm)	Copper (ppm)	Sample	Arsenic (ppm)	Copper (ppm)
1	26	45	1	20	160
2	16	1	2	17	100
3	34	132	3	10	1
4	41	98	4	1	124
5	88	213	5	14	220
6	50	88	6	18	124
7	37	152	7	48	250
8	451	227	8	11	86
9	20	44	9	20	1
10	115	136	10	1	225
11	1	188			
12	60	164			
13	12	229			
14	76	43			
15	340	123			
16	83	139			
17	109	77			
18	1,234	205			
19	17	1			
20	1	123			
21	39	168			
22	235	341			
23	408	32			
24	47	326			
25	119	158			
26	155	160			
27	17	117			
28	1	165			
29	560	190			
30	100	201			

5.2.3 Background Evaluation

The main background issues at this site:

- Is copper a COPC at this site?
- What is the background level of arsenic?

Copper

As Table 5-9 indicates, copper is present in both site and background samples; however, the concentration differences do not appear to be significant. In this case, the Slippage and two-sample t tests were conducted to ensure the statistical similarity between site and background copper concentrations. These tests were selected for the following reasons:

- The operational history of the facility did not indicate any excessive use or releases of copper at the site; as a result, the null hypothesis (H_0) was defined as the following: copper is not a COPC.
- The descriptive statistics of soil copper measurements (Table 5-9) do not indicate a substantial difference between site and background concentrations.

TABLE 5-9. Descriptive statistics

	Site Data		Background Data		Combined Data	
Parameter	Arsenic (ppm)	Copper (ppm)	Arsenic (ppm)	Copper (ppm)	Arsenic (ppm)	Copper (ppm)
Mean	150	143	16	129	116	139
Standard Deviation	250	83	13	87	224	83
Minimum	1	1	1	1	1	1
Maximum	1,234	341	48	250	1,234	341
Median	55	145	16	124	38	137

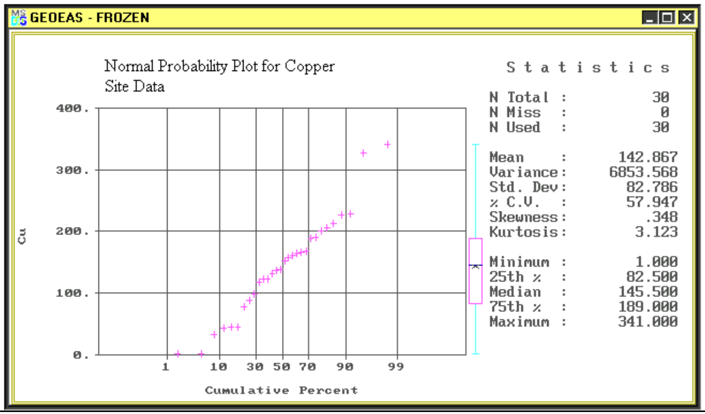
- The combined site and background datasets appear to be derived from a single normally distributed population, as demonstrated by their normal probability plot and by the Shapiro-Wilk test for normality (Figure 5-10).
- The assessment of differences among extreme site and background copper concentrations was conducted using the Slippage test, which is a nonparametric test. This test does not require any specific distributional assumptions about the extreme site and background concentrations (Table 5-10).
- The difference between central tendencies (i.e., mean concentrations) of site and background datasets is assessed by the two-sample t test. This test was selected due to the nearly normal distribution of the observed measurements (Table 5-11).

As displayed in Tables 5-10 and 5-11, both tests indicate that differences between the site and background copper concentrations are not statistically significant. Therefore, copper was not identified as a COPC.

Arsenic

Combining site and background arsenic concentrations on a probability plot resulted in a highly segmented plot (Figure 5-11). The Shapiro-Wilk normality test was applied to the site, background, and combined arsenic datasets (Figure 5-11). These test results indicate that none of the arsenic datasets can be considered as a single normally distributed population. The probability plot (Figure 5-11) indicates the presence of multiple populations

Site Data

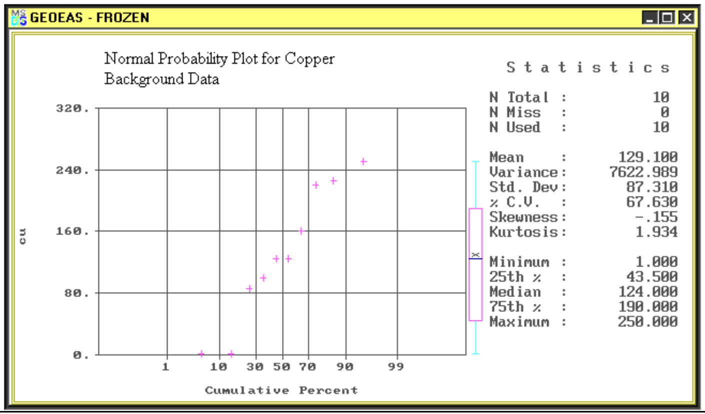


```
S t a t i s t i c s
N Total  :        30
N Miss   :         0
N Used   :        30

Mean     :   142.867
Variance:   6853.568
Std. Dev:     82.786
% C.V.   :    57.947
Skewness:       .348
Kurtosis:      3.123

Minimum  :     1.000
25th %   :    82.500
Median   :   145.500
75th %   :   189.000
Maximum  :   341.000
```

Shapiro-Wilk Test for Normality:	W-Statistic	n	Critical $W_{0.05}$
Copper	0.963	30	0.939

$W>W_{0.05}$: Normality of site copper data is not rejected

Background Data

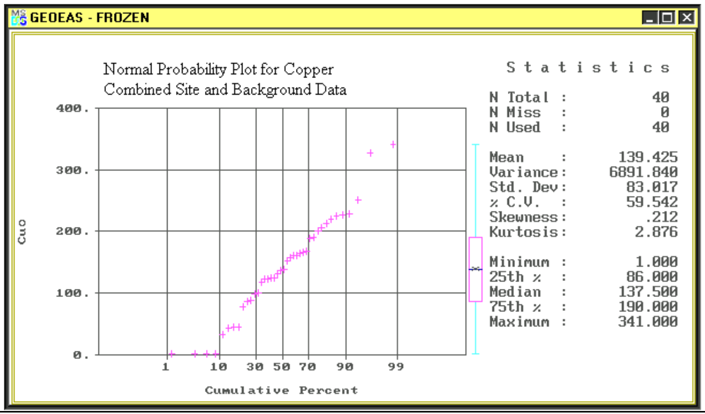


```
S t a t i s t i c s
N Total  :        10
N Miss   :         0
N Used   :        10

Mean     :   129.100
Variance:   7622.989
Std. Dev:     87.310
% C.V.   :    67.630
Skewness:      -.155
Kurtosis:      1.934

Minimum  :     1.000
25th %   :    43.500
Median   :   124.000
75th %   :   190.000
Maximum  :   250.000
```

Shapiro-Wilk Test for Normality:	W-Statistic	n	Critical $W_{0.05}$
Copper	0.930	10	0.869

$W>W_{0.05}$: Normality of background copper data is not rejected

Combined Data



```
S t a t i s t i c s
N Total  :        40
N Miss   :         0
N Used   :        40

Mean     :   139.425
Variance:   6891.840
Std. Dev:     83.017
% C.V.   :    59.542
Skewness:       .212
Kurtosis:      2.876

Minimum  :     1.000
25th %   :    86.000
Median   :   137.500
75th %   :   190.000
Maximum  :   341.000
```

Shapiro-Wilk Test for Normality	W-Statistic	n	Critical $W_{0.05}$
Copper	0.969	40	0.949

$W>W_{0.05}$: Normality of combined site and background copper data is not rejected

FIGURE 5-10. Exploratory data analysis for copper
($W_{0.05}$ is the critical value at 95% confidence limit from Table C-2)

TABLE 5-10. Slippage test results for copper

Site Data	Background Data
341	250
326	225
229	220
227	160
213	124
205	124
201	100
190	86
188	1
168	1
165	
164	
160	
158	
152	
139	
136	
132	
123	
123	
117	
98	
88	
77	
45	
44	
43	
32	
1	
1	

H_o = Copper is not a COPC.
α = 5%.
$K = 2 < K_c = 10$.
Result = Copper is not a COPC.

TABLE 5-11. Student's two-sample t-test results for copper

Parameter	Site Data	Background Data
Mean	143	129
Variance	6,858	7,621
Observations	30	10
Pooled Variance	7,039	
Hypothesized Mean Difference	0	
df	38	
t-Stat	0.449	
t-Critical (one-tail at α = 5%)	1.686	

H_o = Copper is not a COPC.
α = 5%.
t-Stat = 0.449 < t-Critical = 1.686.
Result = Copper is not a COPC.

(i.e., background versus impacted). As discussed in Section 2.2.3, the observed inflection points can be used as preliminary cutoffs between various populations.

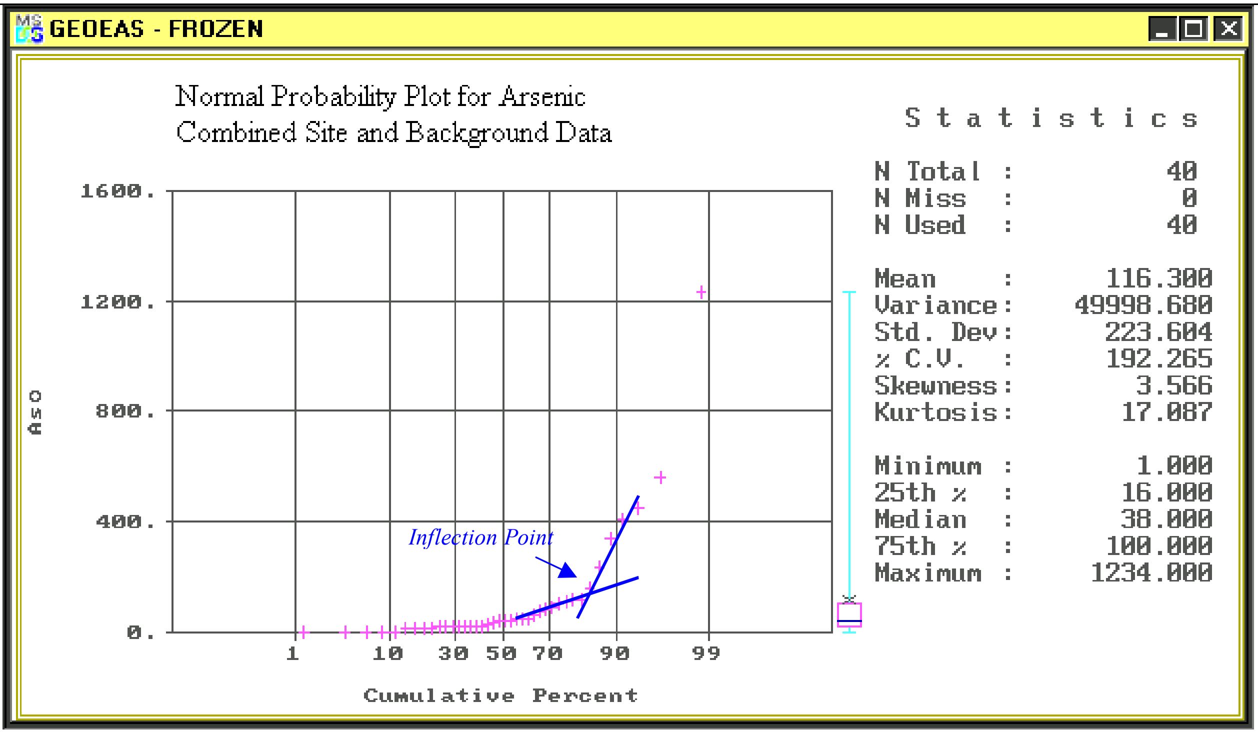
Arsenic was identified as a COPC due to the significant difference between the site and background mean concentrations, as illustrated in Table 5-9, and due to observed multiple populations seen in the probability plots. This conclusion is consistent with the site knowledge derived from the operational history of the facility, where past arsenic releases were reported.

5.2.4 Summary of the Comparative Method Case Study

The background analysis for this case study assisted in attaining the following results:

- ❑ Copper was established as a soil background constituent, and thus was not considered in subsequent risk calculations.
- ❑ The background level of arsenic was negotiated with regulators based on the observed values, which substantially limited the delineation efforts associated with the arsenic-impacted soil at the site.
- ❑ Given the alternative computed ranges of arsenic background levels at the site, negotiation efforts concerning a favorable site-specific industrial risk-based level were substantially minimized.

Combined Site and Background



```
S t a t i s t i c s

N Total  :       40
N Miss   :        0
N Used   :       40

Mean     :  116.300
Variance: 49998.680
Std. Dev:   223.604
% C.V.   :  192.265
Skewness:     3.566
Kurtosis:    17.087

Minimum  :    1.000
25th %   :   16.000
Median   :   38.000
75th %   :  100.000
Maximum  : 1234.000
```

Site Data

Shapiro-Wilk Test for Normality

	W-Statistic	df	Significance
Arsenic	0.961	30	0.412*

*Not considered normally distributed

Background Data

Shapiro-Wilk Test for Normality

	W-Statistic	df	Significance
Arsenic	0.934	10	0.476*

*Not considered normally distributed

Combined Site and Background

Shapiro-Wilk Test for Normality

	W-Statistic	df	Significance
Arsenic	0.963	40	0.350*

*Not considered normally distributed

FIGURE 5-11. Exploratory data analysis for arsenic

APPENDIX A: GEOCHEMISTRY FOR BACKGROUND ANALYSIS

A.1 Metals in Rocks and Minerals

A.1.1 Rocks and their Composition

Rocks are classified as igneous, sedimentary, or metamorphic according to mode of origin (see Box A-1), and on the basis of characteristic groupings or assemblages of particular minerals. For example, granite consists primarily of alkalic feldspar (potassium- and sodium-aluminosilicates) and quartz (silicon dioxide), with small amounts of muscovite, biotite, and hornblende; whereas basalt is composed primarily of calcic plagioclase (calcium-aluminosilicates) and pyroxene minerals, with or without olivine (pyroxene and olivine are rich in iron and magnesium).

A.1.2 Minerals

A mineral is a natural chemical element or compound that has a specific chemical composition, usually with a characteristic crystalline structure. The crystalline structure of a mineral reflects an orderly, repetitive arrangement of atoms and specific ratios of constituent elements. Figure A-1 shows halite, or common table salt, a simple mineral formed by ionic bonding between metal (sodium) atoms and nonmetal (chlorine) atoms. The ratio of sodium atoms to chlorine atoms in halite is 1:1; however, the atomic weight of sodium is less than two-thirds that of chlorine, resulting in a sodium/chloride weight ratio of approximately 0.65.

Arsenopyrite, or iron arsenide sulfide, is one example of a mineral that may contribute to elevated

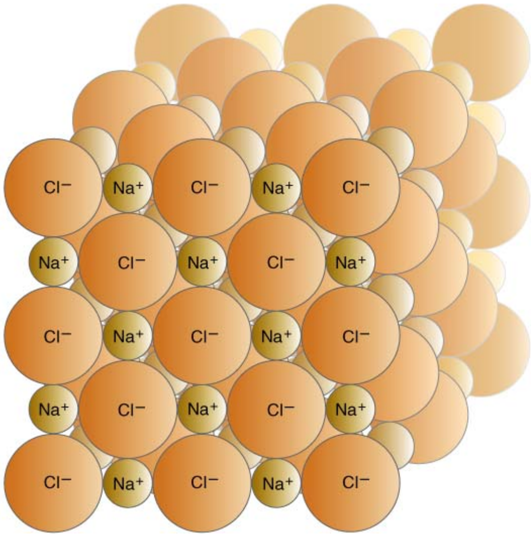


FIGURE A-1. Ionic bonding of metal and nonmetal atoms in halite (common table salt)

BOX A-1. Rock types

Rocks are the solid aggregates of mineral grains that form the earth's crust. Although some rocks (e.g., limestone and quartz sandstone) contain just one mineral, most rocks contain groupings or assemblages of multiple types of mineral. Rocks are classified into the following three fundamental groups, based on mode of origin:

Igneous rocks are formed when minerals crystallize from molten rock that is either extruded at the earth's surface (lava) or intruded deeper within the crust (magma). Depending on the temperatures and pressures that exist at the time of crystallization, igneous rocks can be coarse-grained (e.g., granite), fine-grained (e.g., basalt), or glassy (e.g., obsidian).

Sedimentary rocks are typically stratified and may consist of fragments of older rocks (e.g., gravel, sand), broken shells, rounded mineral grains, secondary minerals such as clays, or chemical precipitates.

Metamorphic rocks are formed when sedimentary or igneous rocks are altered by heat and/or pressure, and are classified according to mineral assemblage and texture. Thermally metamorphosed rocks occur at the edges of igneous intrusions, and are characterized by secondary minerals such as andalusite and garnet. The intense pressures and high temperatures that exist deep within the crust can alter the mineralogy of rocks over wide areas, forming platy minerals such as mica, and other high-pressure minerals (e.g., staurolite). These regionally metamorphosed rocks are often exposed when rocks are uplifted and eroded to form mountain ranges.

background arsenic concentrations. The chemical formula for arsenopyrite, FeAsS, indicates that iron, arsenic, and sulfur atoms combine in a 1:1:1 ratio. The iron/arsenic weight ratio is approximately 0.75. Table A-1 lists a few of the approximately 500 arsenic-containing minerals that have been found in nature. The elemental ratios can be determined from the chemical formulas.

Ten chemical elements—oxygen, silicon, aluminum, iron, calcium, sodium, potassium, magnesium, titanium, and phosphorus—account for more than 99 percent of the mass of the earth's crust (Alloway, 1990). Elements classified as heavy metals (Table A-2) account for a very small percentage of the total mass of the crust; these elements most commonly occur as trace constituents of the principal rock-forming minerals. Concentrations of individual heavy metals in crustal rocks are typically less than 0.01 percent (100 mg/kg) and rarely more than 0.1 percent (1,000 mg/kg). (However, heavy metals may be highly concentrated in the ore minerals such as galena, the principal ore of lead, and cinnabar, the principal ore of mercury.) In igneous rocks, heavy metals become incorporated into minerals at trace concentrations by substituting for cations (positively charged ions) of one of the major elements at the time of crystallization. This substitution is governed by the ionic charge, ionic radius, and electronegativity of the major element and the trace element replacing it (Alloway, 1990).

TABLE A-1. Arsenic-containing minerals

Mineral	Chemical Formula
Arsenopyrite (iron arsenide sulfide)	FeAsS
Conichalcite (hydrous calcium copper arsenate)	$CaCu(AsO_4)(OH)$
Erythrite (hydrated cobalt arsenate)	$Co_3(AsO_4)_2{\cdot}8H_2O$
Mimetite (lead arsenate chloride)	$Pb_5(AsO_4)_3Cl$
Orpiment (arsenic sulfide)	As_2S_3
Scorodite (hydrated iron arsenate)	$FeAsO_4{\cdot}2H_2O$

TABLE A-2. Selected naturally occurring heavy metals

Name	Symbol	Atomic Weight (grams/mole)
Antimony	Sb	121.76
Arsenic	As	74.92
Cadmium	Cd	112.41
Chromium	Cr	51.99
Cobalt	Co	58.93
Copper	Cu	63.55
Gold	Au	196.97
Lead	Pb	207.20
Manganese	Mn	54.94
Mercury	Hg	200.59
Molybdenum	Mo	95.94
Nickel	Ni	58.69
Selenium	Se	78.96
Silver	Ag	107.87
Tin	Sn	118.71
Thallium	Tl	204.38
Tungsten	W	183.85
Uranium	U	238.03
Vanadium	V	50.94
Zinc	Zn	65.39

Table A-3 lists the average concentrations of heavy metals in various igneous and sedimentary rocks.

Different types of rocks can have very different average metals concentrations and ranges. Arsenic concentrations, for example, vary between approximately 1 and 900 mg/kg, depending on rock type. Arsenic and other heavy metals may occur naturally at high concentrations in fine-grained sedimentary rocks such as shale. Chromium and nickel often occur at very high concentrations in ultramafic igneous rocks (e.g., dunite, peridotite, serpentinite). An important feature is that concentrations of certain metals, such as chromium and nickel, tend to vary proportionally, regardless of rock type: i.e., even though chromium and nickel concentrations vary over wide ranges depending on the rock type, the concentration ratios of these two metals tend to be relatively constant.

A.1.3 Elemental Distribution and Association

As a result of the geochemical properties of the elements, certain types of rocks and the soils developed on them contain distinctive groups of metals and other elements (i.e., associated elements). Patterns of elemental distribution and association can be very useful for background analysis. Early in the twentieth century, the geochemist V.M. Goldschmidt developed a classification system based on the distribution of metals in the geological environment. Goldschmidt's system

TABLE A-3. Mean concentrations of heavy metals in unaltered rocks (mg/kg)

Name	Earth's Crust	Igneous Rocks			Sedimentary Rocks		
		Ultramafic	Mafic	Granitic	Limestone	Sandstone	Shales
Antimony	0.2	0.1	0.2	0.2	0.3	0.05	1.5
Arsenic	1.5	1	1.5	1.5	1	1	13 (1–900)
Cadmium	0.1	0.12	0.13	0.09	0.028	0.05	0.22 (<240)
Chromium	100	2,980	200	4	11	35	90
Cobalt	20	110	35	1	0.1	0.3	19
Copper	50	42	90	13	5.5	30	39
Gold	0.004	0.003	0.003	0.002	0.002	0.003	0.0025
Lead	14	14	3	24	5.7	10	23
Manganese	950	1,040	1,500	400	620	460	850
Mercury	0.05	0.004	0.01	0.08	0.16	0.29	0.18
Molybdenum	1.5	0.3	1	2	0.16	0.2	2.6
Nickel	80	2,000	150	0.5	7	9	68
Selenium	0.05	0.13	0.05	0.05	0.03	0.01	0.5
Silver	0.07	0.06	0.1	0.04	0.12	0.25	0.07
Thallium	0.6	0.0005	0.08	1.1	0.14	0.36	1.2
Tin	2.2	0.5	1.5	3.5	0.5	0.5	6
Tungsten	1	0.1	0.36	1.5	0.56	1.6	1.9
Uranium	2.4	0.03	0.43	4.4	2.2	0.45	3.7
Vanadium	160	40	250	72	45	20	130
Zinc	75	58	100	52	20	30	120

Source: Alloway (1990, Table 3.2).

(Table A-4) classifies metals based on their affinity for iron (siderophiles), sulfur (chalcophiles), and silicates (lithophiles). Because aluminum is a major component of many silicate minerals, the lithophile metals also exhibit an affinity for aluminum. These classifications reflect commonly observed patterns of elemental distribution and association.

As a result of elemental (i.e., geochemical) association, concentration ratios of certain metals in rocks and soils are often restricted to relatively narrow ranges. This phenomenon is very useful in background analysis, as discussed in Section 3.2.

The electronic structure relationships reflected in the periodic table (see Figure A-2 and Box A-2) can be used to predict the distribution and association of metals and other elements in geologic materials. Metals with the ability to substitute for each other in the crystalline structure of minerals are commonly associated in the geological environment. In addition to the electrical charge relationships reflected in the organization of the periodic table, ionic radius also is an important factor in elemental association—metals with similar ionic radii are commonly associated due to their ability to substitute for each other.

TABLE A-4. Goldschmidt's classification system for metals

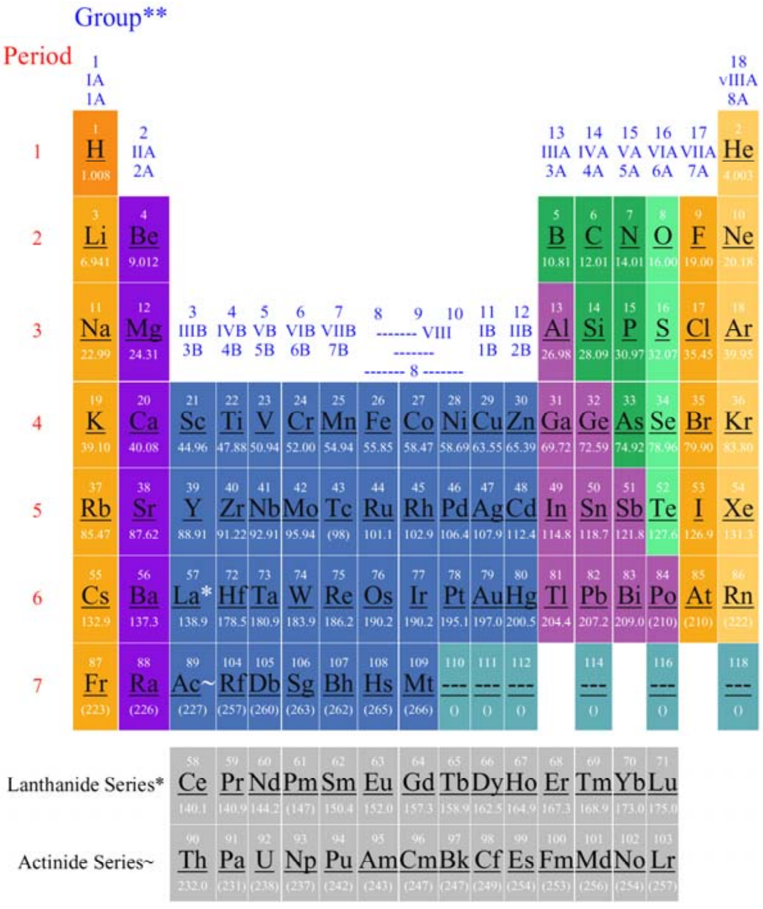
Category	Affinity for	Metal
Siderophiles	Iron	Cobalt Nickel Gold (Lead)[(a)] (Arsenic)[(a)]
Chalcophiles	Sulphur (normally occur in sulphide deposits)	Copper Silver (Gold)[(a)] Zinc Cadmium Mercury Lead Arsenic Antimony Selenium Thallium
Lithophiles	Silicates[(b)]	Vanadium Chromium Manganese Uranium

(a) (Metal) primarily belongs in another group, but has some characteristics that relate it to this group.

(b) Aluminum is a major component of many silicate minerals.

Igneous Rocks

Igneous rocks are common parent rocks for soil formation, and are the original sources of the geologic materials that form most sedimentary and metamorphic rocks. Therefore, the factors involved in elemental association in igneous rocks influence elemental association in many soils. The composition of the magma from which igneous rocks are crystallized (along with pressure and temperature conditions) determines which minerals are formed. *Felsic* magmas—magmas rich in potassium, sodium silica, and aluminum—crystallize into minerals that make up rocks such as granite. *Mafic* magmas—magmas rich in magnesium, iron, and calcium (as well as silica and aluminum)—crystallize into minerals that make up rocks such as basalts. *Ultramafic* magmas crystallize into minerals that make up rocks such as serpentinite.



Group**

Period	1 IA 1A	2 IIA 2A	3 IIIB 3B	4 IVB 4B	5 VB 5B	6 VIB 6B	7 VIIB 7B	8 VIII 8	9 VIII 8	10 VIII 8	11 IB 1B	12 IIB 2B	13 IIIA 3A	14 IVA 4A	15 VA 5A	16 VIA 6A	17 VIIA 7A	18 VIIIA 8A
1	1 H 1.008																	2 He 4.003
2	3 Li 6.941	4 Be 9.012											5 B 10.81	6 C 12.01	7 N 14.01	8 O 16.00	9 F 19.00	10 Ne 20.18
3	11 Na 22.99	12 Mg 24.31											13 Al 26.98	14 Si 28.09	15 P 30.97	16 S 32.07	17 Cl 35.45	18 Ar 39.95
4	19 K 39.10	20 Ca 40.08	21 Sc 44.96	22 Ti 47.88	23 V 50.94	24 Cr 52.00	25 Mn 54.94	26 Fe 55.85	27 Co 58.47	28 Ni 58.69	29 Cu 63.55	30 Zn 65.39	31 Ga 69.72	32 Ge 72.59	33 As 74.92	34 Se 78.96	35 Br 79.90	36 Kr 83.80
5	37 Rb 85.47	38 Sr 87.62	39 Y 88.91	40 Zr 91.22	41 Nb 92.91	42 Mo 95.94	43 Tc (98)	44 Ru 101.1	45 Rh 102.9	46 Pd 106.4	47 Ag 107.9	48 Cd 112.4	49 In 114.8	50 Sn 118.7	51 Sb 121.8	52 Te 127.6	53 I 126.9	54 Xe 131.3
6	55 Cs 132.9	56 Ba 137.3	57 La* 138.9	72 Hf 178.5	73 Ta 180.9	74 W 183.9	75 Re 186.2	76 Os 190.2	77 Ir 190.2	78 Pt 195.1	79 Au 197.0	80 Hg 200.5	81 Tl 204.4	82 Pb 207.2	83 Bi 209.0	84 Po (210)	85 At (210)	86 Rn (222)
7	87 Fr (223)	88 Ra (226)	89 Ac~ (227)	104 Rf (257)	105 Db (260)	106 Sg (263)	107 Bh (262)	108 Hs (265)	109 Mt (266)	110 --- 0	111 --- 0	112 --- 0		114 --- 0		116 --- 0		118 --- 0
Lanthanide Series*			58 Ce 140.1	59 Pr 140.9	60 Nd 144.2	61 Pm (147)	62 Sm 150.4	63 Eu 152.0	64 Gd 157.3	65 Tb 158.9	66 Dy 162.5	67 Ho 164.9	68 Er 167.3	69 Tm 168.9	70 Yb 173.0	71 Lu 175.0		
Actinide Series~			90 Th 232.0	91 Pa (231)	92 U (238)	93 Np (237)	94 Pu (242)	95 Am (243)	96 Cm (247)	97 Bk (247)	98 Cf (249)	99 Es (254)	100 Fm (253)	101 Md (256)	102 No (254)	103 Lr (257)		

** Groups are noted by 3 notation conventions.

Source: LANL (2000)

FIGURE A-2. Periodic table of the elements

Some of the factors that influence elemental distribution and association in igneous rocks are summarized below (from Krauskopf and Bird, 1995):

- ❑ Cations with large ionic radii and low electric charges, including rubidium, cesium, barium, lead, and thallium, tend to substitute for potassium, and therefore are concentrated in the minerals that form felsic rocks rather than those that form mafic rocks. These elements are therefore associated (i.e., commonly found together) in felsic rocks.
- ❑ Cations with smaller radii and relatively high charges, including uranium, thorium, boron, beryllium, molybdenum, tungsten,

BOX A-2. Column groups in the periodic table

Elements in the same column of the periodic table (Figure A-2) are closely related in their electronic structures; these elements form ions with the same electrical charge, resulting in very similar bonding characteristics. Metallic elements in the same column (e.g., calcium, magnesium, and the other column IIA elements) therefore tend to form analogous mineral structures when combined with nonmetallic elements in the same column (e.g., chlorine, bromine, and the other column VIIA elements). The chemical properties and geological behavior of common elements can be used to predict the properties and behavior of less common elements found in the same column of the periodic table. Similarities are particularly strong among the metallic elements in columns toward the left side of the table, and among the nonmetallic elements on the right. The transition elements in the middle of the table are less regular in their relationships, and therefore it is more difficult to accurately predict their chemical behavior. However, the relationships reflected by the periodic table can still be useful for predicting chemical properties and behavior.

niobium, tantalum, tin, and zirconium, also tend to be concentrated and associated in the minerals that form felsic rocks.

- Many elements with ions of intermediate radii, especially metals of the transition groups, substitute readily for iron and magnesium, and therefore are concentrated and associated in mafic and ultramafic rocks. Manganese, vanadium, and titanium tend to be concentrated in mafic rocks. Chromium, nickel, and cobalt tend to be strongly enriched along with magnesium in ultramafic rocks.

- Among the chalcophile elements (see Table A-4), a few substitute to some extent for major cations in silicate structures (lead and thallium substitute for potassium, zinc for iron and magnesium, and magnesium and bismuth for calcium). However, the chalcophile elements tend to accumulate in the residual solutions left after crystallization of the silicates, and are commonly concentrated and associated in sulfide ore deposits.

- Some minor elements have ionic radii and chemical properties that are so similar to major elements that they tend to occur only in the same minerals as their more common relatives. For example, gallium is always present in aluminum minerals, and very seldom becomes sufficiently segregated to appear in a mineral of its own. Other associated pairs are rubidium-potassium, hafnium-zirconium, and cadmium-zinc.

Sedimentary Rocks

Although most of the earth's crust consists of igneous and metamorphic rocks, sedimentary rocks account for approximately 75 percent of the rocks exposed on the earth's surface, and therefore are the most common parent rocks for soil formation. Sedimentary rocks are formed by lithification (consolidation into rock) of sediments. These sediments can include fragments of igneous rocks and resistant primary minerals, secondary minerals such as clays, or chemical precipitates such as calcium carbonate.

The physical and chemical redistribution that occurs during sedimentary processes can concentrate elements within rocks composed of only one mineral, or simple assemblages of a few minerals. For example, silicon is concentrated in pure quartz (silicon dioxide) sandstone; aluminum is concentrated in bauxite; iron is concentrated in sedimentary oxides, carbonates, and silicates; calcium is concentrated in limestone and gypsum; and high concentrations of sodium and potassium occur in evaporites (sediments deposited when salts precipitate from surface waters).

Physical processes can lead to separation of minerals based on density (i.e., sorting by gravity) and resistance to dissolution and disintegration. Placer ore deposits, including deposits of gold, platinum, monazite, and zircon, are formed by physical enrichment.

Chemical processes including precipitation and sorption are very important causes of mineral segregation in sedimentary rocks. Preferential precipitation due to differences in solubility and redox conditions may segregate minerals and chemical elements in layered sedimentary rocks. For example, manganese is precipitated in sedimentary environments under oxidizing conditions. Uranium and vanadium may be precipitated under reducing conditions, because these elements tend to be less soluble in their lower oxidation states.

Metal concentrations in sedimentary rocks depend on the mineralogy of the sediments, the metal concentrations in the water in which the sediments were deposited, and the sorptive properties of the mineral grains. Sandstones, which consist primarily of quartz grains, usually have very low metal concentrations because trace metals do not readily substitute for silicon atoms in the crystal matrix, and the sand grains have very low sorptive capacities. Fine-grained sedimentary rocks tend to be enriched in metals due to adsorption of cations to clay and fine-grained organic matter. For example, shale, the sedimentary rock that forms when clay is lithified, tends to have relatively high concentrations of heavy metals. Black shales, which contain both clay and organic particles, tend to have even higher concentrations of heavy metals due to the additional sorptive capacity provided by the organic matter. Sedimentary rocks containing organic matter tend to be enriched in heavy metals including vanadium, molybdenum, nickel, cobalt, arsenic, and copper (Krauskopf and Bird, 1995). Table A-5 lists average concentrations of metals and other minor elements in shales, sandstones, and carbonate rocks.

Metamorphic Rocks

Chemical elements also can be redistributed and segregated during metamorphism. However, unless reactive fluids move through fractures and pore networks during a process called *metasomatic mass transfer,* elemental redistribution is generally limited. During metasomatic mass transfer, minerals are altered as elements are transferred from the reactive fluids. This process may form secondary minerals enriched in major elements such as hydrogen, potassium, sodium, calcium, silicon, and iron, and minor elements such as boron, lithium, chlorine, fluorine, sulfur, and tin (Krauskopf and Bird, 1995).

A.2 Weathering

Soils are formed by the gradual breakdown of the rocks that form the earth's crust through a process known as weathering. Exposure to the atmosphere and the hydrosphere (i.e., surface water and groundwater) allows physical and chemical weathering processes to gradually decompose solid rock. Biological activity also contributes to soil formation, particularly after breakdown of rocks has been initiated by physical and chemical weathering processes.

A.2.1 Physical Weathering

Physical weathering refers to the mechanical disintegration of rock by natural forces. Physical weathering alone does not alter the chemical composition of minerals. The most powerful and fastest-acting physical weathering agent is frost action, the freezing and thawing of water. When water freezes and thaws within cracks and pores, it exerts extreme pressures that eventually break up rock. Temperature change causes differential expansion and contraction of rock and exerts pressures that eventually lead to disintegration. Unloading occurs when igneous or metamorphic rocks formed under pressure deep within the earth are uplifted and exposed by erosion. Release of the confining pressure causes outward expansion and results in breaking away of sheets of rock in a process known as exfoliation. Mechanical grinding occurs as glaciers move downslope, and when rocks are transported in rivers and steams. High relief and rainstorms also facilitate physical weathering.

A.2.2 Chemical Weathering

Chemical weathering is the most important process in soil formation. During chemical weathering, the chemical elements that make up the primary rock-forming minerals are mobilized, resulting in physical disintegration of the rock, formation of new (secondary) minerals, and changes in the overall

TABLE A-5. Average concentrations of minor elements in shales, sandstones, and carbonate rocks (mg/kg)

Element	Shales	Sandstones	Carbonates
Barium[(a)]	600	10–100	10
Lithium[(a)]	60	15	5
Rubidium[(a)]	140	60	3
Sironium[(a)]	400	20	610
Bromine	5	1	6.2
Cerium	70	92	11.5
Chlorine	180	10	150
Cobalt	20	0.3	0.1
Chromium	100	35	11
Copper	50	1–10	4
Fluorine	600	270	330
Gallium	25	12	4
Germanium	1.5	0.8	0.2
Iodine	2	1.7	1.2
Manganese	850	10–100	1,100
Nickel	80	2	20
Lead	20	7	9
Scandium	15	1	1
Thorium	12	1.7	1.7
Titanium	4,600	1,500	400
Uranium	3.5	0.5	2.2
Vanadium	130	20	20
Yttrium	35	40	30
Zinc	90	16	20
Zirconium	180	220	19
Arsenic[(b)]	10	1	1
Boron[(b)]	100	35	20
Molybdenum[(b)]	2	0.2	0.4
Phosphorus[(b)]	750	170	400
Selenium[(b)]	0.6	0.1	0.1

Source: Krauskopf and Bird (1995, Table 20-5).
(a) Ionic potential < 2.5.
(b) Ionic potential > 9.5.

chemical composition. The overall chemical composition of weathered rock can be radically different from that of the parent rock.

Water plays a key role in chemical weathering and soil formation. Chemical weathering reactions include dissolution, hydration, oxidation, and acid titration, all of which occur in the presence of water. Acid titration, the most important weathering process, dissolves and mobilizes metallic cations within minerals and replaces them with hydrogen ions. The hydrogen ions are supplied by natural carbonic and organic acids associated with water flowing over and through the rock. Removal of metallic cations alters the primary rock-forming minerals, forming secondary minerals that are depleted in metals. Because aluminum is relatively stable compared to the other metals within the primary minerals, cation replacement tends to convert the primary minerals to aluminosilicates (e.g., clay minerals). Metals removed from the primary minerals are transported in solution until they precipitate as metal oxides and hydroxides, or are adsorbed onto the surface of clay and organic particles.

Chemical weathering reaction rates depend on the amount of water in the environment, and are directly proportional to temperature. For example, chemical weathering occurs relatively quickly in humid tropical environments and slowly in cold dry environments. Variations in the resistance of the minerals that make up the rock and their grain sizes also affect weathering rates. Table A-6 shows estimated rates of chemical weathering for different climates and rock types. Biochemical reactions also contribute to soil formation as the inorganic chemicals in weathered rock are metabolized by plants and animals, and organic compounds are added to the broken and altered rock and mineral particles.

TABLE A-6. Effect of climate on chemical weathering

Rock Type	Climate	Estimated years required to weather 1 mm of fresh rock to a kaolinitic saprolite
Felsic	Tropical semi-arid	65–200
	Tropical humid	20–70
	Temperate humid	41–250
	Cold humid	35
Metamorphic	Temperate humid	33
Mafic	Temperate humid	68
	Tropical humid	40
Ultramafic	Tropical humid	21–35

Source: Krauskopf and Bird (1995, Table 13-1).

Mineral alteration and the dissolution of metals can be illustrated by examining the effects of weathering on two common rock-forming silicate minerals: orthoclase (potassium feldspar) and olivine (iron and magnesium silicates). When orthoclase is subjected to chemical weathering, potassium is removed in solution, and the secondary minerals are clay (aluminosilicates) and silica (silicon dioxide). Secondary clay minerals then may be enriched in metals due to their high sorption capacities. Figure A-3 shows the chemical reactions involved in the weathering of orthoclase.

During chemical weathering of olivine, magnesium and some of the iron are removed in solution, yielding as secondary minerals hematite (iron oxide), limonite (a mixture of hydrated iron oxides and hydroxides), and silica. Table A-7 shows the chemical weathering products of orthoclase, olivine, and the other common silicate minerals.

$$\overset{\text{Water}}{H_2O} \overset{\textit{plus}}{+} \overset{\text{carbon dioxide}}{CO_2} \overset{\textit{yields}}{\rightarrow} \overset{\text{carbonic acid}}{H_2CO_3} \overset{\textit{yields}}{\rightarrow} \overset{\text{hydrogen ion}}{H^+} \overset{\textit{plus}}{+} \overset{\text{bicarbonate ion}}{(HCO_3)^-}$$

$$\overset{\text{Potassium feldspar}}{2K(AlSi_3O_8)} \overset{\textit{plus}}{+} \overset{\text{hydrogen ions}}{2H^+} \overset{\textit{plus}}{+} \overset{\text{water}}{H_2O} \overset{\textit{yields}}{\rightarrow} \overset{\text{clay}}{Al_2Si_2O_5(OH)_4} \overset{\textit{plus}}{+} \overset{\text{potassium ions}}{2K^+} \overset{\textit{plus}}{+} \overset{\text{silica}}{4SiO_2}$$

FIGURE A-3. Chemical reactions involved in the weathering of orthoclase

TABLE A-7. Chemical weathering products of common rock-forming silicate minerals

	Important Decomposition Products		
Mineral	**Composition**	**Minerals**	**Others**
Quartz	SiO_2	Quartz grains	Some silica in solution
Feldspars			
Orthoclase (or K-spar)	$K(AlSi_3O_8)$	Clay Silica	Potassium carbonate (soluble) Some silica in solution
Albite (sodium plagioclase)	$Na(AlSi_3O_8)$	Clay	Some silica in solution
Anorthite (calcium plagioclase)	$Ca(Al_2Si2O_8)$	Silica Calcite	Sodium and calcium carbonates (soluble)
Ferromagnesians			
Biotite Augite Hornblende	Fe, Mg, Ca silicates of Al	Clay Hematite Limonite Silica Calcite	Calcium and magnesium carbonates (soluble) Some silica in solution
Olivine	$(Fe, Mg)_2SiO_4$	Hematite Limonite Silica	Iron and magnesium carbonates (soluble) Some silica in solution

Source: Judson et al. (1987, Table 5.1).

Although metals are mobilized and redistributed during the chemical weathering process, the concentration ratios of certain metals in soil often reflect the ratios that existed in the parent rocks: soils tend to "inherit" the geochemical profile of the parent rocks. However, in some cases, the transport and redistribution of metals that occurs during weathering may significantly alter the geochemical profile. Because the clay minerals formed by chemical weathering are aluminosilicates, and clay minerals have a strong affinity for metallic cations, metal concentrations in soil containing large quantities of clay tend to be a function of aluminum concentrations.

A.3 Soil Composition and Geochemical Characteristics

A.3.1 Soil Components

Soil is composed of four basic components:

- Inorganic (i.e., mineral) material (derived from the parent rock)
- Organic material
- Air
- Water.

The abundance of each component and its importance in the soil system varies from one soil to another.

The mineral fraction, which includes remnants of the original parent rock and the secondary minerals formed through weathering, typically makes up the greatest portion of the overall soil mass (more than 90 percent). The physical and chemical characteristics of the soil therefore depend greatly on the characteristics of the original rock. Mineral grain sizes in most soils are predominantly in the sand, silt, and clay ranges. Hard minerals derived from the parent rock end up as large, durable sand grains; softer minerals are more readily weathered down to smaller silt- and clay-size particles, and to individual chemical elements (which may be dissolved and mobilized as ions). The relative proportions of sand, silt, and clay determine the texture of the soil.

The organic fraction of soil is composed of the plants and animals (including microorganisms) that cycle chemical nutrients within the soil, the waste products released from plants and animals, and their dead remains. Organic matter generally accounts for 2 to 6 percent of the overall soil volume in the upper horizons. Humification occurs when microorganisms break down (decay) organic matter and return it to inorganic (mineral) form. Plants then can recycle the inorganic chemicals and form new organic compounds. Plants extract mineral nutrients, primarily metals, from the soil by a process called chelation, in which a hydrogen ion is exchanged for a metallic ion. High moisture content and moderate ambient temperatures tend to result in relatively high organic carbon concentrations, whereas high ambient temperatures and low moisture content tend to result in low organic carbon concentrations.

The water fraction, along with the soil air, fills the pore spaces between the minerals and organic compounds in the soil. As water flows through the pore spaces, chemical elements are dissolved and transported along with fine soil particles, contributing to continued chemical weathering, leaching, and mineral alteration.

The air fraction fills pore space not occupied by water. Oxygen and carbon dioxide in the soil air are transferred to the dissolved phase. Oxygen and carbon dioxide are used by soil organisms, and affect redox conditions and pH within the soil. Dissolution of carbon dioxide forms carbonic acid, and thus decreases the pH. Redox conditions and pH strongly influence metal transport and sorption within soils.

A.3.2 Soil Profiles

As soils are formed by the physical, chemical, and biochemical mechanisms described above, horizontal subdivisions or “horizons” are developed (Figure A-4). Soil horizons are contrasting layers formed by the vertical redistribution of minerals, chemical elements, and organic compounds. The surface (A) horizon contains loose mineral particles mixed with organic material. The A horizon also is known as the zone of leaching because chemicals including salts, iron oxides, and clay minerals are transported downward by percolating water. Chemicals leached from the A horizon are deposited within the B horizon, the zone of accumulation. The B horizon typically is enriched with calcium carbonate, iron oxides, clays, and other aluminum minerals. Below the B horizon, the effects of weathering become less pronounced; the soil grades into partially decomposed parent rock, the C horizon, which in turn is underlain by parent rock.

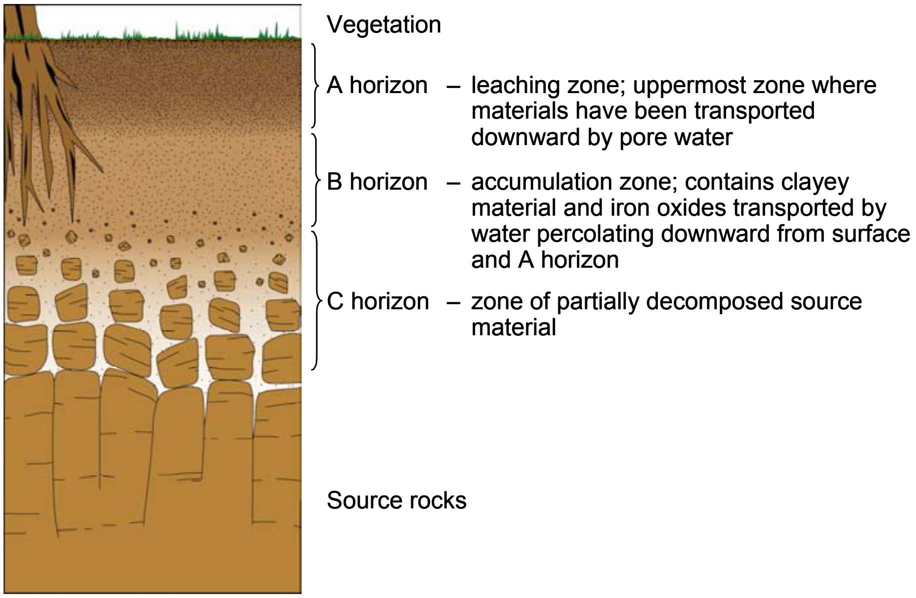


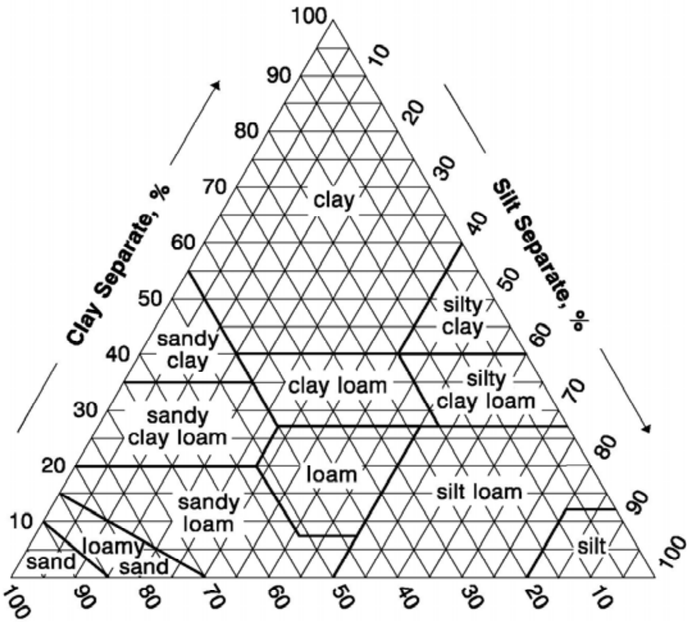
FIGURE A-4. Soil profile

Because most chemical releases occur above ground surface, the A horizon is often the subject of background analysis. Silver, arsenic, cadmium, copper, mercury, lead, antimony, and zinc tend to be concentrated in the A horizon due to cycling through vegetation, atmospheric deposition, and adsorption by organic matter. Information regarding the plant communities that occur onsite and differences in vegetation density can be useful to identify geochemical relationships for background analysis, and provide evidence to support conclusions of the analysis. Technical information regarding the effects of vegetation on the distribution of heavy metals in soil is available in Kabata-Pendias (2001) and Adriano et al. (1997)

The B horizon typically is enriched in clays and hydrous oxides containing elements such as aluminum, iron, gallium, magnesium, nickel, scandium, titanium, vanadium, and zirconium, which have been transported downward from the overlying A horizon.

A.3.3 Soil Texture

Soil texture is determined by the relative amounts of three groups of soil particles: sand, silt, and clay. Sand particles can be seen by the naked eye. A microscope is required to see individual silt particles, whereas individual clay particles are so small that they are visible only with an electron microscope. Soil texture has a great effect on the distribution of naturally occurring metals in soil. Very fine-grained soils, particularly soils with a high percentage of clay-size particles, have a high capacity to absorb metals from solutions. The U.S. EPA (1995) defines soils and sediments as mineral and naturally occurring organic materials with particle sizes less than 2 mm (sand, silt and clay), and notes that, because of their high surface area/mass ratios, these fine-grained particles have a greater affinity for inorganic chemicals (from both natural and anthropogenic sources) than coarser-grained materials. Figure A-5 shows a grain-size classification system.



Source: USDA NRCS (1998)

FIGURE A-6. USDA guide for textural classification in soil families

Coarse soils contain relatively large amounts of sand; loam soils contain similar amounts of sand-, silt-, and clay-size particles; and silt soils have the texture of flour. The USDA soil classification diagram (Figure A-6) can be used to classify soils based on the relative percentage of each grain size. For example, as shown on Figure A-6, the USDA system classifies a soil with 20 percent silt, 30 percent clay, and 50 percent sand as a sandy clay loam.

The Unified Soil Classification System (USCS) is commonly used for engineering and geotechnical applications (ASTM, 2001). The USCS gives each soil group a two-letter designation based on the percentages of gravel, sand, silt, and clay (Figure A-7).

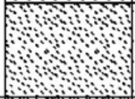
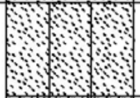
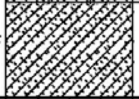
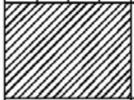
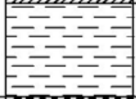
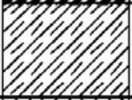
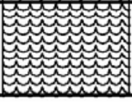
Extremely fine-grained (colloidal) soil particles, including both clay minerals and organic matter, have greater sorption capacity (ability to attract ions) than large particles such as sand. Therefore, metal concentrations in soil tend to be inversely proportional to grain size. Because adsorption is a surface phenomenon, the rate and extent of adsorption increase as the surface area of the sorptive medium increases. As shown in Figure A-8, surface area increases as mineral grains are divided into smaller particles. Fine-grained media (e.g., clays) therefore have high surface area/mass ratios and relatively greater sorption capacities than media

Class term		mm
Granules		4
		2
Sand	very coarse	
	coarse	1
	medium	0.50
	fine	0.25
	very fine	0.125
		0.062
Silt	coarse	0.031
	medium	0.016
	fine	0.008
	very fine	
		0.004
Clay		

Source: Tucker (1991, Table 2.1)

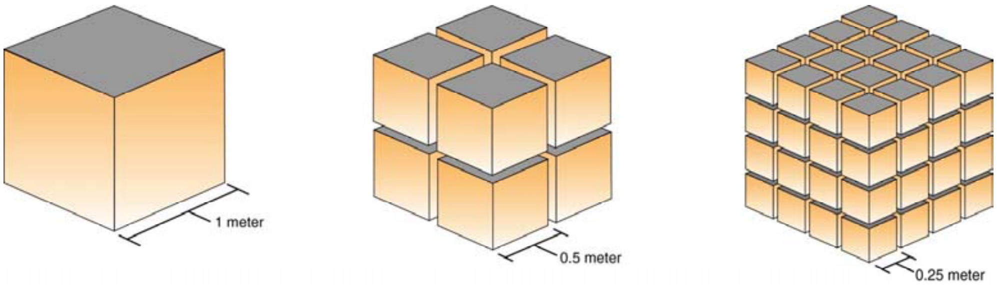
FIGURE A-5. Soil and sediment particle sizes

SOIL CLASSIFICATION CHART

MAJOR DIVISIONS			SYMBOLS		TYPICAL DESCRIPTIONS
			GRAPH	LETTER	
COARSE GRAINED SOILS MORE THAN 50% OF MATERIAL IS LARGER THAN NO. 200 SIEVE SIZE	GRAVEL AND GRAVELLY SOILS MORE THAN 50% OF COARSE FRACTION RETAINED ON NO. 4 SIEVE	CLEAN GRAVELS (LITTLE OR NO FINES)		GW	WELL-GRADED GRAVELS, GRAVEL - SAND MIXTURES, LITTLE OR NO FINES
				GP	POORLY-GRADED GRAVELS, GRAVEL - SAND MIXTURES, LITTLE OR NO FINES
		GRAVELS WITH FINES (APPRECIABLE AMOUNT OF FINES)		GM	SILTY GRAVELS, GRAVEL - SAND - SILT MIXTURES
				GC	CLAYEY GRAVELS, GRAVEL - SAND - CLAY MIXTURES
	SAND AND SANDY SOILS MORE THAN 50% OF COARSE FRACTION PASSING ON NO. 4 SIEVE	CLEAN SANDS (LITTLE OR NO FINES)		SW	WELL-GRADED SANDS, GRAVELLY SANDS, LITTLE OR NO FINES
				SP	POORLY-GRADED SANDS, GRAVELLY SAND, LITTLE OR NO FINES
		SANDS WITH FINES (APPRECIABLE AMOUNT OF FINES)		SM	SILTY SANDS, SAND - SILT MIXTURES
				SC	CLAYEY SANDS, SAND - CLAY MIXTURES
FINE GRAINED SOILS MORE THAN 50% OF MATERIAL IS SMALLER THAN NO. 200 SIEVE SIZE	SILTS AND CLAYS	LIQUID LIMIT LESS THAN 50		ML	INORGANIC SILTS AND VERY FINE SANDS, ROCK FLOUR, SILTY OR CLAYEY FINE SANDS OR CLAYEY SILTS WITH SLIGHT PLASTICITY
				CL	INORGANIC CLAYS OF LOW TO MEDIUM PLASTICITY, GRAVELLY CLAYS, SANDY CLAYS, SILTY CLAYS, LEAN CLAYS
				OL	ORGANIC SILTS AND ORGANIC SILTY CLAYS OF LOW PLASTICITY
	SILTS AND CLAYS	LIQUID LIMIT GREATER THAN 50		MH	INORGANIC SILTS, MICACEOUS OR DIATOMACEOUS FINE SAND OR SILTY SOILS
				CH	INORGANIC CLAYS OF HIGH PLASTICITY
				OH	ORGANIC CLAYS OF MEDIUM TO HIGH PLASTICITY, ORGANIC SILTS
HIGHLY ORGANIC SOILS				PT	PEAT, HUMUS, SWAMP SOILS WITH HIGH ORGANIC CONTENTS

NOTE: DUAL SYMBOLS ARE USED TO INDICATE BORDERLINE SOIL CLASSIFICATIONS

FIGURE A-7. Unified Soil Classification System



One 1-m cube
Total surface area = $6\ \text{m}^2$

Eight 0.5-m cubes
Total surface area = $12\ \text{m}^2$

Sixty-four 0.25-m cubes
Total surface area = $24\ \text{m}^2$

As a cube is divided into smaller units, the volume remains unchanged, but the overall surface area increases.

FIGURE A-8. Relationship of volume, particle size, and surface area

with low surface area/mass ratios (e.g., sands). In addition to large surface area/mass ratios, clays and organic colloids tend to be highly charged relative to their surface areas. Under normal conditions (pH = 6 to 8), the distribution of electric charge within a clay particle is such that the surface layer is negatively charged. The positively charged metallic ions released from minerals during weathering are strongly attracted to the negatively charged surfaces. Soils with high concentrations of clay minerals and/or organic colloids are therefore likely to have relatively high metal concentrations. Silts tend to have moderate sorption capacity and metal concentrations, whereas sands and gravels have low sorption capacity and relatively low metal concentrations.

A.3.4 Effects of Soil and Soil Water Chemistry on Metal Distribution

Chemical properties of soil and soil water, including TOC concentration, pH, redox potential, and ionic strength strongly influence the distribution of metals in soil. Soils with a high percentage of organic material (high TOC) will tend to have higher metal concentrations due to the high sorptive capacity of the organic particles. Low pH (acid) conditions tend to dissolve and mobilize metals, whereas high pH (basic) conditions favor metal precipitation, resulting in increased metals concentrations in the soil.

Soil pH also affects the electric charge on the surface of mineral particles. The point of zero charge (PZC) is the pH value at which the mineral surface effectively has no net charge. At pH values less than the PZC, a mineral surface has a net positive charge; at pH values greater than the PZC, net surface charge is negative. The PZC phenomenon is especially important for clay minerals. PZC pH values for the clay minerals kaolinite and montmorillonite, the iron oxide goethite, and quartz are listed in Table A-8.

TABLE A-8. Point of zero charge pH values

Mineral	pH Values Corresponding to Point of Zero Charge
Montmorillonite	2–3
Kaolinite	4.6
Goethite	7.3–7.8
Quartz	2–3

As a result of their large surface areas and low PZC pH values, clays have a very high sorptive capacity. Within the normal soil pH range (6 to 8), clay particles have a strongly negative net surface charge, and therefore act as “sinks” for the positively charged metal ions.

Redox conditions within the soil also affect the solubility and transport of metals. Under reducing conditions, most metallic ions tend to remain in solution, and therefore are relatively mobile. Under oxidizing conditions, metallic ions are oxidized (i.e., they lose electrons) and tend to combine with nonmetallic ions (e.g., oxygen); these combined ions form compounds that precipitate out of solution. For example, under oxidizing conditions, the manganese ion Mn^{2+} tends to lose electrons and form Mn^{4+}, which combines readily with oxygen and precipitates as MnO_2.

Ionic strength also affects the tendency of metals to precipitate out of solution. A solution with high ionic strength has a high concentration of dissolved salts. Salts dissociate into ions in solution. The major cations formed when common salts dissociate are potassium, sodium, and calcium. The major anions are chloride, sulfate, carbonate, and bicarbonate. Comparison of the solubilities of minerals in pure water versus water with a high salt content indicates that increased salt content leads to increased mineral solubility. Therefore, if the ionic strength of pore water in soil is high, metallic ions will tend to remain in solution. Conversely, if ionic strength is low, metals will tend to precipitate out of solution, resulting in higher concentrations in the soil.

A.3.5 Metal Concentrations in Soil

Table A-9 lists "generic" background metal concentration ranges for various types of surface soils found in the United States. If metal concentrations exceed the ranges shown in these tables, a chemical release may have occurred.

As shown in Table A-9, background copper concentrations can range between approximately 3 and 300 mg/kg, and the concentration of lead in various soils ranges from 10 to 70 mg/kg. The average concentration range for lead in urban areas is significantly higher than in rural areas, primarily due to combustion of leaded fuels. This is one example of the effect of anthropogenic processes on background metal concentrations.

The type of parent rock is only one of the factors that control metal concentrations in soils. Weathering, biologically mediated chemical reactions, and other natural geochemical processes can significantly enrich or deplete the concentrations of certain metals. Metal concentrations tend to be highest in soils that contain high percentages of clay and organic material (e.g., clay and clay-loamy soils, organic light [or rich] soils). This is a result of the high capacity of clay and organic matter to adsorb metallic ions. The strong affinity of certain metals for organic carbon and clay particles suggests that, in some cases, it may be advantageous to normalize metal concentrations to the organic carbon (i.e., TOC) or clay content of each soil sample before implementing the background analysis methods described in this document. Characteristics of site soils (e.g., clay and TOC content), and the affinity of the specific target metals for organic carbon and clay, should be considered to determine whether normalization is appropriate. Details of the effects of TOC and clay content on the distribution of various metals in soil are discussed in Alloway (1990) and Kabata-Pendias (2001).

A.3.6 Elemental Correlation

In rocks of the same type, certain groups of elements are often highly correlated (i.e., their concentration ratios tend to be relatively constant). In addition, although their concentrations may vary considerably among different rock types, certain elements still tend to be correlated. For example, as shown on Figure A-9, nickel/chromium concentration ratios tend to be relatively constant among rocks classified as granite, shale, clay, and basalt. Table A-10 lists average chromium and nickel concentrations and concentration ratios for a number of different rock types. Elemental correlation such as that observed for nickel and chromium is common for many pairs and groups of metals, and therefore provides a very powerful tool for background analysis. If correlation exists, background ranges of the metals can be defined based on the relatively constant ratios of one metal to another. If a metal is found at an elevated concentration that does not fit the background ratio, a chemical release should be suspected.

A landmark paper published in 1988 by Florida Department of Environmental Regulation (now Florida Department of Environmental Protection) describes a method for evaluating background metal concentrations in estuarine sediments, and provides guidelines for distinguishing between natural sediments and sediments that have been impacted by a chemical release (FDER, 1988). The FDEP method is based on the observation that

TABLE A-9. Generic background concentrations of inorganics in U.S. surface soils (mg/kg, except where noted)

Soil Type	Aluminum		Antimony		Arsenic		Barium		Beryllium		Cadmium		Chromium	
	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean
Alluvial soils	—	—	—	—	2.1–22.0	8.2	200–1,500	660	1–3	1.6	—	—	15–100	55
Chernozems and dark prairie soils	—	—	—	—	1.9–23	8.8	100–1,000	595	<1–3	1.5	—	—	15–150	55
Clay and clay-loamy soils	—	—	—	—	1.7–27.0	7.7	150–1,500	535	<1–15	1.9	—	—	20–100	55
Forest soils	—	—	—	—	1.5–16.0	6.5	150–2,000	505	1–3	1.9	0.5–1.5	.73	15–150	55
Light desert soils	—	—	—	—	1.2–18.0	6.4	300–2,000	835	<1–7	2.1	—	—	10–200	60
Light loamy soils	—	—	—	—	0.4–31.0	7.3	70–1,000	555	1–3	1.7	—	—	10–100	55
Loess & soils on silt deposits	—	—	—	—	1.9–16.0	6.6	200–1,500	675	1–3	1.7	—	—	10–100	55
Organic light (or rich*) soils	—	—	—	—	<0.1–48.0	5.0	10–700	265	<1–1.5	1.2	—	—	1–100	20
Silty prairie soils	—	—	—	—	2.0–12.0	5.6	200–1,500	765	1–1.5	1.4	—	—	20–100	50
Soils on glacial till and drift	—	—	—	—	2.1–12.0	6.7	300–1,500	765	<1–2	1.6	—	—	30–150	80
Soils over granites and gneisses	7.2–8.2%	—	—	—	0.7–15.0	3.6	300–1,500	785	1–2	1.6	—	—	10–100	45
Soils over limestones and calcareous rocks	0.43–1.3%	—	—	—	1.5–21.0	7.8	150–1,500	520	1–2	1.6	—	—	5–150	50
Various soils	0.45–10%	—	—	—	1.0–93.2	7.0	70–3,000	560	<1–5	1.6	0.41–0.57		7–1500	50

Soil Type	Cobalt		Copper		Iron		Lead		Manganese		Mercury		Nickel	
	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean
Alluvial soils	3–20	9.0	5–50	27	—	—	10–30	18	150–1,500	405	0.02–0.15	0.05	7–50	19.0
Chernozems and dark prairie soils	3–15	7.5	10–70	27	—	—	10–70	19	100–2,000	600	0.02–0.53	0.10	7–70	19.5
Clay and clay-loamy soils	3–30	8.0	7–70	29	—	—	10–70	22	50–2,000	580	0.01–0.09	0.13	5–50	20.5
Forest soils	5–20	10.0	7–150	17	—	—	10–50	20	150–1,500	645	0.02–0.14	0.06	7–100	22.0
Light desert soils	3–20	10.0	5–100	24	—	—	10–70	23	150–1,000	360	0.02–0.32	0.06	7–150	22.0
Light loamy soils	3–30	7.5	3–70	25	—	—	<10–50	20	50–1,000	480	0.01–0.06	0.07	5–200	22.0
Loess & soils on silt deposits	3–30	11.0	7–100	25	—	—	10–30	19	50–1,500	525	0.01–0.38	0.08	5–30	17.0
Organic light (or rich*) soils	3–10	6.0	1–100	15	—	—	10–50	24	7–1,500	260	0.01–4.60	0.28	5–50	12.0
Silty prairie soils	3–15	7.5	10–50	20	—	—	10–30	21	200–1,000	430	0.02–0.06	0.04	<5–50	16.0
Soils on glacial till and drift	5–15	7.5	15–50	21	—	—	10–30	17	200–700	475	0.02–0.36	0.07	10–30	18.0
Soils over granites and gneisses	3–15	6.0	7–70	24	1.4–2.7%	—	10–50	21	150–1,000	540	0.01–0.14	0.06	<5–50	18.5
Soils over limestones and calcareous rocks	3–20	9.5	7–70	21	0.4–1%	—	10–50	22	70–2,000	470	0.01–0.5	0.08	<5–70	18.0
Various soils	3–50	10.5	3–300	26		—	<10–70	26	20–3,000	490	0.02–1.50	0.17	<5–150	18.5

TABLE A-9. Generic background concentrations of inorganics in U.S. surface soils (mg/kg, except where noted) (continued)

Soil Type	Selenium		Silver		Thallium		Vanadium		Zinc	
	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean
Alluvial soils	<0.1–2.0	0.5	—	—	—	—	30–150	79	20–108	58.5
Chernozems and dark prairie soils	<0.1–1.2	0.4	0.44–0.93	—	—	—	30–150	92	20–246	83.5
Clay and clay-loamy soils	<0.1–1.9	0.5	—	—	—	—	20–150	87	20–220	67
Forest soils	<0.1–1.6	0.4	—	—	—	—	15–200	85	25–155	45.7
Light desert soils	<0.1–1.1	0.5	—	—	—	—	30–150	93	25–150	52.5
Light loamy soils	0.02–1.2	0.33	—	—	—	—	20–150	77	20–118	55
Loess & soils on silt deposits	0.02–0.7	0.26	—	—	—	—	20–300	102	20–109	58.5
Organic soils	<0.1–1.5	0.3	2–5	—	—	—	<7–150	38	<5–108	34
Silty prairie soils	<0.1–1.0	0.3	—	—	—	—	50–150	87	30–88	54.3
Soils on glacial till and drift	0.2–0.8	0.4	—	—	—	—	30–200	95	47–131	64
Soils over granites and gneisses	<0.1–1.2	0.4	—	—	0.6–2.3	—	50–200	100	30–125	73.5
Soils over limestones and calcareous rocks	0.1–1.4	0.19	—	—	0.01–0.14	—	10–150	72	10–106	50
Various soils	<0.1–4.0	0.31	0.03–0.09	—	0.02–2.8	—	0.7–98	—	13–300	73.5

Source: ORNL (1999), based on Kabata-Pendias and Pendias (1984).
— = no data

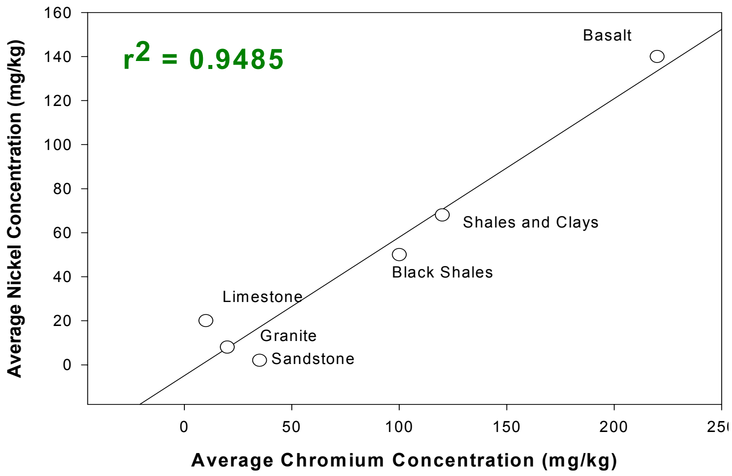


FIGURE A-9. Correlation plot of chromium vs. nickel concentrations

the concentrations of certain metals in natural (uncontaminated) sediments tend to be directly proportional to aluminum concentrations. Metals data from a wide range of estuarine sites around the State of Florida were evaluated, and the natural metal/aluminum relationships were quantified by linear regression. The scatter plots presented in Figure A-10 (FDER, 1988) illustrate the relatively strong association with aluminum observed for some metals in natural Florida sediments. The approximately linear relationships exhibited by some of the scatter plots indicate relatively constant concentration ratios for certain metals with respect to aluminum. Concentrations of a suspected COPC metal that tend to fit a linear relationship with respect to aluminum are most likely within the background range, whereas elevated concentrations that deviate from the relationship may represent contamination. Aluminum was chosen as a reference element for several reasons: it is the most abundant naturally occurring metal, it is highly refractory (i.e., resistant to weathering), and its concentrations are generally not influenced by chemical releases.

The FDEP background analysis method can be extended to soils, and reference metals other than aluminum can be used. In some environments, iron can be used as a reference metal. For example, in the State of Hawaii, iron is ubiquitous in soils formed by the weathering of volcanic rocks (basalts). The primary iron-containing minerals are altered to clays and iron oxides. Because these secondary minerals have high sorption capacities, heavy metal concentrations in Hawaiian soil often show a high degree of correlation with iron concentrations. Elemental correlation is central to geochemical association analysis—Step 1 of the Geochemical Method (Section 3.2).

A.3.7 Enrichment

Weathering and other natural processes can cause enrichment of metals in soils relative to their parent rocks; therefore, possible enrichment processes should be carefully considered before concluding that elevated metal concentrations are the result of a chemical release. The enrichment processes that form ore deposits can result in very high natural metals concentrations in rocks and soils (e.g., quartz vein and placer metal deposits). However, ore deposits are relatively rare; therefore, the enrichment that occurs when metals are redistributed by weathering and soil formation processes are of more importance for background analysis. The most significant enrichment processes are summarized as follows:

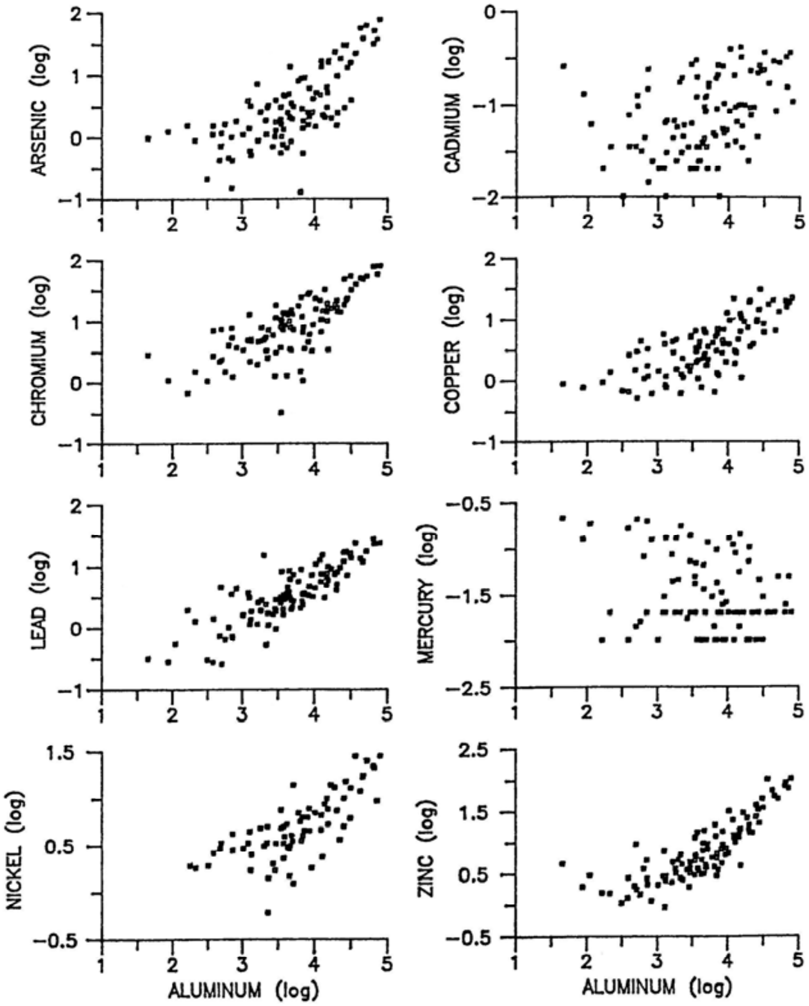
- ❑ ***Sedimentary processes***. Fine-grained sedimentary rocks (e.g., shale and black shale) tend to be enriched in metals due to adsorption of metallic cations to clay minerals and fine-grained organic matter.

TABLE A-10. Chromium/nickel concentrations in different rocks

	Chromium		Nickel		Average Cr/Ni Concentration Ratio
Name	**Average**	**Range**	**Average**	**Range**	
Basaltic igneous	220	40–600	140	45–410	1.6
Granitic igneous	20	2–90	8	2–20	2.5
Shales and clays	120	30–590	68	20–250	1.8
Black shales	100	26–1,000	50	10–500	2.0
Limestone	10	—	20	—	0.5
Sandstone	35	—	2	—	17.5

Source: Alloway (1990, Table 7-2).

- ***Mineral alteration during weathering***. The secondary minerals that replace primary minerals during weathering tend to be aluminosilicates (predominantly clays) and therefore contain high concentrations of aluminum.
- ***Sorption***. Fine-grained soils containing high concentrations of clay minerals or organic matter tend to contain high concentrations of metals due to sorption of metallic ions. These translocated metals may accumulate in the A horizon (sorption to organic matter) or B horizon (sorption to clay minerals).
- ***Biological reactions***. Metal uptake by vegetation and adsorption by organic matter produced by plants and animals may result in high concentrations of certain metals in the A horizon.
- ***Precipitation***. The metallic cations removed from primary minerals are transported in solution and precipitate as metal oxides and hydroxides under favorable chemical conditions (e.g., redox and pH). These translocated metals may accumulate in the B horizon.



Source: FDER (1988, Figure 7)

FIGURE A-10. Example aluminum scatter plots

Enrichment can be quantified and used to evaluate background concentration ranges in soil. Use of enrichment factors in the Geochemical Method is described in Section 3.3.

A.4 References

Adriano, D.C., Z.-S. Chen, and S.-S. Yang, eds. 1997. *Biogeochemistry of Trace Metals.* Hanover, NH: I.K. Iskander, Cold Regions Research and Engineering Laboratory. September.

Alloway, B.J. 1990. *Heavy Metals in Soils.* Glasgow: Blackie and Son.

American Society for Testing and Materials (ASTM). 2001. *Classification D2487-00 Standard Classification of Soils for Engineering Purposes (Unified Soil Classification System).* West Conshohocken, PA.

Florida Department of Environmental Regulation (FDER). 1988. *A Guide to the Interpretation of Metal Concentrations in Estuarine Sediments.* Eds. S.J. Schropp and H.L. Windom. Tallahassee, FL. April.

Judson, S., M.E. Kauffman, and L.D. Leet. 1987. *Physical Geology.* 7th ed. New York: Prentice-Hall.

Kabata-Pendias, A., and H. Pendias. 1984. *Trace Elements in Soils and Plants.* Boca Raton, FL: CRC Press.

Krauskopf, K.B., and D.K. Bird. 1995. *Introduction to Geochemistry.* 3rd ed. New York: McGraw-Hill.

Los Alamos National Laboratory (LANL). 2000. *Periodic Table of the Elements.* Los Alamos, NM: LANL, Chemistry Division.

Oak Ridge National Laboratory (ORNL). 1999. "Risk Assessment Information System: Inorganic Soil Background Selection." *Risk Information Web Server*. Prepared by DOE's ORNL. Updated March 5. Available at: http://risk.lsd.ornl.gov/homepage/bg_search.shtml.

Tucker, M.E. 1991. *Sedimentary Petrology, An Introduction to the Origin of Sedimentary Rocks.* 2nd ed. Oxford: Blackwell Scientific Publications.

United States Department of Agriculture, Natural Resources Conservation Service (USDA NRCS). 1998. *Field Book for Describing and Sampling Soils.* Lincoln, NE: National Soil Survey Center.

United States Environmental Protection Agency (U.S. EPA). 1995. *Determination of Background Concentrations of Inorganics in Soils and Sediments at Hazardous Waste Sites.* EPA Engineering Forum Issue paper. EPA/540/S-96/500. Washington, DC: Office of Solid Waste and Emergency Response. December.

APPENDIX B: SUPPLEMENTARY STATISTICAL TESTS

B.1 Statistical Tests of Normality

B.1.1 Shapiro-Wilk W Test

The Shapiro-Wilk W test is highly recommended for testing whether data have a normal distribution. It also may be used to test for a lognormal distribution, if the data are first transformed by computing the natural logarithm of each datum. The W test is recommended in several U.S. EPA guidance documents (U.S. EPA, 1992a and 2000) and in many statistical texts (Gilbert, 1987; Conover, 1998). It is available in many software packages including GRITS/STAT (U.S. EPA, 1992b) and DataQUEST (U.S. EPA, 1997). The W test has been shown to have more power than other tests to detect when data are not from a normal or lognormal distribution. The W test should be conducted in conjunction with constructing normal and lognormal probability plots (Section 2.2.3) in order to visually confirm whether the normal or lognormal distribution is an acceptable fit to the data. The W test:

- ☐ Requires the use of a table of coefficients (Table C-10) and critical values (Table C-11).
- ☐ Can only be conducted if the number of samples is less than or equal to 50 because the table of critical values (Table C-11) does not extend beyond n = 50. For larger datasets, the D'Agostino test or approximations of the Shapiro-Wilk W test can be performed.
- ☐ Is somewhat tedious to compute by hand, but it is easily conducted using the DataQUEST software.
- ☐ Should not be used if the dataset contains nondetects.
- ☐ May not have sufficient power to detect non-normality if the underlying distribution is only slightly different than the normal distribution or if the number of data in the dataset is too small.

Table B-1 shows the power of the W test to detect a lognormal distribution in the data, rather than a normal distribution. This table was assembled using computer simulations for which 1,000 datasets of *N* measurements each were generated from lognormal distributions with various degrees of skewness (long tail towards high concentrations). Values of the power are provided in Table B-1 for various numbers of samples (N = 10 to 100) and lognormal distribution shapes, as indicated by the coefficient of variation (CV, which is the standard deviation divided by the mean). The CVs range from 0.1 to 1.3. Lognormal distributions that are only slightly asymmetric will have a small CV, whereas highly skewed (asymmetrical) lognormal distributions have large CV.

The computations needed to conduct the W test and an example are provided in Box B-1.

B.1.2 D'Agostino Test

The D'Agostino test (D'Agostino, 1971) may be used to test if the measurements are from a normal distribution. This test can be used when the number of samples exceeds 50. D'Agostino (1971) showed that the performance of the test compares favorably with other tests. The assumptions and their verifications for applying the W test also apply to this test. The D'Agostino test:

- ☐ Cannot be conducted if n < 50 or n > 1,000.
- ☐ Requires the use of a special table of critical values (Table C-12).
- ☐ Is tedious to compute by hand.
- ☐ Cannot be conducted if the dataset contains nondetects.
- ☐ May not have large power to detect non-normality if the underlying distribution is only slightly different than the normal distribution or if the number of data in the dataset is small.

The computations needed to conduct the D'Agostino test and an example are provided in Box B-2.

B.1.3 Other Tests

A large number of other statistical tests can be used to test hypotheses about which probability distribution best fits a dataset. These tests are commonly called goodness-of-fit tests. A thorough summary of the scientific literature on this topic, with many examples provided, is in D'Agostino and Stephens (1986); this book is suitable for someone who has some training in statistics. U.S. EPA (2000) provides descriptions of several tests, most of which can be conducted using the DataQUEST software (U.S. EPA, 1997).

U.S. EPA (2000) recommends the use of the W test if the number of samples is less than 50, and either the Filliben statistic or the studentized range test if the number of samples equals or is greater than 50. The Filliben test (Filliben, 1975) is not illustrated in this document because it is closely related to the W test and is difficult to compute by hand, although it is easily computed using DataQUEST software (U.S. EPA, 1997). U.S. EPA (2000) recommends the studentized range test except when the data appear to be lognormally distributed. The test, illustrated in U.S. EPA (2000), is simpler to compute than the W test and critical values needed for the test are available for sample sizes (*n*) up to 1,000.

If several goodness-of-fit tests are applied to the same dataset, the test results may differ. Under such conditions, the results of the most appropriate tests with respect to the investigated dataset, in conjunction with probability plots, should be considered as the basis for subsequent decisions.

B.2 Descriptive Summary Statistics for Datasets with Large Numbers of Nondetects

Descriptive summary statistics of datasets with large numbers of nondetects can be computed by either using the Cohen method, calculating a trimmed mean, or calculating a Winsorized mean and standard deviation. These methods are defined and their assumptions, advantages, and disadvantages are listed in Table B-2. Examples of computing the median, trimmed mean, the Winsorized mean and standard deviation are illustrated in Box B-3. The Cohen method for computing the mean and standard deviation of a normally distributed set of data that contains nondetects is explained and illustrated in Box B-4.

If more than 50% of the measurements in the dataset are nondetects, the loss of information is too great for descriptive statistics to provide much insight into the location and shape of the underlying distribution of measurements. The only descriptive statistics that might be possible to compute are pth percentiles for values of *p* that are greater than the proportion of nondetects present in the sample and when no nondetects are greater than the k(n+1)th largest datum, where *k* is defined in Box 2-4 of the main document.

U.S. EPA (2000) cautions that no general procedures exist for the statistical analyses of censored datasets that can be used in all applications of statistical analysis. For this reason, U.S. EPA guidelines should be implemented cautiously. U.S. EPA (2000) also suggests the data analyst should consult a statistician for the most appropriate way to statistically evaluate or analyze a dataset that contains nondetects.

Akritas et al. (1994, pp. 221–242) provide a review of the statistical literature that deals with the statistical analysis of censored environmental datasets. Further review is provided by Helsel and Hirsch (1992).

B.3 Statistical Tests for Outliers

This section provides detailed description of various statistical tests for determining outliers. The assumptions, advantages, and disadvantages of each test are provided in Table B-3. The procedures for conducting the Dixon extreme value test, the Discordance test, and the Walsh test, with an example for each, are provided in Boxes B-5, B-6, and B-7, respectively. The Rosner test is described in Box B-8 and illustrated in Box B-9.

B.4 References

Akritas, M.G., T.F. Ruscitti, and G.P. Patil. 1994. Statistical Analysis of Censored Environmental Data. In *Handbook of Statistics, Volume 12, Environmental Statistics*, eds. G.P. Patil and C.R. Rao. New York: North-Holland.

Conover, W.J. 1998. *Practical Nonparametric Statistics*. 3rd ed. Wiley Series in Probability and Statistics, Applied Probability and Statistics Section. New York: Wiley & Sons.

D'Agostino, R.B. 1971. An Omnibus Test of Normality for Moderate and Large Size Samples. *Biometrika* 58:341–348.

D'Agostino, R.B., and M.A. Stephens. 1986. *Goodness-of-Fit Techniques*. New York: Marcel Dekker.

Filliben, J.J. 1975. The Probability Plot Correlation Coefficient Test for Normality. *Technometrics* 17:111–117.

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring.* New York: Van Nostrand Reinhold.

Helsel, D.R., and R.M. Hirsch. 1992. *Statistical Methods in Water Resources*. New York: Elsevier.

United States Environmental Protection Agency (U.S. EPA). 1992a. *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities – Addendum to Interim Final Guidance.* Washington, DC: Office of Solid Waste, Permits and State Programs Division. June.

United States Environmental Protection Agency (U.S. EPA). 1992b. *User Documentation: GRITS/STAT, v4.2.* EPA/625/11-91/002. Cincinnati: Center for Environmental Research Information.

United States Environmental Protection Agency (U.S. EPA). 1997. *Data Quality Evaluation Statistical Toolbox (DataQUEST), User's Guide, EPA QA/G-9D, QA96 Version.* EPA/600/R-96/085. Washington, DC: Office of Research and Development. December.

United States Environmental Protection Agency (U.S. EPA). 2000. *Guidance for Data Quality Assessment: Practical Methods for Data Analysis. EPA QA/G-9, QA 00 Update.* EPA/600/R-96/084. Washington, DC: Office of Environmental Information. July.

TABLE B-1. Power of the W test to reject the null hypothesis of a normal distribution when underlying distribution is lognormal

	Power of W Test for Simulated Test Conditions												
Sample (N)	**CV= 0.1**	**CV= 0.2**	**CV= 0.3**	**CV= 0.4**	**CV= 0.5**	**CV= 0.6**	**CV= 0.7**	**CV= 0.8**	**CV= 0.9**	**CV= 1.0**	**CV= 1.1**	**CV= 1.2**	**CV= 1.3**
10	0.059	0.077	0.109	0.179	0.225	0.273	0.342	0.403	0.452	0.487	0.508	0.537	0.565
15	0.080	0.096	0.177	0.242	0.376	0.462	0.534	0.626	0.677	0.701	0.746	0.777	0.811
20	0.054	0.117	0.232	0.346	0.496	0.599	0.684	0.746	0.825	0.851	0.893	0.925	0.923
25	0.081	0.185	0.299	0.434	0.562	0.741	0.817	0.86	0.887	0.930	0.961	0.964	0.970
30	0.066	0.206	0.371	0.513	0.698	0.791	0.876	0.891	0.959	0.973	0.978	0.986	0.992
35	0.077	0.192	0.348	0.603	0.746	0.831	0.903	0.967	0.970	0.983	0.987	0.993	0.998
40	0.101	0.219	0.459	0668	0.826	0.903	0.957	0.972	0.992	0.996	0.998	0.997	0.996
60	0.135	0.349	0.608	0.832	0.946	0.972	0.995	0.997	0.999	1	0.999	1	1
70	0.112	0.363	0.706	0.883	0.961	0.989	0.999	1	1	1	1	1	1
80	0.127	0.396	0.732	0.921	0.987	0.997	0.999	0.999	1	1	1	1	1
90	0.171	0.448	0.79	0.941	0.992	0.999	1	1	1	1	1	1	1
100	0.156	0.551	0.811	0.970	0.993	0.999	1	1	1	1	1	1	1

TABLE B-2. Descriptive statistics when 15% to 50% of the dataset are nondetects

Method	Assumptions	Advantages	Disadvantages
Median (when *n* is an odd or an even integer): Determine the median as illustrated in Box 2-4 of main document.	• The largest nondetect is less than the median of the entire dataset (detects + nondetects); that is, there are no nondetects in the upper 50% of the measurements	• A simple procedure	• The median cannot be determined if the assumption is not true.
100p% Trimmed Mean: Determine the percentage (100p%) of measurements below the DL. Discard the largest np measurements and the smallest np measurements. Compute the arithmetic mean on the $n(1 - 2p)$ remaining measurements.	• All nondetects have the same DL. • All detects are larger than the DL • The number of nondetects is no more than np. • The underlying distribution of measurements is symmetric (not skewed). • $0 < p < 0.50$.	• Trimmed mean is not affected by outliers that have been trimmed from the dataset.	• Cannot be used if the assumptions are not true.
Winsorized Mean ($\bar{x}_w$): If n′ nondetects are in the lower tail of a dataset with *n* measurements (including nondetects): • Replace the n′ nondetects by the next *largest* detected datum. • Also replace the n′ largest measurements by the next *smallest* measurement. • Obtain the Winsorized Mean, $\bar{x}_w$, by computing the arithmetic mean of the resulting set of *n* measurements	• All nondetects have the same DL. • All detects are larger than the DL. • The underlying distribution of the measurements is symmetric (not skewed).	• Winsorized mean is not affected by outliers that are among the largest measurements.	• Cannot be used if the assumptions are not true.
Winsorized Standard Deviation (s_w) If n′ nondetects are in the lower tail of a dataset with *n* measurements (detects plus nondetects): • Replace the n′ nondetects by the next *largest* detected datum. • Also replace the n′ largest measurements by the next *smallest* measurement. • Compute the standard deviation, *s*, of the new set of *n* measurements. • Compute $s_w = [s(n - 1)]/(v - 1)$ where $v = n-2n'$ is the number of measurements not replaced during the Winsorization process.	• All nondetects have the same DL. • All detects are greater than the DL. • The underlying distribution of the measurements is symmetric (not skewed). • The quantity v must be greater than 1.	• If the measurements are normally distributed, then confidence intervals for the mean can be computed using the method in Gilbert (1987, p. 180).	• Cannot be used if the assumptions are not true.
Cohen Method for the Mean and Standard Deviation: See Box B-5.	• All nondetects have the same DL. • The underlying distribution of the measurements is normal. • Measurements obtained are representative of the underlying normal distribution.	• Has good performance if the underlying assumptions are valid and if the number of samples is sufficiently large.	• The assumptions must be valid.
pth Sample Percentile: See Box 2-4.	• All nondetects have the same DL. • All detects are greater than the DL. • The computed value of k (see Box 2-4) must be larger than the number of nondetects plus 1.	• Provides an estimate of the value that is exceeded by 100(1 – p)% of the underlying population	• Cannot be computed when the assumption on k is not valid.

Sources: Gilbert (1987); U.S. EPA (2000).

TABLE B-3. Assumptions, advantages, and disadvantages of outlier tests

Statistical Test	Assumptions	Advantages	Disadvantages
Dixon Test	• $n \leq 25$ • Measurements are representative of the underlying population. • The measurements without the suspect outlier are normally distributed; otherwise, consult a statistician. • Test can be used to test for either one suspect large outlier or one suspect small outlier. The latter case is not considered here as it is not of interest for determining COPCs.	• Simple to compute by hand. • The test is available in the DataQUEST software (U.S. EPA, 1997).	• Test should be used for only one suspected outlier. Use the Rosner test if multiple suspected outliers are present. • Must conduct a test for normality on the dataset after deleting the suspect outlier and before using the Dixon test.
Discordance Test	• $3 < n \leq 50$ • Measurements are representative of underlying population. • The measurements without the suspected outlier are normally distributed; otherwise, consult a statistician. • Test can be used to test that the largest measurement is a suspected outlier, or the smallest measurement is a suspected outlier. The latter case is not considered here as it is not of interest for determining COPCs.	• Simple to compute by hand. • The test is available in the DataQUEST software (U.S. EPA, 1997).	• Test can be used for only one suspected outlier. Use the Rosner test if there are multiple suspected outliers. • Must conduct a test for normality on the dataset after deleting the suspect outlier and before using the Discordance test.
Walsh Test	• $n > 60$ • Measurements are representative of the underlying population. • Test can be used to test that the largest *r* measurements or the smallest *r* measurements are suspected outliers. The latter case (discussed in U.S. EPA, 2000) is not considered here as it is not of interest for determining COPCs.	• Can test for 1 or more outliers. • The measurements need not be normally distributed. • Need not conduct a test for normality before using the test. • The test is available in the DataQUEST software (U.S. EPA, 1997).	• Must have $n > 60$ to conduct the test. • The test can only be performed for the $\alpha = 0.05$ and 0.10 significance levels, and the α level used depends on n: the $\alpha = 0.05$ level can only be used if $n > 220$ and the $\alpha = 0.10$ level can only be used if $60 < n \leq 220$. • Test calculations are more complex than for the Dixon test or the Discordance test. • The number of identified suspected outliers, *r*, are accepted or rejected as a group rather than one at a time.
Rosner Test	• $n \geq 25$ • Measurements are representative of underlying population. • The measurements without the suspected outliers are normally distributed; otherwise, consult a statistician.	• Can test for up to 10 outliers. • The test is available in the DataQUEST software (U.S. EPA, 1997).	• Must conduct a test for normality after deleting the suspected outliers and before using Rosner test. • Computations are more complex than for Dixon test or the Discordance test.

BOX B-1. Procedure for conducting the Shapiro-Wilk W test

Select the significance level, α, desired for the test, where $0 < \alpha < 0.5$. That is, select the tolerable probability, α, that the W test will declare that the measurements in the dataset are not from a normal distribution when in fact they are from a normal distribution.

- Compute the arithmetic mean of the *n* data: $\bar{x} = (x_1 + x_2 + \ldots + x_n) / n$
- Compute the denominator *d* of the W test statistic using the *n* data and $\bar{x}$:

$$d = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2$$

- Order the *n* data from smallest to largest. Denote these "sample order statistics" by $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, where $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$.
- Compute *k*, where $k = n/2$ if *n* is an even integer and $k = (n - 1)/2$ if *n* is an odd integer
- Turn to Table C-10 to obtain the coefficients $a_1, a_2, \ldots, a_k$ for the value of *n*.
- Compute the W test statistic

$$W = \{ a_1(x_{(n)} - x_{(1)}) + a_2(x_{(n-1)} - x_{(2)}) + \ldots + a_k(x_{(n-k+1)} - x_{(k)})\}^2 / d$$

- Conclude that the dataset is not normally distributed if the value of W is less than the critical value given in Table C-11 for the selected significance level α.

Example:

- Suppose we select $\alpha = 0.05$
- Suppose there are $n = 10$ measurements in the dataset:

1.20, 0.13, 1.69, 1.05, 1.12, 0.45, 2.06, 0.60, 0.76, 1.37

- The arithmetic mean of these data is

$$\bar{x} = (1.2 + 0.13 + 1.69 + 1.05 + 1.12 + 0.45 + 2.06 + 0.60 + 0.76 + 1.37) / 10$$
$$= 1.04$$

- The denominator *d* of the W test statistic using the *n* data and $\bar{x}$ is:

$$d = (1.2 - 1.04)^2 + (0.13 - 1.04)^2 + \ldots + (1.37 - 1.04)^2 = 3.05$$

- Order the $n = 10$ measurements from smallest to largest to obtain:

0.13, 0.45, 0.60, 0.76, 1.05, 1.12, 1.20, 1.37, 1.69, 2.06

- Compute $k = n/2 = 10/2 = 5$ because *n* is an even integer.
- In Table C-10 we find that the $k = 5$ coefficients are

$$a_1 = 0.5739,\ a_2 = 0.3291,\ a_3 = 0.2141,\ a_4 = 0.1224,\ a_5 = 0.0399$$

- Therefore, the computed W statistic is:

$$W = \{0.5739(2.06 - 0.13) + 0.3291(1.69 - 0.45) + 0.2141(1.37 - 0.60) + 0.1224(1.20 - 0.76) + 0.0399(1.12 - 1.05)\}^2 / 3.05$$
$$= 0.989$$

The critical value from Table C-11 for $n = 10$ and $\alpha = 0.05$ is 0.842. Therefore, because 0.989 is not less than 0.842, we conclude the measurements appear to be normally distributed. The data do not provide convincing evidence the distribution of the measurements is not normal.

BOX B-2. Procedure for conducting the D'Agostino Test

- Select the significance level, α, desired for the test, where $0 < \alpha < 0.5$.
- Compute $s = \{[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2] / n\}^{1/2}$
- Order the *n* data from smallest to largest. Denote these sample order statistics by

$$x_{(1)}, x_{(2)}, \ldots, x_{(n),} \text{ where } x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$$

- Compute $D = \{[1 - 0.5(n+1)]x_{(1)} + [2 - 0.5(n+1)]x_{(2)} + \ldots + [n - 0.5(n+1)]x_{(n)}\} / n^2 s$
- Compute $Y = (D - 0.282094) / (0.02998598 / n^{1/2})$
- Conclude the data are not from a normal distribution, if Y is less than the critical value $Y_{\alpha/2}$ or greater than the critical value $Y_{1-\alpha/2}$, that are found in Table C-12 for each value of n.

Example (from Gilbert, 1987, p. 161):

- Suppose we select $\alpha = 0.05$
- Suppose n = 115 and the computed value of *s* is

$$\{[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2] / 115\}^{1/2} = 0.4978$$

- Then the value of $n^2 s$, the denominator of *D*, is $(115)^2(0.4978) = 6{,}583$
- As $0.5(n+1) = 0.5(116) = 58$, and using the sample order statistics $x_{[i]}$, the numerator of D equals

$$\{[1–58]x_{(1)} + [2–58]x_{(2)} + \ldots + [115 - 58]x_{(115)}\} = 1{,}833.3$$

- Therefore, $D = 1{,}833.3 / 6{,}583 = 0.2785$
- Therefore, $Y = (0.2785 - 0.282094) / (0.02998798 / 115^{1/2}) = -1.29$
- From Table C-12, we find using linear interpolation that $Y_{0.025} = -2.522$ and $Y_{0.975} = 1.339$.
- Because -1.29 is not less than -2.522 and not larger than 1.339, we cannot conclude that the measurements are not normally distributed.

BOX B-3. Examples of computing the median, trimmed mean, and Winsorized mean and standard deviation using a dataset that contains nondetects

The following examples use this dataset of 12 measurements (after being ordered from smallest to largest): <0.15, <0.15, <0.15, 0.18, 0.25, 0.26, 0.27, 0.36, 0.50, 0.62, 0.63, 0.79. Note three nondetects are in this dataset, but each one has the same detection limit, 0.15. If multiple detection limits are present, consult a statistician for the best way to summarize the data.

Median

The median of the dataset is (0.26 + 0.27) / 2 = 0.265. Note the nondetects do not have any impact on computing the median because fewer than half of the data were nondetects.

100p% Trimmed Mean

The percentage of nondetect measurements is 100(3/12) = 25%. Therefore, we set p = 0.25 and compute the 25% trimmed mean (25% of *n* is 3). Discard the smallest 0.25(12) = 3 and largest 3 measurements, that is, discard the three nondetects and the measurements 0.62, 0.63, 0.79. Compute the arithmetic mean on the remaining 6 measurements: Trimmed Mean = (0.18 + 0.25 + 0.26 + 0.27 + 0.36 + 0.50) / 6 = 0.30. This estimate is valid if the underlying distribution of measurements is symmetric. If the distribution is not symmetric, this trimmed mean is a biased estimate.

Winsorized Mean

Replace the three nondetects by the next largest detected datum, which is 0.18. Replace the three largest measurements by the next smallest measurement, which is 0.50. Compute the arithmetic mean of the new set of 12 data: 0.18, 0.18, 0.18, 0.18, 0.25, 0.26, 0.27, 0.36, 0.50, 0.50, 0.50, 0.50.

$$\bar{x}_w = (0.18 + 0.18 + 0.18 + 0.18 + 0.25 + 0.26 + 0.27 + 0.36 + 0.50 + 0.50 + 0.50 + 0.50) / 12 = 0.32$$

This estimate is valid if the underlying distribution of measurements is symmetric. If the distribution is not symmetric, this Winsorized mean is a biased estimate.

Winsorized Standard Deviation

Replace the three nondetects by the next largest detected datum, which is 0.18. Replace the three largest measurements by the next smallest measurement, which is 0.50. Compute the standard deviation, *s*, of the new set of 12 data:

$$s = [(0.18 - 0.32)^2 + (0.18 - 0.32)^2 + (0.18 - 0.32)^2 + (0.18 - 0.32)^2 + (0.25 - 0.32)^2 + (0.26 - 0.32)^2 + (0.27 - 0.32)^2 + (0.36 - 0.32)^2 + (0.50 - 0.32)^2 + (0.50 - 0.32)^2 + (0.50 - 0.32)^2 + (0.50 - 0.32)^2] / 11$$

$$= 0.1416$$

Compute $v = n - 2n' = 12 - 2(3) = 6$

Compute the Winsorized Standard Deviation:

$$s_w = [s(n - 1)]/(v - 1) = [0.1416(11)] / 5 = 0.31$$

This estimate is valid if the underlying distribution of measurements is symmetric. If the distribution is not symmetric, this Winsorized standard deviation is a biased estimate.

BOX B-4. Cohen method for computing the mean and variance of a censored dataset (U.S. EPA, 2000; Gilbert, 1987, p. 182)

- Let the single detection limit be denoted by DL. Let $x_1, x_2, \ldots, x_n$ denote the *n* measurements in the dataset, including those that are less than DL. Let k be the number out of *n* that are greater than the DL.
- Compute h = (n-k)/n, which is the fraction of the *n* measurements that are below the DL.
- Compute the arithmetic mean of the k measurements that exceed the DL as follows:

$$\bar{x}_c = (x_1 + x_2 + \ldots + x_k) / k$$

where $x_1, x_2, \ldots$, and x_k are all the measurements > DL.

- Compute the following statistic using the k measurements that exceed the DL:

$$s_c^2 = [(x_1 - \bar{x}_c)^2 + (x_2 - \bar{x}_c)^2 + \ldots + (x_k - \bar{x}_c)^2] / k$$

- Compute $G = s_c^2 / (\bar{x}_c - DL)^2$
- Obtain the value of λ from Table C-13 for values of h and γ. Use linear interpolation in the table if necessary.
- Compute the Cohen mean and variance as follows:

$$\text{Cohen Mean} = \bar{x}_c - \lambda(\bar{x}_c - DL)$$
$$\text{Cohen Variance} = s_c^2 + \lambda(\bar{x}_c - DL)^2$$

- Cohen Standard Deviation is the square root of Cohen Variance.

Example:

- n = 25 measurements of a chemical in soil were obtained. One detection limit was equal to 36. Five measurements were reported as <36 (ND). The data obtained were:

<36, <36, <36, <36, <36, 49, 49, 59, 61, 62, 62, 65, 65, 65, 70, 72, 80, 80, 99, 99, 104, 110, 140, 142, 144.

- Compute h = (25 – 20)/25 = 0.20 = fraction of the 25 measurements that are below the detection limit
- Compute the arithmetic mean of the 20 measurements that exceed the detection limit:

$$\bar{x}_c = (49 + 49 + 59 + \ldots + 142 + 144) = 83.85$$

- Compute $s_c^2 = [(49 - 83.85)^2 + (49 - 83.85)^2 + (59 - 83.85)^2 + \ldots + (142 - 83.85)^2 + (144 - 83.85)^2] / 20$

$$= 882.63$$

- Compute $G = 882.63 / (83.85 - 36)^2 = 0.385$
- From Table C-13, we find by linear interpolation between γ = 0.35 and γ = 0.40 for h = 0.20 that λ = 0.291.
- Therefore, Cohen mean and variance are:

$$\text{Cohen Mean} = 83.85 - 0.291(83.85 - 36) = 69.9$$

$$\text{Cohen Variance} = 882.63 + 0.291(83.85 - 36)^2 = 1{,}548.9$$

- Cohen Standard Deviation = $(1548.9)^{1/2} = 39.4$

BOX B-5. Procedure for conducting the Dixon extreme value test (U.S. EPA, 2000)

- Let $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ be the *n* measurements in the dataset after they have been listed in order from smallest to largest. The parentheses around the subscripts indicate the measurements are ordered from smallest to largest.
- $x_{(n)}$ (the largest measurement) is suspected of being an outlier.
- Perform test for normality on $x_{(1)}$ through $x_{(n-1)}$.
- Specify the tolerable decision error rate, α (significance level), desired for the test. α may only be set equal to 0.01, 0.05 or 0.10 for the Dixon test.
- Compute C

$$\begin{aligned} C &= [x_{(n)} - x_{(n-1)}] / [x_{(n)} - x_{(1)}] \text{ if } 3 \le n \le 7 \\ &= [x_{(n)} - x_{(n-1)}] / [x_{(n)} - x_{(2)}] \text{ if } 8 \le n \le 10 \\ &= [x_{(n)} - x_{(n-2)}] / [x_{(n)} - x_{(2)}] \text{ if } 11 \le n \le 13 \\ &= [x_{(n)} - x_{(n-2)}] / [x_{(n)} - x_{(3)}] \text{ if } 14 \le n \le 25 \end{aligned}$$

  If C exceeds the critical value in Table C-14 for the specified *n* and α, then declare that $x_{(n)}$ is an outlier and should be investigated further.

Example: Suppose the ordered dataset is 34, 50, 52, 60, 62. Suppose we wish to test if 62 is an outlier from an assumed normal distribution for the n = 5 data. Perform a test for normality on the data 34, 50, 52, 60. Note that any test for normality will have little ability to detect nonnormality on the basis of only 4 data values. (See Section B.1 for statistical methods of testing the normality assumption.) Suppose α is selected to be 0.05, that is, we want no more than a 5% chance the test will incorrectly declare the largest observed measurement is an outlier. Compute C = (62 – 60)/(62 – 34) = 0.071. Determine the test critical value from Table C-14. The critical value is 0.642 when n = 5 and α = 0.05. As 0.071 is less than 0.642, the data do not indicate the measurement 62 is an outlier from an assumed normally distribution.

BOX B-6. Procedure for conducting the Discordance outlier test (U.S. EPA, 2000)

- Let $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ be the *n* measurements in the dataset after they have been listed in order from smallest to largest.
- $x_{(n)}$ (the largest measurement) is suspected of being an outlier.
- Specify the tolerable decision error rate, α (significance level) desired for the test. α may be specified to be 0.01 or 0.05 for the Discordance outlier test.
- Compute the sample arithmetic mean, $\bar{x}$, and the sample standard deviation, s.
- Compute $D = [x_{(n)} - \bar{x}] / s$
- If D exceeds the critical value from Table C-15 for the specified *n* and α, $x_{(n)}$ is an outlier and should be further investigated.

Example: Suppose the ordered dataset is 34, 50, 52, 60, 62, and it is decided to test if 62 is an outlier from an assumed normal distribution for the data. Suppose α is selected to be 0.05. Using the n = 5 data, we compute $\bar{x}$ = 51.6 and s = 11.08. Therefore, D = (62 – 51.6) / 11.08 = 0.939. The critical value from Table C-15 for n = 5 and α = 0.05 is 1.672. As 0.939 is less than 1.672, the data do not indicate the measurement 62 is an outlier from an assumed normal distribution.

BOX B-7. Procedure for conducting the Walsh outlier test (U.S. EPA, 2000)

- Let $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ denote *n* measurements in the dataset after they have been listed in order from smallest to largest. Do not apply the test if $n < 60$. If $60 < n \le 220$, then use $\alpha = 0.10$. If $n > 220$, then use $\alpha = 0.05$.
- Identify the number of possible outliers, *r*, where *r* can equal 1.
- Compute: $c = [(2n)^{1/2}]$, $k = r + c$, $b^2 = 1/\alpha$,

$$a = (1 + b\{(c - b^2)/(c - 1)\}^{1/2}) / (c - b^2 - 1)$$

where [ ] indicates rounding the value to the largest possible integer (that is, 3.24 becomes 4).

- The Walsh test declares that the r largest measurements are outliers (with a α level of significance) if

$$x_{(n+1-r)} - (1 + a)x_{(n-r)} + ax_{(n+1-k)} > 0$$

Example: Suppose $n = 70$ and that $r = 3$ largest measurements are suspected outliers. The significance level $\alpha = 0.10$ must be used because $60 < n \le 220$. That is, we must accept a probability of 0.10 the test will incorrectly declare that the three largest measurements are outliers.

- Compute $c = [(2 \times 70)^{1/2}] = 12$
  $k = 3 + 12 = 15$
  $b^2 = 1 / 0.10 = 10$
  $a = 1 + 3.162\{(12 - 10) / (12 - 1)\}^{1/2}\} / (12 - 10 - 1) = 2.348$
- $x_{(n+1-r)} = x_{(70+1-3)} = x_{(68)}$ is the 68th largest measurement (two measurements are larger)
  $x_{(n-r)} = x_{(70-3)} = x_{(67)}$ is the 67th largest measurement
  $x_{(n+1-k)} = x_{(70+1-15)} = x_{(56)}$ is the 56th largest measurement
- Order the 70 measurements from smallest to largest. Suppose $x_{(68)} = 83$, $X_{(67)} = 81$, and $x_{(56)} = 20$.
- Compute $x_{(n+1-r)} - (1+a)x_{(n-r)} + ax_{(n+1-k)} = 83 - (1+2.348)81 + 2.348(20) = -141.22$ which is smaller than 0. Therefore, the Walsh test indicates that the $r = 3$ largest measurements are not outliers.

BOX B-8. Procedure for conducting the Rosner outlier test (U.S. EPA, 2000)

STEP 1:

- Select the desired significance level α, that is, the tolerable probability that the Rosner test will falsely declare that outliers are present.
- Let $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ denote *n* measurements in the dataset after they have been listed in order from smallest to largest, where $n \geq 25$.
- Identify the maximum number of possible outliers, denoted by r.

STEP 2:

- Set i = 0 and use the following formulas to compute the sample arithmetic mean, labeled $\bar{x}^{(0)}$, and $s^{(0)}$ using all *n* measurements. Determine the measurement that is farthest from $\bar{x}^{(0)}$ and label it $y^{(0)}$.

$$\bar{x}^{(i)} = (x_1 + x_2 + \ldots + x_{n-i}) / (n - i)$$

$$s^{(i)} = \{[(x_1 - \bar{x}^{(i)})^2 + (x_2 - \bar{x}^{(i)})^2 + \ldots + (x_{n-i} - \bar{x}^{(i)})^2] / (n - i)\}^{1/2}$$

- Delete $y^{(0)}$ from the dataset of *n* measurements and compute (using i = 1 in the above formulas) the sample arithmetic mean, labeled $\bar{x}^{(1)}$, and $s^{(1)}$ on the remaining n–1 measurements. Determine the measurement that is farthest from $\bar{x}^{(1)}$ and label it $y^{(1)}$.
- Delete $y^{(1)}$ from the dataset and compute (using i = 2 in the above formulas) the sample arithmetic mean, labeled $\bar{x}^{(2)}$, and $s^{(2)}$ on the remaining n-2 measurements.
- Continue using this process until the r largest measurements have been deleted from the dataset.
- The values of $\bar{x}^{(0)}$, $\bar{x}^{(1)}$, …, $s^{(0)}$, $s^{(1)}$, … are computed using the above formulas.

STEP 3:

- To test if there are r outliers in the dataset compute

$$R_r = [\,|y^{(r-1)} - \bar{x}^{(r-1)}|\,] / s^{(r-1)}$$

- Determine the critical value λ_r from Table C-16 for the values of *n*, *r*, and α.
- If R_r exceeds λ_r, conclude *r* outliers are in the dataset.
- If not, test if r –1 outliers are present. Compute

$$R_{r-1} = [\,|y^{(r-2)} - \bar{x}^{(r-2)}|\,] / s^{(r-2)}$$

- Determine the critical value λ_{r-1} from Table C-16 for the values of *n*, r – 1 and α.
- If R_{r-1} exceeds λ_{r-1}, conclude r –1 outliers are in the dataset.
- Continue on in this way until either it is determined that either a certain number of outliers are present or that no outliers exist.

BOX B-9. Example of the Rosner outlier test

STEP 1: Consider the following 32 data points (in ppm) listed in order from smallest to largest: 2.07, 40.55, 84.15, 88.41, 98.84, 100.54, 115.37, 121.19, 122.08, 125.84, 129.47, 131.90, 149.06, 163.89, 166.77, 171.91, 178.23, 181.64, 185.47, 187.64, 193.73, 199.74, 209.43, 213.29, 223.14, 225.12, 232.72, 233.21, 239.97, 251.12, 275.36, and 395.67.

A normal probability plot of the data identified four potential outliers: 2.07, 40.55, 275.36 and 395.67. Moreover, a normal probability plot of the dataset after excluding the four suspect outliers provided no evidence that the data are not normally distributed.

STEP 2: First use the formulas in Step 2 of Box B-9 to compute $\bar{x}^{(0)}$ and $s^{(0)}$ using the entire dataset. Using subtraction, it was found that 395.67 was the farthest data point from $\bar{x}^{(0)}$, so $y^{(0)} = 395.67$. Then 395.67 was deleted from the dataset and $\bar{x}^{(1)}$ and $s^{(1)}$ are computed on the remaining data. Using subtraction, it was found that 2.07 was the farthest value from $\bar{x}^{(1)}$, so $y^{(1)} = 2.07$. This value then was dropped from the data and the process was repeated to determine $\bar{x}^{(2)}$, $s^{(2)}$, $y^{(2)}$ and $\bar{x}^{(3)}$, $s^{(3)}$, $y^{(3)}$. These values are summarized in the table.

i	$\bar{x}^{(i)}$	$s^{(i)}$	$y^{(i)}$
0	169.92	73.95	395.67
1	162.64	62.83	2.07
2	167.99	56.49	40.55
3	172.39	52.18	275.36

STEP 3: To apply the Rosner test, first test if 4 outliers are present. Compute

$$R_4 = |y^{(3)} - \bar{x}^{(3)}| \;/\; s^{(3)} = |275.36 - 172.39| \;/\; 52.18 = 1.97$$

Suppose we want to conduct the test at the $\alpha = 0.05$ level, that is, we can tolerate a 5% chance of the Rosner test falsely declaring 4 outliers. In Table C-16, we find $\lambda_4 = 2.89$ when $n = 32$, $r = 4$ and $\alpha = 0.05$. As $R_4 = 1.97$ is less than 2.89, we conclude that 4 outliers are not present. Therefore, test if 3 outliers are present. Compute

$$R_3 = |y^{(2)} - \bar{x}^{(2)}| \;/\; s^{(2)} = |40.55 - 167.99| \;/\; 56.49 = 2.26$$

In Table C-16 we find $\lambda_3 = 2.91$ when $n = 32$, $r = 3$ and $\alpha = 0.05$. Because $R_4 = 2.26$ is less than 2.91, we conclude that 3 outliers are not present. Therefore, test if 2 outliers are present. Compute

$$R_2 = |y^{(1)} - \bar{x}^{(1)}| \;/\; s^{(1)} = |2.07 - 162.64| \;/\; 62.83 = 2.56$$

In Table C-16, we find $\lambda_2 = 2.92$ for $n = 32$, $r = 2$ and $\alpha = 0.05$. As $R_2 = 2.56$ is less than 2.92, we conclude that 2 outliers are not present in the dataset. Therefore, test if 1 outlier is present. Compute

$$R_1 = |y^{(0)} - \bar{x}^{(0)}| \;/\; s^{(0)} = |395.67 - 169.92| \;/\; 73.95 = 3.05$$

In Table C-16 we find $\lambda_1 = 2.94$ for $n = 32$, $r = 1$and $\alpha = 0.05$. Because $R_1 = 3.05$ is greater than 2.94, we conclude at the $\alpha = 0.05$ significance level that 1 outlier is present in the dataset. Therefore, the measurement 395.67 is considered to be a statistical outlier, and it will be further investigated to determine if it is an error or a valid data value.

APPENDIX C: COMPARATIVE METHOD REFERENCE TABLES

TABLE C-1. Cumulative standard normal distribution (values of the probability ϕ corresponding to the value z_ϕ of a standard normal random variable)

	α									
z_ϕ	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5674	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

TABLE C-2. Critical values (*K*) for the Slippage test for α = 0.01

	Number of Site Measurements, n																								
Number of Background Measurements, m	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
1	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
2	/	/	/	/	/	/	/	/	/	/	/	/	13	14	15	16	17	18	19	20	21	22	23	23	24
3	/	/	/	/	/	/	7	8	9	10	11	11	12	13	14	15	15	16	17	18	18	19	20	21	22
4	/	/	/	/	5	6	7	8	8	9	10	10	11	12	12	13	14	14	15	16	16	17	18	19	19
5	/	/	/	4	5	6	6	7	8	8	9	9	10	11	11	12	12	13	14	14	15	15	16	17	17
6	/	/	/	4	5	5	6	6	7	8	8	9	9	10	10	11	11	12	12	13	14	14	15	15	16
7	/	/	3	4	5	5	6	6	7	7	8	8	9	9	10	10	10	11	11	12	12	13	13	14	14
8	/	/	3	4	4	5	5	6	6	7	7	8	8	8	9	9	10	10	11	11	12	12	12	13	13
9	/	/	3	4	4	5	5	5	6	6	7	7	8	8	8	9	9	10	10	10	11	11	12	12	12
10	/	/	3	4	4	4	5	5	6	6	6	7	7	7	8	8	9	9	9	10	10	11	11	11	12
11	/	/	3	3	4	4	5	5	5	6	6	6	7	7	7	8	8	9	9	9	10	10	10	11	11
12	/	/	3	3	4	4	4	5	5	5	6	6	6	7	7	7	8	8	8	9	9	9	10	10	10
13	/	2	3	3	4	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8	9	9	9	10	10
14	/	2	3	3	4	4	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8	9	9	9	9
15	/	2	3	3	3	4	4	4	5	5	5	5	6	6	6	7	7	7	7	8	8	8	9	9	9
16	/	2	3	3	3	4	4	4	4	5	5	5	6	6	6	6	7	7	7	7	8	8	8	8	9
17	/	2	3	3	3	4	4	4	4	5	6	6	6	6	6	6	6	7	7	7	7	8	8	8	8
18	/	2	3	3	3	3	4	4	4	5	5	5	5	5	6	6	6	6	7	7	7	7	8	8	8
19	/	2	3	3	3	3	4	4	4	4	5	5	5	6	6	6	6	6	6	7	7	7	7	8	8
20	/	2	3	3	3	3	4	4	4	4	5	5	5	5	5	6	6	6	6	7	7	7	7	7	8
21	/	2	3	3	3	3	4	4	4	4	4	5	5	5	5	6	6	6	6	6	7	7	7	7	7
22	/	2	3	3	3	3	3	4	4	4	4	5	5	5	5	5	6	6	6	6	6	7	7	7	7
23	/	2	2	3	3	3	3	4	4	4	4	4	5	5	5	5	5	6	6	6	6	6	7	7	7
24	/	2	2	3	3	3	3	4	4	4	4	4	5	5	5	5	5	6	6	6	6	6	6	7	7
25	/	2	2	3	3	3	3	4	4	4	4	4	4	5	5	5	5	5	6	6	6	6	6	6	7
26	/	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5	5	6	6	6	6	6	6
27	/	2	2	3	3	3	3	3	4	4	4	4	4	4	5	5	5	5	5	6	6	6	6	6	6
28	/	2	2	3	3	3	3	3	4	4	4	4	4	4	5	5	5	5	5	5	6	6	6	6	6
29	/	2	2	3	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5	5	5	6	6	6	6
30	/	2	2	3	3	3	3	3	3	4	4	4	4	4	4	5	5	5	5	5	5	6	6	6	6
31	/	2	2	3	3	3	3	3	3	4	4	4	4	4	4	5	5	5	5	5	5	5	6	6	6
32	/	2	2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5	5	5	5	5	6	6
33	/	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4	5	5	5	5	5	5	6	6	6
34	/	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5	5	5	5	6	6
35	/	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5
36	/	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5
37	/	2	2	2	3	3	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5	5	5	5	5
38	/	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5	5	5	5	5
39	/	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5	5	5	5	5
40	/	2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5	5	5	5
41	/	2	2	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5	5	5	5
42	/	2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5	5	5	5
43	/	2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	5	5	5	5
44	/	2	2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5	5	5
45	/	2	2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	5	5	5
46	/	2	2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	5	5	5
47	/	2	2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	5	5
48	/	2	2	2	2	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	5	5
49	/	2	2	2	2	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	5
50	/	2	2	2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	5

TABLE C-2. Critical values (*K*) for the Slippage test for α = 0.01 (continued)

Number of Site Measurements, n

Number of Background Measurements, m

	26	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**
1	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
2	25	26	27	28	29	30	31	32	32	33	34	35	36	37	38	39	40	41	41	42	43	44	45	46	47
3	22	23	24	25	26	26	27	28	29	30	30	31	32	33	33	34	35	36	37	37	38	39	40	40	41
4	20	21	21	22	23	23	24	25	25	26	27	27	28	29	30	30	31	32	32	33	34	34	35	35	36
5	18	18	19	20	20	21	22	22	23	23	24	25	25	26	26	27	28	28	29	29	30	31	31	32	32
6	16	17	17	18	18	19	19	20	21	21	22	22	23	23	24	24	25	25	26	26	27	28	28	29	29
7	15	15	16	16	17	17	18	18	19	19	20	20	21	21	22	22	23	23	24	24	25	25	26	26	26
8	14	14	15	15	16	16	16	17	17	18	18	19	19	19	20	20	21	21	22	22	23	23	23	24	24
9	13	13	14	14	14	15	15	16	16	16	17	17	18	18	18	19	19	20	20	20	21	21	22	22	22
10	12	12	13	13	13	14	14	15	15	15	16	16	16	17	17	18	18	18	19	19	19	20	20	21	21
11	11	12	12	12	13	13	13	14	14	14	15	15	15	16	16	16	17	17	18	18	18	19	19	19	20
12	11	11	11	12	12	12	13	13	13	14	14	14	15	15	15	16	16	16	17	17	17	17	18	18	18
13	10	10	11	11	11	12	12	12	13	13	13	14	14	14	14	15	15	15	16	16	16	17	17	17	17
14	10	10	10	11	11	11	11	12	12	12	13	13	13	13	14	14	14	15	15	15	15	16	16	16	17
15	9	10	10	10	10	11	11	11	11	12	12	12	13	13	13	13	14	14	14	14	15	15	15	15	16
16	9	9	9	10	10	10	10	11	11	11	11	12	12	12	13	13	13	13	14	14	14	14	15	15	15
17	9	9	9	9	10	10	10	10	11	11	11	11	12	12	12	12	12	13	13	13	13	14	14	14	14
18	8	9	9	9	9	9	10	10	10	10	11	11	11	11	12	12	12	12	12	13	13	13	13	14	14
19	8	8	8	9	9	9	9	10	10	10	10	10	11	11	11	11	12	12	12	12	13	13	13	13	13
20	8	8	8	8	9	9	9	9	9	10	10	10	10	11	11	11	11	11	12	12	12	12	12	13	13
21	8	8	8	8	8	9	9	9	9	9	10	10	10	10	10	11	11	11	11	11	12	12	12	12	12
22	7	8	8	8	8	8	9	9	9	9	9	9	10	10	10	10	10	11	11	11	11	11	12	12	12
23	7	7	8	8	8	8	8	8	9	9	9	9	9	10	10	10	10	10	11	11	11	11	11	11	12
24	7	7	7	8	8	8	8	8	8	9	9	9	9	9	9	10	10	10	10	10	11	11	11	11	11
25	7	7	7	7	8	8	8	8	8	8	9	9	9	9	9	9	10	10	10	10	10	10	11	11	11
26	7	7	7	7	7	7	8	8	8	8	8	8	9	9	9	9	9	9	10	10	10	10	10	10	11
27	7	7	7	7	7	7	7	8	8	8	8	8	8	9	9	9	9	9	9	10	10	10	10	10	10
28	6	7	7	7	7	7	7	7	8	8	8	8	8	8	8	9	9	9	9	9	9	10	10	10	10
29	6	6	7	7	7	7	7	7	7	8	8	8	8	8	8	9	9	9	9	9	9	9	10	10	10
30	6	6	6	7	7	7	7	7	7	7	8	8	8	8	8	8	8	9	9	9	9	9	9	9	10
31	6	6	6	6	7	7	7	7	7	7	7	8	8	8	8	8	8	8	9	9	9	9	9	9	9
32	6	6	6	6	6	7	7	7	7	7	7	7	8	8	8	8	8	8	8	9	9	9	9	9	9
33	6	6	6	6	6	6	7	7	7	7	7	7	7	8	8	8	8	8	8	8	9	9	9	9	9
34	6	6	6	6	6	6	6	7	7	7	7	7	7	7	8	8	8	8	8	8	8	8	9	9	9
35	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	8	8	8	8	8	8	8	8	9	9
36	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	8	8	8	8	8	8	8	8	9
37	5	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	8	8	8	8	8	8	8	8
38	5	5	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	8	8	8	8	8	8	8
39	5	5	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	8	8	8	8	8	8
40	5	5	5	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	7	8	8	8	8
41	5	5	5	5	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	7	8	8	8
42	5	5	5	5	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	7	7	8	8
43	5	5	5	5	5	5	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	7	7	8
44	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	7	7
45	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	7
46	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7
47	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7
48	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7
49	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7
50	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7

TABLE C-3. Critical values (*K*) for the Slippage test for $\alpha = 0.05$

Number of Site Measurements, n

Number of Background Measurements, m	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
1	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	20	21	22	23	24	25
2	/	/	/	/	5	6	7	8	9	9	10	11	12	13	13	14	15	16	16	17	18	19	20	20	21
3	/	/	/	4	5	5	6	7	7	8	9	9	10	11	11	12	12	13	14	14	15	16	16	17	18
4	/	/	3	4	4	5	5	6	6	7	8	8	9	9	10	10	11	11	12	12	13	13	14	14	15
5	/	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	9	10	10	11	11	12	12	13	13
6	/	2	3	3	4	4	4	5	5	6	6	6	7	7	8	8	8	9	9	10	10	10	11	11	12
7	/	2	3	3	3	4	4	5	5	5	6	6	6	7	7	7	8	8	8	9	9	9	10	10	11
8	/	2	3	3	3	4	4	4	5	5	5	6	6	6	6	7	7	7	8	8	8	9	9	9	10
9	/	2	2	3	3	3	4	4	4	5	5	5	5	6	6	6	7	7	7	7	8	8	8	9	9
10	/	2	3	3	3	3	4	4	4	4	5	5	5	5	6	6	6	6	7	7	7	8	8	8	8
11	/	2	2	3	3	3	3	4	4	4	4	5	5	5	5	6	6	6	6	7	7	7	7	8	8
12	/	2	2	3	3	3	3	3	4	4	4	4	5	5	5	5	6	6	6	6	6	7	7	7	7
13	/	2	2	2	3	3	3	3	4	4	4	4	4	5	5	5	5	6	6	6	6	6	7	7	7
14	/	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5	6	6	6	6	6	7
15	/	2	2	2	3	3	3	3	3	4	4	4	4	4	4	5	5	5	5	5	6	6	6	6	6
16	/	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5	5	6	6	6	6
17	/	2	2	2	2	3	3	3	3	3	4	4	4	4	4	4	5	5	5	5	5	5	6	6	6
18	/	2	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4	5	5	5	5	5	5	5	6
19	/	2	2	2	2	3	3	3	3	3	3	3	4	4	4	4	4	4	5	5	5	5	5	5	5
20	1	2	2	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5	5	5	5
21	1	2	2	2	2	2	3	3	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5	5	5
22	1	2	2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5	5	5
23	1	2	2	2	2	2	2	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5	5
24	1	2	2	2	2	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5
25	1	2	2	2	2	2	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5
26	1	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5
27	1	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4
28	1	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4
29	1	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4
30	1	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4
31	1	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4
32	1	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4
33	1	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4
34	1	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4
35	1	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4
36	1	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4
37	1	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4
38	1	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4
39	1	1	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	4	4
40	1	1	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	4
41	1	1	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	4
42	1	1	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3
43	1	1	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3
44	1	1	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3
45	1	1	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3
46	1	1	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3
47	1	1	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3
48	1	1	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3
49	1	1	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3
50	1	1	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3

TABLE C-3. Critical values (*K*) for the Slippage test for α = 0.05 (continued)

Number of Site Measurements, n

Number of Background Measurements, m

	26	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**
1	26	27	28	29	30	31	32	33	34	35	36	37	38	39	39	40	41	42	43	44	45	46	47	48	49
2	22	23	23	24	25	26	26	27	28	29	30	30	31	32	33	33	34	35	36	37	37	38	39	40	40
3	18	19	19	20	21	21	22	23	23	24	24	25	26	26	27	28	28	29	30	30	31	31	32	33	33
4	15	16	17	17	18	18	19	19	20	20	21	21	22	22	23	23	24	24	25	26	26	27	27	28	28
5	14	14	14	15	15	16	16	17	17	18	18	18	19	19	20	20	21	21	22	22	23	23	23	24	24
6	12	12	13	13	14	14	14	15	15	16	16	16	17	17	18	18	18	19	19	20	20	20	21	21	21
7	11	11	12	12	12	13	13	13	14	14	14	15	15	15	16	16	16	17	17	18	18	18	19	19	19
8	10	10	11	11	11	12	12	12	13	13	13	13	14	14	14	15	15	15	16	16	16	17	17	17	17
9	9	9	10	10	10	11	11	11	11	12	12	12	13	13	13	13	14	14	14	15	15	16	16	16	16
10	9	9	9	9	10	10	10	10	11	11	11	11	12	12	12	12	13	13	13	14	14	14	14	15	15
11	8	8	9	9	9	9	9	10	10	10	10	11	11	11	11	12	12	12	12	13	13	13	13	14	14
12	8	8	8	8	8	9	9	9	9	10	10	10	10	10	11	11	11	11	12	12	12	12	12	13	13
13	7	7	8	8	8	8	8	9	9	9	9	9	10	10	10	10	11	11	11	11	11	12	12	12	12
14	7	7	7	7	8	8	8	8	8	9	9	9	9	9	10	10	10	10	10	11	11	11	11	11	12
15	7	7	7	7	7	7	8	8	8	8	9	9	9	9	9	9	9	10	10	10	10	10	11	11	11
16	6	6	7	7	7	7	7	7	8	8	8	8	8	9	9	9	9	9	9	10	10	10	10	10	10
17	6	6	6	7	7	7	7	7	7	8	8	8	8	8	8	8	9	9	9	9	9	9	10	10	10
18	6	6	6	6	6	7	7	7	7	7	7	8	8	8	8	8	8	8	9	9	9	9	9	9	10
19	6	6	6	6	6	6	7	7	7	7	7	7	7	8	8	8	8	8	8	8	9	9	9	9	9
20	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	8	8	8	8	8	8	8	9	9	9
21	5	5	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	8	8	8	8	8	8	8	9
22	5	5	5	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	8	8	8	8	8	8
23	5	5	5	5	5	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	8	8	8	8
24	5	5	5	5	5	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	7	8	8
25	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7	8
26	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7	7	7
27	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7
28	4	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	7	7	7	7	7
29	4	4	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	7	7	7
30	4	4	4	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	7	7
31	4	4	4	4	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	6
32	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6	6
33	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6	6
34	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6	6	6
35	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6
36	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6	6	6
37	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6	6
38	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	6	6
39	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	6
40	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5
41	4	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5
42	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5
43	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5
44	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5
45	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5
46	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5
47	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5
48	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5
49	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5
50	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5

TABLE C-4. Values of *r*, *k*, and α for the Quantile test when α is approximately equal to 0.01

Number of Reference (Background) Measurements, m	Number of Site Measurements, n																			
	5	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**	**65**	**70**	**75**	**80**	**85**	**90**	**95**	**100**
5	r, k α		11,1 0.008	13,13 0.015	16,16 0.014	19,19 0.013	22,22 0.013	25,25 0.013	28,28 0.012											
10		6,6 0.005	7,7 0.013	9,9 0.012	11,11 0.011	13,13 0.010	14,14 0.014	16,16 0.013	18,18 0.012	19,19 0.015	21,21 0.014	23,23 0.013	25,25 0.012	26,26 0.015	28,28 0.014	30,30 0.013				
15	3.3 0.009	7,6 0.007	6,6 0.008	7,7 0.012	8,8 0.014	10,10 0.009	11,11 0.011	12,12 0.013	13,13 0.014	15,15 0.011	16,16 0.012	17,17 0.013	18,18 0.014	19,19 0.015	21,21 0.012	22,22 0.013	23,23 0.014	24,24 0.015	26,26 0.013	27,27 0.013
20	6,4 0.005	4,4 0.008	5,5 0.009	6,6 0.010	7,7 0.011	8,8 0.011	9,9 0.011	10,10 0.011	11,11 0.011	12,12 0.011	13,13 0.011	14,14 0.012	15,15 0.012	16,16 0.012	17,17 0.012	18,18 0.012	19,19 0.012	19,19 0.015	20,20 0.015	21,21 0.015
25	4,3 0.009	7,5 0.012	4,4 0.015	5,5 0.013	6,6 0.011	7,7 0.010	8,8 0.009	9,9 0.009	9,9 0.014	10,10 0.012	11,11 0.011	12,12 0.011	12,12 0.015	13,13 0.014	14,14 0.013	15,15 0.012	16,16 0.011	16,16 0.014	17,17 0.014	18,18 0.013
30	4,3 0.006	3,3 0.012	4,4 0.009	5,5 0.007	6,6 0.006	6,6 0.012	7,7 0.010	8,8 0.008	8,8 0.013	9,9 0.011	10,10 0.009	10,10 0.013	11,11 0.011	12,12 0.010	12,12 0.013	13,13 0.012	14,14 0.011	14,14 0.014	15,15 0.012	15,15 0.015
35	2,2 0.013	3,3 0.008	4,4 0.006	4,4 0.014	5,5 0.010	6,6 0.007	6,6 0.012	7,7 0.009	7,7 0.014	8,8 0.011	9,9 0.009	9,9 0.013	10,10 0.010	10,10 0.014	11,11 0.011	11,11 0.015	12,12 0.012	13,13 0.011	13,13 0.013	14,14 0.012
40	2,2 0.008	3,3 0.008	7,5 0.013	4,4 0.007	5,5 0.006	5,5 0.012	6,6 0.008	6,6 0.013	7,7 0.009	7,7 0.013	8,8 0.010	8,8 0.014	9,9 0.011	9,9 0.014	10,10 0.011	10,10 0.014	11,11 0.012	11,11 0.014	12,12 0.012	12,12 0.014
45	2,2 0.008	6,4 0.008	3,3 0.013	4,4 0.007	4,4 0.014	5,5 0.008	5,5 0.014	6,6 0.009	6,6 0.013	7,7 0.009	7,7 0.013	8,8 0.009	8,8 0.012	9,9 0.009	9,9 0.012	10,10 0.009	10,10 0.012	10,10 0.015	11,11 0.012	11,11 0.014
50		4,3 0.013	3,3 0.010	4,4 0.005	4,4 0.010	5,5 0.006	5,5 0.010	5,5 0.015	6,6 0.009	6,6 0.013	7,7 0.009	7,7 0.012	8,8 0.009	8,8 0.011	8,8 0.014	9,9 0.011	9,9 0.013	10,10 0.010	10,10 0.012	10,10 0.015
55		4,3 0.010	3,3 0.008	7,5 0.013	4,4 0.008	4,4 0.014	5,5 0.007	5,5 0.011	6,6 0.007	6,6 0.010	6,6 0.014	7,7 0.009	7,7 0.012	8,8 0.008	8,8 0.010	8,8 0.013	9,9 0.009	9,9 0.012	9,9 0.014	10,10 0.011
60		4,3 0.008	3,3 0.007	3,3 0.014	4,4 0.006	4,4 0.011	5,5 0.006	5,5 0.009	5,5 0.013	6,6 0.007	6,6 0.010	6,6 0.014	7,7 0.009	7,7 0.011	7,7 0.014	8,8 0.010	8,8 0.012	8,8 0.015	9,9 0.010	9,9 0.013
65		4,3 0.007	3,3 0.006	3,3 0.012	6,5 0.006	4,4 0.009	4,4 0.013	5,5 0.007	5,5 0.010	5,5 0.014	6,6 0.008	6,6 0.011	6,6 0.014	7,7 0.009	7,7 0.011	7,7 0.014	8,8 0.009	8,8 0.011	8,8 0.014	9,9 0.010
70		2,2 0.014	6,4 0.008	3,3 0.010	7,5 0.013	4,4 0.007	4,4 0.011	5,5 0.005	5,5 0.008	5,5 0.011	5,5 0.015	6,6 0.008	6,6 0.011	6,6 0.014	7,7 0.009	7,7 0.011	7,7 0.013	8,8 0.009	8,8 0.011	8,8 0.013
75		2,2 0.013	4,3 0.014	3,3 0.008	3,3 0.014	4,4 0.006	4,4 0.009	4,4 0.013	5,5 0.006	5,5 0.009	5,5 0.012	6,6 0.007	6,6 0.009	6,6 0.011	6,6 0.014	7,7 0.009	7,7 0.011	7,7 0.013	8,8 0.008	8,8 0.010
80		2,2 0.011	4,3 0.012	3,3 0.007	3,3 0.012	6,5 0.006	4,4 0.008	4,4 0.011	5,5 0.005	5,5 0.007	5,5 0.010	5,5 0.013	6,6 0.007	6,6 0.009	6,6 0.012	6,6 0.014	7,7 0.008	7,7 0.010	7,7 0.013	7,7 0.015
85		2,2 0.010	4,3 0.010	3,3 0.006	3,3 0.011	7,5 0.013	4,4 0.006	4,4 0.009	4,4 0.013	5,5 0.006	5,5 0.008	5,5 0.011	5,5 0.014	6,6 0.008	6,6 0.010	6,6 0.012	6,6 0.014	7,7 0.008	7,7 0.010	7,7 0.012
90			4,3 0.009	3,3 0.005	3,3 0.009	3,3 0.014	4,4 0.005	4,4 0.008	4,4 0.011	5,5 0.005	5,5 0.007	5,5 0.009	5,5 0.012	5,5 0.015	6,6 0.008	6,6 0.010	6,6 0.012	6,6 0.014	7,7 0.008	7,7 0.010
95			4,3 0.008	6.4 0.008	3.3 0.008	3.3 0.013	6.5 0.005	4.4 0.007	4.4 0.010	4.4 0.013	5.5 0.006	5.5 0.008	5.5 0.010	5.5 0.013	6.6 0.007	6.6 0.008	6.6 0.010	6.6 0.012	6.6 0.014	7.7 0.008
100			4,3 0.007	4,3 0.014	3,3 0.007	3,3 0.011	7,5 0.013	4,4 0.006	4,4 0.008	4,4 0.011	4,4 0.015	5,5 0.007	5,5 0.009	5,5 0.011	5,5 0.013	6,6 0.007	6,6 0.009	6,6 0.010	6,6 0.012	6,6 0.014

TABLE C-5. Values of *r*, *k*, and α for the Quantile test when α is approximately equal to 0.025

Number of Site Measurements, n

Number of Reference (Background) Measurements, m

	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
5	r, k α		9,9 0.030	12,12 0.024	15,15 0.021	17,17 0.026	20,20 0.024	22,22 0.028	25,25 0.025											
10		7,6 0.029	6,6 0.028	8,8 0.022	9,9 0.029	11,11 0.024	12,12 0.029	14,14 0.025	15,15 0.029	17,17 0.025	18,18 0.029	20,20 0.026	21,21 0.029	23,23 0.026	24,24 0.029	26,26 0.026	27,27 0.029			
15	11,5 0.030	6,5 0.023	5,5 0.021	6,6 0.024	7,7 0.026	8,8 0.027	9,9 0.028	10,10 0.029	11,11 0.030	13,13 0.022	14,14 0.023	15,15 0.023	16,16 0.024	17,17 0.025	18,18 0.025	19,19 0.026	21,21 0.021	21,21 0.027	22,22 0.027	23,23 0.027
20	8,4 0.023	3,3 0.030	4,4 0.026	5,5 0.024	6,6 0.022	7,7 0.020	12,11 0.021	13,12 0.024	9,9 0.028	10,10 0.026	11,11 0.024	12,12 0.023	13,13 0.022	13,13 0.029	14,14 0.027	15,15 0.026	16,16 0.025	17,17 0.024	17,17 0.029	18,18 0.028
25	2,2 0.023	8,5 0.027	6,5 0.021	7,6 0.023	5,5 0.025	6,6 0.020	10,9 0.026	7,7 0.027	8,8 0.023	13,12 0.027	9,9 0.027	10,10 0.024	11,11 0.022	11,11 0.028	12,12 0.025	13,13 0.023	13,13 0.028	14,14 0.025	15,15 0.023	15,15 0.028
30	6,3 0.026	6,4 0.026	9,6 0.026	4,4 0.021	7,6 0.029	5,5 0.026	9,8 0.024	6,6 0.029	7,7 0.023	12,11 0.021	8,8 0.025	9,9 0.021	9,9 0.027	10,10 0.023	10,10 0.029	11,11 0.025	11,11 0.030	12,12 0.026	13,13 0.023	13,13 0.027
35	7,3 0.030	4,3 0.030	3,3 0.023	6,5 0.020	4,4 0.026	10,8 0.022	5,5 0.027	9,8 0.024	6,6 0.027	7,7 0.020	7,7 0.027	8,8 0.021	8,8 0.027	9,9 0.022	9,9 0.027	10,10 0.022	10,10 0.027	11,11 0.022	11,11 0.027	12,12 0.023
40	3,2 0.029	4,3 0.022	8,5 0.028	11,7 0.025	6,5 0.028	4,4 0.030	10,8 0.026	5,5 0.027	9,8 0.023	6,6 0.026	10,9 0.028	7,7 0.024	12,11 0.020	8,8 0.023	8,8 0.029	9,9 0.022	9,9 0.027	10,10 0.021	10,10 0.026	11,11 0.021
45	3,2 0.023	8,4 0.029	6,4 0.030	3,3 0.026	8,6 0.021	4,4 0.023	7,6 0.025	5,5 0.020	5,5 0.028	9,8 0.023	6,6 0.024	10,9 0.026	7,7 0.022	7,7 0.027	8,8 0.020	8,8 0.025	8,8 0.030	9,9 0.023	9,9 0.027	10,10 0.021
50		2,2 0.025	6,4 0.022	3,3 0.021	11,7 0.027	6,5 0.026	4,4 0.026	7,6 0.028	5,5 0.021	5,5 0.028	9,8 0.022	6,6 0.023	6,6 0.029	7,7 0.020	7,7 0.025	12,11 0.020	8,8 0.022	8,8 0.026	13,12 0.027	9,9 0.023
55		2,2 0.022	4,3 0.029	8,5 0.028	3,3 0.028	8,6 0.021	4,4 0.020	4,4 0.029	10,8 0.021	5,5 0.022	5,5 0.028	9,8 0.022	6,6 0.023	6,6 0.028	10,9 0.029	7,7 0.023	7,7 0.027	12,11 0.023	8,8 0.023	8,8 0.027
60		14,5 0.022	4,3 0.024	8,5 0.021	3,3 0.023	11,7 0.029	6,5 0.024	4,4 0.023	7,6 0.023	10,8 0.024	5,5 0.023	5,5 0.029	9,8 0.022	6,6 0.022	6,6 0.027	10,9 0.027	7,7 0.021	7,7 0.025	7,7 0.030	8,8 0.021
65		6,3 0.028	7,4 0.021	6,4 0.025	10,6 0.025	3,3 0.029	8,6 0.021	6,5 0.029	4,4 0.026	7,6 0.026	10,8 0.026	5,5 0.023	5,5 0.029	9,8 0.022	6,6 0.021	6,6 0.026	10,9 0.026	7,7 0.020	7,7 0.024	7,7 0.028
70		6,3 0.024	2,2 0.029	6,4 0.021	8,5 0.028	3,3 0.025	13,8 0.026	6,5 0.023	4,4 0.022	4,4 0.028	7,6 0.028	10,8 0.027	5,5 0.024	5,5 0.029	9,8 0.022	6,6 0.021	6,6 0.025	6,6 0.029	10,9 0.030	7,7 0.022
75		11,4 0.022	2,2 0.026	4,3 0.028	8,5 0.022	3,3 0.022	9,6 0.028	8,6 0.021	6,5 0.027	4,4 0.024	7,6 0.023	7,6 0.030	10,8 0.029	5,5 0.024	5,5 0.029	9,8 0.021	6,6 0.021	6,6 0.024	6,6 0.028	10,9 0.028
80		7,3 0.028	2,2 0.024	4,3 0.024	6,4 0.028	10,6 0.024	3,3 0.027	13,8 0.027	6,5 0.023	4,4 0.020	4,4 0.026	7,6 0.024	10,8 0.023	5,5 0.020	5,5 0.025	5,5 0.029	9,8 0.021	6,6 0.020	6,6 0.024	6,6 0.027
85		3,2 0.029	2,2 0.021	4,3 0.021	6,4 0.023	8,5 0.028	3,3 0.023	9,6 0.030	8,6 0.020	6,5 0.026	4,4 0.022	4,4 0.028	7,6 0.026	10,8 0.024	5,5 0.021	5,5 0.025	5,5 0.029	9,8 0.021	6,6 0.020	6,6 0.023
90			5,3 0.020	11,5 0.027	9,5 0.023	8,5 0.023	3,3 0.021	3,3 0.028	13,8 0.028	6,5 0.022	6,5 0.029	4,4 0.024	4,4 0.029	7,6 0.028	10,8 0.026	5,5 0.022	5,5 0.025	5,5 0.030	9,8 0.021	9,8 0.025
95			10,4 0.029	2,2 0.029	4,3 0.028	6,4 0.029	10,6 0.023	3,3 0.025	11,7 0.026	8,6 0.020	6,5 0.025	4,4 0.021	4,4 0.026	7,6 0.024	7,6 0.029	10,8 0.027	5,5 0.022	5,5 0.026	5,5 0.030	9,8 0.021
100			6,3 0.029	2,2 0.027	4,3 0.025	6,4 0.025	8,5 0.028	3,3 0.022	3,3 0.029	13,8 0.028	6,5 0.022	6,5 0.028	4,4 0.023	4,4 0.027	7,6 0.025	10,8 0.022	10,8 0.028	5,5 0.022	5,5 0.026	5,5 0.030

TABLE C-6. Values of *r*, *k*, and α for the Quantile test when α is approximately equal to 0.05

Number of Reference (Background) Measurements, m	Number of Site Measurements, n																			
	5	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**55**	**60**	**65**	**70**	**75**	**80**	**85**	**90**	**95**	**100**
5	**r, k** **α**		8,8 0.051	10,10 0.057	13,13 0.043	15,15 0.048	17,17 0.051	19,19 0.054	21,21 0.056											
10		4,4 0.043	5,5 0.057	14,12 0.045	8,8 0.046	9,9 0.052	10,10 0.058	12,12 0.046	13,13 0.050	14,14 0.054	15,15 0.057	17,17 0.049	18,18 0.052	19,19 0.055	20,20 0.057	21,21 0.059	23,23 0.053			
15	2,2 0.053	3,3 0.052	4,4 0.050	5,5 0.048	6,6 0.046	7,7 0.045	8,8 0.044	9,9 0.043	9,9 0.060	10,10 0.057	11,11 0.055	12,12 0.054	13,13 0.052	14,14 0.051	15,15 0.050	16,16 0.049	16,16 0.058	17,17 0.057	18,18 0.056	19,19 0.055
20	9,4 0.040	8,5 0.056	6,5 0.040	4,4 0.053	5,5 0.043	9,8 0.052	6,6 0.056	7,7 0.048	8,8 0.043	8,8 0.057	9,9 0.051	10,10 0.046	10,10 0.057	11,11 0.052	12,12 0.048	12,12 0.057	13,13 0.053	14,14 0.049	14,14 0.057	15,15 0.054
25	6,3 0.041	6,4 0.043	3,3 0.046	6,5 0.052	4,4 0.055	5,5 0.041	5,5 0.059	6,6 0.046	11,10 0.042	7,7 0.050	8,8 0.042	8,8 0.053	9,9 0.045	9,9 0.055	10,10 0.048	11,11 0.042	11,11 0.050	11,11 0.058	12,12 0.052	12,12 0.060
30	3,2 0.047	2,2 0.058	10,6 0.052	3,3 0.058	11,8 0.045	4,4 0.056	8,7 0.045	5,5 0.054	6,6 0.040	6,6 0.053	7,7 0.041	7,7 0.052	8,8 0.042	8,8 0.051	9,9 0.042	9,9 0.050	9,9 0.059	10,10 0.049	10,10 0.057	11,11 0.049
35	8,3 0.046	2,2 0.045	6,4 0.058	3,3 0.043	6,5 0.041	4,4 0.040	4,4 0.057	8,7 0.043	5,5 0.051	9,8 0.052	6,6 0.047	6,6 0.058	7,7 0.043	7,7 0.053	8,8 0.041	8,8 0.049	8,8 0.057	9,9 0.046	9,9 0.053	10,10 0.044
40	4,2 0.055	5,3 0.048	4,3 0.057	10,6 0.059	3,3 0.053	6,5 0.048	4,4 0.043	4,4 0.058	8,7 0.042	5,5 0.048	9,8 0.047	6,6 0.042	6,6 0.051	11,10 0.042	7,7 0.045	7,7 0.053	8,8 0.041	8,8 0.048	8,8 0.055	9,9 0.043
45	4,2 0.045	9,4 0.047	2,2 0.059	8,5 0.052	3,3 0.042	8,6 0.041	6,5 0.054	4,4 0.045	4,4 0.058	8,7 0.041	5,5 0.046	5,5 0.057	9,8 0.056	6,6 0.047	6,6 0.055	11,10 0.046	7,7 0.047	7,7 0.054	8,8 0.041	8,8 0.047
50		6,3 0.052	2,2 0.050	6,4 0.051	12,7 0.050	3,3 0.049	8,6 0.049	6,5 0.059	4,4 0.047	4,4 0.059	8,7 0.041	5,5 0.045	5,5 0.054	9,8 0.051	6,6 0.043	6,6 0.050	6,6 0.058	7,7 0.042	7,7 0.048	7,7 0.054
55		3,2 0.059	2,2 0.043	4,3 0.056	8,5 0.058	3,3 0.041	5,4 0.041	6,5 0.046	9,7 0.042	4,4 0.048	4,4 0.059	8,7 0.040	5,5 0.043	5,5 0.052	9,8 0.048	6,6 0.040	6,6 0.047	6,6 0.054	11,10 0.043	7,7 0.043
60		3,2 0.052	5,3 0.052	4,3 0.046	6,4 0.059	3,3 0.035	3,3 0.047	8,6 0.043	6,5 0.051	9,7 0.046	4,4 0.049	4,4 0.059	13,10 0.052	5,5 0.042	5,5 0.050	5,5 0.058	9,8 0.054	6,6 0.044	6,6 0.050	6,6 0.056
65		3,2 0.045	5,3 0.043	2,2 0.053	6,4 0.048	10,6 0.050	3,3 0.040	3,3 0.053	6,5 0.041	6,5 0.055	4,4 0.042	4,4 0.050	4,4 0.060	13,10 0.052	5,5 0.041	5,5 0.048	5,5 0.055	9,8 0.051	6,6 0.041	6,6 0.047
70		8,3 0.057	9,4 0.048	2,2 0.047	4,3 0.055	8,5 0.050	5,4 0.041	3,3 0.046	3,3 0.057	6,5 0.045	6,5 0.058	4,4 0.043	4,4 0.051	4,4 0.060	13,10 0.051	5,5 0.041	5,5 0.047	5,5 0.054	9,8 0.048	9,8 0.057
75		8,3 0.049	6,3 0.056	2,2 0.043	4,3 0.047	6,4 0.054	10,6 0.053	3,3 0.040	3,3 0.051	8,6 0.044	6,5 0.049	9,7 0.041	4,4 0.044	4,4 0.052	5,5 0.060	13,10 0.051	8,7 0.047	5,5 0.046	5,5 0.052	5,5 0.058
80		4,2 0.059	6,3 0.048	5,3 0.053	2,2 0.055	6,4 0.046	8,5 0.055	5,4 0.042	3,3 0.045	3,3 0.055	6,5 0.041	6,5 0.052	9,7 0.043	4,4 0.045	4,4 0.053	7,6 0.058	13,10 0.051	8,7 0.046	5,5 0.045	5,5 0.051
85		4,2 0.054	3,2 0.058	5,3 0.047	2,2 0.050	4,3 0.054	4,3 0.048	10,6 0.056	5,4 0.049	3,3 0.049	3,3 0.059	6,5 0.044	6,5 0.055	9,7 0.046	4,4 0.046	4,4 0.053	7,6 0.059	10,8 0.060	8,7 0.045	5,5 0.044
90			3,2 0.053	5,3 0.041	2,2 0.046	6,4 0.059	6,4 0.051	8,5 0.058	5,4 0.042	3,3 0.044	3,3 0.053	8,6 0.045	6,5 0.047	6,5 0.058	4,4 0.041	4,4 0.047	4,4 0.054	7,6 0.059	10,8 0.060	8,7 0.045
95			3,2 0.048	9,4 0.048	2,2 0.042	2,2 0.056	4,3 0.059	8,5 0.050	10,6 0.058	5,4 0.048	3,3 0.048	3,3 0.056	6,5 0.041	6,5 0.050	9,7 0.040	4,4 0.042	4,4 0.048	4,4 0.054	7,6 0.059	10,8 0.059
100			3,2 0.044	6,3 0.057	5,3 0.054	2,2 0.052	4,3 0.053	6,4 0.056	10,6 0.049	5,4 0.043	3,3 0.043	3,3 0.051	3,3 0.059	6,5 0.044	6,5 0.053	9,7 0.042	4,4 0.043	4,4 0.049	4,4 0.055	7,6 0.059

TABLE C-7. Values of *r*, *k*, and α for the Quantile test when α is approximately equal to 0.10

Number of Site Measurements

Number of Reference (Background) Measurements, m

	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
5	**r, k** α		7,7 0.083	8,8 0.116	10,10 0.109	12,12 0.104	14,14 0.100	15,15 0.117	17,17 0.112											
10		3,3 0.105	4,4 0.108	5,5 0.109	6,6 0.109	7,7 0.109	8,8 0.109	9,9 0.109	10,10 0.109	11,11 0.109	12,12 0.109	13,13 0.109	14,14 0.109	15,15 0.109	16,16 0.109	17,17 0.109	18,18 0.109			
15	9,4 0.098	10,6 0.106	3,3 0.112	4,4 0.093	5,5 0.081	5,5 0.117	6,6 0.102	7,7 0.092	7,7 0.118	8,8 0.106	9,9 0.098	9,9 0.118	10,10 0.109	11,11 0.101	11,11 0.118	12,12 0.110	13,13 0.104	13,13 0.118	14,14 0.111	15,15 0.106
20	3,2 0.091	2,2 0.103	5,4 0.093	3,3 0.115	4,4 0.085	4,4 0.119	5,5 0.093	10,9 0.084	6,6 0.099	7,7 0.083	7,7 0.102	8,8 0.088	8,8 0.105	9,9 0.092	9,9 0.107	10,10 0.095	10,10 0.108	11,11 0.098	11,11 0.110	12,12 0.100
25	4,2 0.119	7,4 0.084	8,5 0.112	3,3 0.080	3,3 0.117	4,4 0.080	4,4 0.107	8,7 0.108	5,5 0.101	10,9 0.088	6,6 0.096	6,6 0.114	7,7 0.093	7,7 0.108	8,8 0.091	8,8 0.104	8,8 0.117	9,9 0.100	9,9 0.112	10,10 0.098
30	4,2 0.089	5,3 0.089	2,2 0.106	14,8 0.111	3,3 0.088	3,3 0.119	9,7 0.116	4,4 0.100	8,7 0.093	5,5 0.088	5,5 0.106	6,6 0.080	6,6 0.095	6,6 0.110	7,7 0.087	7,7 0.100	7,7 0.113	8,8 0.092	8,8 0.103	8,8 0.115
35	5,2 0.109	3,2 0.119	2,2 0.086	6,4 0.119	5,4 0.091	3,3 0.093	3,3 0.120	9,7 0.112	4,4 0.094	4,4 0.114	8,7 0.107	5,5 0.094	5,5 0.110	6,6 0.081	6,6 0.094	6,6 0.107	6,6 0.120	7,7 0.094	7,7 0.105	7,7 0.116
40	5,2 0.087	3,2 0.098	5,3 0.119	2,2 0.107	12,7 0.109	5,4 0.102	3,3 0.097	6,5 0.100	9,7 0.109	4,4 0.090	4,4 0.107	8,7 0.097	5,5 0.086	5,5 0.099	5,5 0.112	6,6 0.082	6,6 0.093	6,6 0.104	6,6 0.116	7,7 0.089
45	6,2 0.103	3,2 0.082	5,3 0.094	2,2 0.091	6,4 0.115	7,5 0.086	5,4 0.112	3,3 0.100	6,5 0.101	9,7 0.107	4,4 0.087	4,4 0.102	4,4 0.117	8,7 0.107	5,5 0.091	5,5 0.103	5,5 0.115	6,6 0.083	6,6 0.093	6,6 0.103
50		7,3 0.083	9,4 0.115	7,4 0.097	2,2 0.108	10,6 0.112	5,4 0.090	3,3 0.084	3,3 0.103	6,5 0.102	9,7 0.105	4,4 0.084	4,4 0.098	4,4 0.112	8,7 0.099	5,5 0.084	5,5 0.095	5,5 0.105	5,5 0.116	6,6 0.083
55		4,2 0.109	3,2 0.114	5,3 0.114	2,2 0.095	6,4 0.112	14,8 0.111	5,4 0.098	3,3 0.088	3,3 0.105	6,5 0.103	9,7 0.104	4,4 0.082	4,4 0.095	4,4 0.107	4,4 0.120	8,7 0.107	5,5 0.088	5,5 0.098	5,5 0.108
60		4,2 0.095	3,2 0.100	5,3 0.097	2,2 0.084	2,2 0.109	8,5 0.119	5,4 0.082	5,4 0.105	3,3 0.091	3,3 0.106	6,5 0.103	9,7 0.102	4,4 0.081	4,4 0.092	4,4 0.103	4,4 0.115	8,7 0.100	5,5 0.083	5,5 0.092
65		4,2 0.084	3,2 0.089	5,3 0.082	7,4 0.090	2,2 0.097	6,4 0.110	12,7 0.113	5,4 0.089	5,4 0.111	3,3 0.093	3,3 0.108	6,5 0.104	9,7 0.101	7,6 0.084	4,4 0.090	4,4 0.100	4,4 0.110	8,7 0.094	8,7 0.107
70		5,2 0.115	7,3 0.101	9,4 0.106	5,3 0.112	2,2 0.088	2,2 0.109	8,5 0.114	7,5 0.081	5,4 0.096	3,3 0.083	3,3 0.096	3,3 0.109	6,5 0.104	9,7 0.101	7,6 0.082	4,4 0.088	4,4 0.097	4,4 0.107	4,4 0.117
75		5,2 0.103	7,3 0.088	3,2 0.111	5,3 0.098	7,4 0.101	2,2 0.099	2,2 0.119	10,6 0.117	5,4 0.083	5,4 0.102	3,3 0.085	3,3 0.098	3,3 0.110	6,5 0.105	9,7 0.100	7,6 0.081	4,4 0.086	4,4 0.095	4,4 0.104
80		5,2 0.093	4,2 0.116	3,2 0.101	5,3 0.086	7,4 0.086	2,2 0.091	2,2 0.109	8,5 0.110	14,8 0.110	5,4 0.089	5,4 0.107	3,3 0.088	3,3 0.099	3,3 0.111	6,5 0.105	6,5 0.120	9,7 0.116	4,4 0.084	4,4 0.093
85		4,2 0.106	4,2 0.106	3,2 0.092	9,4 0.117	5,3 0.111	2,2 0.083	2,2 0.101	2,2 0.118	10,6 0.112	7,5 0.084	5,4 0.094	5,4 0.111	3,3 0.090	3,3 0.101	3,3 0.112	6,5 0.105	6,5 0.119	9,7 0.114	4,4 0.083
90			4,2 0.097	3,2 0.085	3,2 0.119	5,3 0.099	7,4 0.095	2,2 0.093	2,2 0.109	8,5 0.108	12,7 0.114	5,4 0.083	5,4 0.099	3,3 0.082	3,3 0.092	3,3 0.102	3,3 0.112	6,5 0.105	6,5 0.119	9,7 0.113
95			4,2 0.089	7,3 0.100	3,2 0.110	5,3 0.089	7,4 0.084	2,2 0.086	2,2 0.102	2,2 0.117	10,6 0.108	14,8 0.117	5,4 0.088	5,4 0.103	3,3 0.084	3,3 0.094	3,3 0.103	3,3 0.113	6,5 0.106	6,5 0.118
100			4,2 0.082	7,3 0.090	3,2 0.102	5,3 0.080	5,3 0.109	2,2 0.080	2,2 0.095	2,2 0.110	6,4 0.118	12,7 0.109	7,5 0.086	5,4 0.093	5,4 0.108	3,3 0.086	3,3 0.095	3,3 0.104	3,3 0.114	6,5 0.106

TABLE C-8. Critical values, w_{α}, for the WRS test (n = the number of site measurements; m = the number of background measurements)

n	α	m																		
		2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
2	0.05	0	0	0	1	1	1	2	2	2	2	3	3	4	4	4	4	5	5	5
	0.10	0	1	1	2	2	2	3	3	4	4	5	5	5	6	6	7	7	8	8
3	0.05	0	1	1	2	3	3	4	5	5	6	6	7	8	8	9	10	10	11	12
	0.10	1	2	2	3	4	5	6	6	7	8 1	9	10	11	11	12	13	14	15	16
4	0.05	0	1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	18	19
	0.10	1	2	4	5	6	7	8	10	11	12	13	14	16	17	18	19	21	22	23
5	0.05	1	2	3	5	6	7	9	10	12	13	14	16	17	19	20	21	23	24	26
	0.10	2	3	5	6	8	9	11	13	14	16	18	19	21	23	24	26	28	29	31
6	0.05	1	3	4	6	8	9	11	13	15	17	18	20	22	24	26	27	29	31	33
	0.10	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	35	37	39
7	0.05	1	3	5	7	9	12	14	16	18	20	22	25	27	29	31	34	36	38	40
	0.10	2	5	7	9	12	14	17	19	22	24	27	29	32	34	37	39	42	44	47
8	0.05	2	4	6	9	11	14	16	19	21	24	27	29	32	34	37	40	42	45	48
	0.10	3	6	8	11	14	17	20	23	25	28	31	34	37	40	43	46	49	52	55
9	0.05	2	5	7	10	13	16	19	22	25	28	31	34	37	40	43	46	49	52	55
	0.10	3	6	10	13	16	19	23	26	29	32	36	39	42	46	49	53	56	59	63
10	0.05	2	5	8	12	15	18	21	25	28	32	35	38	42	45	49	52	56	59	63
	0.10	4	7	11	14	18	22	25	29	33	37	40	44	48	52	55	59	63	67	71
11	0.05	2	6	9	13	17	20	24	28	32	35	39	43	47	51	55	58	62	66	70
	0.10	4	8	12	16	20	24	28	32	37	41	45	49	53	58	62	66	70	74	79
12	0.05	3	6	10	14	18	22	27	31	35	39	43	48	52	56	61	65	69	73	78
	0.10	5	9	13	18	22	27	31	36	40	45	50	54	59	64	68	73	78	82	87
13	0.05	3	7	11	16	20	25	29	34	38	43	48	52	57	62	66	71	76	81	8
	0.10	5	10	14	19	24	29	34	39	44	49	54	59	64	69	75	80	85	90	95
14	0.05	4	8	12	17	22	27	32	37	42	47	52	57	62	67	72	78	83	88	93
	0.10	5	11	16	21	26	32	37	42	48	53	59	64	70	75	81	86	92	98	103
15	0.05	4	8	13	19	24	29	34	40	45	51	56	62	67	73	78	84	89	95	101
	0.10	6	11	17	23	28	34	40	46	52	58	64	69	75	81	87	93	99	105	111
16	0.05	4	9	15	20	26	31	37	43	49	55	61	66	72	78	84	90	96	102	108
	0.10	6	12	18	24	30	37	43	49	55	62	68	75	81	87	94	100	107	113	120
17	0.05	4	10	16	21	27	34	40	46	52	58	65	71	78	84	90	97	103	110	116
	0.10	7	13	19	26	32	39	46	53	59	66	73	80	86	93	100	107	114	121	128
18	0.05	5	10	17	23	29	36	42	49	56	62	69	76	83	89	96	103	110	117	124
	0.10	7	14	21	28	35	42	49	56	63	70	78	85	92	99	107	114	121	129	136
19	0.05	5	11	18	24	31	38	45	52	59	66	73	81	88	95	102	110	117	124	131
	0.10	8	15	22	29	37	44	52	59	67	74	82	90	98	105	113	121	129	136	144
20	0.05	5	12	19	26	33	40	48	55	63	70	78	85	93	101	108	116	124	131	139
	0.10	8	16	23	31	39	47	55	63	71	79	87	95	103	111	120	128	136	144	152

TABLE C-9. Critical values for the two-sample t test

Degrees of Freedom	**1 − α**								
	.70	**.75**	**.80**	**.85**	**.90**	**.95**	**.975**	**.99**	**.995**
1	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.536	0.691	0.866	1.074	1.34	1.753	2.131	2.602	2.947
16	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.531	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.531	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704
60	0.527	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660
120	0.526	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617
∞	0.524	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576

TABLE C-10. Coefficients a_k for the Shapiro-Wilk W test for normality

k\n	2	3	4	5	6	7	8	9	10
1	0.7071	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.5868	0.5739
2	—	0.0000	0.1677	0.2413	0.28D6	0.3031	0.3164	0.3244	0.3291
3	—	—	—	0.0000	0.0875	0.1401	0.1743	0.1976	0.2141
4	—	—	—	—	—	0.0000	0.0561	0.0947	0.1224
5	—	—	—	—	—	—	—	0.0000	0.0399

k\n	11	12	13	14	15	16	17	18	19	20
1	0.5601	0.5475	0.5359	0.5251	0.5150	0.5056	0.4968	0.4886	0.4808	0.4734
2	0.3315	0.3325	0.3325	0.3318	0.3306	0.3290	0.3273	0.3253	0.3232	0.3211
3	0.2260	0.2347	0.2412	0.2460	0.2495	0.2521	0,2540	0.2553	0.2561	0.2565
4	0.1429	0.1506	0.1707	0.1802	0.1876	0.1939	0.1988	0.2027	0.2059	0.2085
5	0.0695	0.0922	0.1099	0.1240	0.1353	0.1447	0.1524	0.1587	0.1641	0.1686
6	0.0000	0.0303	0 0539	0.0727	0.0880	0.1005	0.1109	0.1197	0.1271	0.1334
7	—	—	0.0000	0.0240	0.0433	0.0593	0.0725	0.0837	0.0932	0.1013
8	—	—	—	—	0.0000	0.0196	0.0359	0.0496	0.0612	0.0711
9	—	—	—	—	—	—		0.0163	0.0303	0.0422
10	—	—	—	—	—	—	—	—	0.0000	0.0140

k\n	21	22	23	24	25	26	27	28	29	30
1	0.4643	0.4590	0.4542	0.4493	0.4450	0.4407	0.4366	0.4328	0.4291	0.4254
2	0.3185	0.3156	0.3126	0.3098	0.3069	0.3043	0.3018	0.2992	0.2968	0.2944
3	0.2578	0.2571	0.2563	0.2554	0.2543	0.2533	0.2522	0.2510	0.2499	0.2487
4	0.2119	0.2131	0.2139	0.2145	0.2148	0.2151	0.2152	0.2151	0.2150	0.2148
5	0.1736	0.1764	0.1787	0.1007	0.1822	0.1836	0.1840	0.1857	0.1864	0.1870
6	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563	0.1584	0.1601	0.1616	0.1630
7	0.1092	0.1150	0.1201	0.1245	0.1263	0.1316	0.1346	0.1372	0.1395	0.1415
8	0.0804	0.0878	0.0941	0.0997	0.1046	0.1089	0.1128	0.1162	0.1192	0.1219
9	0.0530	0.0618	0.0696	0.0764	0.0823	0.0876	0.0923	0.0965	0.1002	0.1036
10	0.0263	0.0368	0.0459	0.0539	0.0610	0.0672	0.0728	0.0778	0.0822	0.0862
11	0.0000	0.0122	0.0228	0.0321	0.0403	0.0476	0.0540	0.0598	0.0650	0.0697
12	—	—	0.0000	0.0107	0.0200	0.0284	0.0358	0.0424	0.0483	0.0537
13	—	—	—	—	0.0000	0.0094	0.0178	0.0253	0.0320	0.0381
14	—	—	—	—	—	—	0.0000	0.0084	0.0159	0.0227
15	—	—	—	—	—	—	—	—	0.0000	0.0076

TABLE C-11. Critical values for the Shapiro-Wilk W test for normality

n	$W_{0.01}$	$W_{0.02}$	$W_{0.05}$	$W_{0.10}$	$W_{0.50}$
3	0.753	0.756	0.767	0.789	0.859
4	0.687	0.707	0.748	0.792	0.935
5	0.686	0.715	0.762	0.806	0.927
6	0.713	0.743	0.788	0.826	0.927
7	0.730	0.760	0.803	0.838	0.928
8	0.749	0.778	0.818	0.851	0.932
9	0.764	0.791	0.829	0.859	0.935
10	0.781	0.806	0.842	0.869	0.938
11	0.792	0.817	0.850	0.876	0.940
12	0.805	0.828	0.859	0.883	0.943
13	0.814	0.837	0.866	0.889	0.945
14	0.825	0.846	0.874	0.895	0.947
15	0.835	0.855	0.881	0.901	0.950
16	0.844	0.863	0.887	0.906	0.952
17	0.851	0.869	0.892	0.910	0.954
18	0.858	0.874	0.897	0.914	0.956
19	0.863	0.879	0.901	0.917	0.957
20	0.868	0.886	0.905	0.920	0.969
21	0.873	0.884	0.908	0.923	0.960
22	0.878	0.892	0.911	0.926	0.961
23	0.881	0.895	0.914	0.928	0.962
24	0.884	0.898	0.916	0.930	0.963
25	0.886	0.901	0.918	0.931	0.964
26	0.891	0.904	0.920	0.933	0.965
27	0.894	0.906	0.923	0.935	0.965
28	0.896	0.908	0.924	0.936	0.966
29	0.898	0.910	0.926	0.937	0.966
30	0.900	0.912	0.927	0.939	0.967
31	0.902	0.914	0.929	0.940	0.967
32	0.904	0.915	0.930	0.941	0.968
33	0.906	0.917	0.931	0.942	0.968
34	0.908	0.919	0.933	0..943	0.969
35	0.910	0.920	0.934	0.944	0.969
36	0.912	0.922	0.935	0.945	0.970
37	0.914	0.924	0.936	0.946	0.970
38	0.916	0.925	0.938	0.947	0.971
39	0.917	0.927	0.939	0.948	0.971
40	0.919	0.928	0.940	0.949	0.972
41	0.920	0.929	0.941	0.950	0.972
42	0.922	0.930	0.942	0.951	0.972
43	0.923	0.932	0.943	0.951	0.973
44	0.924	0.933	0.944	0.952	0.973
45	0.926	0.934	0.945	0.953	0.973
46	0.927	0.935	0.945	0.953	0.974
47	0.928	0.936	0.946	0.954	0.974
48	0.929	0.937	0.947	0.954	0.974
49	0.929	0.937	0.947	0.955	0.974
50	0.930	0.938	0.947	0.955	0.974

TABLE C-12. Critical values for the D’Agostino test for normality (values of Y such that 100p% of the distribution of Y is less than Y_p)

n	$Y_{0.005}$	$Y_{0.01}$	$Y_{0.025}$	$Y_{0.05}$	$Y_{0.10}$	$Y_{0.90}$	$Y_{0.95}$	$Y_{0.975}$	$Y_{0.99}$	$Y_{0.995}$
50	–3.949	–3.442	–2.757	–2.220	–1.661	0.759	0.923	1.038	1.140	1.192
60	–3.846	–3.360	–2.699	–2.179	–1.634	0.807	0.986	1.115	1.236	1.301
70	–3.762	–3.293	–2.652	–2.146	–1.612	0.844	1.036	1.176	1.312	1.388
80	–3.693	–3.237	–2.613	–2.118	–1.594	0.874	1.076	1.226	1.374	1.459
90	–3.635	–3.100	–2.580	–2.095	–1.579	0.899	1.109	1.268	1.426	1.518
100	–3.584	–3.150	–2.552	–2.075	–1.566	0.920	1.137	1.303	1.470	1.569
150	–3.409	–3.009	–2.452	–2.004	–1.520	0.990	1.233	1.423	1.623	1.746
200	–3.302	–2.922	–2.391	–1.960	–1.491	1.032	1.290	1.496	1.715	1.853
250	–3.227	–2.861	–2.348	–1.926	–1.471	1.060	1.328	1.545	1.779	1.927
300	–3.172	–2.816	–2.316	–1.906	–1.456	1.080	1.357	1.528	1.826	1.983
350	–3.129	–2.781	–2.291	–1.888	–1.444	1.096	1.379	1.610	1.863	2.026
400	–3.094	–2.753	–2.270	–1.873	–1.434	1.108	1.396	1.633	1.893	2.061
450	–3.064	–2.729	–2.253	–1.861	–1.426	1.119	1.411	1.652	1.918	2.090
500	–3.040	–2.709	–2.239	–1.850	–1.419	1.127	1.423	1.668	1.938	2.114
550	–3.019	–2.691	–2.226	–1.841	–1.413	1.135	1.434	1.682	1.957	2.136
600	–3.000	–2.676	–2.215	–1.833	–1.408	1.141	1.443	1.694	1.972	2.154
650	–2.984	–2.663	–2.206	–1.826	–1.403	1.147	1.451	1.704	1.986	2.171
700	–2.969	–2.651	–2.197	–1.820	–1.399	1.152	1.458	1.714	1.999	2.185
750	–2.956	–2.640	–2.189	–1.814	–1.395	1.157	1.465	1.722	2.010	2.199
800	–2.944	–2.630	–2.182	–1.809	–1.392	1.161	1.471	1.730	2.020	2.211
850	–2.933	–2.621	–2.176	–1.804	–1.389	1.165	1.476	1.737	2.029	2.221
900	–2.923	–2.613	–2.170	–1.800	–1.386	1.168	1.481	1.743	2.037	2.231
950	–2.914	–2.605	–2.164	–1.796	–1.383	1.171	1.485	1.749	2.045	2.241
1000	–2.906	–2.599	–2.159	–1.792	–1.381	1.174	1.489	1.754	2.052	2.249

TABLE C-13. Values of the parameter λ for the Cohen estimates of the mean and variance of normally distributed datasets that contain nondetects

	h											
γ	**.01**	**.02**	**.03**	**.04**	**.05**	**.06**	**.07**	**.08**	**.09**	**.10**	**.15**	**.20**
00	.010100	.020400	.030902	.041583	.052507	.063625	.074953	.08649	.09824	.11020	.17342	.24268
05	.010551	.021294	.032225	.043350	.054670	.066159	.077909	.08983	.10197	.11431	.17925	.25033
10	.010950	.022082	.033398	.044902	.056596	.068483	.080563	.09285	.10534	.11804	.18479	.25741
15	.011310	.022798	.034466	.046318	.058356	.070586	.083009	.09563	.10845	.12148	.18985	.26405
20	.011642	.023459	.035453	.047829	.059990	.072539	.085280	.09822	.11135	.12469	.19460	.27031
25	.011952	.024076	.036377	.048858	.061522	.074372	.087413	.10065	.11408	.12772	.19910	.2762
30	.012243	.024658	.037249	.050018	.062969	.076106	.089433	.10295	.11667	.13059	.20338	.2819
35	.012520	.025211	.038077	.051120	.064345	.077736	.091355	.10515	.11914	.13333	.20747	.2873
40	.012784	.025738	.038866	.052173	.065660	.079332	.093193	.10725	.12150	.13595	.21129	.2925
45	.013036	.026243	.039624	.053182	.066921	.080845	.094958	.10926	.12377	.13847	.21517	.2976
50	.013279	.026728	.040352	.054153	.068135	.082301	.096657	.11121	.12595	.14090	.21882	.3025
55	.013513	.027196	.041054	.055089	.069306	.083708	.098298	.11208	.12806	.14325	.22225	.3072
60	.013739	.027849	.041733	.055995	.070439	.085068	.099887	.11490	.13011	.14552	.22578	.3118
65	.013958	.028087	.042391	.056874	.071538	.086388	.10143	.11666	.13209	.14773	.22910	.3163
70	.014171	.028513	.043030	.057726	.072505	.087670	.10292	.11837	.13402	.14987	.23234	.3206
75	.014378	.029927	.043652	.058556	.073643	.088917	.10438	.12004	.13590	.15196	.23550	.32489
80	.014579	.029330	.044258	.059364	.074655	.090133	.10580	.12167	.13775	.15400	.23858	.32903
85	.014773	.029723	.044848	.060153	.075642	.091319	.10719	.12225	.13952	.15599	.24158	.33307
90	.014967	.030107	.045425	.060923	.075606	.092477	.10854	.12480	.14126	.15793	.24452	.33703
95	.015154	.030483	.045989	.061676	.077549	.093611	.10987	.12632	.14297	.15983	.24740	.34091
1.00	.015338	.030850	.046540	.062413	.078471	.094720	.11116	.12780	.14465	.16170	.25022	.34471

						h						
γ	**.25**	**.30**	**.35**	**.40**	**.45**	**.50**	**.55**	**.60**	**.65**	**.70**	**.80**	**.90**
.00	.31862	.4021	.4941	.5961	.7096	.8388	.9808	1.145	1.336	1.561	2.176	3.283
.05	.32793	.4130	.5066	.6101	.7252	.8540	.9994	1.166	1.358	1.585	2.203	3.314
.10	.33662	.4233	.5184	.6234	.7400	.8703	1.017	1.185	1.379	1.608	2.229	3.345
.15	.34480	.4330	.5296	.6361	.7542	.8860	1.035	1.204	1.400	1.630	2.255	3.376
.20	.35255	.4422	.5403	.6483	.7673	.9012	1.051	1.222	1.419	1.651	2.280	3.405
.25	.35993	.4510	.5506	.6600	.7810	.9158	1.067	1.240	1.439	1.672	2.305	3.435
.30	.36700	.4595	.5604	.6713	.7937	.9300	1.083	1.257	1.457	1.693	2.329	3.464
.35	.37379	.4676	.5699	.6821	.8060	.9437	1.098	1.274	1.475	1.713	2.353	3.492
.40	.38033	.4735	.5791	.6927	.8179	.9570	1.113	1.290	1.494	1.732	2.376	3.520
.45	.38665	.4831	.5880	.7029	.8295	.9700	1.127	1.306	1.511	1.751	2.399	3.547
.50	.39276	.4904	.5967	.7129	.8408	.9826	1.141	1.321	1.528	1.770	2.421	3.575
.55	.39679	.4976	.6061	.7225	.8517	.9950	1.155	1.337	1.545	1.788	2.443	3.601
.60	.40447	.5045	.6133	.7320	.8625	1.007	1.169	1.351	1.561	1.806	2.465	3.628
.65	.41008	.5114	.6213	.7412	.8729	1.019	1.182	1.368	1.577	1.824	2.486	3.654
.70	.41555	.5180	.6291	.7502	.8832	1.030	1.195	1.380	1.593	1.841	2.507	3.679
.75	.42090	.5245	.6367	.7590	.8932	1.042	1.207	1.394	1.608	1.851	2.528	3.705
.80	.42612	.5308	.6441	.7676	.9031	1.053	1.220	1.408	1.624	1.875	2.548	3.730
.85	.43122	.5370	.6515	.7781	.9127	1.064	1.232	1.422	1.639	1.892	2.568	3.754
.90	.43622	.5430	.6586	.7844	.9222	1.074	1.244	1.435	1.653	1.908	2.588	3.779
.95	.44112	.5490	.6656	.7925	.9314	1.085	1.255	1.448	1.668	1.924	2.607	3.803
1.00	.44592	.5548	.6724	.8005	.9406	1.095	1.287	1.461	1.882	1.940	2.626	3.827

TABLE C-14. Critical values for the Dixon extreme value test for outliers

	Level of Significance α		
n	**0.10**	**0.05**	**0.01**
3	0.886	0.941	0.988
4	0.679	0.765	0.889
5	0.557	0.642	0.780
6	0.482	0.560	0.698
7	0.434	0.507	0.637
8	0.479	0.554	0.683
9	0.441	0.512	0.635
10	0.409	0.477	0.597
11	0.517	0.576	0.679
12	0.490	0.546	0.642
13	0.467	0.521	0.615
14	0.492	0.546	0.641
15	0.472	0.525	0.616
16	0.454	0.507	0.595
17	0.438	0.490	0.577
18	0.424	0.475	0.561
19	0.412	0.462	0.547
20	0.401	0.450	0.535
21	0.391	0.440	0.524
22	0.382	0.430	0.514
23	0.374	0.421	0.505
24	0.367	0.413	0.497
25	0.360	0.406	0.489

TABLE C-15. Critical values for the Discordance test for outliers

	Level of Significance	
n	**0.01**	**0.05**
3	1.155	1.153
4	1.492	1.463
5	1.749	1.672
6	1.944	1.822
7	2.097	1.938
8	2.221	2.032
9	2.323	2.110
10	2.410	2.176
11	2.485	2.234
12	2.550	2.285
13	2.607	2.331
14	2.659	2.371
15	2.705	2.409
16	2.747	2.443
17	2.785	2.475
18	2.821	2.504
19	2.854	2.532
20	2.884	2.557
21	2.912	2.580
22	2.939	2.603
23	2.963	2.624
24	2.987	2.644
25	3.009	2.663
26	3.029	2.681
27	3.049	2.698
28	3.068	2.714
29	3.085	2.730
30	3.103	2.745
31	3.119	2.759
32	3.135	2.773
33	3.150	2.786
34	3.164	2.799
35	3.178	2.811
36	3.191	2.823
37	3.204	2.835
38	3.216	2.846
39	3.228	2.857
40	3.240	2.866
41	3.251	2.877
42	3.261	2.887
43	3.271	2.896
44	3.282	2.905
45	3.292	2.914
46	3.302	2.923
47	3.310	2.931
48	3.319	2.940
49	3.329	2.948
50	3.336	2.956

TABLE C-16. Approximate critical values for the Rosner test for outliers

n	r	α	
		0.05	**0.01**
25	1	2.82	3.14
	2	2.80	3.11
	3	2.78	3.09
	4	2.76	3.06
	5	2.73	3.03
	10	2.59	2.85
26	1	2.84	3.16
	2	2.82	3.14
	3	2.80	3.11
	4	2.78	3.09
	5	2.76	3.06
	10	2.62	2.89
27	1	2.86	3.18
	2	2.84	3.16
	3	2.82	3.14
	4	2.80	3.11
	5	2.78	3.09
	10	2.65	2.93
28	1	2.88	3.20
	2	2.86	3.18
	3	2.84	3.16
	4	2.82	3.14
	5	2.80	3.11
	10	2.68	2.97
29	1	2.89	3.22
	2	2.88	3.20
	3	2.86	3.18
	4	2.84	3.16
	5	2.82	3.14
	10	2.71	3.00
30	1	2.91	3.24
	2	2.89	3.22
	3	2.88	3.20
	4	2.86	3.18
	5	2.84	3.16
	10	2.73	3.03
31	1	2.92	3.25
	2	2.91	3.24
	3	2.89	3.22
	4	2.88	3.20
	5	2.86	3.18
	10	2.76	3.06
46	1	3.09	3.45
	2	3.09	3.44
	3	3.08	3.43
	4	3.07	3.41
	5	3.06	3.40
	10	3.00	3.34
32	1	2.94	3.27
	2	2.92	3.25
	3	2.91	3.24
	4	2.89	3.22
	5	2.88	3.20
	10	2.78	3.09
33	1	2.95	3.29
	2	2.94	3.27
	3	2.92	3.25
	4	2.91	3.24
	5	2.89	3.22
	10	2.80	3.11
34	1	2.97	3.30
	2	2.95	3.29
	3	2.94	3.27
	4	2.92	3.25
	5	2.91	3.24
	10	2.82	3.14
35	1	2.98	3.32
	2	2.97	3.30
	3	2.95	3.29
	4	2.94	3.27
	5	2.92	3.25
	10	2.84	3.16
36	1	2.99	3.33
	2	2.98	3.32
	3	2.97	3.30
	4	2.95	3.29
	5	2.94	3.27
	10	2.86	3.18
37	1	3.00	3.34
	2	2.99	3.33
	3	2.98	3.32
	4	2.97	3.30
	5	2.95	3.29
	10	2.88	3.20
38	1	3.01	3.36
	2	3.00	3.34
	3	2.99	3.33
	4	2.98	3.32
	5	2.97	3.30
	10	2.91	3.22
70	1	3.26	3.62
	2	3.25	3.62
	3	3.25	3.61
	4	3.24	3.60
	5	3.24	3.60
	10	3.21	3.57
39	1	3.03	3.37
	2	3.01	3.36
	3	3.00	3.34
	4	2.99	3.33
	5	2.98	3.32
	10	2.91	3.24
40	1	3.04	3.38
	2	3.03	3.37
	3	3.01	3.36
	4	3.00	3.34
	5	2.99	3.33
	10	2.92	3.25
41	1	3.05	3.39
	2	3.04	3.38
	3	3.03	3.37
	4	3.01	3.36
	5	3.00	3.34
	10	2.94	3.27
42	1	3.06	3.40
	2	3.05	3.39
	3	3.04	3.38
	4	3.03	3.37
	5	3.01	3.36
	10	2.95	3.29
43	1	3.07	3.41
	2	3.06	3.40
	3	3.05	3.39
	4	3.04	3.38
	5	3.03	3.37
	10	2.97	3.30
44	1	3.08	3.43
	2	3.07	3.41
	3	3.06	3.40
	4	3.05	3.39
	5	3.04	3.38
	10	2.98	3.32
45	1	3.09	3.44
	2	3.08	3.43
	3	3.07	3.41
	4	3.06	3.40
	5	3.05	3.39
	10	2.99	3.33
250	1	3.67	4.04
	5	3.67	4.04
	10	3.66	4.03

TABLE C-16. Approximate critical values for the Rosner test for outliers (continued)

		α	
n	**r**	**0.05**	**0.01**
47	1	3.10	3.46
	2	3.09	3.45
	3	3.09	3.44
	4	3.08	3.43
	5	3.07	3.41
	10	3.01	3.36
48	1	3.11	3.46
	2	3.10	3.46
	3	3.09	3.45
	4	3.09	3.44
	5	3.08	3.43
	10	3.03	3.37
49	1	3.12	3.47
	2	3.11	3.46
	3	3.10	3.46
	4	3.09	3.45
	5	3.09	3.44
	10	3.04	3.38
50	1	3.13	3.48
	2	3.12	3.47
	3	3.11	3.46
	4	3.10	3.46
	5	3.09	3.45
	10	3.05	3.39
60	1	3.20	3.56
	2	3.19	3.55
	3	3.19	3.55
	4	3.18	3.54
	5	3.17	3.53
	10	3.14	3.49

		α	
n	**r**	**0.05**	**0.01**
80	1	3.31	3.67
	2	3.30	3.67
	3	3.30	3.66
	4	3.29	3.66
	5	3.29	3.65
	10	3.26	3.63
90	1	3.35	3.72
	2	3.34	3.71
	3	3.34	3.71
	4	3.34	3.70
	5	3.33	3.70
	10	3.31	3.68
100	1	3.38	3.75
	2	3.38	3.75
	3	3.38	3.75
	4	3.37	3.74
	5	3.37	3.74
	10	3.35	3.72
150	1	3.52	3.89
	2	3.51	3.89
	3	3.51	3.89
	4	3.51	3.88
	5	3.51	3.88
	10	3.50	3.87
200	1	3.61	3.98
	2	3.60	3.98
	3	3.60	3.97
	4	3.60	3.97
	5	3.60	3.97
	10	3.59	3.96

		α	
n	**r**	**0.05**	**0.01**
300	1	3.72	4.09
	5	3.72	4.09
	10	3.71	4.09
350	1	3.77	4.14
	5	3.76	4.13
	10	3.76	4.13
400	1	3.80	4.17
	5	3.80	4.17
	10	3.80	4.16
450	1	3.84	4.20
	5	3.83	4.20
	10	3.83	4.20
500	1	3.86	4.23
	5	3.86	4.23
	10	3.86	4.22

APPENDIX D: FURTHER READINGS

D.1 Printed Matter

American Society for Testing and Materials (ASTM). 1994. D5549-94e1. West Conshohocken, PA.

American Society for Testing and Materials (ASTM). 1994. *Standard Guide for Selection of Simulation Approaches in Geostatistical Site Investigations.* D5924-96e1. West Conshohocken, PA.

American Society for Testing and Materials (ASTM). 1996. *Standard Guide for Analysis of Spatial Variation in Geostatistical Site Investigations.* D5922-96e1. West Conshohocken, PA.

American Society for Testing and Materials (ASTM). 1996. *Standard Guide for Selection of Kriging Methods in Geostatistical Site Investigations.* D5923-96e1. West Conshohocken, PA.

Connor, J.J., H.T. Shacklette, R.J. Ebens, J.A. Erdman, A.T. Miesch, R.R. Tidball, and H.A. Tourtelot. 1975. *Background Geochemistry of Some Rocks, Soils, Plants, and Vegetables in the Conterminous United States.* U.S. Geological Survey Professional Paper 574-F. Washington, DC.

California Environmental Protection Agency (Cal/EPA), Department of Toxic Substances Control (DTSC). 1992. *Supplemental Guidance for Human Health Multimedia Risk Assessments of Hazardous Waste Sites and Permitted Facilities.* State of California, Office of the Science Advisor. July.

California Environmental Protection Agency (Cal/EPA), Department of Toxic Substances Control (DTSC). 1994. *Preliminary Endangerment Assessment Guidance Manual.* State of California. January.

California Environmental Protection Agency (Cal/EPA), Department of Toxic Substances Control (DTSC). 1996. *Determining Ambient Concentrations of Metals for Risk Assessments at Hazardous Waste Sites and Permitted Facilities.* Interim final policy. State of California. June.

Davidson, J.R. Jr., N.L. Hassig, J.E. Wilson, and R.O. Gilbert. 2001. *Visual Sample Plan Version 1.0 User's Guide.* PNNL-13490. Richland, WA: Pacific Northwest National Laboratory. March.

Dixon, W.J. 1953. Processing Data for Outliers. *Biometrics* 9:74–89.

Entz, T., and C. Chang. 1991. Evaluation of Soil Sampling Schemes for Geostatistical Analyses: A Case Study for Soil Bulk Density. *Canadian Journal of Soil Science* 71:165–176.

Fleishhauer, H.L., and N. Korte. 1990. Formulation of Cleanup Standards for Trace Elements with Probability Plots. *Environmental Management* 14(1): 95–105.

Gehan, E.A. 1965. A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-Censored Samples. *Biometrika* 52:203–223.

Gilbert, R.O., J.R. Davidson, Jr., J.E. Wilson, and B.A. Pulsipher. 2001. *Visual Sample Plan (VSP) Models and Code Verification.* PNNL-13450. Richland, Wash.: Pacific Northwest National Laboratory.

Gogolak, C.V., G.E. Powers, and A.M. Huffert. 1997. *A Nonparametric Statistical Methodology for the Design and Analysis of Final Status Decommissioning Surveys.* NUREG-1505. Washington, DC: U.S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, Division of Regulatory Applications.

Hamblin, W.K. 1985. *The Earth's Dynamic Systems.* 4th ed. Minneapolis: Burgess.

Lepeltier, C. 1969. A Simplified Statistical Treatment of Geochemical Data by Graphical Representation. *Economic Geology* 64: 538–550.

Moore, D.S., and G.P. McCabe. 1999. *Introduction to the Practice of Statistics.* 3rd ed. New York: W.H. Freeman.

Reyment, R.A., and K.G. Jöreskog. 1993. *Applied Factor Analysis in the Natural Sciences.* Cambridge, England: Cambridge University Press.

Rosenbaum, S. 1954. Tables for a Nonparametric Test of Location. *Annals of Mathematical Statistics* 25:146–150.

Rosner, B. 1983. Percentage Points for a Generalized ESD Many-Outlier Procedure. *Technometrics* 25:165–172.

Shapiro, S. and M.B. Wilk. 1965. An Analysis of Variance Test for Normality. *Biometrika* 52:591–611.

Sinclair, A.J. 1974. Selection of Threshold Values in Geochemical Data Using Probability Graphs. *J. Geochem. Explor.* 3:129–149.

Thompson, S.K., and G.A. Seber. 1996. *Adaptive Sampling*. New York: Wiley.

United States Air Force (USAF). 1996. *Operable Unit 3 Remedial Investigation Report, Andersen Air Force Base, Guam.* Final. USAF-204-R. USAF Installation Restoration Program. December. Revised March 1997.

United States Environmental Protection Agency (U.S. EPA). 1985. *National Oil and Hazardous Substances Pollution Contingency Plan. Final Rule.* 50 Federal Register 47912. November.

United States Environmental Protection Agency (U.S. EPA). 1989. *Statistical Methods for Evaluating the Attainment of Cleanup Standards, Vol. 1: Soils and Solid Media.* EPA 230/02-89-042. Washington, DC: Office of Policy, Planning, and Evaluation.

Walsh, J.E. 1958. Large Sample Nonparametric Rejection of Outlying Observations. *Annals of the Institute of Statistical Mathematics* 10:223–232.

Wen, X., E.H. DeCarlo, and Y.H. Li. 1997. Interelement Relationships in Ferromanganese Crusts from the Central Pacific Ocean: Their Implication for Crust Genesis. *Marine Geology* 136:277–297.

Zhou, X., S. Gao, and S.L. Hui. 1997. Methods for Comparing the Means of Two Independent Log-Normal Samples. *Biometrics* 53:1129–1135.

D.2 Internet Resources

Department of Energy pages on the DQO process, case studies, software, and training: URL: <http://dqo.pnl.gov/> (Pacific Northwest National Laboratory) URL: <http://www.hanford.gov/dqo/> (Hanford site)

U.S. EPA Office of Environmental Information (formerly, in part, Center for Environmental and Information Statistics). Information on environmental quality, status, and trends: URL: <http://www.epa.gov/oei/quality.htm>

U.S. EPA Quality System DQO guidance documents: URL: <http://www.epa.gov/quality1/>

Glossary of Geology Terms – U.S. Geological Survey: URL: <http://geology.wr.usgs.gov/docs/usgsnps/misc/glossaryAtoC.html#A>

Glossary of Soil Science Terms – Soil Science Society of America: URL: <http://www.soils.org/sssagloss/>

Hanford's Environmental Restoration Project – Bechtel Hanford, Inc.: URL: <http://www.bhi-erc.com>

HyperStat Online, an introductory-level hypertext statistics book: URL: <http://davidmlane.com/hyperstat/index.html>

The Internet Glossary of Statistical Terms – The Animated Software Co.: URL: <http://www.animatedsoftware.com/statglos/statglos.htm>

The "new statistics" of resampling (drawing repeated samples from the given data, or population suggested by the data) – Resampling Stats: URL: <http://www.resample.com>

NFESC Environmental Services and BRAC page – numerous links to guidance/policy documents and technical information: URL: <http://erb.nfesc.navy.mil/>

Statistical courses, software, and additional information: URL: <http://www.statistics.com/>

Statistics Glossary – Centre for Applied Statistics, Lancaster University, U.K.: URL: <http://www.cas.lancs.ac.uk/glossary_v1.1/main.html>

D.3 Computer Software

Visual Sample Plan (sampling design software) from Department of Energy: URL: <http://dqo.pnl.gov/VSP/Index.htm>

ELIPGRID-PC (Hot Spot Probability Calculations): Upgraded Version. Davidson, J.R. 1995. ORNL/TM-13103. Oak Ridge, Tenn.: Oak Ridge National Laboratory. December. URL: <http://dqo.pnl.gov/software/elipgrid.htm>

Data Quality Objectives Decision Error Feasibility Trials (DQO/DEFT). 1994. EPA QA/G-4D. EPA 600/R-96/056. Washington, DC: Office of Research and Development. URL: <http://www.epa.gov/nerlesd1/databases/deft/abstract.htm>

GEOPACK (geostatistical software). United States Environmental Protection Agency. 1990. EPA/600/8-90/004. Washington, DC: Environmental Sciences Division. URL: <http://www.epa.gov/nerlesd1/databases/geo-pack/abstract.htm>

Geostatistical Environmental Assessment Software (Geo-EAS) 1.2.1. United States Environmental Protection Agency. 1988. EPA/600/4-88/033a. Washington, DC: Environmental Sciences Division. URL: <http://www.epa.gov/nerlesd1/databases/geo-eas/abstract.htm>

MARSSIM (Multi-Agency Radiation Survey and Site Investigation Manual) software tools. United States Environmental Protection Agency. URL: <http://www.epa.gov/radiation/marssim/tools.htm>

SCOUT: A Data Analysis Program (a univariate and multivariate data analysis tool). United States Environmental Protection Agency. 1993. Washington, DC: Office of Research and Development, Environmental Monitoring Systems Laboratory. URL: <http://www.epa.gov/nerlesd1/databases/scout/abstract.htm>

SigmaPlot and *SPSS.* Chicago: SPSS Inc. URL: <www.spss.com>

StatView 5.0. Cary, N.C.: SAS Institute, Inc. URL: <www.statview.com>