NAVAL FACILITIES ENGINEERING COMMAND
Washington, DC 20374-5065

NFESC
User's Guide
UG-2054-ENV

GUIDANCE FOR ENVIRONMENTAL BACKGROUND ANALYSIS

Volume II: Sediment

Prepared by

Battelle Memorial Institute
Earth Tech, Inc.
NewFields, Inc.

April 2003

Approved for public release; distribution is unlimited.

♻ Printed on recycled paper

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 property uses have resulted in actual or suspected chemical releases. Each volume in this series is devoted to a particular medium. Volume I of this series was released in April 2002, and focuses on background analyses of chemicals in soils. This volume focuses on analytical methods and procedures that can be used to identify background chemicals in the sediment medium (whether from anthropogenic or natural sources), and estimate the chemical concentration ranges that represent site-specific background conditions.

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 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.

Soil and sediment background analyses have many similarities. However, for sediment background analyses, special consideration must be given to unique features of the aquatic and terrestrial environment that affect the distribution of sediment chemicals, including: the hydrodynamic, ecological, and biological characteristics of the sediment basin; potential sources of contamination within the sediment basin and the upstream watershed; and the lateral, vertical, and temporal distribution of the investigated sediments (Section 2). The background analysis techniques presented in this volume focus primarily on evaluation of sediment metal concentration data. However, these techniques can be applied to any location-specific quantitative data, including organic chemical concentration data, and the ecological and biological measurements gathered during a typical sediment investigation.

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.
- ❑ Navy Policy on Sediment Site Investigation and Response Action: DON, 2002.
- ❑ Implementation Guide for Assessing and Managing Contaminated Sediments at Navy Facilities: DON NAVFAC, 2003.

ACKNOWLEDGMENTS

This document was prepared by Battelle Memorial Institute, Earth Tech, Inc., and Newfields, Inc. for the Naval Facilities Engineering Command (NAVFAC). Numerous people from all three organizations contributed to this document.

NAVFAC's Risk Assessment Workgroup (RAW) and the RAW Background Subgroup made significant contributions to this document. In particular, several of NAVFAC's employees provided technical input, review, and direction throughout the project. Also, this guidance has gone through a peer review by independent experts in the field of environmental statistics.

Furthermore, this comprehensive document incorporates significant information from five previous Naval Facilities Engineering Command (NAVFAC) documents*: Procedural Guidance for Statistically Analyzing Environmental Background Data* (1998), *Handbook for Statistical Analysis of Environmental Background Data* (DON, 1999a), *Protocol for Background Evaluation* presented in Appendix C of the *Site Management Plan Update for the Pearl Harbor Naval Complex* (DON, 2000c), *Navy Policy on Sediment Site Investigation and Response Action* (DON, 2002), and the recently published *Implementation Guide for Assessing and Managing Contaminated Sediments at Navy Facilities* (DON, 2003). Thus, authors and contributors to these three documents are noted.

Finally, although this guide is not intended to supersede these previous documents, it does serve as a compendium of the earlier guidance, provides additional instruction, and contains up-to-date information.

EXECUTIVE SUMMARY

This guidance document provides instructions for characterizing sediment background conditions at Navy sites or adjacent areas of a sediment basin where past property uses have resulted in actual or suspected chemical releases. Background analysis is necessary to identify sediment 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 site-specific sediment background conditions.

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. In some areas, unacceptable risks may be associated with sediment 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 United States Environmental Protection Agency (U.S. EPA) technical guidance. Similarly, according to the February 2002 Navy sediment policy document *Navy Policy on Sediment Investigations and Response Actions*, all sediment investigations and response actions, including background investigations, must be scientifically defensible, technically feasible, risk-based, cost-effective, and directly linked to a specific Navy CERCLA/RCRA site.

The background analysis techniques presented in this document 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.*

Sediment background analysis is an integral component of most Navy sediment site investigations. Existing data that may be relevant to background should be reviewed and assessed during the initial phase of an investigation. The geological, ecological, biological, and hydrodynamic characteristics of the investigated sediment basin, as well as operational history, physical characteristics, and chemical characteristics of the site, should be evaluated to: (1) identify the conditions that may affect the spatial (both lateral and vertical) and temporal distributions of impacted and background sediments, and (2) develop a list of chemicals of potential concern (COPCs). 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 sediment background conditions are collected. After the additional data have been collected and incorporated into the dataset, the reported chemical concentrations should be compared to appropriate screening criteria in order to determine which chemicals should be carried forward for background analysis.

This document presents step-by-step instructions for sediment background analysis and a case study to illustrate application of the methods. For many sediment basins and target chemicals, background chemical concentration ranges can be estimated by the spatial analysis and probability plotting techniques presented in Exploratory Data Analysis (Section 2.2). These analyses consider the location, depth, and physical characteristics of the investigated sediments, including particle size and organic carbon content. Biological and ecological data also can be used to augment the exploratory analyses and assess potentially impacted areas. If the initial analyses do not yield a technically defensible and reliable estimate of the sediment background concentration range for a particular target chemical, the project team should identify appropriate methods for further analysis (Section 2.3). Detailed instructions for analysis using the Geochemical Method (Section 3) and the Comparative Method (Section 4) are presented; a case study (Section 5) illustrates the application of both methods.

The Geochemical Method uses techniques based on geochemical principles to estimate sediment background metal concentration ranges. Chemical concentrations in sediments are controlled by the chemical composition of the parent rocks and the geochemical processes that occur during soil formation, sediment transport, and sediment deposition; therefore, certain metals tend to occur together in natural sediments, and metal/metal concentration ratios often are restricted to relatively narrow ranges. These natural geochemical and geological relationships can be characterized and used to estimate the chemical concentration ranges that represent background conditions.

The Comparative Method is based on comparison of chemical concentration data from the potentially impacted area to data from a background/reference area. The objective of this method is to determine whether concentrations of the target chemicals at the investigation site are statistically similar to reference area concentrations. Sediment in the reference area must be physically, geochemically, ecologically, and anthropogenically similar to sediment in the potentially impacted area. An understanding of the hydrodynamic regime of the sediment basin is necessary to identify suitable sediment background areas. In estuarine environments such as bays, river mouths, and lagoons, the distribution of impacted sediments may not fit either the linear model or the radial concentration gradient pattern commonly observed at contaminated terrestrial soil sites. In these environments, chemical, biological, and ecological data should be used to investigate candidate areas away from the source area before they can be considered as background areas.

CONTENTS

FIGURES

TABLES

BOXES

ACRONYMS AND ABBREVIATIONS

ACF	advanced chemical fingerprinting
AF	absolute frequency
AOC	area of concern
ARAR	applicable or relevant and appropriate requirement
ASCE	American Society of Civil Engineers
ASTM	American Society for Testing and Materials
AVS	acid volatile sulfide
BERA	baseline ecological risk assessment
BRAC	Base Realignment and Closure (Act)
Cal/EPA	California Environmental Protection Agency
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
CF	cumulative frequency
cm	centimeter
CNO	Chief of Naval Operations
COPC	chemical of potential concern
^{137}Cs	cesium-137
CSM	conceptual site model
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)
EB	estuarine bathymetry
EFA	Engineering Field Activity
EFD	Engineering Field Division
ER-L	effects range–low
ER-M	effects range–median
ESI	Environmental Sensitivity Index
FDEP	Florida Department of Environmental Protection
FDER	Florida Department of Environmental Regulation
GIS	Geographic Information System
H_a	alternative hypothesis
H_o	null hypothesis
HS	hydrographic survey

IR	Installation Restoration
IR CDQM	Installation Restoration Chemical Data Quality Manual
J	estimated concentration
LANL	Los Alamos National Laboratory
MDEQ	Michigan Department of Environmental Quality
MERA	Michigan Environmental Response Act
mg/kg	milligram(s) per kilogram
NAPP	National Aerial Photography Program
NCP	National Oil and Hazardous Substances Pollution Contingency Plan
ND	nondetect
NFA	no further action
NOAA	National Oceanic and Atmospheric Administration
NOS	National Ocean Service
NPDES	National Pollutant Discharge Elimination System
NQ	not qualified
NRC	Nuclear Regulatory Commission
NRCS	Natural Resources Conservation Service
OERR	Office of Emergency and Remedial Response (of U.S. EPA)
ORR	Office of Response and Restoration (of NOS)
OSWER	Office of Solid Waste and Emergency Response (of U.S. EPA)
OU	operable unit
PACDIV	Pacific Division Naval Facilities Engineering Command
PAH	polycyclic aromatic hydrocarbon
PCB	polychlorinated biphenyl
POI	point of interest
ppb	part(s) per billion
ppm	part(s) 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
RL	reporting limit
SAIC	Science Applications International Corporation
SAP	Sampling and Analysis Plan
SARA	Superfund Amendments and Reauthorization Act
SWDIV	Southwest Division Naval Facilities Engineering Command
SWMU	solid waste management unit

TOC	total organic carbon
U	nondetect
UCL	upper confidence limit
UJ	nondetect estimated
USACE	United States Army Corps of Engineers
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
WCSD	Watershed Contaminated Source Document
WDOE	Washington State Department of Ecology
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.
Ambient Chemical Concentrations	Total concentrations of both naturally occurring chemicals and anthropogenic chemicals not related to specific point sources or site releases.
Background Area	See *Reference Area.*
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 of suspected contamination.
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 a perfect linear relationship with positive slope exists 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 a perfect linear relationship with negative slope exists 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	A scientific and statistical data evaluation to determine if environmental investigation data 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 that 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.
Detection Limit	The minimum concentration of an analyte that can be measured within a given matrix and reported with a 99% confidence that the analyte concentration is greater than zero.
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 soil or sediment. 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 or sediment sample divided by this ratio in source rocks.
Exploratory Data Analysis	A statistical and graphic procedure for examining data in order to describe the main distributional features of measured data.
Facility	See *Installation.*
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 graphically displaying the characteristics of a distribution of items in a given population or sample. In a histogram, each measure is usually represented by a single block 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.
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.
Installation	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 parts or all of the investigated sediment basin. Also referred to as a *facility.*
Inter-Quartile Range (IQR)	A measure of the spread of or dispersion within a dataset. The IQR is the 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 a dataset (totaling the various individual results and dividing by the number of results involved).
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	Concentrations of naturally occurring chemicals in environments that have not been influenced by human activity.
Nondetects	Measurements reported by the analytical laboratory as below either the detection limit or the reporting 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 generally are 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 rule for describing the probability measures associated with all the values of a random variable. For a discrete random variable, the probability distribution is described in terms of a probability mass function, which is a list of probabilities associated with each of the possible values of the discrete random variable. For continuous random variables, the probability distribution is described in terms of a 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.
Reference Area	An area within a sediment basin where detected chemicals are attributed to natural or anthropogenic background sources only. Also referred to as *background area.* Background or reference areas usually are located upstream of the impacted portions of the sediment basin.
Regression	A set of techniques to characterize the manner in which one of the measures changes as the other measure changes.
Reporting Limit	A project- and laboratory-specific numerical threshold value used for reporting analytical data. Laboratory measurements below this value are reported as the numerical threshold value followed by "U" (nondetect). This value is typically one to five times the detection limit, depending on the analytical method and matrix. The detection limit can vary considerably from sample to sample because of matrix effects. Ideally, the reporting limit will not change, and will be set high enough to account for matrix effects, yet low enough to meet project-specific DQOs.
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.
Sediment	Any materials deposited at the bottom of water bodies, such as oceans, rivers, lakes, harbors, and storm drains.
Sediment Basin	A hydrological feature such as a lake, estuary, bay, harbor, or other water body in which sediments are deposited.
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, usually being 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%).
Site	A zone designated for investigation because of actual, suspected, or potential chemical releases. A site usually consists of both impacted and background areas. Site-specific field data are used to evaluate the extent of each area.
Skewness	A measure of asymmetry of the distribution of the sample data values.
Standard Deviation(s) (r^2)	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	The set of environmental space/time units within spatial and temporal 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.
Watershed	The area that drains into a sediment basin. A watershed may contain uplands and wetlands, as well as more than one site.

1. INTRODUCTION

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

Sediments are defined as materials deposited on the bottom of water bodies, such as a lakes, seas, rivers, harbors, storm drains, or other sediment basins. Surface water (e.g., rivers, streams) transports sediments from their sources—primarily weathered rocks and soils within the watershed—to sediment basins. Chemicals from both natural and anthropogenic sources may become adsorbed to sediment particles and accumulate in the pore-water within the sediment.

Contaminated sediments pose a particularly important and complex problem for the Navy. Most Navy facilities are located within marine or estuarine sediment basins and watersheds along the coastline. A sediment basin may be impacted by Navy chemical releases, as well as by releases associated with a wide variety of non-Navy activities. Because contaminants in the aquatic environment can be transported long distances from many potential sources, it can be particularly difficult to identify contaminant sources, and determine whether the Navy is responsible for cleanup.

In addition to contaminants associated with Navy and non-Navy chemical releases, anthropogenic background chemicals associated with nonpoint sources may be ubiquitous within a sediment basin. Natural background chemicals also may be present at relatively high concentrations. The Navy is not responsible for cleanup of natural or anthropogenic background chemicals; therefore, at most Navy sediment investigation sites, background analysis will be required to define the extent of site-related contamination and assess the need for cleanup action.

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. Navy policy on sediment investigations and response actions provides guidelines for ensuring that all sediment investigations and response actions are scientifically defensible, technically feasible, risk-based, and cost-effective.

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 the 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 must be discussed and documented in the risk characterization sections of 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 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 because of 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; however, they must not be within the background range. Cleanup efforts should be limited to chemicals associated with a site-related release that may pose unacceptable risks to human health or the environment.

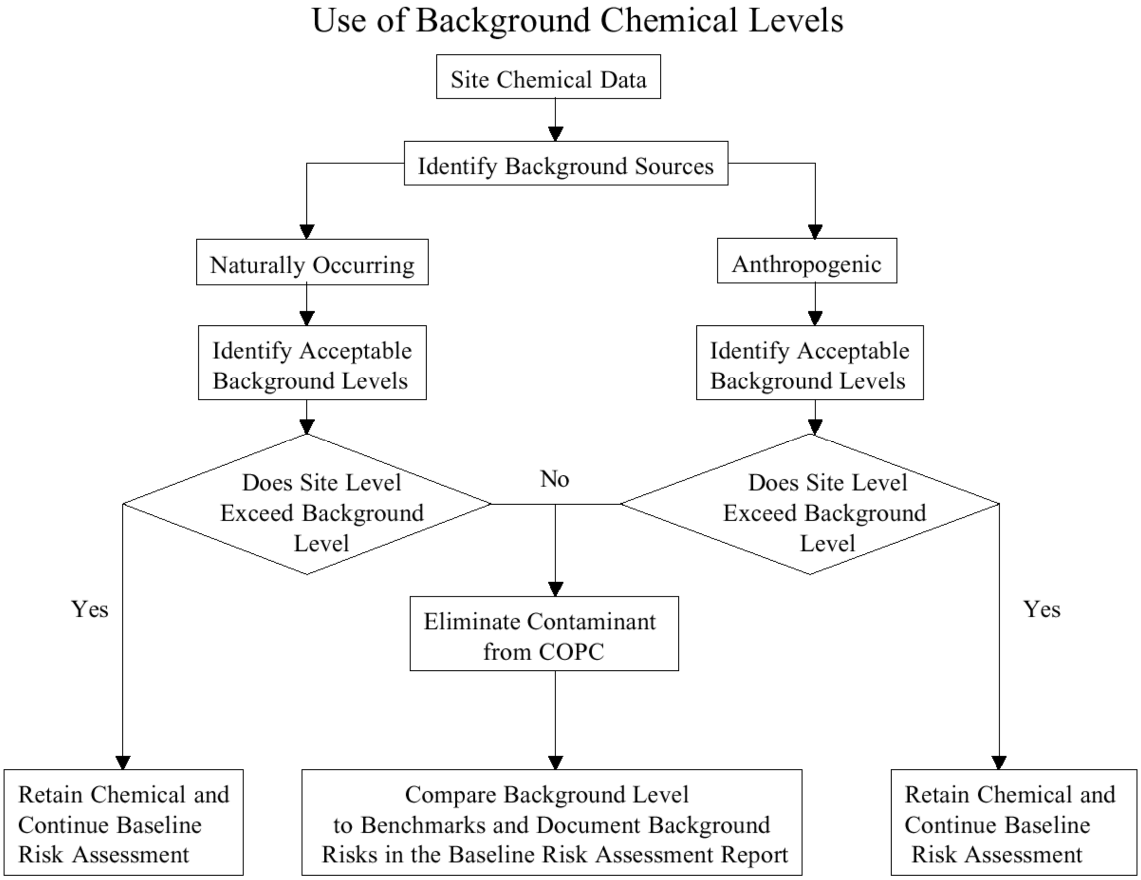


FIGURE 1-1. Navy policy on use of background chemical levels (DON, 2000a)

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 by 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 both policies, 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).

1.1.4 Navy Sediment Policy

The *Navy Policy on Sediment Site Investigation and Response Action* (DON, 2002) (i.e., the Navy sediment policy) specifies requirements for sediment investigations and response actions conducted under the IR Program. The Navy sediment policy requires that:

- ❑ All sediment investigations and response actions must be scientifically defensible, technically feasible, risk-based, and cost-effective.

- ❑ All sediment investigations and response actions must be directly linked to a specific Navy CERCLA/RCRA site.

- ❑ All potential Navy and non-Navy contaminant sources within the watershed must be identified to determine whether the Navy is solely responsible for the contamination.

- ❑ If non-Navy contaminant sources are identified, the project team shall prepare a Watershed Contaminated Source Document (WCSD).
- ❑ The Navy shall not clean up contamination from a non-Navy source unless the Navy has contributed to the risk associated with the sediments. The Navy will not clean up a site before the source is contained. Any potential for recontamination by non-Navy sources shall be documented.
- ❑ Sediment site investigations should be planned in accordance with the seven-step data quality objectives (DQO) framework, and a conceptual site model (CSM) must be developed to define the dynamics of the sediment site.
- ❑ All sediment investigations and response actions shall be consistent with Navy polices on risk assessment and background chemical levels.
- ❑ Sediment cleanup goals shall be based on site-specific information and risk-based criteria or background conditions.
- ❑ If monitoring is required to confirm the effectiveness of a response action, a monitoring plan with exit strategies shall be developed before collecting the first monitoring sample.

Watershed Contaminated Source Document (WCSD)

According to the Navy sediment policy (DON, 2002), if a sediment site is potentially impacted by non-Navy contaminant sources, a WCSD must be prepared before investigations or response actions are initiated. The WCSD is a summary report (usually 2 to 10 pages) that identifies both Navy and non-Navy contaminant sources that may impact the sediment basin. The primary objective of the WCSD is to document the existence of all non-Navy sources that discharge chemicals to the watershed or sediment basin. The document should include a pictorial CSM that identifies the potential contaminant sources, transport mechanisms, exposure routes, and receptors. In addition to its function in the WCSD, the CSM is a dynamic tool for investigation and response action planning, organizing site knowledge, identifying data gaps, and evaluating alternatives for site assessment or cleanup during various phases of sediment investigations and response actions. Findings of the WCSD should be used to ensure that all Navy sediment investigations or response actions are linked to a specific Navy CERCLA/RCRA site (as required by the Navy sediment policy). The WCSD should be submitted to the appropriate regulatory agencies for review.

1.1.5 Navy Sediment Assessment and Management Guidance

In conjunction with the Navy sediment policy, Naval Facilities Engineering Command has developed the *Implementation Guide for Assessing and Managing Contaminated Sediments at Navy Facilities* (DON, 2003) to provide sediment-specific technical information for remedial project managers and project teams responsible for assessment and management of contaminated Navy sediment sites. The implementation guide presents a consistent and effective approach to site characterization, risk assessment, remedial option evaluation, long-term monitoring, and site closeout. The use of background data in risk assessments and implementation of the *Navy Interim Final Policy on the Use of Background Chemical Levels* (DON, 2000a) are also discussed.

1.2 Scope of Sediment Background Analysis Guidance Document

This guidance document presents techniques for evaluating chemical data and sediment characteristics to distinguish between sediments that are impacted by a site-related chemical release and those that are not impacted by a site-related release. Both natural processes (e.g., deposition of naturally occurring metallic minerals in fluvial sediments) and anthropogenic processes (e.g., deposition of chemicals from internal combustion engine exhaust and highway runoff) may result in elevated concentrations of various chemicals—including hazardous substances—in otherwise nonimpacted sediments. 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 in a sediment basin, sediment samples are analyzed for chemicals that may have been released as a result of site activities. Because chemicals associated with a known or suspected Navy release may also be derived from background sources (i.e., natural geochemical sources and anthropogenic sources not associated with a site-related release), *background analysis should be conducted early in the site investigation process*. This will ensure that only sediments that have been impacted by a Navy chemical release are targeted for cleanup. Failure to distinguish between concentrations associated with a site-related chemical release and background concentrations may lead investigators to establish 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. In addition, because sediment deposition is an ongoing process, background chemicals present in other areas of the basin or watershed most likely will be transported back into the cleanup area.

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, occur naturally in all sediments and may be 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.

Background conditions for organic chemicals such as polycyclic aromatic hydrocarbons (PAHs) can be evaluated by the statistical and graphical techniques described in the main text of this document, or by the advanced chemical fingerprinting (ACF) methods described in Appendix A. Background analysis by ACF involves distinguishing the organic compounds associated with a site-related chemical release from those attributable to background sources. Appendix A focuses on the application of ACF to background analysis of PAHs in sediments.

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. An understanding of the geological, geochemical, and hydrological processes that control the occurrence and concentrations of naturally occurring chemicals in sediments is also essential. Therefore, this guidance document also provides guidance for evaluating the physical and chemical processes that control metal concentrations in sediments, including the geochemical characteristics of the aquatic environment that affect the redistribution of metals.

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

- ❑ ***Exploratory Data Analysis*** (Section 2.2) computes the summary statistics of available sediment chemical concentrations, investigates their cumulative population distributions, and evaluates their spatial distributions to estimate background ranges for the target chemicals. For this purpose, the available data are graphed and plotted to facilitate visual inspection of the results.
- ❑ ***Geochemical Background Analysis*** (i.e., the Geochemical Method) (Section 3) is based on the geochemical association relationships commonly observed among naturally occurring background chemicals, or correlations between background chemicals and parameters such as grain size or total organic carbon (TOC) content. The Geochemical Method can distinguish between background concentrations and concentrations that represent a release by using various statistical techniques to evaluate these natural chemical relationships

and identify anomalies that may represent contamination.

- ***Comparative Statistical Analysis*** (i.e., the Comparative Method) (Section 4) compares chemical concentrations detected in sediments to the chemical concentrations ranges that exist at "reference areas"—i.e., areas that have not been impacted by site-related chemical releases. For this method to be successful, sediments at the reference area must have physical, geochemical, biological, and anthropological characteristics similar to native sediments in the investigation 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 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; background analysis is therefore an integral part of the environmental assessment, decision-making, and cleanup process.

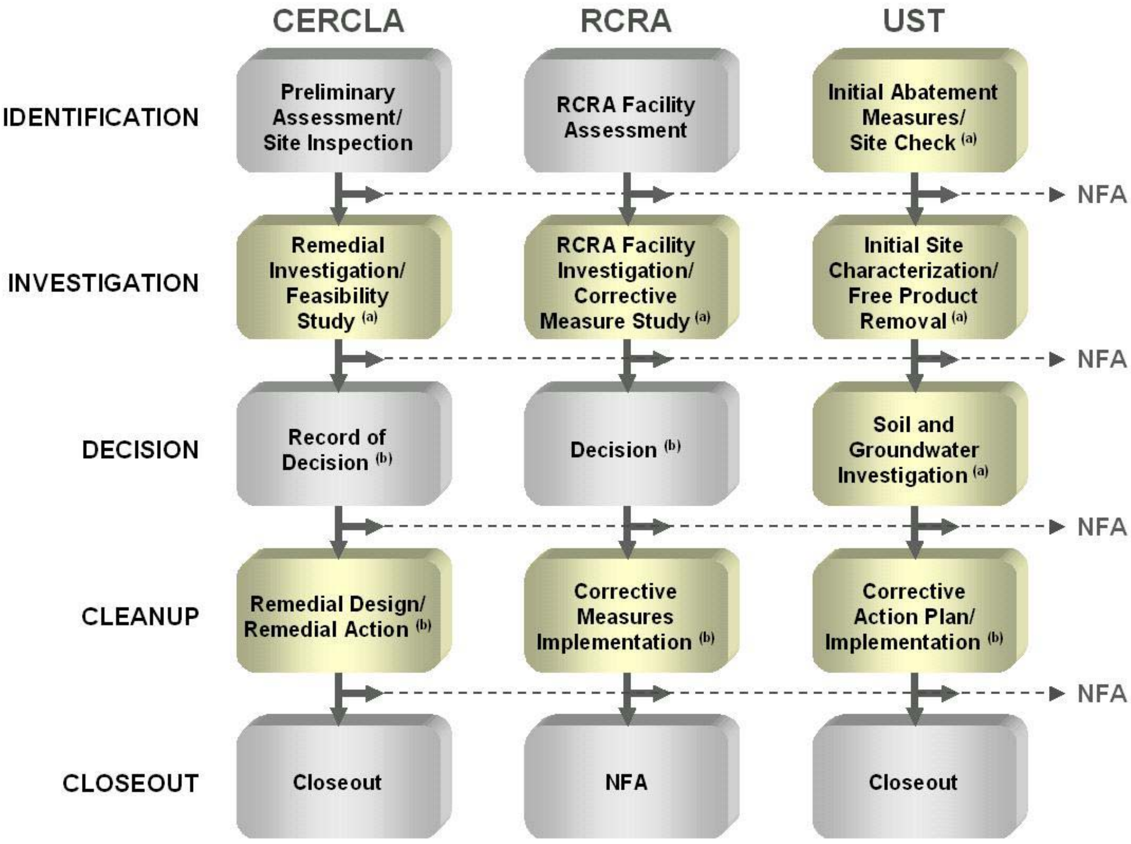
Figure 1-2 illustrates the roles of background analysis in environmental investigation and restoration activities conducted under the three primary regulatory frameworks (CERCLA, RCRA, and the UST Program). Background analysis should be incorporated into the site identification, investigation, decision, 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 sediment 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 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 action for a particular site. In addition, by



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
NFA	no further action

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

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 Installation Restoration Program, and describes the procedures the Navy has developed to ensure compliance with these laws and regulations. According to the *IR 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 *IR 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 CERCLA/NCP, require protection of human health and the environment at sites where CERCLA hazardous substances have been released. To comply with 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. Some 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.

1.5.1 Site, Sediment Basin, and Background Areas

The following terminology is used throughout this document:

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 describe 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 Useability 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. 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:*** 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—background should instead 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 and sediments compared to background concentrations. These procedures are based on the approach employed by the State of Michigan Department of Environmental Quality (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.

- ❑ *Site* describes a zone designated for investigation because of 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.

- ❑ *Sediment basin* refers to a hydrological feature such as a lake, estuary, bay, harbor, or other water body in which sediments are deposited.

- ❑ *Watershed* refers to the geographic area that drains into the sediment basin. A watershed may contain uplands and wetlands, as well as more than one site.

- ❑ *Background area* or *reference area* refers to an area within a sediment basin where the detected chemicals are attributed to natural or anthropogenic background sources only. These background or reference areas are usually located upstream of

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. This regional guidance 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) in order 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 Florida Department of Environmental Regulation):*** FDER published a technical document titled *A Guide to the Interpretation of Metal Concentrations in Estuarine Sediments* (FDER, 1988). This publication 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 (Section 3). 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 are elevated as a result of a chemical release. *Operational Memorandum No. 15* (MDEQ, 1993) to MERA established default background concentrations based on a comprehensive background 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.

the impacted portions of the sediment basin.

- ❑ *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 different sites, and may cover a portion of the sediment basin or the entire sediment basin.

The concept of impacted versus background 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 environmental media are attributable to the natural geological or hydrogeological characteristics of the area. These chemicals have not been altered by human activity. Some examples include organic compounds derived from natural oil seeps, metals from bedrock, PAHs generated by forest fires, and naturally occurring asbestos.

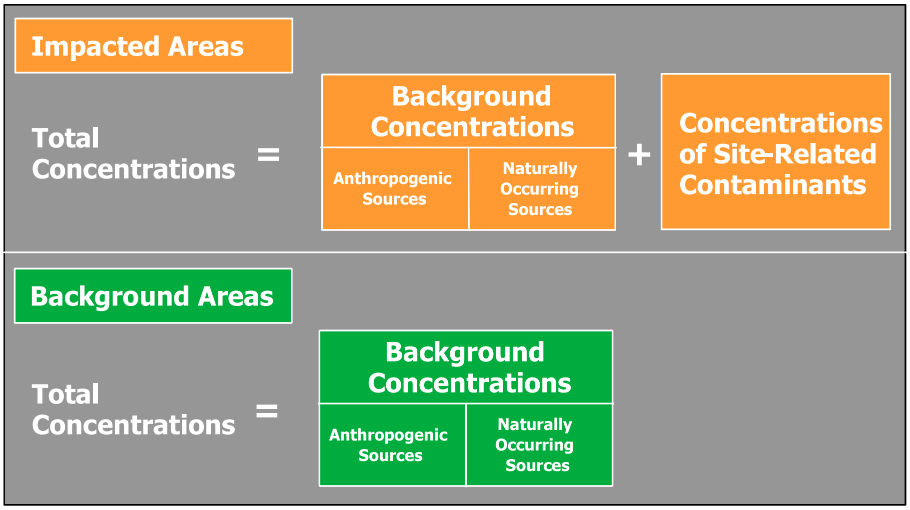
- ❑ ***Anthropogenic Chemicals:*** Synthetic or natural substances that have been released to the environment as a result of human activities, but are not related to specific activities conducted at a site. These chemicals are usually ubiquitous in the environment, and may impact very large areas. Anthropogenic 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, dichlorodiphenyltrichloroethane [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 those 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 detected routinely during sampling and analysis. Unless background conditions are accounted for in one of the following two ways, project teams may



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

FIGURE 1-3. Concept of impacted versus background areas

unnecessarily remediate chemicals and areas where no releases have occurred:

- ***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, 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. (Note that the term "background level" as defined in this document does not necessarily correspond to chemical-specific criteria that may be established for purposes such as compliance with United States Army Corps of Engineers [USACE] permits for discharge of dredged material into waters of the United States and ocean waters.)

1.6 Background Analysis Components

The scope of the background investigation must be consistent with the Navy background and sediment policies (DON, 2000a and 2002). The flowchart shown in Figure 1-4 displays the sequence of steps that should be followed to quantify background concentration ranges for chemicals in sediment at a typical Navy installation. 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.

The Figure 1-4 flowchart illustrates the full background analysis sequence for a typical sediment investigation. As indicated by the decision questions shown on the flowchart, certain steps or methods can be bypassed or combined depending on site conditions and project-specific requirements.

TABLE 1-1. Examples of sediment background concentration ranges and ecological risk-based screening criteria

Background Chemical	Background Concentration Range in Sediment (mg/kg)	Marine Sediment ER-L[(a)] (mg/kg)
Cadmium[(b)]	0.1–10.0	1.2
Chromium[(c)]	0.1–200	81
Lead[(c)]	0.025–80	46.7
Mercury[(b)]	0.025–0.25	0.15

(a) Buchman (1999). The effects range–low (ER-L) value represents the concentration at which toxic effects may begin to be observed in sensitive species.
(b) Canova (1999).
(c) FDER (1988).

Notes:

Numeric prefixes correspond to sediment guidance document section numbers.

COPC = chemical of potential concern

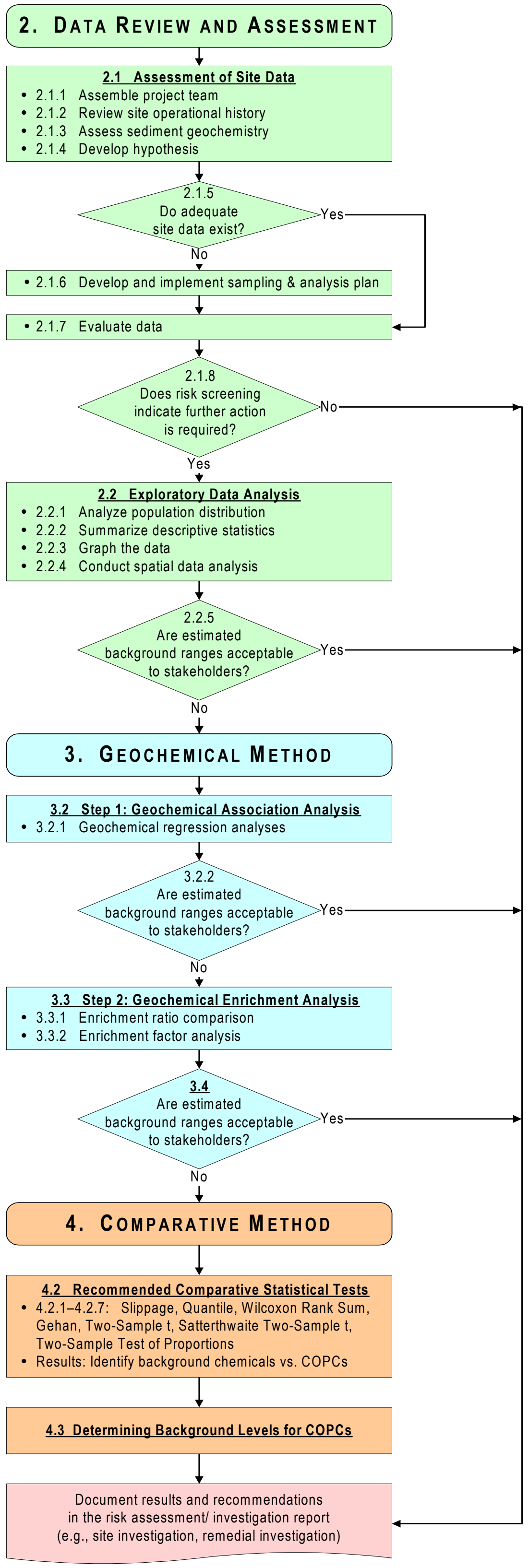


FIGURE 1-4. Typical background data evaluation process for inorganic constituents in sediment

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 risk-based criteria. Chemicals with representative exposure concentrations below their risk-based criteria do not require background analysis and can be screened out. (Chemicals without risk-based screening criteria cannot be screened out at this stage.) The remaining chemicals should be subjected to exploratory statistical and spatial data analysis in order 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 in following subsections.

1.6.2 Methods for Further Analysis

If exploratory statistical and spatial data analyses do not yield satisfactory background range estimates, then the data should be analyzed further using 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 sediment. Because the method does not require reference area data for comparison, it can be used when it is not possible to identify a reference area. It often is very difficult or impossible to identify a suitable sediment reference area, and sediment sampling/analysis costs can be substantial; therefore, as indicated by the sequence depicted in the Figure 1-4 flowchart, the Geochemical Method usually is preferred over the Comparative Method. The 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 relationships 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 the fundamental 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 datasets representing potentially impacted sites to reference area datasets in order to determine whether the potentially impacted site concentrations and reference area concentrations are statistically similar. For this purpose, adequate reference-area sampling data are required in addition to the potentially impacted data. Comparative tests are categorized as parametric or nonparametric. Parametric tests are based on specific distributional assumptions, (e.g., normality of mean concentrations), whereas nonparametric tests require no such assumptions. Each test is designed to assess specific aspects of the investigated data. For example, certain tests are designed to evaluate the similarity of extreme site and background concentrations, whereas others are designed to assess central tendencies (median or mean) of the observed concentrations.

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

Sediment in the reference area must be physically, geochemically, ecologically, and anthropogenically similar to sediment in the potentially impacted area. An understanding of the hydrodynamic regime of the sediment basin is necessary to identify suitable sediment background areas. In estuarine environments such as bays, river mouths, and lagoons, the distribution of impacted sediments may not fit the linear model (lower concentration upstream versus higher concentration downstream), or the radial concentration gradient pattern commonly observed at contaminated terrestrial soil sites. In these environments, chemical, biological, and ecological data should be used to investigate candidate areas away from the source area before they can be considered background areas.

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 watershed/sediment basin history and setting.
- ❑ 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 geochemical and anthropogenic information.
- ❑ Descriptions of the sediment basin and reference area datasets.
- ❑ Rationale to demonstrate that the datasets are adequate for background analysis.
- ❑ 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) and the Navy sediment policy (DON, 2002), the background analysis report must address the following issues:

- ❑ Chemicals that may have been released at a site or within the watershed 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 Unique Aspects of Sediment Site Characterization and Background Analysis

As noted in the Navy sediment policy (DON, 2002), many Navy installations are located along water bodies that are impacted by a wide variety of contaminant sources, such as municipal storm drains, urban and agricultural runoff, and private industrial facilities. In addition, chemicals can be transported long distances in the aquatic environment. Therefore, the spatial and temporal patterns of sediment contamination are typically more complex than soil contamination. As a result, many of the methods and procedures used for sediment investigations and response actions are different from those used for soil investigations and response actions. Box 1-4 describes some of the differences between sediment and soil investigations and background analyses.

Extreme care must go into the planning and design of sediment investigations and any subsequent

BOX 1-4. Site investigation and background analysis: Sediment vs. soil

Sediment and soil site investigations and background analyses share the same fundamental objectives: to identify COPCs, define the extent of site-related contamination, and assess risks to human and ecological receptors. However, sediment investigations are usually more complicated than soil investigations, primarily because the physical, chemical, and biological characteristics of the aquatic environment are more complex and dynamic than those of the terrestrial environment. Some of the distinguishing features of sediment investigations are:

1. **Multiple Background Sources:** Watersheds often contain many nonpoint sources that contribute a wide range of anthropogenic background chemicals to sediment basins (e.g., in runoff from agricultural lands, highways, urban areas, and industrial areas). Therefore, considerable effort may be required to assess anthropogenic background levels at a sediment basin. Sources of natural background chemicals also may be very diverse. Natural background concentrations in sediments depend on the type of soil and rock in the sediment source area, and processes such as adsorption and sorting that occur during sediment transport and deposition.
2. **Variable Background Conditions:** The distribution of natural background chemicals at a sediment basin may show great variability with depth. This is primarily due to distinct vertical layering of sediments with different mineralogy and grain size. For example, sand layers are likely to have much lower natural metal concentrations than clay layers. Furthermore, unlike typical soil background conditions, the background chemical characteristics of a sediment basin may change with time due to redistribution of sediments by current and wave action, as well as deposition of new sediments. Such variations must be addressed when characterizing background conditions at a sediment basin.
3. **Extent of Site-Related Impacts:** Site-related chemicals can be transported long distances in the aquatic environment. As a result, it is frequently difficult to differentiate between site-related chemicals and background chemicals in sediments. For example, in estuarine environments such as bays, river mouths, and lagoons, the distribution of impacted sediments may not fit the linear model or the radial concentration gradient pattern commonly observed at contaminated terrestrial soil sites. In these environments, chemical, biological, and ecological data should be used to investigate the extent of impacted areas and background areas. Furthermore, due to the direct contact between sediment and water, concentrations of volatile and soluble chemicals are usually very low. Therefore, sediment COPCs tend to be chemicals that accumulate in the solid phase, such as metals and hydrophobic organic compounds.
4. **Ecological Risk-Driven Investigations:** Humans do not inhabit sediment sites, and generally have very little exposure to the sediment environment; therefore, ecological risk is typically the dominant concern. Detailed information regarding the ecological receptors that inhabit or use the site, site-specific exposure pathways, and the food chain may not be available. Because many animals can move freely between the site and other areas, exposure frequency and duration may be difficult to assess. In addition, sediment sites are not as accessible as soil sites; therefore, the aquatic ecosystem may not be well defined. Because of these factors, sediment risk assessments usually are considerably more complex than soil risk assessments.
5. **Risk-Based Sediment Criteria:** Risk-based screening criteria for protection of aquatic ecological receptors generally are not as well defined as they are for human and terrestrial ecological receptors, and may not be available for many chemicals. In addition, different regulators and stakeholders may identify different screening criteria depending on their objectives. For example, criteria designed to guide environmental cleanup of a sediment site will not be the same as criteria for dredging.
6. **Sediment Investigation Difficulties:** Sediment investigations generally require considerably more effort and cost than soil investigations. Procedures for accessing sampling areas, locating sampling points, and collecting representative samples are time consuming and usually require specialized equipment.

response actions. The Navy sediment policy (DON, 2002) concludes:

> "Careful thought must go into the planning and design of investigations and the response actions for sediments. Source identification, Conceptual Site Models, problem formulation and DQOs must be utilized in the characterization of the site. Identification of all potential sources, both Navy and non-Navy, is essential to the decision-

> making process. The remedial alternative selected shall be risk-based and the Navy source should be contained before the commencement of the sediment remediation."

Ecological exposure pathways at a sediment site are more likely to be complete than human exposure pathways; therefore, ecological risks are usually the primary focus of sediment investigations. A general CSM of a typical sediment basin ecosystem, showing potential exposure pathways and ecological receptors, is presented in Figure 1-5. A site-specific CSM should be developed for each investigation site, and site-specific ecological receptors, exposure pathways, and exposure point concentrations should be evaluated during the risk assessment process. Chemicals that are not screened out during the initial risk screening steps should be carried forward for background analysis.

The distribution of chemicals in sediments and the surrounding aqueous environment is controlled by many factors, including geographical, geological, geochemical, biological, meteorological, and hydrodynamic conditions. Knowledge of sediment components, depositional environments, the sediment profile, and sediment texture is required to evaluate the effect of these factors on the distribution of chemicals in the sediment environment.

1.7.1 Sediment Components

Sediments consist of three basic components:

- ❑ Inorganic material (primarily minerals derived from the parent or source rocks)
- ❑ Organic material
- ❑ Porewater.

The abundance of each component and its importance in the sediment system varies vertically and horizontally within a sediment basin.

The inorganic fraction, which includes remnants of the original parent rock and secondary minerals formed through weathering, typically makes up the greatest portion of the overall sediment mass. Therefore, the physical and chemical characteristics of sediments depend on the characteristics of the weathered rock and soil from which they are derived. Mineral grain sizes in most sediment deposits are 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 sediment (see Section 1.7.4).

The organic fraction is composed of plants and animals (including microorganisms) that cycle chemical nutrients within the sediment, the waste products released from aquatic plants and animals, and their dead remains.

The water fraction fills the pore space between sediment grains. Dissolved chemicals are transported through the pore spaces. Dissolved oxygen and carbon dioxide are used by sediment organisms, and affect redox conditions and pH within sediment deposits. Redox conditions and pH strongly influence metal transport and sorption in the sediment environment (see Section 2.1.3).

1.7.2 Depositional Environments

Sediment characteristics, particularly the grain size distribution (i.e., texture—see Section 1.7.4), are controlled primarily by the physical configuration and hydrodynamic characteristics of the depositional environment. The distribution of sediment chemicals is strongly influenced by texture; therefore, the distribution of both background and anthropogenic chemicals depends on the depositional environment. Coarse sediments (e.g., sands) are deposited in relatively high-energy environments such as beaches and riverbeds, whereas fine sediments do not settle out until they reach lower-energy (e.g., offshore) areas. Background metal concentrations therefore tend to increase when moving from high- to low-energy depositional environments (e.g., out from the shoreline).

Depositional environments and sediment textures encountered in rivers, bays, estuaries (the most common locations for Navy facilities), and other common hydrologic features are described below:

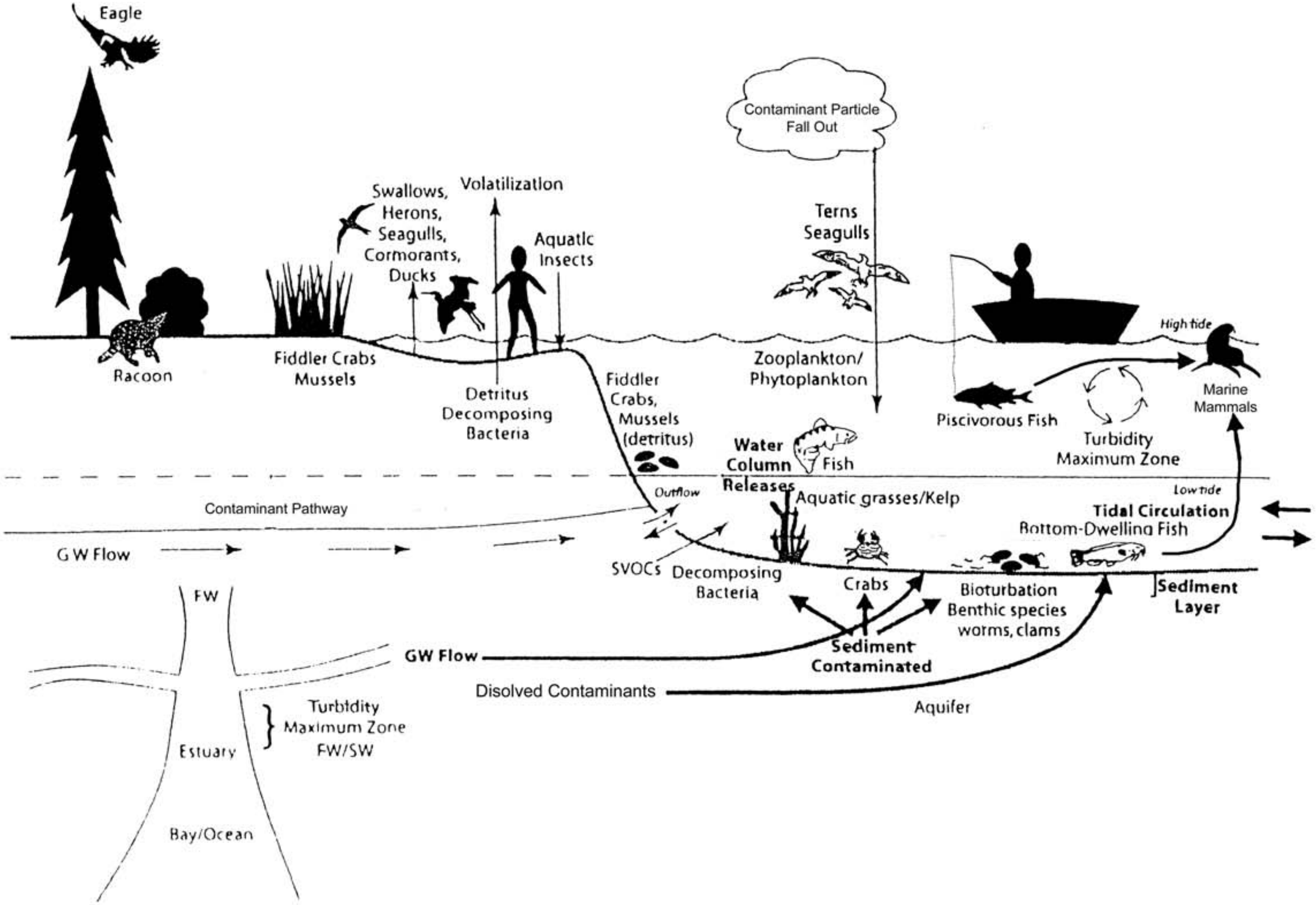


FIGURE 1-5. Generic conceptual site model of sediment ecosystem (Source: DON, 2003)

- **Rivers** erode rocks and soil in upland areas and transport the resulting sediments downstream. Coarse sediments including gravel and sand are deposited in the high-energy, upstream sections of a river, while finer sediments such as silt are deposited in the low-energy, downstream sections.

- **Lakes** are low-energy depositional environments fed by rivers and streams. Lake sediments tend to be relatively fine-grained (e.g., clay, silt, and fine sand).

- **Estuaries** are sections of rivers affected by tides or regions near river mouths, where fresh water from the river mixes with the salt water of the sea. Estuarine sediments typically grade from fine sand (upstream) to clayey silt (downstream).

- **Bays** are protected on each side by a headland. Bay sediments are typically relatively fine-grained (e.g., silt) in the interior portions of the bay, and become more coarse-grained (e.g., sand) toward the mouth due to increased wave action.

- **Deltas** are formed by fan-shaped deposits in areas where a river (or smaller stream) flows into a standing body of water, such as a lake or sea. Delta sediments are coarse (e.g., sands) near the river mouth, and become finer (e.g., silts and clays) with distance from the shore.

- **Beaches** extend inland from the water line. The line of permanent vegetation usually defines the inland boundary of a beach. Beach sediments typically are composed of relatively coarse sands and gravels.

- **Tidal flats** are areas of low relief cut by meandering tidal channels that are periodically flooded and drained by tides (usually twice each day). Tidal flat sediments typically consist of laminated or rippled clay, silt, and fine sand. Intense bioturbation by burrowing organisms is common.

- **Barrier islands** are semi-permanent nearshore sand deposits exposed to wave energy and dominated by a marine fauna. Barrier islands are separated from the mainland by a lagoon, and are commonly associated with tidal flat deposits.

- **Lagoons** occur on the landward side of barrier islands or reefs, and also may exist in the center of atolls. Lagoons are protected from ocean waves and contain finer sediment than the beaches (usually silt and mud).

- **Swamps** are low inland areas near the sea, and contain silts, muds, and organic sediments.

1.7.3 Sediment Profile

Layers in the vertical sediment profile may vary significantly in composition, texture, chemical, and biological characteristics. Changes in sediment sources and the hydrodynamic environment can cause distinct layering. Temporal changes in land-use patterns associated with increased urbanization or agricultural activities may result in layers with very different compositions and textures, and different concentrations of both anthropogenic and natural background chemicals. Natural and artificial environmental changes within a sediment basin or watershed also may result in distinct layering. For example, changes in terrestrial or aquatic vegetation patterns may produce layers with different organic carbon concentrations. It should also be noted, however, that many contaminated Navy sediment sites are located in quiescent areas such as the interior portions of a bay or estuary, where sediments tend to be fine-grained and homogeneous, with little or no stratification.

Sediment accumulation or burial occurs when the deposition rate exceeds the erosion rate. Sediment deposition rates vary widely among marine environments. In deep-sea environments, deposition rates may be as low as 1 cm per thousand years, whereas rates greater than 30 cm per thousand years are common in continental margins and other nearshore environments (Kennett, 1982). Compaction decreases pore space and reduces the volume of the water fraction as older sediments are buried beneath more recent deposits.

Sediment burial does not prevent redistribution of the associated chemicals. Physical and biological processes can mix surface and deeper sediments, and may expose formerly buried sediments at the surface. Sediment deposition rates, sediment erosion and removal, and vertical mixing of surface and deeper sediments must be assessed to fully understand a sediment site or basin. After initial deposition, sediments may be remobilized and transported to other locations by waves, currents, or other flow mechanisms. As shown in Figure 1-6, natural and anthropogenic chemicals that occur in the solid phase may be redistributed by flowing water or by organisms.

Biological activities, such as bioturbation by burrowing organisms as shown in Figure 1-6, can play a significant role in redistributing sediments and their associated chemicals. Bioturbation causes vertical mixing of sediment layers and increases pore space. Increased pore space increases the volume of the water fraction and thus promotes transfer of chemicals between the aqueous and solid phases. Ingestion and defecation of fine sediment particles by aquatic and benthic organisms increase the organic content of the particle surfaces, and therefore affect chemical partitioning between the solid and aqueous phases (see Section 2.1.3). Under normal conditions, intense bioturbation does not extend to depths greater than 10 cm below the sediment-water interface; however, on the continental shelf, bioturbation may extend to depths as great as 4 m (Kennett, 1982).

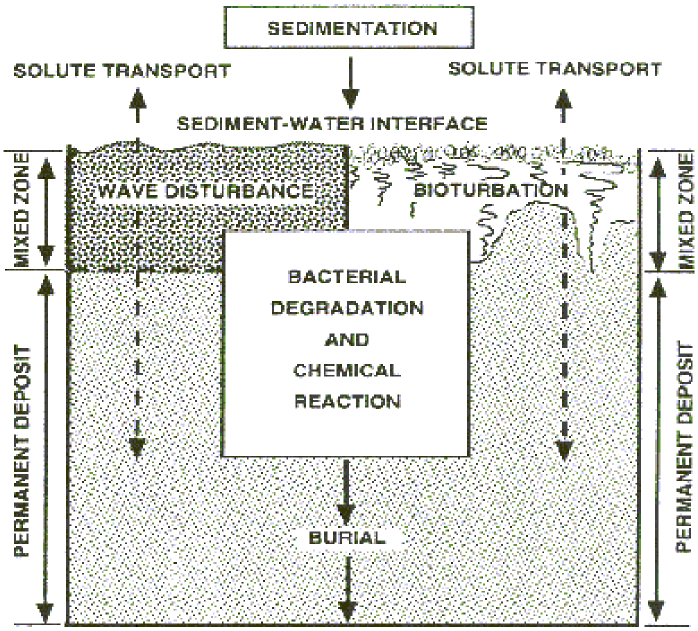


FIGURE 1-6. Sediment schematic diagram: Major chemical fate and transport processes

Sediment layer ages (i.e., time since deposition) and accumulation rates can be estimated by various methods, such as monitoring changes in the bottom profile and measuring the decay of radioisotopes (e.g., lead-210, cesium-137) that are incorporated into sediments at the time of deposition (USGS, 1998). Age dating and chemical analysis of sediment core samples may reveal variations in background conditions. These methods also can indicate the chemical concentrations associated with sediment layers that pre-date site-related activities, and therefore may be useful to estimate background concentration ranges. Tools that can be used to characterize sediment layers and sediment basins are listed in Table 1-2.

1.7.4 Sediment Texture

Sediment texture is determined by the relative amounts of three groups of 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.

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 1-7 shows a grain-size classification system.

The USDA soil classification diagram (Figure 1-8; USDA NRCS, 1998) can

TABLE 1-2. Parameters and data for characterization of sediments and sediment basins

Parameter	Data Source
Sediment age, sediment accumulation rate,	Radioisotope profiles (lead-210, cesium-137); precision stake survey using sonar techniques
Biologic activity and degree of vertical mixing of surface and subsurface sediment	Benthic and epibenthic fauna site data; radioisotope profiles
Wave erosion	Model from wind data from government or other sources; direct site measurements using pressure sensor
Erosion by tidal currents	Government sources (e.g., NOAA); direct site measurements using current meter
Erosion due to vessel activity	Coast Guard
Sediment removal or erosion due to other anthropogenic activity (e.g., dredging, construction)	Site records and interviews
Sediment transport	Various predictive sediment transport models
Sediment type	Sediment grain-size data
Bedforms	Side scan sonar
Bathymetry	NOAA charts, site-specific bathymetric mapping data
Sediment stratigraphy and geometry	Sediment core data; acoustic imaging
Geotechnical properties (in situ density and porosity, strength)	Geotechnical testing of site samples

NOAA = National Oceanic and Atmospheric Administration.

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 1-7. Sediment and soil particle sizes

be used to classify soils or sediments based on the relative percentage of each grain size. For example, as shown in the diagram, the USDA system classifies soil or sediment with 10% silt, 40% clay, and 50% sand as sandy clay.

The Unified Soil Classification System (USCS) is most commonly used to classify soils for engineering and geotechnical applications, but also can be used to classify sediments (ASTM, 2003). The USCS gives each group a two-letter designation based on the percentages of gravel, sand, silt, and clay (Figure 1-9).

Texture has a very significant effect on the distribution of naturally occurring metals in soil and sediment because metal concentrations tend to be inversely proportional to grain size. Metal

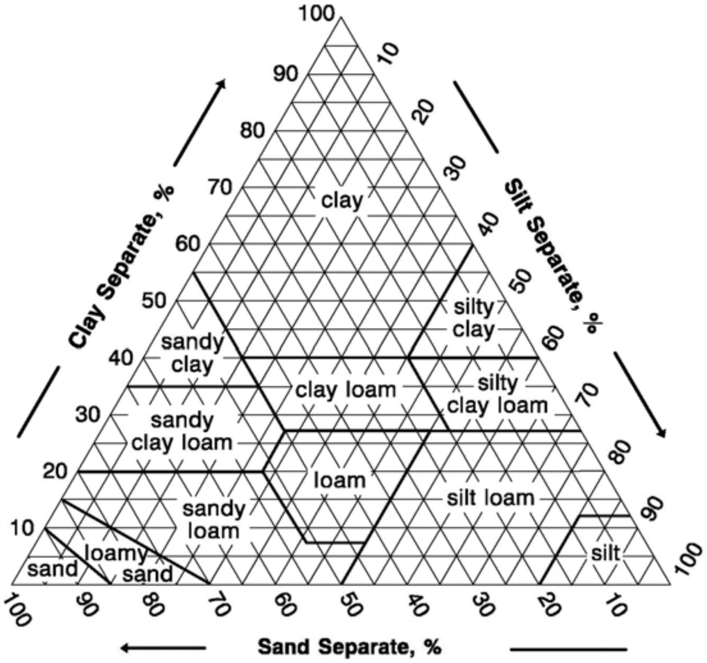
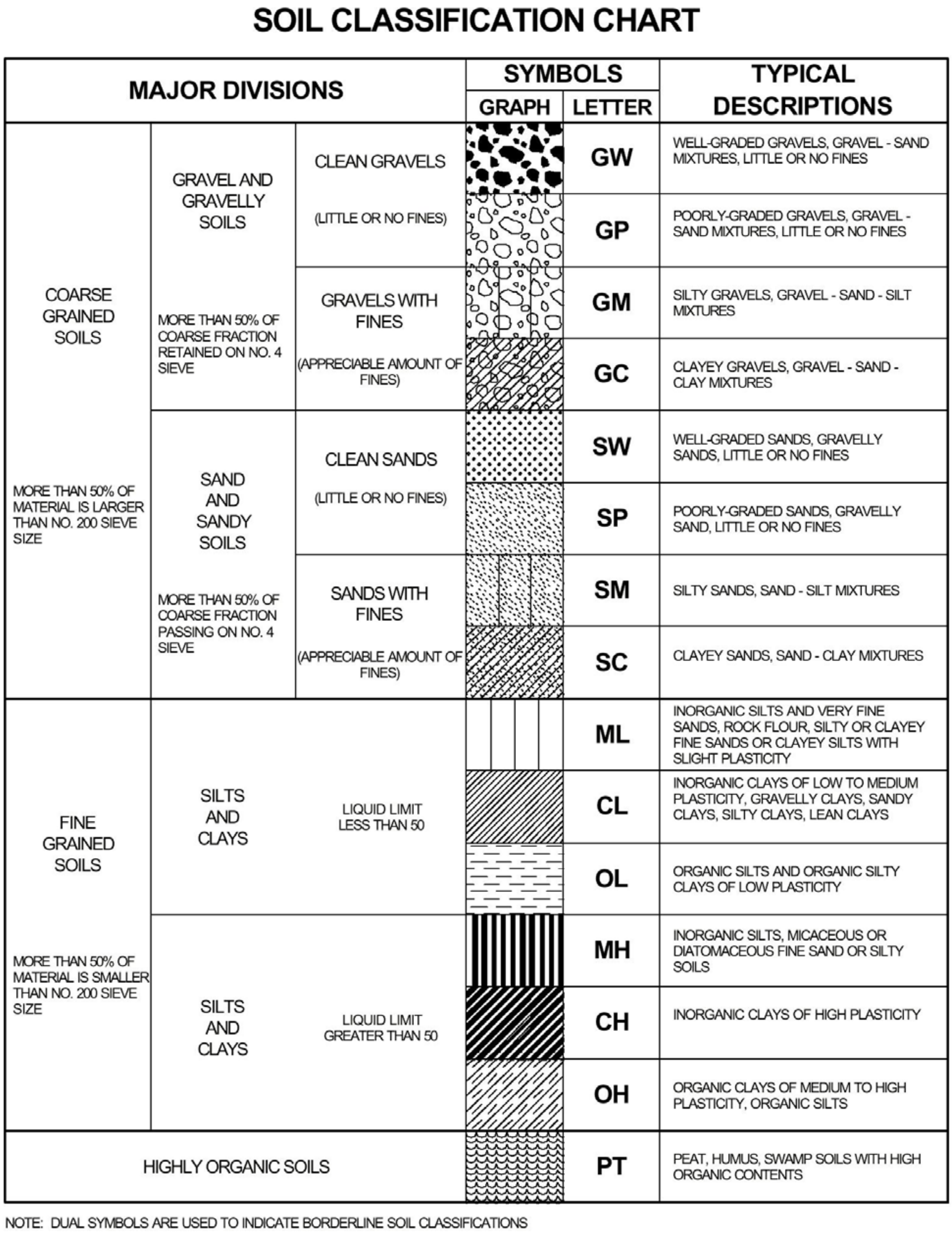


FIGURE 1-8. USDA guide for textural classification



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 1-9. Unified Soil Classification System

concentrations depend on the sorptive capacity of the soil or sediment particles. The minerals that form fine-grained sediments often contain relatively high metal concentrations within their crystalline structures. Also, fine-grained soils and sediments, particularly those with a high percentage of clay-size particles and organic colloids, have greater sorption capacity (i.e., ability to attract and retain metallic ions) than coarse-grained soils and sediments such as sands and gravels. Conversely, the minerals that form coarse-grained sediments generally contain low metal concentrations within their crystalline structures. (For example, most sand deposits consist primarily of particles of a single, nonmetallic mineral—quartz, i.e., silicon dioxide.)

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 1-10, 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 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. Positively charged metallic ions released from minerals during weathering, soil formation, and sediment transport are strongly attracted to the negatively charged surfaces. Soils and sediments with high concentrations of clay minerals and/or organic colloids therefore are likely to have 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.

1.8 References

American Society for Testing and Materials (ASTM). 2003. *D2488-00 Standard Practice for Description and Identification of Soils (Visual-Manual Procedure)*. West Conshohocken, PA.

Buchman, M.F. 1999. *NOAA Screening Quick Reference Tables*. NOAA HAZMAT Report 99-1. Seattle, WA: National Oceanic and Atmospheric Administration, Coastal Protection and Restoration Division.

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.

Canova, J.L. 1999. "Elements in South Carolina Inferred Background Soil and Sediment Samples." *South Carolina Geology, 41*: 11-25.

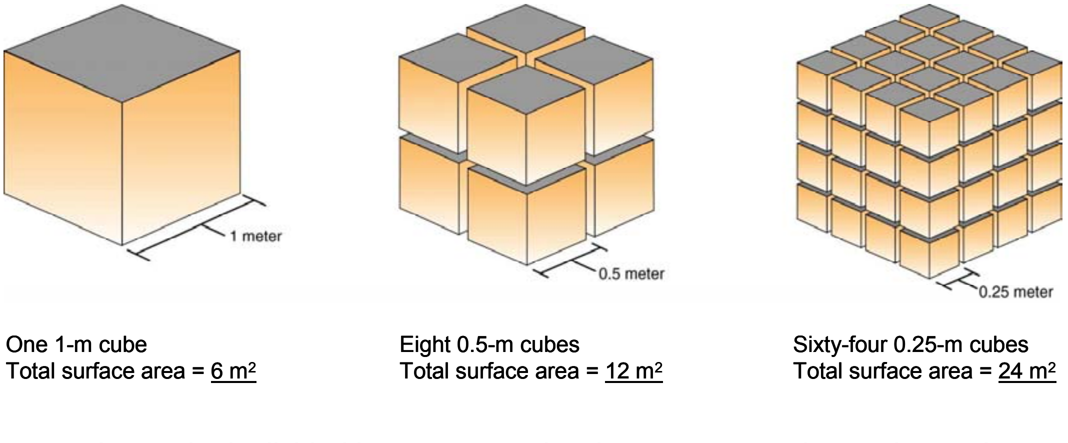


FIGURE 1-10. Relationship of volume, particle size, and surface area

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.

Department of the Navy (DON). 2002. *Navy Policy on Sediment Investigation and Response Action.* Memo from Chief of Naval Operations (N45) (William G. Mattheis, Deputy Director, Environmental Protection, Safety, and Occupational Health Division), to: Commander, Naval Facilities Engineering Command. 5090, Ser. N453E/2U589601. February 8.

Department of the Navy (DON). 2003. *Final Implementation Guide for Assessing and Managing Contaminated Sediments at Navy Facilities*. Naval Facilities Engineering Command. February.

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.

Kennett, J.P. 1982. *Marine Geology*. Englewood Cliffs, N.J.: Prentice Hall.

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.

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). 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." Washington, DC: 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*. Boston. Prepared by U.S. EPA Region 1, Boston, MA.

United States Geological Survey (USGS). 1998. *Short-Lived Isotopic Chronometers, A Means of Measuring Decadal Sedimentary Dynamics.* FS-73-98.

2. DATA REVIEW AND ASSESSMENT

As a component of a site/watershed environmental investigation, a background analysis 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 naturally occurring metals (e.g., arsenic) have several salient characteristics that can complicate the remedy decision process:

- ❑ They have low risk-based criteria (i.e., their background levels may exceed risk-based criteria).
- ❑ They often are 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 characteristics and ensuring that background analyses are scientifically accurate and technically defensible for appropriate decision-making. Sediment sampling costs can be very substantial (significantly greater than soil sampling costs); therefore, review and assessment of existing data is particularly important for sediment investigations. 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 be repetitive. 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 regarding response actions, and has minimized the probability of decision errors.

2.1 Assessment of Site Data

Background analysis should be based on a well-defined DQO decision-making framework. This framework is particularly useful when answering 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, 2000d). 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 assure 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, 2000c), which consists of the following steps:

1. Review 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 sediment 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 should be measured, such as aluminum, iron, magnesium, and silicon (for clay sediments), and/or calcium (for carbonate sediments). These supplementary data then should be combined with the existing site data.

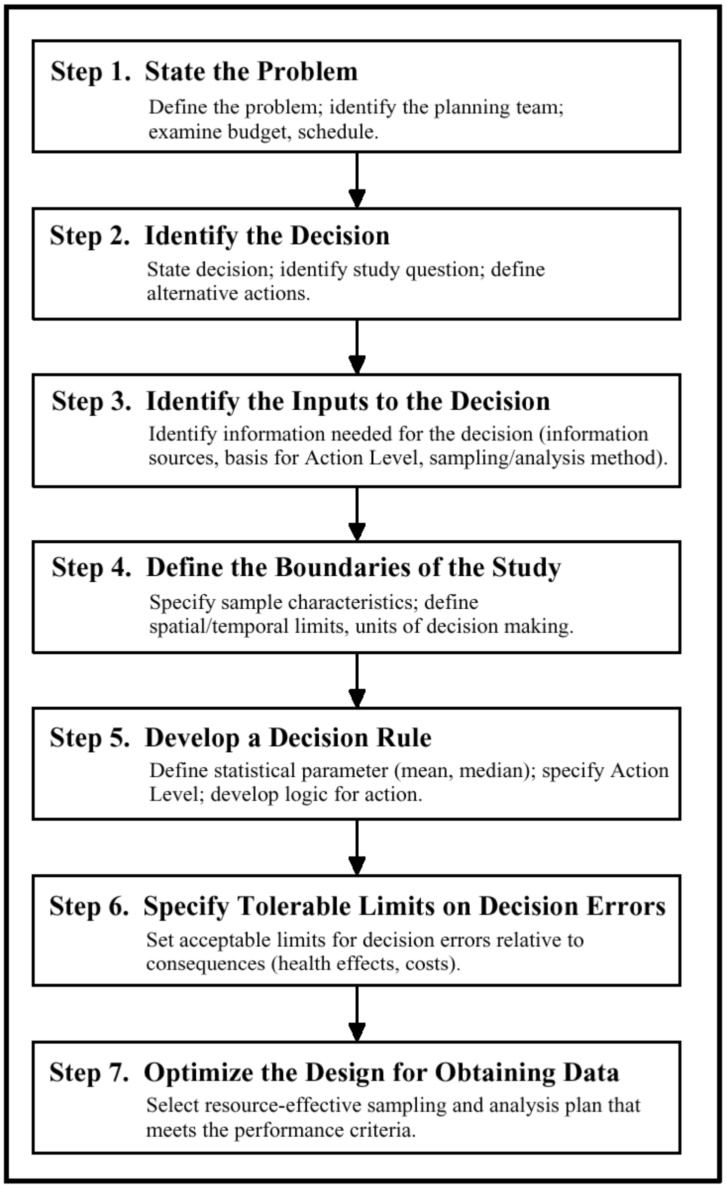


FIGURE 2-1. Seven-step DQO planning process (Source: U.S. EPA, 2000d)

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 can 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, identifying appropriate analytical methods, and evaluating site-specific elemental relationships.
- ***Geologists/Sedimentologists/Hydrogeologists.*** Geologists, sedimentologists, and hydrogeologists provide expertise in characterizing regional/site geology, identifying site sediment types, determining the adequacy of data coverage, and developing site background conceptual models.

- ***Marine or Freshwater Biologists.*** Biologists provide expertise in identifying ecological concerns such as potential receptors and pathways for exposure to COPCs and background chemicals.

- ***Toxicologists/Risk Assessors.*** Toxicologists/risk assessors provide expertise in characterizing site-specific risks to human health and the environment 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 experts in 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/Watershed Operational History and Conditions

A well-planned investigation requires information on the operational history of sites located within the sediment basin and watershed, and their physical, biological, and chemical characteristics. This information is necessary 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

As noted in Section 1.1.4, if a sediment site is potentially impacted by non-Navy contaminant sources, then a WCSD must be prepared prior to initiating any Navy sediment investigation. Preparation of the WCSD requires a review of the following aspects of the operational history of the site, sediment basin, and watershed area:

- Potential or known chemical or hazardous substance discharge points or areas at the site, within the sediment basin, or in the watershed

- Drainage channels and conduits that may transport sediments and chemicals from impacted areas within the watershed to the sediment basin

- Storage of hazardous substances or chemicals at the site, within the sediment basin, or at upstream areas within the watershed

- Transport, handling, or use of hazardous substances or chemicals at the site, within the sediment basin, or at upstream areas within the watershed.

The primary sources of operational history information are property owners and operators of sites within the sediment basin and watershed. Previous environmental or regulatory investigations that have compiled information directly applicable to the evaluation of background conditions at the sites will be of particular interest. If possible, the following information also should be acquired:

- History of the industrial, commercial, or residential uses of the site, sediment basin, and watershed

- Types of materials that were stored, handled, manufactured, or disposed of at the site, sediment basin, or watershed

- Locations of activities involving potential COPCs and potential releases

- Locations of historical filling activities and shoreline modifications

- Maps showing the topography, surface water drainage patterns, and locations of storm drains and sewer outfalls in the sediment basin and watershed area

- Maps and “as-built” diagrams of current and former structures related to storage, handling, manufacture, or disposal of suspected COPCs

- Locations of visible signs of potential COPC release, e.g., soil/sediment staining, discoloration, odor, and/or stressed ecological receptors (note that visible evidence of COPC release to sediments is commonly absent)

- Other potential sources of site history information, particularly former site employees or area residents.

Other sources of operational history information include the following: 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), National Oceanic and Atmospheric Administration (NOAA), and National Ocean Service (NOS) maps, and other maps of the sites and surrounding area; historical aerial photographs of the site and surrounding area; National Pollutant Discharge Elimination System (NPDES) permits and monitoring data; and employees or residents of neighboring properties. 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, 2003i] and Phase II [ASTM, 2003j] site assessments).

Physical Setting

The sediment background chemicals that occur at a sediment basin and the fate and transport of COPCs depend on physical characteristics such as the topography, geology, and hydrology of the sediment basin and watershed. Therefore, these physical characteristics should be evaluated carefully during the data assessment phase.

Topographic and Geographic Information

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

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

- ☐ Terrain conditions, surface drainage patterns, and the types of rock and soil that occur within the watershed (to identify potential chemical sources and transport pathways)
- ☐ Sediment basin water depth and bottom contour data (to evaluate the depositional environment)
- ☐ Ocean current and wind information (to evaluate sediment transport within the sediment basin)
- ☐ Locations of existing sampling points (to evaluate the spatial distribution of detected chemicals)
- ☐ Known or potential disposal or release locations (e.g., sewer outfalls) within the sediment basin and watershed (to identify potential chemical sources)
- ☐ Other pertinent information, including property boundaries, right of ways, and utility corridors.

Depending on the location of the sediment basin and watershed, the following maps, charts, and aerial photographs may be available:

- ☐ ***USGS Quadrangle ("Quad") Maps.*** Quads are 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 1:24,000 (1 inch = 2,000 feet) to 1:250,000 (1 inch = 20,833 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, 2002). The availability of large-scale maps will vary. Digital versions of many USGS quads are available in tiled format on the TerraServer Web site, a joint venture of the USGS and Microsoft Corporation (www.terraserver.microsoft.com).
- ☐ ***Geologic Survey Maps.*** Geologic maps that indicate the surface rock and soil types that occur in different areas are available for some of the areas covered by USGS quads.

- ***USDA Soil Conservation Survey Maps.*** United States Department of Agriculture soil surveys and maps are useful to identify the soils within a watershed that may be contributing sediment to the basin.

- ***Photogrammetric Aerial Survey Maps (i.e., Aerial Photographs).*** Aerial photographs of the land within a watershed may show features not illustrated on USGS quads (e.g., stressed vegetation). 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. High-resolution 9 × 9 inch photographs are available from both the NOS and the National Aerial Photography Program (NAPP). The NOS photographs are at scales between 1:10,000 and 1:50,000; currently more than 500,000 such photos showing the shoreline and offshore features exist in the NOS archives, dating from 1945 to the present year. NAPP is an interagency effort coordinated by the USGS to acquire new aerial photographic coverage of the conterminous United States every 5 to 7 years (approximately 90 percent coverage is currently available). The NAPP photographs are at a scale of 1:40,000; each covers the equivalent of one-quarter of a standard USGS 7.5-minute quad map. The NAPP images are also available in digital format as digital orthophoto quadrangles (DOQs), at a resolution of 1 meter per pixel. The DOQs are orthorectified so they can be read and measured as maps. Tiled versions of downsampled DOQs (linked to digitized USGS quad maps) can be freely viewed and downloaded from the TerraServer Web site (www.terraserver.microsoft.com).

- ***NOAA Nautical Charts.*** Although NOAA nautical charts are intended primarily for use as navigational tools, they also can serve as base maps for environmental investigations. NOAA charts depict the location of the shoreline, minimum water depths, and other useful information.

- ***NOS Coastal Survey Maps.*** Coastal survey maps accurately depict the shoreline, and shoreline features such as rocks and tidal flats and range in scale from 1:5,000 to 1:40,000. About 2,000 current coastal survey maps exist that cover the shoreline of the United States and its territories.

- ***NOS Environmental Sensitivity Index (ESI) Maps.*** ESI maps are a product of the Hazardous Materials Response Division of the NOS Office of Response and Restoration (ORR). ESI maps are plotted on 7.5-minute USGS quads, and contain three types of information: shoreline classification (i.e., sensitivity to oil releases), human-use resources, and biological resources. Collections of these maps (ESI atlases) have been prepared for most of the United States shoreline, including Alaska and the Great Lakes.

- ***NOS Estuarine Bathymetry (EB) Maps.*** EB maps depict estuary bottom topography based on available depth sounding data.

- ***NOS Hydrographic Survey (HS) Maps.*** HS maps define the topographic configuration of the bottoms of water bodies and adjacent land areas.

- ***Historical/Archaeological Maps and Photographs.*** Historical and archaeological maps and photographs can identify land areas and water bodies 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 sediment basin data. 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.

Regional and Site-Specific Rock, Soil, and Sediment Information

The distribution of background chemicals within a sediment basin depends upon the types of soil and rock that occur within the watershed surrounding the sediment basin (see Section 2.1.3). Therefore, the physical, geochemical, and biological characteristics of the rock and soil that occur in the watershed should be evaluated carefully. Accurate field geological observations and descriptions of rocks and soil are essential to plan a technically defensible background analysis strategy.

Concentrations of naturally occurring metals frequently are closely related to sediment mineralogy and particle size. Therefore, the absence of site-specific information on sediment characteristics 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 sediments. In this document, the term “sediment type” is used to differentiate sediments based on characteristics including texture (e.g., sand, silt, clay), and composition (e.g., mineralogy, organic content).

Sources of geologic, soil, and sediment 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 United States 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.

A qualified sedimentologist, geologist, or geotechnical engineer should evaluate sediment samples in the field to assess sediment mineralogy and texture. Sediment texture should be classified according to USDA NRCS, USCS, or ASTM standards (see Section 1.7.4). Sand, silt, and clay percentages should be estimated for each sample. In addition, to accurately quantify sediment characteristics, geotechnical testing and general chemistry analysis 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:

- ☐ Soil/sediment type (ASTM, 2003b)
- ☐ Particle size (ASTM, 2003d)
- ☐ Density (ASTM, 2003c)
- ☐ Moisture content (ASTM, 2003a).

The following general chemistry analyses also are strongly recommended:

- ☐ Soil/sediment pH (U.S. EPA [2003] SW-846 Method 9045)

- ☐ Oxidation-reduction (redox) conditions (American Public Health Association [1998] Standard Method 2580A)

- ☐ Cation exchange capacity of soil/sediments (U.S. EPA [2003] SW-846 Method 9081)

- ☐ TOC content (U.S. EPA [2003] SW-846 Method 9060).

Sediment Profiles

In some cases, the sediment vertical profile also should be investigated, particularly if background or contaminant concentrations are expected to exhibit significant temporal trends. Such trends can be interpreted through age-dating of sediment cores (see Box 2-1).

BOX 2-1. Age-dating of sediment cores in background analysis

In most background analyses, the concentration ranges of the investigated chemicals are assumed to be independent of time. As a result, sampling associated with background analysis usually is limited to surficial sediments (the upper 4 to 6 inches is typically the horizon of interest), and the resulting data represent recent depositional and chemical conditions. Sediment deposition, however, is an ongoing process, and sediment samples from different depth intervals tend to represent discrete historical periods.

In some sediment basins, background concentrations may vary significantly with depth, or among sediment layers with different characteristics. These chemical concentration trends reflect temporal variations in the watershed and sediment basin. For example, concentrations of anthropogenic background chemicals tend to be higher in the upper sediment horizons due to expansion of agricultural and urban areas in the watershed, and the associated increases in runoff and chemical loading. The temporal pattern of background concentrations can be established by comparing sediment concentrations at different depths. Anthropogenic background concentrations within the sediment profile usually show very gradual increases. In contrast, a chemical release will usually cause a distinct increase in COPC concentrations in sediments deposited at or shortly after the time of the release.

Sediment ages (i.e., time since deposition) at various depth intervals can be evaluated by a process known as sediment core age-dating. Sediment ages are most accurately estimated when reliable estimates of annual sediment accumulation rates are available. For example, during a USGS investigation of sediment cores from Coralville Reservoir, Iowa, annual accumulation rates were used to estimate the year of deposition for each sediment depth interval (Kalkhoff and Van Metre, 1997). The concentration profile of organochlorine compounds also was evaluated, and correlated with the sediment age data. Elevated organochlorine concentrations were detected at depths between 40 and 50 inches below the sediment surface, in sediment layers deposited during the 1960s and 1970s. The investigation concluded that the organochlorines are anthropogenic background chemicals associated with the widespread use of organochlorine pesticides in the Iowa River watershed during the 1960s and 1970s.

Although reliable sediment accumulation rate estimates are not available for most sediment basins, sediment profiles can be age-dated if suitable time-dependent signature indicators exist. Examples of such indicators include radioisotopes (e.g., lead-210, cesium-137 [^{137}Cs]) that are incorporated into sediments at the time of deposition (USGS, 1998). During a USGS investigation of sediment profiles at Lake Mead, Nevada, sediment ages were estimated by evaluating ^{137}Cs activities in sediment core samples (Covay and Beck, 2001). The activity of ^{137}Cs, a byproduct of atmospheric nuclear weapons testing, is a relevant signature indicator at Lake Mead due to its proximity to the Nevada Test Site. Significant ^{137}Cs concentrations first appeared in the atmosphere in about 1952, as a result of atmospheric nuclear weapons testing. Atmospheric ^{137}Cs concentrations peaked in about 1964. The USGS analyzed samples of each sediment layer to quantify concentrations of organochlorine and semivolatile compounds, and the activity of ^{137}Cs. The resulting concentration and age profiles were used to assess temporal trends in anthropogenic background concentrations. Two major assumptions are required to estimate accumulation rates from sediment age profiles: that deposition has occurred at a constant rate without interruption, and that the sediment column is undisturbed. Therefore, sediment sampling sites proposed for age dating should be evaluated to determine whether these assumptions are justified.

The location of the sediment basin, and its geochemical, hydrodynamic, and ecological characteristics, determine which signature indicators are most appropriate. Investigators have used a wide range of indicators. Radiocarbon concentrations in shells buried at different depths have proven to be a useful age-dating indicator (Mangerud, 1972). Diatom (microscopic algae that form a durable siliceous frustule) concentration profiles have been used to assess long-term environmental changes (Huvane and Cooper, 1999). These indicators have proven effective in paleoecological and paleoclimatic investigations; however, the profiles may not have the resolution required to adequately evaluate recent background trends.

Biological Characteristics

COPC and background chemicals may significantly affect the ecology of a sediment site; biological data therefore should be evaluated thoroughly during the data review and assessment phase. Because sediment site investigations typically include ecological investigations and risk assessments, the data needed to evaluate the biological characteristics of a sediment site usually are collected in conjunction with the physical and chemical data needed to evaluate background conditions.

Biological data should be assessed to determine whether a sediment area is suitable to serve as a reference site for background analysis by the Comparative Method. Biological data also can be used in the Comparative Method (see Section 4.1.1), if the data can be expressed as quantitative, location-specific measurements, and can be separated into distinct datasets for statistical comparison (i.e., potentially impacted data vs. background data).

The biological characteristics of a sediment site or basin usually are evaluated through sediment toxicity and bioaccumulation testing, and investigation of the benthic community structure.

Sediment Toxicity

The potential toxicity of chemicals in sediments is evaluated by conducting bioassays—laboratory tests in which test organisms are exposed to sediment samples from the site. Bioassay testing quantifies toxicity in terms of endpoints such as survival, growth, and reproduction. The test organisms normally should be representative of benthic or epibenthic organisms that inhabit the site, and testing methods should approximate actual site conditions. The 10-day amphipod toxicity test is the most common bioassay used to evaluate sediment toxicity. Sediment bioassay test methods and interpretation are described in U.S. EPA/USACE (1991, 1998) and U.S. EPA (1994a, 1994b).

Bioaccumulation

Chemicals incorporated directly into organisms through ingestion or direct exposure to sediments may bioaccumulate in the tissues of other organisms as they move up the food chain. Therefore, toxic effects may intensify as tissue concentrations increase through bioaccumulation. Hydrophobic organic chemicals generally have higher bioaccumulation potential than hydrophilic organic chemicals. Metals with high aqueous solubilities tend to be more bioavailable, and therefore have higher bioaccumulation potential. The bioaccumulation potential of sediment chemicals typically is quantified by 28-day flowthrough tests conducted in the laboratory using worms or clams as test organisms. Bioaccumulation potential also can be measured by in situ studies (e.g., caged fish testing). Sediment bioaccumulation testing and interpretation is described in U.S. EPA/USACE (1991) and U.S. EPA (2000a).

Benthic Community

Benthic invertebrates are usually the first organisms to be adversely affected by sediment chemicals; therefore, the effects of sediment chemicals on an ecosystem can be evaluated by investigating the structure of the benthic community. Parameters used to assess the benthic community include species diversity, abundance, and biomass. In addition to sediment chemicals, many other environmental factors, including food availability, water quality (e.g., temperature, depth, light, salinity), seasonal cycles, and predation will affect the benthic community. Therefore, benthic investigations should be designed, conducted, and interpreted by an experienced benthic biologist. Benthic data can be useful to screen and identify nonimpacted reference sites for background analysis by the Comparative Method. Benthic sampling and analysis methods are described in U.S. EPA (1987, 2000b).

Target Chemicals

After evaluating the operational history of potentially contaminated sites within the watershed and sediment basin, and the geochemical and anthropogenic characteristics of the potentially impacted sediments, 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 2-2).

Background Areas (Reference Areas)

Initial assessment of background conditions may include identification of potential reference areas. Areas of sediment that are likely to be free of site-related impacts (e.g., areas located upstream of the investigation site and suspected chemical release sites) should be evaluated for this purpose. Reference areas also must have physical, geochemical, biological, and anthropogenic characteristics similar to the impacted portions of the sediment basin.

Existing regional background data for sediments with characteristics similar to those of site sediments may be useful for screening purposes (e.g.,

BOX 2-2. Identifying target analytes

COPCs – Select based on Navy operational history and data from previous investigations.

Examples:

- Arsenic, chromium, and mercury (ship maintenance and building, aerial fallout, sewage effluent, fungicides)
- Copper, lead, tin, and zinc (old paint, marine antifoulants)
- Antimony, copper, and lead (firing ranges)
- Arsenic and pesticides (wood preservatives, pesticide equipment rinsing, agricultural runoff)
- Chromium (metal plating, alloys)
- Lead (leaded gasoline, battery 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)

during site inspections), and to support background analysis conclusions. If suitable reference areas do not exist within the same sediment basin, it may be necessary to identify other watersheds/basins that are free of significant contamination and have physical, geochemical, biological, and anthropogenic characteristics similar to those of the investigation site. The operational history and historical aerial photographs of a watershed and sediment basin, as well as hydrographic and geologic maps, can provide valuable information for defining potential background areas (U.S. EPA, 1992a).

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

❑ ***Identify pertinent operations and historical uses of the properties surrounding the sediment basin.*** Current operations and past uses or activities at locations upstream of the potential background area should be examined to evaluate known or potential contaminant sources, locations of past releases, or locations of contaminant disposal.

❑ ***Compare physical characteristics.*** The topographic, geographic, geologic, hydrologic, and anthropogenic characteristics of the potential background area should be compared to the corresponding characteristics of the potentially impacted portion of the sediment basin. Rock types, soil types, fill areas, and land use practices within the watershed should be identified to evaluate potential chemical sources and transport pathways.

❑ ***Assess the hydrodynamic regime.*** To locate potential background areas, an understanding of the hydrodynamic regime of the sediment basin is required. After evaluating the hydrodynamic regime and identifying a potential background area, chemical, biological, and ecological data should be used to investigate the area before it can be considered a suitable background (or reference) area.

❑ ***Formulate a CSM.*** Information on known or potential contaminant sources should be integrated with the potential fate and transport pathways to develop a CSM.

The process described above provides a technically defensible basis for identifying appropriate background areas. This process is especially critical if background datasets are needed in the analysis.

Box 2-3 provides an overview of the types of information and data that should be assembled and

BOX 2-3. Assembling operational, physical, and biological data

Reviewing the operational history and the physical, biological, and chemical characteristics of a sediment basin and its watershed often involves assembling and reviewing a large amount of information and data:

- **Operational history.** Much information regarding the operational history and chemical characteristics of a sediment basin and watershed—including potential contaminant sources, chemical transport mechanisms, exposure routes, and receptors—will be gathered during preparation of the WCSD (Section 1.1.4). Property owners and operators of sites and facilities within the watershed are valuable sources of information. Results of previous environmental investigations at the sediment basin and neighboring or related sites are of primary interest.
- **Physical setting.** Topographic and geographic information is acquired primarily from maps and aerial photographs. Geologic and soil survey maps, sediment profiles, and careful field observation of local geology provide the soil and rock type information essential for evaluating sources of natural background chemicals within a watershed. NOAA, NOS, and other charts of the marine and freshwater environment provide the information required to assess hydrodynamic characteristics and depositional environments within a sediment basin.
- **Biological characteristics.** The biological data needed to support a sediment background evaluation are typically collected during site assessment phase, along with the data needed to evaluate risks to ecological receptors. The biological characteristics of a sediment basin are usually evaluated through sediment toxicity and bioaccumulation testing, and investigation of the benthic community structure.

Evaluating the mass of information collected during the data review and assessment phase requires careful coordination by the project team. Compiling the data in a GIS can be very helpful for managing and evaluating the data. For investigations requiring review and assessment of extremely large quantities of information and data, it may be beneficial to apply data mining techniques. These techniques can reveal previously unknown patterns and relationships within the data through a combination of machine learning, statistical analysis, predictive modeling, and database technology. The large volumes of data and relationships revealed by data mining techniques can be illustrated with statistical and visual models such as decision trees and 3-D graphics.

reviewed before applying the background analysis techniques presented in this document.

2.1.3 Assess Sediment Geochemistry

The naturally occurring chemicals that occur in soils and sediments originate in the rocks that form the earth's crust. Therefore, the types and concentrations of soil and sediment background chemicals depend primarily on the composition of the parent rocks and their component minerals. Geochemical processes transport and redistribute these naturally occurring chemicals, particularly metals, in the geologic environment. Therefore, geochemical knowledge is essential to understand the origins of natural background chemicals in both soil and sediment, and is particularly useful to differentiate between background metals and metals that may be associated with a chemical release. The following sections present an overview of the important geological and geochemical principles relevant to background analysis.

Metals in Rocks and Minerals

This section presents an overview of basic geology and commonly observed patterns of metal distribution in rocks and minerals. These fundamental concepts are very useful for background analysis and other investigations that require knowledge of the geochemistry of natural occurring metals in soils and sediments.

Rocks and Their Composition

Rocks are classified as igneous, sedimentary, or metamorphic according to mode of origin (see Box 2-4), 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,

BOX 2-4. 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 typically are 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 often are exposed when rocks are uplifted and eroded to form mountain ranges.

with or without olivine (pyroxene and olivine are rich in iron and magnesium).

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 2-2 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.

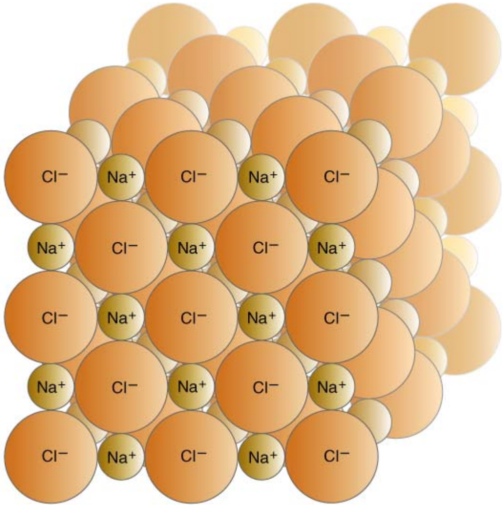


FIGURE 2-2. Ionic bonding of metal and nonmetal atoms in halite (common table salt)

Arsenopyrite, or iron arsenide sulfide, is one example of a mineral that may contribute to elevated 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 2-1 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,

TABLE 2-1. 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)	$FeAsO_4 \cdot 2H_2O$

titanium, and phosphorus—account for more than 99% of the mass of the earth's crust (Alloway, 1990). Elements classified as heavy metals (Table 2-2) 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 typically are less than 0.01% (100 mg/kg) and rarely more than 0.1% (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 2-2. Selected naturally occurring heavy metals

Name	Symbol	Atomic Weight (grams/mole)	Name	Symbol	Atomic Weight (grams/mole)
Antimony	Sb	121.76	Molybdenum	Mo	95.94
Arsenic	As	74.92	Nickel	Ni	58.69
Cadmium	Cd	112.41	Selenium	Se	78.96
Chromium	Cr	51.99	Silver	Ag	107.87
Cobalt	Co	58.93	Tin	Sn	118.71
Copper	Cu	63.55	Thallium	Tl	204.38
Gold	Au	196.97	Tungsten	W	183.85
Lead	Pb	207.20	Uranium	U	238.03
Manganese	Mn	54.94	Vanadium	V	50.94
Mercury	Hg	200.59	Zinc	Zn	65.39

Table 2-3 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

TABLE 2-3. 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).

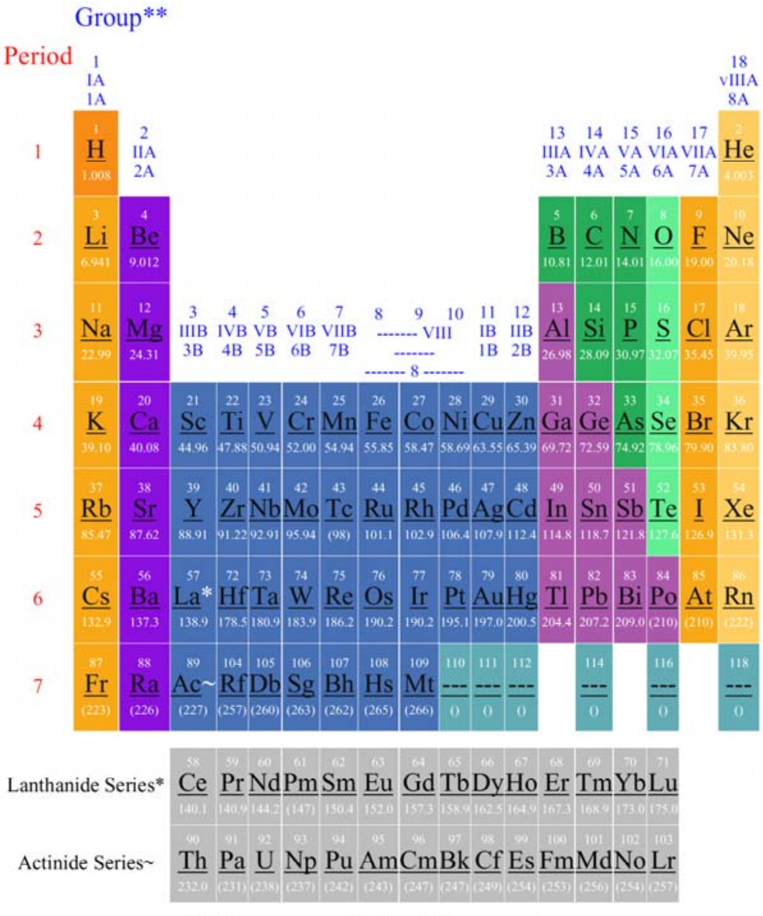
nickel, tend to vary proportionally, regardless of rock type; for example, even though chromium and nickel concentrations vary over wide ranges depending on the rock type, the concentration ratios of these two metals tend to be relatively constant.

Elemental Distribution and Association

As a result of the geochemical properties of the elements, certain types of rocks and the soils and sediments derived from 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.

The electronic structure relationships reflected in the periodic table (see Figure 2-3 and Box 2-5) 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 is also an important factor in elemental association—metals with similar ionic radii are commonly associated due to their ability to substitute for each other.

Early in the twentieth century, the geochemist V.M. Goldschmidt developed a classification system to explain the distribution of chemical elements in the earth's crust and atmosphere. Goldschmidt's system classifies the elements based on their affinity for iron (siderophiles), sulfur (chalcophiles),



Group**

Period	1 IA 1A	2 IIA 2A	3 IIIB 3B	4 IVB 4B	5 VB 5B	6 VIB 6B	7 VIIB 7B	8	9 ------- VIII ------- ------- 8 -------	10	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 2-3. Periodic table of the elements

BOX 2-5. Column groups in the periodic table

Elements in the same column of the periodic table (Figure 2-3) 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. Therefore, it is more difficult to accurately predict their chemical behaviors; however, the relationships reflected by the periodic table can still be used for this purpose.

silicates (lithophiles), and atmospheric gases (atmophiles).

Goldschmidt's classification system reflects commonly observed patterns of elemental distribution and association of metals in rocks, soils, and sediments (see Table 2-4). Because aluminum is a major component of many silicate minerals, the lithophile metals also exhibit an affinity for aluminum.

Igneous Rocks

Igneous rocks are common parent rocks for soil and sediment formation, and are the original sources of the geologic materials that form most sedimentary and metamorphic rocks. Therefore, the factors that control elemental association in igneous rocks strongly influence elemental association in many soils and sediments. 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.

TABLE 2-4. 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.

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 commonly found together (associated) in felsic rocks.
- Cations with smaller radii and relatively high charges, including uranium, thorium, boron, beryllium, molybdenum, tungsten, 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.
- Some of the chalcophile elements (see Table 2-4) substitute to some extent for major cations in silicate structures (e.g., lead and thallium substitute for potassium, and zinc substitutes for iron and magnesium). 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% 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. Erosion of previously formed sedimentary rocks is also a very important source of new sediments.

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 the 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). Soils and sediments formed by weathering and erosion of these sedimentary rocks also will have elevated metal concentrations. Table 2-5 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 generally is 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).

Weathering

Soils and sediments are formed by the gradual breakdown of the rocks that form the earth's crust,

TABLE 2-5. Average concentrations of minor elements in shales, sandstones, and carbonate rocks (mg/kg)

Element	Shales	Sandstones	Carbonates	Element	Shales	Sandstones	Carbonates
Barium(a)	600	10–100	10	Nickel	80	2	20
Lithium(a)	60	15	5	Lead	20	7	9
Rubidium(a)	140	60	3	Scandium	15	1	1
Strontium(a)	400	20	610	Thorium	12	1.7	1.7
Bromine	5	1	6.2	Titanium	4,600	1,500	400
Cerium	70	92	11.5	Uranium	3.5	0.5	2.2
Chlorine	180	10	150	Vanadium	130	20	20
Chromium	100	35	11	Yttrium	35	40	30
Cobalt	20	0.3	0.1	Zinc	90	16	20
Copper	50	1–10	4	Zirconium	180	220	19
Fluorine	600	270	330	Arsenic(b)	10	1	1
Gallium	25	12	4	Boron(b)	100	35	20
Germanium	1.5	0.8	0.2	Molybdenum(b)	2	0.2	0.4
Iodine	2	1.7	1.2	Phosphorus(b)	750	170	400
Manganese	850	10–100	1,100	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.

through a process known as weathering. Exposure to the atmosphere and the hydrosphere (i.e., surface and groundwater) allows physical and chemical weathering processes to gradually decompose solid rock. Biological activity also contributes to soil and sediment formation, particularly after breakdown of rocks has been initiated by physical and chemical weathering processes.

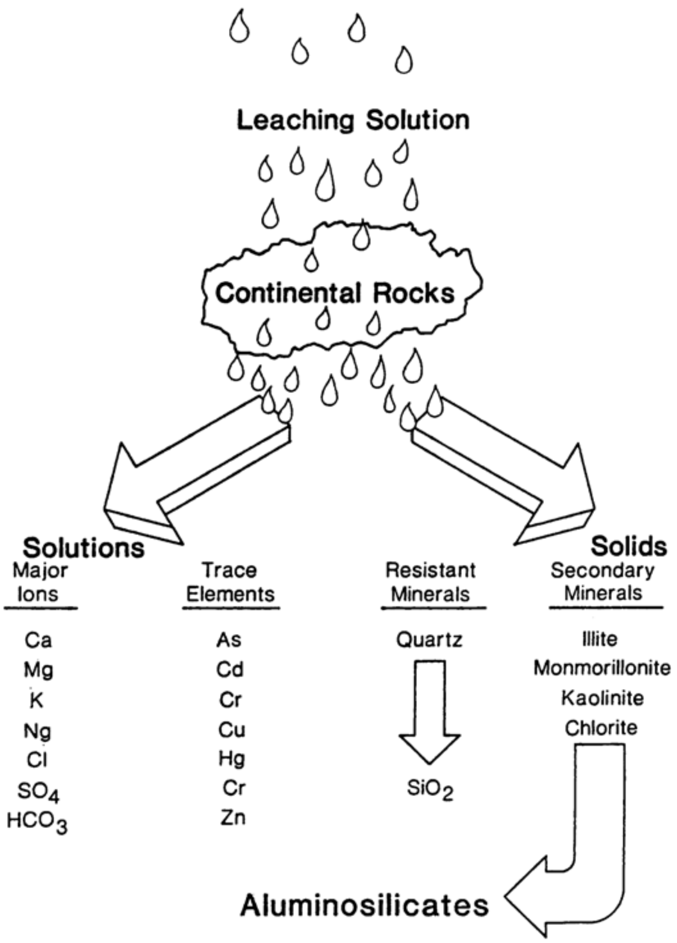
Physical Weathering

Physical weathering refers to the mechanical disintegration of rock by natural forces. 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 streams. High relief and rainstorms also facilitate physical weathering.

Chemical Weathering

Chemical weathering, the most important process in soil and sediment formation, mobilizes the chemical elements that make up rock-forming minerals. This process, shown in Figure 2-4, results in the physical disintegration of the rock, formation of new (secondary) minerals, and changes in 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. 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



Source: FDER (1988, Figure 1).

FIGURE 2-4. Conceptual model of the chemical weathering process

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 2-6 shows estimated rates of chemical weathering for different climates and rock types. Biochemical reactions also contribute to soil and sediment formation, as plants and animals metabolize the inorganic chemicals in weathered rock, and organic compounds are added to the broken and altered rock and mineral particles.

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 aluminum silicate) 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 capacity. Figure 2-5 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 2-7 shows the chemical weathering products of orthoclase, olivine, and the other common silicate minerals.

Although metals are mobilized and redistributed during the chemical weathering process, the concentration ratios of certain metals in soil and sediment often reflect the ratios that

TABLE 2-6. 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).

$$\begin{array}{ccccccccc} \text{Water} & \textit{plus} & \text{carbon dioxide} & \textit{yields} & \text{carbonic acid} & \textit{yields} & \text{hydrogen ion} & \textit{plus} & \text{bicarbonate ion} \\ H_2O & + & CO_2 & \rightarrow & H_2CO_3 & \rightarrow & H^+ & + & (HCO_3)^- \end{array}$$

$$\begin{array}{ccccccccccc} \text{Potassium feldspar} & \textit{plus} & \text{hydrogen ions} & \textit{plus} & \text{water} & \textit{yields} & \text{clay} & \textit{plus} & \text{potassium ions} & \textit{plus} & \text{silica} \\ 2K(AlSi_3O_8) & + & 2H^+ & + & H_2O & \rightarrow & Al_2Si_2O_5(OH)_4 & + & 2K^+ & + & 4SiO_2 \end{array}$$

FIGURE 2-5. Chemical reactions involved in the weathering of orthoclase

existed in the parent rocks: soils and sediments 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 soils and sediments that contain large quantities of clay often tend to be a function of aluminum concentrations.

Effects of Sediment and Water Chemistry on Metal Distribution

Chemical properties of sediment particles and the surrounding aqueous phase (i.e., water above the sediment/water interface and within the pore space) strongly influence the distribution of metals. The most important properties are:

- ❑ Aqueous solubility of the metal of concern
- ❑ pH of the aqueous phase
- ❑ Oxidation-reduction (redox) potential of the aqueous phase
- ❑ Ionic strength of the aqueous phase
- ❑ Affinity of the metal for organic carbon
- ❑ TOC concentrations
- ❑ Acid volatile sulfide (AVS) concentrations.

TABLE 2-7. 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).

Metals with high solubilities will tend to remain mobile in the aqueous phase, whereas low-solubility metals will accumulate in sediments as they precipitate out of solution and sorb to fine-grained sediment particles. Note, however, that aqueous solubility is highly dependent upon characteristics of the aqueous phase (e.g., pH, redox potential, ionic strength).

Low pH (acid) conditions tend to dissolve and mobilize metals, whereas high pH (basic) conditions tend to favor metal precipitation, resulting in increased metal concentrations in the sediment relative to the water. pH also affects the electric charge on the surface of sediment particles. The point of zero charge (PZC) is the pH value at which the particle surface effectively has no net charge. At pH values less than the PZC, a particle 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 2-8. As a result of their large surface areas and low PZC pH values, clays have a very high sorptive capacity. Within the normal 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 aqueous environment also affect the solubility and transport of metals. Under reducing conditions (as in anoxic sediments), many 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 often 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.

TABLE 2-8. Point of zero charge pH values

Mineral	pH Values Corresponding to Point of Zero Charge
Montomorillonite	2–3
Kaolinite	4.6
Goethite	7.3–7.8
Quartz	2–3

Ionic strength also affects the tendency of metals to remain in 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 (e.g., seawater) indicates that increased salt content leads to increased mineral solubility. Therefore, if the ionic strength of the water above and within the sediment pore space is high (as in seawater), metallic ions will tend to remain in solution. Conversely, if ionic strength is low (as in fresh water), metals will tend to precipitate out of solution, resulting in higher concentrations in the sediment.

Sediments with a high percentage of organic material (high TOC) will tend to have high metal concentrations because the organic particles have a high sorptive capacity. It also should be noted that high TOC concentrations may tend to limit the toxicity of metals to aquatic and benthic organisms because sorbed metals are not as bioavailable as dissolved metals.

The presence of sulfide ions in water also has been identified as a factor that affects the partitioning of metals between the solid and aqueous phases, and controls the toxicity of heavy metals to benthic organisms. Sulfide ions, which are produced in sediments through the breakdown of organic matter, form insoluble compounds when they react with metals. High sulfide concentrations (measured as AVS) therefore cause metals to precipitate out of solution and accumulate in the solid phase (where they are relatively nontoxic) (Di Toro et al., 1990, 1991).

Metal Concentration Ranges in Sediment

Natural geochemical processes can cause certain metals to become highly enriched in some environments. Table 2-9 lists background metal concentration ranges for several types of sediments and soils found in the United States (i.e., southern

TABLE 2-9. Average range and low- to no-effect levels of selected inorganics in sediments and soils (mg/kg unless otherwise noted)

Media and Source	Ag	As	Ba	Cd	Cr	Cu	Fe%	Hg	Mn	Ni	Pb	Zn
Sediments[a,b]												
Nonpolluted, Great Lakes[a] (USACE, 1977)	—	<3	<20	—	<25	<25	<1.7	<1.0	<300	<20	<40	<90
No effect level (Persaud et al., 1989)[b]	—	4.0	—	0.6	22	15	2.0	0.1	400	15	23	65
Effects range low, marine sediments (Long et al., 1995)[b]	1.0	8.2	—	1.2	81	34	—	0.15	—	20.9	46.7	150
No adverse biological effects, marine sediments (Washington State Department of Ecology [WDOE], 1991)[b]	6.1	57		5.1	260	390	—	0.41	—	—	450	410
Control sediments, Southern California (Lee et al., 1989)[a]	0.06–2.0	3–15	—	0.001–2	6.5–40	2.8–30	—	<1.0	—	<20.0	<10.0	<70.0
Control sediments, Puget Sound (Lee et al., 1989)[a]	1.2	3–15	—	3.1–18.3	20.9	10–50	—	0.02–0.12	—	13.0	8	—
Control sediments, Yaquina Bay (Lee et al., 1989)[a]	0.55	—	—	0.47	19.3	6.3	—	—	—	14.5	5.5	26.3
No effect threshold, freshwater sediments[b] (Environment Canada, 1992)	—	3.0	—	0.2	55	28	—	0.05	—	35	23	100
Lowest effect level, freshwater sediments (Persaud et al., 1992)	0.5	6	—	0.6	26	16	—	0.2	460	16	31	120
Threshold effect levels for freshwater sediments (Environment Canada, 1994)	—	5.9	—	0.596	37.3	35.7	—	0.174	—	18.0	35.0	123.1
Threshold effect levels for marine sediments (Environment Canada, 1994)	0.73	7.24	—	0.676	52.3	18.7	—	0.13	—	15.9	30.2	124
Effects range low, freshwater (Ingersoll et al., 1995)	—	13	—	0.70	39	41	20	—	730	24	55	110
Soils (control values)[a]												
Average and common range in natural soils (summarized in Shields, 1988)	0.05 0.01–5	5 0.1–40	430 100–3,500	0.06 0.01–7	100 5–3,000	30 2–100	— —	0.11 0.01–0.8	600 100–4,000	40 5–1,000	10 2–200	50 10–300
Average concentration in earth's crust (Merck, 1989)	—	0.5	500	0.1–0.2	100–300	70	5	0.5	850	180	20	200
Average concentration in earth's crust (CRC Press, 1992)	0.07	1.8	425	0.2	100	55	5.63	0.08	950	75	12.5	70
Relative abundance in soils (Martin and Whitfield, 1983)	0.05	6.0	—	0.35	70	34	4.0	—	1,000	50	35	90

(a) Nonpolluted and control values represent natural background.

(b) No-effect refers to no measurable impact to benthic organisms when exposed to sediments with stated levels of metals.

California, Puget Sound, and Yaquina Bay control sediments, and Great Lakes nonpolluted values). If metal concentrations exceed the upper levels of the background ranges shown in this table, a chemical release should be suspected; however, given the wide variations in natural metal concentrations that can occur in sediments, background analysis will be necessary to confirm this hypothesis. The table also lists commonly used risk-based screening criteria for sediments. As the table indicates, it is not unusual for metal concentrations in background sediments to exceed risk-based screening criteria.

TABLE 2-10. Nickel/chromium concentrations in different rocks

	Nickel		Chromium	
Rock Type	Average	Range	Average	Range
Basaltic igneous	140	45–410	220	40–60
Granitic igneous	8	2–20	20	2–90
Shales and clays	68	20–250	120	30–590
Black shales	50	10–500	100	26–1,000
Limestone	20	—	10	—
Sandstone	2	—	35	—

Source: Alloway (1990, Table 7-2).

Elemental Correlation

As noted previously, concentration ratios of certain groups of metals in naturally occurring rocks, soils, and sediments often are restricted to relatively narrow ranges. These natural relationships are very useful for evaluating background metal concentration ranges, and are central to geochemical association analysis—Step 1 of the Geochemical Method (Section 3.2).

Table 2-10 lists typical nickel and chromium concentrations for a number of different rock types. Although their concentrations may vary considerably among different rock types, certain elements nevertheless tend to be strongly correlated. Nickel and chromium are examples of metals that are commonly correlated in rocks, soils, and sediments. For example, as shown in Figure 2-6, 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 powerful tool for background analysis. 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.

The use of elemental correlation as a tool for evaluating background metal concentration ranges in sediments is well documented in the geochemical literature. 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). 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 (Figure 2-7) (FDER, 1988). The use of aluminum as a reference element also is described in Hanson et al. (1993), Schropp et al. (1990), and Bertine and Goldberg (1977).

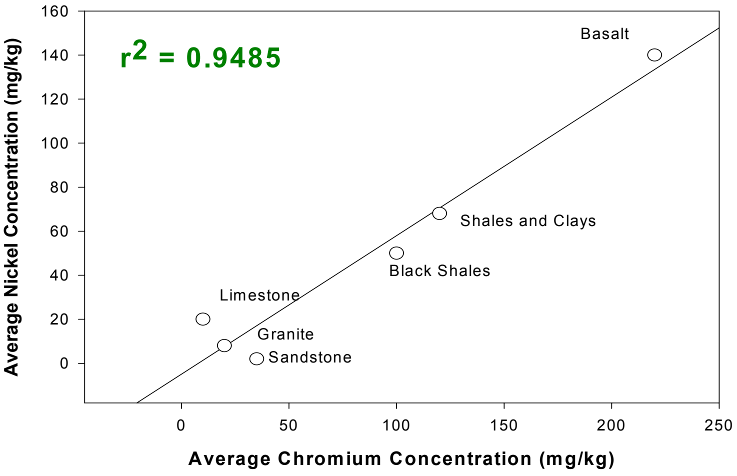
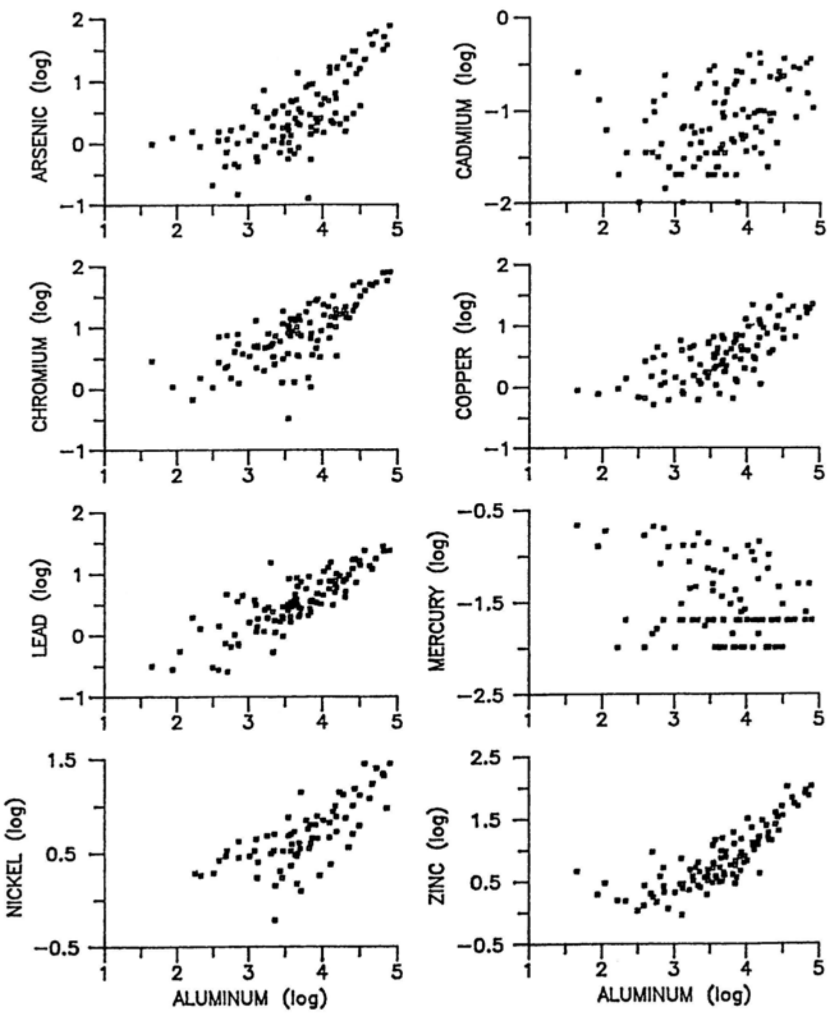


FIGURE 2-6. Correlation plot of nickel vs. chromium concentrations



Source: FDER (1988, Figure 7)

FIGURE 2-7. FDEP scatter plots

Aluminum is a very useful reference element for background analysis because concentrations of certain metals in natural (uncontaminated) sediments often are directly proportional to aluminum concentrations. As noted above, most clay minerals are aluminosilicates, and clay minerals have a strong affinity for metallic cations; therefore, trace metal concentrations in sediments that contain large quantities of clay tend to be strongly correlated with aluminum concentrations. Aluminum is particularly suitable for use as a reference element in background analysis for several reasons: it is the most abundant naturally occurring metal, it is highly refractory (i.e., resistant to weathering), and its concentrations generally are not influenced by chemical releases.

Reference metals other than aluminum also can be used for background analysis. In some environments, iron can be used as a reference metal. For example, in Hawaii, iron is ubiquitous in soils and sediments 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 sediments often show a high degree of correlation with iron concentrations. It should be noted, however, that iron is more reactive than aluminum, and tends to be remobilized in reducing (e.g., anoxic) environments; therefore, iron may not be a suitable reference element for some sediment sites. The use of iron as a reference element is described in Trefry et al. (1985), Daskalakis and O'Connor (1995), and Morse et al. (1993). Other elements, such as lithium (Loring, 1990) and rare earth elements (Olmez et al., 1991), also have been used as reference elements.

A suitable reference element can be identified only after carefully evaluating the relationship between a potential reference element and a suspected COPC (see Section 3). However, selection of a reference element may depend on the geologic characteristics of the region in which the sediment investigation site is located. In general, aluminum appears to be the most useful reference element for sites located on the east coast of the United States, whereas iron or other reference parameters may be more useful for sites located on the west coast (including Alaska and Hawaii). On the east coast, most watersheds drain the Appalachian Mountains and adjacent upland terranes with relatively constant granitic compositions. The minerals that compose granitic rocks weather primarily to clays and other aluminosilicate minerals. Trace metal/aluminum ratios in sediments from different areas of the east coast therefore are relatively constant. The geology of the west coast is more varied (e.g., mafic and ultramafic igneous rocks are more common); therefore, trace metal/aluminum ratios generally are not as constant. As a result, iron, other reference elements, or reference parameters such as grain size and TOC may be more useful in the west.

Correlation with Grain Size and TOC

As noted above, coarse sediments generally have low sorption capacity and a low proportion of metallic mineral particles, whereas fine sediments generally have high sorption capacity and a high proportion of metallic minerals. In addition, sediments with a high percentage of organic material (high TOC) tend to have high metal concentrations because the organic particles have relatively high sorption capacity. Therefore, metal concentrations in sediment tend to be inversely proportional to grain size, and directly proportional to TOC content.

The relationships between metal concentrations and TOC content or grain size have been successfully used for background analysis (Cal/EPA, 1998). Site-specific data should be evaluated carefully to determine whether the relationship is strong enough to be used for background analysis.

It also should be noted that in some situations, sequential extraction or other chemical analysis techniques may be useful to distinguish between contamination and natural background conditions. However, this approach would be successful only when the chemical species or adsorption characteristics of a metal that may be attributable to a chemical release are not the same as those of the naturally occurring forms of the same metal.

2.1.4 Develop Hypothesis

The project team should develop a hypothesis 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 portions of the sediment basin. The following hypotheses should be considered:

- ***Example Hypothesis 1*****:** The targeted chemicals are likely to occur as background chemicals within the watershed and sediment basin, and also may occur as the result of a chemical release, but are likely to have impacted only portions of the sediment basin (i.e., localized or hotspot impact).

- If portions of the sediment basin are impacted by a COPC metal, concentrations in certain areas of the site will be elevated relative to the underlying background concentrations. Consequently, the sediment basin dataset will contain two distinct populations that can usually be distinguished by graphical techniques (see Section 2.2). In addition, as a result of elemental association, concentrations of naturally occurring metals tend to be correlated. If a chemical release has occurred, metal concentration ratios in samples from the impacted areas will tend to deviate from the underlying natural association relationship. Therefore, the Geochemical Method (Section 3) can usually distinguish background concentrations from concentrations that represent contamination. If necessary, however, background also can be evaluated by the Comparative Method (Section 4).

- ***Example Hypothesis 2*****:** The targeted chemicals are likely to occur as background chemicals within the watershed and sediment basin, but also may occur as the result of a chemical release, and, if so, are likely to have impacted the entire areal extent of the sediment basin (i.e., basin-wide impact).

- If the entire investigation 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. (However, a release that covers the entire areal extent of a site is relatively uncommon.) In this case, the sediment basin dataset alone may not be enough to distinguish between background and concentrations that represent a chemical release. If data from a suitable reference (background) area are available, background can be evaluated by the Comparative Method (Section 4). The combined potentially impacted and background datasets also can be used to identify background concentration ranges by the graphical techniques presented in Section 2.2, or by the Geochemical Method (Section 3).

Once a defensible hypothesis has been developed, it should be presented to the stakeholders. After concurrence of all the stakeholders on the appropriateness of the chosen hypothesis, appropriate methods for conducting the subsequent background analyses should be selected.

2.1.5 Determine Whether Adequate Sediment Data Exist

Once quantitative and qualitative watershed and sediment basin information has been compiled and reviewed, the project team must determine whether the existing data are adequate. The graphical and analytical tools presented in this guidance document require mid- to large-size datasets, which typically contain chemical data representing at least 20 to 30 sediment samples collected in both impacted and nonimpacted areas of the basin.

If the available dataset represents only a few samples collected from a small background area, the background analysis methods presented in the following sections are not likely to yield satisfactory results. As discussed in Box 2-6, past experience at Navy sites demonstrates that background analysis based on small datasets should be avoided whenever possible.

If the project team concludes that additional data are needed to proceed with the background analysis, a sampling and analysis plan should be developed within the framework of the DQO process (as described in the following section).

2.1.6 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 Sampling and Analysis Plan (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 2.1.7). As additional data needs become evident (e.g., if it is necessary to expand the sampling area, or to collect data from a reference site), the SAP should be amended or a SAP addendum should be prepared.

For the Comparative Method, the extent of the reference or background area must be clearly defined and justified in the SAP. Sediments in the reference area should be physically, geochemically, biologically, and anthropogenically similar to site sediments. 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 nonproductive 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 sediment data, which typically are collected by judgmental sampling—preferential sampling of locations where contamination is believed likely to occur based on information regarding the site history (areas impacted by outfalls or other discharge points are usually the focus of sediment investigations). In general, under such conditions, additional sampling will focus on areas of the sediment basin 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. The detailed aspects of these techniques are beyond the scope of this guidance; further information is available

BOX 2-6. Analysis based on small background datasets

Occasionally, the available dataset at a given site represents only a few "background" samples (e.g., less than 20 measurements) collected in an area believed to be nonimpacted. These small background datasets often are encountered when stakeholders disagree on the extent of the impacted areas. Such disagreements result in the identification of only a few small locations remote from suspected contaminant sources as "background areas." Due to the small size of such background areas, only a few samples can be collected, and the resulting datasets are too small to reliably characterize background conditions.

If necessary, it may be possible to generate a rough estimate of the background concentration range for a chemical by evaluating a small dataset. In such cases, the background analysis must include a thorough review of site history (review of a chronological series of aerial photographs may indicate which areas are most likely to be nonimpacted). Basin hydrology and sedimentology also should be evaluated, followed by a comprehensive analysis of the available site and background area sediment data (as described in Sections 2 and 3 of this document). Although these procedures can provide evidence to justify selection of the background area and may yield defensible background range estimates for the target chemicals, background range estimates based solely on a small dataset are likely to be questioned as unreliable.

Past experience at Navy sites has demonstrated that relying solely on small background datasets to estimate background concentration ranges often leads to the following problems:

- **Unreliable Statistical Analyses:** When the background dataset is small, it is difficult or impossible to make statistically robust and reliable estimates of background concentration ranges; furthermore, statistical comparison of the background dataset to the potentially impacted population will be problematic and prone to large errors.
- **Unexplainable Large Concentration Ranges:** In many instances, concentration values within a small background dataset are highly variable and cover a very wide range. These situations tend to confuse the analysis and lead to the erroneous labeling of true background concentrations as statistical outliers. This often raises questions as to whether the higher concentrations represent background conditions or indicate the presence of isolated hotspots.
- **Unproductive Negotiations:** Limited background datasets are usually questioned as not representative of actual conditions, leading to complicated negotiations, qualitative discussions, and speculative conjectures. In many cases, the background dataset has become the primary subject of remedial discussions rather than a useful tool for decision-making.

in Matheron (1971), Journel and Huijbregts (1978), Isaaks and Srivastava (1989), USACE (1997), and the American Society of Civil Engineers (ASCE, 1990a,b). Standard guides for application of geostatistical techniques in environmental site investigations are provided by ASTM (2003e-h).

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. Two such designs are simple random sampling and systematic sampling. If systematic sampling is used, sample locations could be at the nodes of a square or triangular grid system that is placed at a random starting place 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 sediment 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. In 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 and other designs are discussed in U.S. EPA (2002a) and Gilbert (1987).

The SAP also should address issues that can arise when sediment data collected at different times are combined. It may be necessary to identify potentially impacted sediment layers or depths by relating sediment ages or accumulation rates to the time of a suspected release. Sampling depths then can be adjusted as necessary to allow datasets to be combined appropriately (see Section 2.1.7).

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. For the Geochemical Method, analytical data may be necessary to characterize the concentration distribution of metals that are expected to represent background conditions. These potential reference metals can be identified on the basis of site-specific sediment characteristics and geochemistry.

The SAP should specify laboratory reporting limits (RLs) for all analytes. The RL is typically one to five times the detection limit (DL), depending on the analytical method and matrix. The DL can vary considerably from sample to sample because of matrix effects. Ideally, the RL will not change, and will be set high enough to account for matrix effects, yet low enough to meet project-specific DQOs. RLs 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 sediments. Because certain chemicals can pose unacceptable risks to ecological receptors at low concentrations, analytical methods with exceptionally low RLs (such as NOAA National Status and Trends methods) may be required. 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 QC sampling).

Because sediment types play a very important role in the distribution of background chemicals, the SAP should specify procedures for field logging of sediment samples and for geotechnical sampling and testing procedures (e.g., particle-size distribution, density, etc.). In addition to sediment 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.

2.1.7 Evaluate Data

After the analytical data are received from the laboratory, the data should be evaluated to ensure 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* (2000c) 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

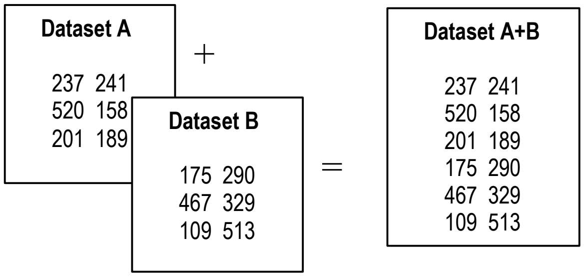
Data should be validated for quality in accordance with Appendix H, Data Validation, of the Navy *Installation Restoration Chemical Data Quality Manual (IR CDQM)* (DON, 1999b). The data validation procedures assign qualifiers to the data that give the end users a qualitative measure of data usability. Data may be assigned the following typical qualifiers:

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

These qualification codes will be assigned to each qualified data point. Combinations of qualifiers, such as nondetect estimated (UJ), 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, sediment samples from the same area may have been collected and analyzed at several different times. 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.



Should Datasets be Combined?

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, they must all be representative of the same target population. Datasets from geochemically, anthropogenically, biologically, and/or physically different sediment types may not be suitable for combination as a single target population. For example, the chemical composition of recently deposited sediments just below the sediment-water interface may be very different from that of earlier, deeper sediments. Under such a condition, combining recent and earlier sediment datasets 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 and Box 4-3). Similarly, combining judgmental/purposeful (often clustered) sampling data with grid sampling data could lead to an unrepresentative site dataset.

Certain classes of organic compounds, e.g., PAHs and polychlorinated biphenyls (PCBs), can be quantified either as individual compounds or as functional groups. Before organic chemical datasets are combined, investigators should verify that the same procedures were used to calculate total concentrations for each dataset.

As a result of depositional processes, chemical concentrations may be relatively uniform over the sediment site area (chemical concentrations at a soil site are more likely to show significant spatial dependence). If good evidence indicates that sediment chemical concentrations are generally uniform, sampling locations will not be a critical issue of concern. However, this assumption should not be made without 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 RLs and measurement biases should 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 2-7).

BOX 2-7. Statistical methods for comparing datasets

Graphic Methods: Histograms, boxplots, and probability plots (Section 2.2.3) of individual datasets are useful to assess 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 (2000c, 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 applied 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 data distributions.

2.1.8 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 (e.g., U.S. EPA preliminary remediation goal [PRGs]). Risk-based screening criteria should be identified in accordance with appropriate guidance, including U.S. EPA human health and ecological risk assessment guidance (1992b, 1997b), the Navy tiered ecological evaluation process (DON, 1999c), and Navy guidance documents such as the *Implementation Guide for Assessing and Managing Contaminated Sediment at Navy Facilities* [DON, 2003]). 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.

As specified by U.S. EPA RAGS protocols and Navy risk assessment policy (DON, 1999c, 2001), maximum concentrations should be used for the initial comparison to screening criteria. If the maximum concentrations of a chemical exceed its risk-based screening level, a representative exposure concentration should be calculated and compared to the risk-based screening level 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, 2000c), representative exposure concentrations are average concentrations over an exposure area. If the dataset is small (less than 20 measurements), risk assessors typically use the maximum detected concentration of a chemical as a conservative measure of the representative exposure concentration. For larger datasets, the representative exposure concentration is expressed as a UCL of the mean concentration.

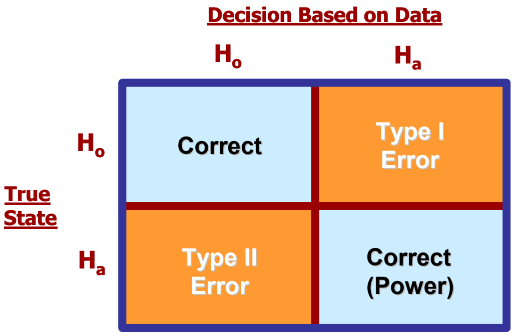
The choice of the appropriate method for computing the UCL depends on the distribution of measured values. For example, if the dataset exhibits lognormal distribution, then lognormal H-statistics (U.S. EPA, 1992b) can be used to compute UCLs. Otherwise, the normality of mean concentration, supported by the central limit theory, can be invoked in order to use Student’s t distribution for UCL computation. For further information on computing UCLs, see U.S. EPA (1992b).

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 2-8). These tools, described below, are used to:

1. Determine the probability distribution of the investigated data
2. Compute descriptive summary statistics of measured values
3. Compute representative exposure concentrations for risk screening
4. Identify potential outliers
5. Determine background ranges.

BOX 2-8. What is a statistical test?



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

A statistical test is a comparison of some data-based quantity (test statistic) with a critical value that is usually 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, denoted by 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 suspect data are not from an underlying normal distribution.

If the test rejects H_o in favor of H_a, then it can be concluded the data do not support the null hypothesis, and therefore should be examined closely to see if they 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, 2000d]).

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 require no specific assumptions regarding the distribution of the data.

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 systematic square or triangular grid sampling design, as discussed in Section 2.1.6.

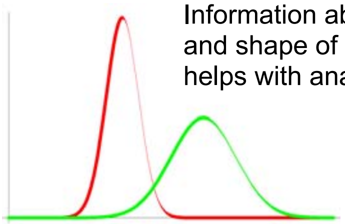
The concentration data associated with each sediment sampling location should be posted on a suitable base map of the site to enable investigators to look for anomalies in the spatial distribution of each suspected COPC. Visual inspection of such spatial plots can reveal potentially impacted areas of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

Samples from clustered and biased locations may not provide data that are representative of the sediment basin. The data also may exhibit spatial correlation, and therefore should not 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.

As noted in Section 2.1.7, clustered and biased datasets can be declustered prior to statistical testing. 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 sediment 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 sediment datasets; these statistics are listed and defined in Box 2-9. 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 2-10.

Datasets with Nondetects

In this document, nondetects are defined as measurements reported by the analytical laboratory as below either the DL or the RL. 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 (2000c) is to replace the nondetects with a surrogate value, e.g., the DL, one-half the DL, or zero. (Note that the difference between these potential surrogate values will be minimal if methods with low DLs [e.g., NOAA National Status and Trends methods] are used.) The descriptive summary statistics in Box 2-9 then may be computed using the (now) full dataset; however, some of the resulting statistics will be biased to some degree. (The median, *p*th 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 (2000c) is to forgo replacing nondetects with a surrogate value. Instead, the mean and standard deviation may 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.2. Alternatively, approaches involving the use of percentiles as descriptive statistics may be applied to datasets with a large number of nondetects.

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 reference locations are to

BOX 2-9. 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.
*p*th 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 *p*th percentile is the *k*th largest measurement in the ordered dataset. If *k* is not an integer, the *p*th 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 (s^2)	The sample variance is the square of the sample standard deviation, i.e., 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 2-10. 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 six 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 $(2.0 + 3.01) / 2 = 2.50$.
*p*th Sample Percentile	Suppose the dataset (after being ordered) is 34, 50, 52, 60, 62, and we want to estimate the 60th percentile, i.e., $p = 0.6$. Now, $k = 0.6\ (5 + 1) = 3.6$. Because *k* is not an integer, we linearly interpolate between the 3rd and 4th largest measurements, i.e., 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 standard deviation 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 (s^2)	The sample variance of the dataset 50, 34, 52, 62, 60 is the square of the sample standard deviation, i.e., variance = $(11.08)^2 = 122.77$.
Coefficient of Variation (CV)	The *CV* for the dataset 50, 34, 52, 62, 60 is $CV = 11.08 / 51.6 = 0.21$.

be used in the background analysis, these data also should be graphically evaluated. Graphic evaluation of the sediment area datasets, including both reference and potentially impacted areas, is used to:

- ❑ Conduct exploratory data analyses to develop hypotheses about possible differences in the means, variances, and shapes for the potentially impacted and reference area distributions
- ❑ Visually depict and communicate differences in the distribution parameters (means, variances, and shapes) for the potentially impacted and/or reference area data distributions
- ❑ Determine whether the potentially impacted 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, 2000c), 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 2-11, were selected because they are easy to run and well suited for background analysis. The methods in Table 2-11 can be performed easily using the DataQUEST (U.S. EPA, 1997a) statistical software.

Histograms

The histogram is a bar graph 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 2-8 and 2-9.

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 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

TABLE 2-11. Summary of selected graphic methods and their features

Method	Description	Method Features	
Histogram	A bar graph 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 depict 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.

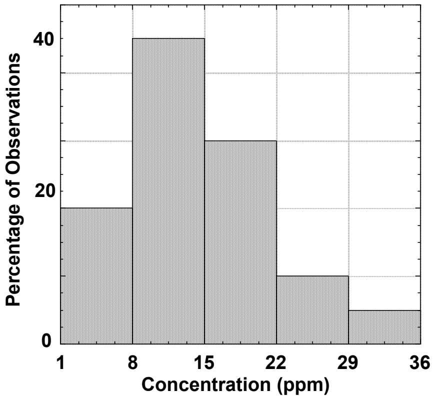


FIGURE 2-8. Example histogram

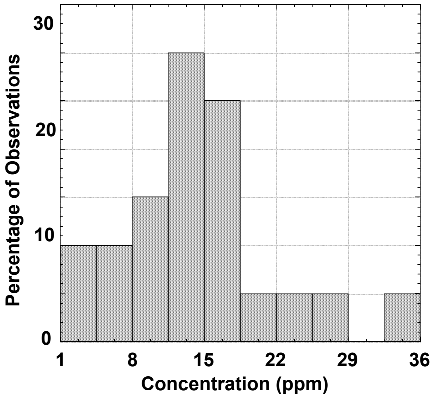


FIGURE 2-9. Example histogram with smaller interval widths

a lognormal or some other skewed distribution. A bimodal histogram may suggest the presence of more than one population in the dataset (e.g., background and contaminant concentration populations). 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 2-11. An example is provided in Box 2-12.

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 2-8 graphs 22 measurements using a concentration range of 7 parts per million (ppm). The histogram in Figure 2-9 is based on the same data as that used for Figure 2-8, but it uses an interval (bar width) of 3.5 ppm rather than 7 ppm. Note that Figure 2-9 gives the impression the data distribution is more skewed to the right (toward larger values) than does Figure 2-8. 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.

Boxplots

The boxplot, sometimes called a box-and-whisker plot, simultaneously displays the full range of the

BOX 2-11. Directions for constructing a histogram (from U.S. EPA, 2000c)

STEP 1: Let $x_1, x_2, \ldots, x_n$ 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 with a length that corresponds to the percentage value computed in Step 3.

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

BOX 2-12. Example: Constructing a histogram (from U.S. EPA, 2000c)

STEP 1: Suppose the following n = 22 measurements (in ppm) of a chemical in sediment 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, i.e., 0 to 5, 5 to 10, 10 to 15, etc. Also, suppose a rule is adopted 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

data, as well as key summary statistics. Figure 2-10 is an example a boxplot of the data listed in Step 1 of Box 2-12. (In this plot, the two <4 values were set equal to 4.)

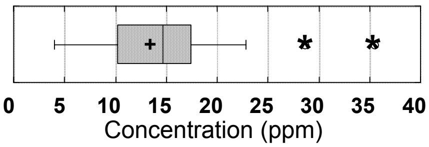


FIGURE 2-10. Example boxplot (box-and-whisker plot)

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 2-9 for definition) indicates the spread of the central 50% of the data, and 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 2-10) may be rotated 90° counterclockwise, so that the box and whiskers are oriented vertically. U.S. EPA (2000c) 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 estimate background concentration ranges.

Probability plots can be constructed by plotting the expected quantiles of the hypothesized distribution on standard graph paper (this method is described in Box 2-13). However, as described in the example shown in Box 2-14, 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 are well fit by the hypothesized distribution being tested. If the hypothesized distribution is the normal distribution, the data values are not transformed and are plotted on a linear scale (y-axis). If the hypothesized distribution is the lognormal distribution, the procedures are the same, except a log-scale is used, or the logarithms of the data are plotted. A lognormally distributed population plotted on a log scale will yield a straight probability curve, as will a normally distributed population plotted on a linear scale. However, when a normally distributed population is plotted on a log scale, the curve will appear convex when viewed from above. Conversely, when a lognormally distributed population is plotted on a linear scale, the curve will appear concave when viewed from above.

BOX 2-13. Directions for constructing a normal probability plot (from U.S. EPA, 2000c)

Let x_1, x_2, …, 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 $x_{(1)}$, $x_{(2)}$, …, $x_{(n')}$, 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, i.e., $AF_1 = 1$. The absolute frequency of value 2 is 1, i.e., $AF_2 = 1$. But the absolute frequency of value 3 is 2, i.e., $AF_3 = 2$, as 3 appears two times in the dataset.

STEP 2: Compute the cumulative frequency (CF), 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 (i.e., 1+1), and the CF for value 3 is 4 (i.e., 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 *i*th 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 2-14. 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

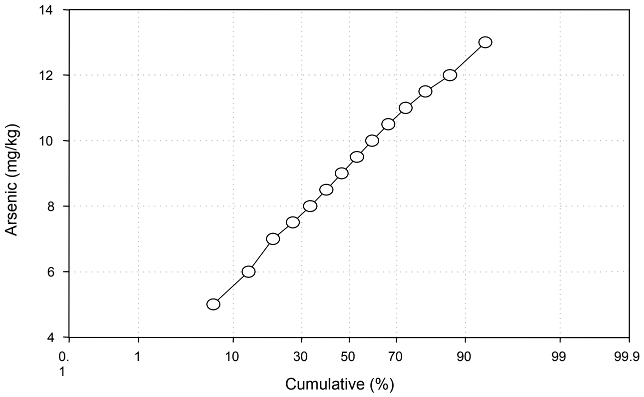


Figure 2-11 is a probability plot constructed to test the null hypothesis that the data have a normal distribution. Note that the x-axis for Figure 2-11 represents cumulative percentages for the standard normal distribution.

If a probability plot does not exhibit a linear pattern for the hypothesized distribution, the 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 on a linear scale 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 when plotted on a linear scale.

The plotted points in Figure 2-11 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 2-12 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 the Shapiro-Wilk W test discussed in Appendix B.1.

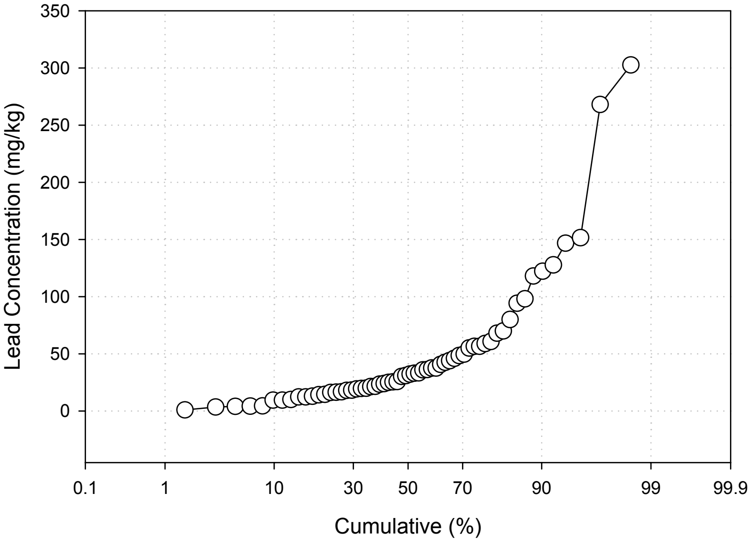


FIGURE 2-11. Example of a probability plot (linear y-axis 100 × cumulative probability on the x-axis)

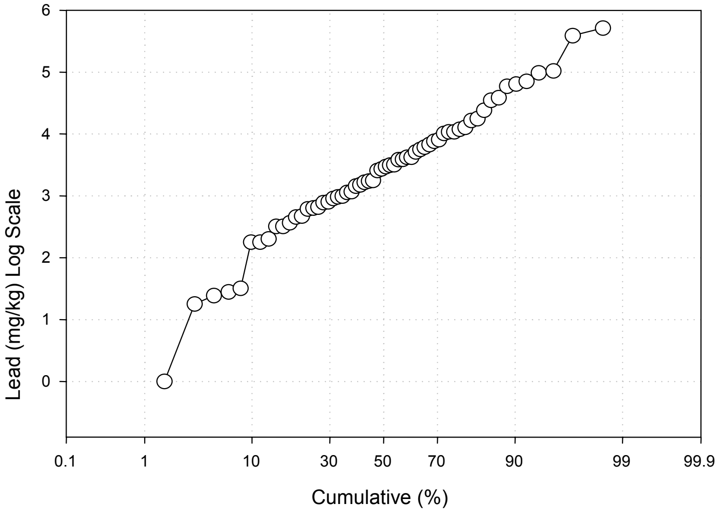


FIGURE 2-12. Example of a log-transformed probability plot

Boxes 2-13 and 2-14 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 2-15). 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. Log-scale plotting is often appropriate even though some of the datasets encountered in background analysis may not be lognormally distributed. For example, if data from a

BOX 2-15. 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 DL 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.

normal population distribution are log-transformed and graphed on a probability plot, it will still be possible to identify outliers.

In background analysis, outliers are always treated as anomalous values. Deletion of such outliers in the subsequent analyses is considered only if other evidence indicates field, laboratory, or hypothesis errors. For example, an outlier value from a reference area should be used in all background computations, unless it is demonstrated that the measured value is affected by field or laboratory errors, or if the targeted area proves to be impacted by processes other than naturally occurring or anthropogenic background sources.

Use of Probability Plots to Identify Background Ranges

Sediment datasets can contain both impacted and background 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. An abrupt change in slope (inflection point) in a probability plot may signify the delimiter value separating two different populations in the investigated dataset. Singh et al. (1994) present a procedure for identifying inflection points as a means to evaluate background ranges.

The probability plotting method for background analysis typically involves one of the following cases:

❑ ***Single Populations:*** A nonsegmented probability plot with no inflection points indicates a single population, e.g., a background population. Note that the gradual curves obtained when a normally distributed population is plotted against a log-scale, and when a lognormally distributed population is plotted against a linear scale, do not contain inflection points.

❑ ***Background Delimiters:*** Segmented probability plots or probability plots with inflection points suggest the existence of multiple subpopulations, including possible outliers. In many cases, an inflection point at which the slope increases, or a break between a segment with a gradual slope followed by a segment with a steeper slope, represents the delimiter between the lowest concentration subpopulation (i.e., the background population) and a higher concentration subpopulation (i.e., potential contamination). This is a conservative approach, because the background range is viewed as a single population, represented by the lowest subpopulation. However, when multiple inflection points are evident on a probability plot, the upper bound of the background range may be considerably higher than the value associated with the lowest inflection point.

❑ ***Multiple Inflection Points:*** As noted in Box 1-4, the background range may be composed of multiple natural or anthropogenic subpopulations (due to factors such as variations in physical characteristics of the sediments [e.g., grain size] or multiple sources of background chemicals). In these situations, the datasets will yield segmented probability plots with multiple inflection points, and the lowest inflection point will not represent the upper bound of the background range. Evidence provided by analysis of univariate and post plots (see Section 2.2.4) should be used to identify the inflection point that represents the background delimiter.

❑ ***Nondelimiting Inflection Points:*** Not all inflection points can be considered background delimiters. Specifically, if the subpopulation above an infection point forms a segment with a more gradual slope than the lower subpopulation, then the inflection point should not be considered a background delimiter. In this case, both segments can be treated as part of the same population.

Background ranges often are used for delineation purposes. For example, the objective of some investigations is to delineate sediments with chemical concentrations above the background range. In this case, the background delimiter for each chemical should be the upper limit of its background range. This type of delimiter is defined as the higher of the maximum background concentration value, or the upper 95th percentile of concentration values within the background subpopulation.

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, 1999a, Section 2.5.6) provides further discussion of the potential limitations of probability plots. For further discussion, including typical outcomes of probability plot analyses, readers are referred to Helsel and Hirsch (1992).

At some sites, "co-contamination" may occur if a chemical release contained both metals and organic compounds. As noted in Box 2-16, 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). The spatial data analysis should be initiated by inspection of concentration data posted on a sediment basin map (a post plot). This can be very useful to evaluate the spatial characteristics of the dataset, and to distinguish 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 sediment. 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, particle size, 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 2-12. By plotting metal concentrations for each variable, the spatial distribution of metals in site sediments 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.

TABLE 2-12. Univariate plot categories and variables

Univariate Plot Category	Variables
Data qualifier	NQ, J, U, UJ
Sampling depth	Surface sediment, deeper sediment
Particle size	Clay, silty clay, 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 thereby 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 Appendix H of the Navy *IR CDQM* (DON 1999b). Data

BOX 2-16. Co-contamination: Organic compounds and metals

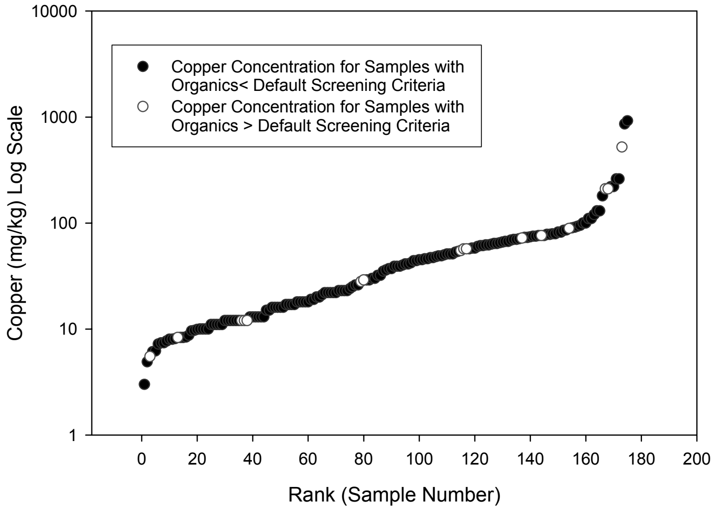
At some sediment investigation sites, the presence of organic contamination at a particular sampling location may indicate that metal contamination also is likely to be present at that location. This "co-contamination" may occur if a chemical release contained both metals and organic compounds (e.g., a release of both organic and inorganic pesticides, or a waste oil release). However, at many sites, there will be little or no correlation between the metal and organic compound distributions. Although some chemical releases contain both metals and organic compounds, the two often are released separately. In addition, 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 section.

With the exception of certain relatively stable compounds such as PCBs, organic compounds are generally less stable and more mobile in the aquatic environment than metals; therefore, any spatial correlation between metals and organic contaminants observed at a sediment basin 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 within sediments. Organic compounds also tend to migrate through the aquatic environment and sediments to a greater extent than metals.

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 (see below). However, it is not necessary to evaluate this relationship to establish background metal concentration ranges. The background distribution of metals in sediment is not affected by an overprint of organic contamination (or any other type of contamination). The procedures described in this section are based on the identification of outliers that do not fit the background concentration population distribution. Such techniques 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 sediment 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 sediments (Section 3) 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 presented below, data points that represent samples with elevated concentrations of organic compounds are marked on the rank plot. Risk-based screening criteria such as NOAA effects range–low (ER-L) or effects range–median (ER-M) values (Buchman, 1999) can be used to define elevated concentrations of organic compounds. The rank plot then can be used to compare the data points corresponding to elevated levels of organic compounds to the rest of the population.

The rank plot shown at right 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.



qualifiers that should be considered during background metals analysis are listed in Table 2-13. Figure 2-13 shows an example data qualifier univariate 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.

3. If most of the data are qualified UJ or U (as illustrated in Figure 2-14), 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 RLs.

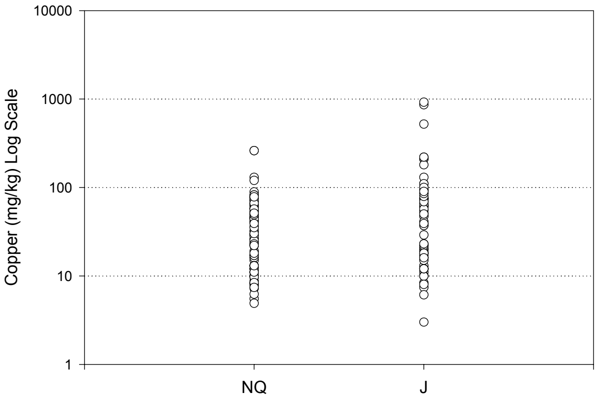


FIGURE 2-13. Univariate plot of copper concentrations vs. data qualifier (noncensored data)

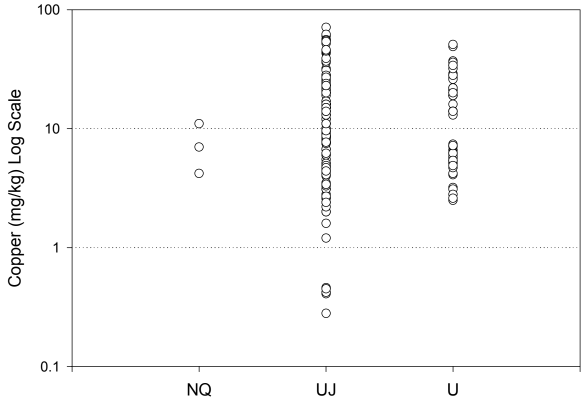


FIGURE 2-14. Univariate plot of copper concentrations vs. data qualifier (censored data)

Procedures for plotting the data qualifier univariate distribution are presented in Box 2-17.

TABLE 2-13. 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 RL or the DL).
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.

BOX 2-17. 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	

Sampling Depth/Sampling Layer Univariate Plot

A sampling depth or sampling layer univariate plot shows the metal concentration ranges that correspond to each sediment sampling depth interval or layer. Figure 2-15 shows an example sampling depth univariate plot for manganese. Sediment sampling depths generally are categorized as follows:

- ❑ "Surface" sediment samples: samples collected from the interval between the water-sediment interface and a depth of approximately 6 inches below the interface; and,
- ❑ "Deeper" sediment samples: any sample collected more than 6 inches below the water-sediment interface. Samples collected beneath the "surface" sample interval are typically identified by the actual depth interval (e.g., 6 to 12 inches below the water-sediment interface).

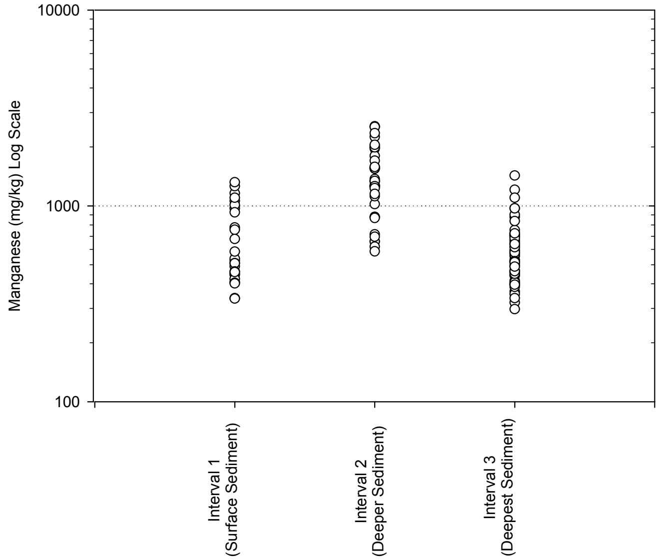


FIGURE 2-15. Univariate plot of manganese concentrations vs. sampling depth

The depth of a particular sediment layer may vary across the site or sediment basin; therefore, if sampling data corresponding to distinct layers are available, it may be more appropriate to segregate and plot the data according to sampling layer instead of sampling depth. Concentrations of both background chemicals and COPCs in different sediment layers may vary significantly due to changes in the depositional environment, sediment sources, and chemical contaminants (see Box 2-1).

By inspecting the sampling depth or sediment layer univariate plot, investigators can reach the following preliminary conclusions:

- ❑ If concentrations in surface sediment are significantly higher than in deeper sediment, relatively recent contaminant deposition should be suspected. It should be noted that evaluation of "recent" contaminant deposition will be dependent on the subaqueous conditions and the sediment deposition or loading rate, which are sediment basin-specific factors. The potential effects of historical or recent dredging also should be considered during analysis of sampling depth univariate plots.
- ❑ If more than two depth intervals or sediment layers are sampled, and an interval or layer with higher concentrations is bracketed between intervals with lower concentrations, contamination of the middle interval should be suspected. The lower concentrations bracketing the higher concentration

may indicate background conditions. For example, Figure 2-15 indicates that manganese concentrations are higher in depth interval 2 than in intervals 1 or 3. These data suggest that the middle interval sediment may be impacted by a manganese release, whereas manganese concentrations in intervals 1 and 3 may represent background conditions.

- ❑ If there are no significant differences between surface sediment and deep sediment metal concentrations, the detected concentrations are likely to be naturally occurring. This can be confirmed by inspecting the other univariate plots, and the probability plot (geochemical association analysis and/or geochemical enrichment analysis also can be used for confirmation).

Procedures for plotting the sampling depth univariate distribution are presented in Box 2-18.

Particle Size Univariate Plot

A particle size univariate plot illustrates the metal concentration ranges that correspond to the different sediments that exist in the sediment basin.

Figure 2-16 shows an example particle size univariate plot for lead. If the concentration ranges for a given metal vary significantly according to particle size, the metal distribution may reflect natural geochemical processes rather than metal contamination. As discussed in

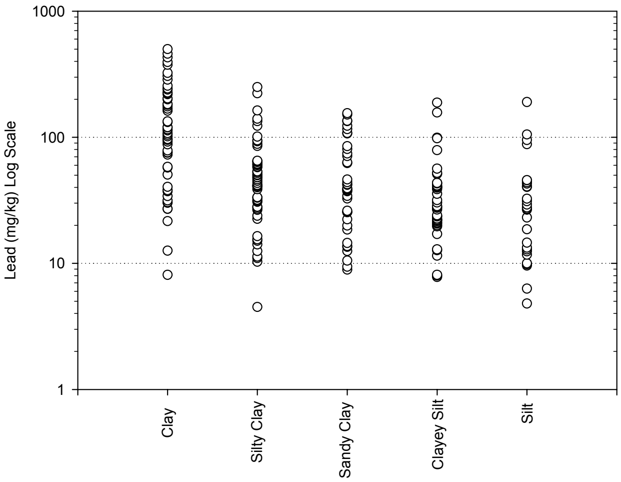


FIGURE 2-16. Univariate plot of lead concentrations vs. particle size (nonimpacted background site)

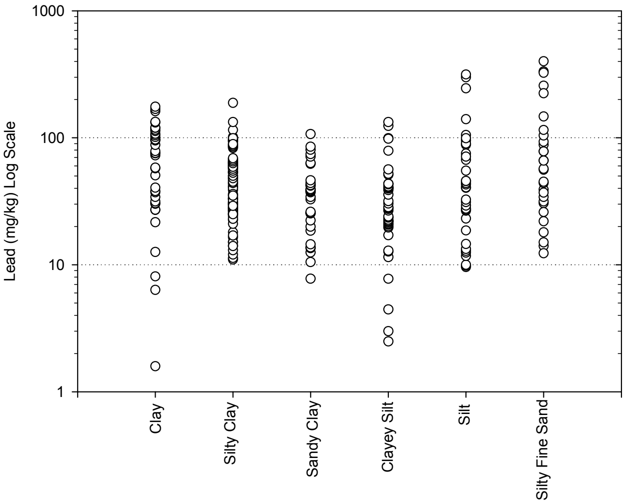


FIGURE 2-17. Univariate plot of lead concentrations vs. particle size (impacted site)

BOX 2-18. 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 Sediment	Deeper Interval Sediment
100	108
10	98
95	96
80	20
20	5
10	9
12	10
95	20

Section 2.1.3, if enrichment is the result of natural geochemical processes, metal concentrations tend to be high in fine-grained sediments such as clays, and lower in relatively coarse-grained sediments such as silts and fine sands. The particle size univariate plot shown in Figure 2-16 indicates that the highest lead concentrations occur in the clay sediment; therefore, the elevated lead concentrations are likely to represent natural background conditions.

A chemical release can impact any type of sediment; therefore, if metal concentration ranges show relatively little variation with particle size, or concentrations are highest in the coarse-grained sediments, contamination should be suspected. For example, Figure 2-17 indicates that relatively high lead concentrations occur in the silt and silty fine sand. Because background concentrations are usually highest in fine-grained sediments, these elevated lead concentrations may be the result of a chemical release, whereas the maximum lead concentration detected in the clay may be useful to estimate the upper bound of the background concentration range.

Procedures for plotting the particle size univariate distribution are presented in Box 2-19.

When evaluating the spatial distribution of sediment types, investigators also should assess the geology and hydrology of the watershed/sediment basin to determine whether sediments in different portions of the basin may be derived from different types of parent rocks or soil, or deposited in radically different hydrodynamic environments. If this is the case, it may be appropriate to identify separate background ranges for each area (see Box 2-20).

Sampling Location Univariate Plot

A sampling location univariate plot illustrates the metal concentration ranges that correspond to different areas of a sediment basin. Concentration anomalies that correspond to individual areas may represent contamination. Figure 2-18 shows an example sampling location univariate plot for tin.

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

- ❑ Each distinct area of the sediment basin has a different operational history; therefore, significant differences in chemical concentrations among the areas may be related to the different activities conducted at sites within the watershed/basin. For example, in Figure 2-18, elevated tin concentrations occur at two areas (AOC-01 and AOC-02). If these areas were near areas used for industrial purposes, outfalls, or other potential contaminant sources, the elevated tin concentrations are likely to represent contamination.
- ❑ If a chemical shows only small concentration differences among areas, the concentrations are likely to represent naturally occurring (background) levels. For example, in Figure 2-18, the maximum detected tin concentrations observed at AOC-03, AOC-04, and AOC-05 are very similar.

Procedures for plotting the sampling location univariate distribution are presented in Box 2-21.

BOX 2-19. Procedures for plotting the particle size univariate distribution

1. Sort the data according to particle size (e.g., clay, silty clay, silt).
2. Enter the data in columns as shown in the table at right.
3. Create a “multiple Y data” scatter plot graph, with particle size 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	Silty Clay	Silt
Analytical Concentrations	100	10	5
	10	12	9
	95	95	10
	80	108	20
	20		30
	85		
	80		

BOX 2-20. Sediments derived from different parent rocks/parent soils or deposited in different aqueous environments: Are multiple background ranges required?

Background chemical concentration ranges in sediments derived from different parent rocks/parent soils or deposited in different hydrodynamic environments can differ widely, even after geochemical redistribution and enrichment. Therefore, investigators should be aware that certain sediment basins might require special consideration (i.e., basins that contain separate and distinct areas of sediments eroded from different types of parent material or deposited in radically different environments). In such cases, it may be necessary to evaluate background separately for each area.

An example of a sediment basin that may require separate analyses is one with multiple lobes, where sediment in one lobe of the basin is derived from weathered volcanic rocks that occur in one portion of the watershed, whereas the rest of the basin contains sediments derived from a coralline limestone formation in a different portion of the watershed. Another example is a basin that contains more than one depositional environment, such as a high-energy nearshore environment where sands are deposited, and a low-energy offshore environment where clays are deposited. At such a site, background metal concentration ranges in the different areas can be very different. If the different sediment types occur in discrete areas, each with a significant areal extent relative to the overall area of the sediment basin, it may be necessary to divide the basin into “subbasins,” and establish separate background ranges for each area.

In some cases, it may be possible to identify separate data populations that represent background concentrations for distinct areas of a sediment basin based on analysis of a comprehensive (i.e., overall sediment basin) 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 watershed/sediment basin, it may be possible to estimate background ranges for each area. However, if the comprehensive analysis does not clearly show separate populations, it may be necessary to segregate the data, and perform a separate analysis for each area.

Probability Plots

Although probability plots can be constructed manually by plotting the data versus quantiles of the hypothesized distribution on standard graph paper (as described in Box 2-13), or by manually plotting cumulative percentages on special probability paper (as shown in Box 2-14), probability plots for background analysis are usually 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 2-19.

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 2-20, 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

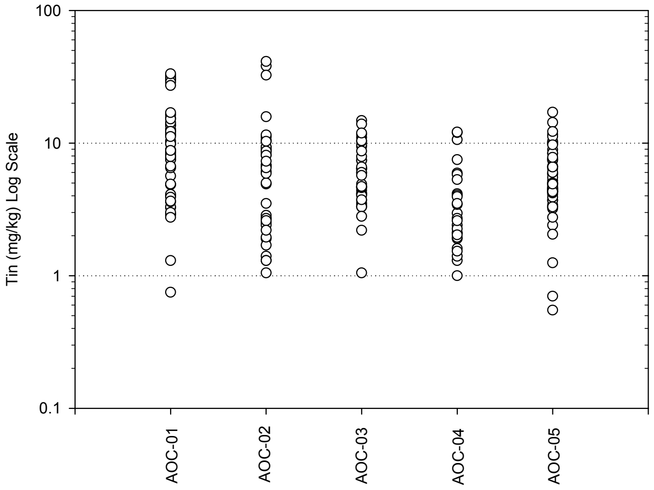


FIGURE 2-18. Univariate plot of tin concentrations vs. sampling location

BOX 2-21. Procedures for plotting the sampling location univariate distribution

1. Sort the data according to sampling location (e.g., AOC-01, AOC-02, AOC-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).

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

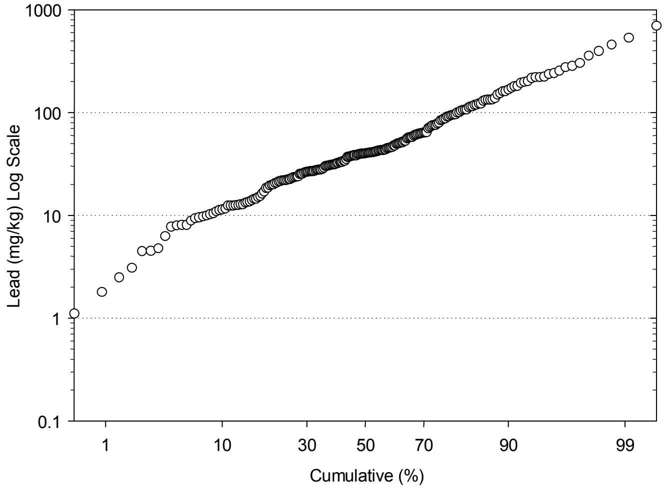


FIGURE 2-19. Log-scale probability plot of lead concentrations (background)

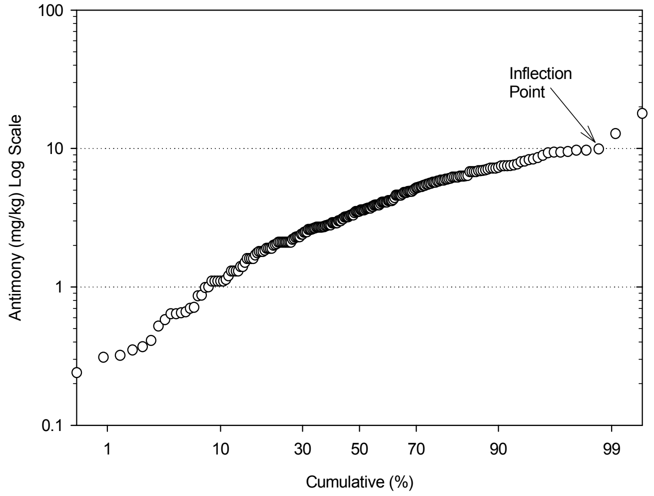


FIGURE 2-20. Log-scale probability plot of antimony concentrations (contaminated)

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. Provided that data clustering or gaps do not confound the analysis, the point at which the slope increases (inflection point) is likely to represent the upper bound of the background concentration range (see Figure 2-20). This hypothesis should be evaluated by inspecting combined plots (described in the following subsection).

Combined Plots

Univariate and probability plots are combined by placing them next to each other on the same page, with equivalent y-axis scales (Figure 2-21). 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 2-22. Decision-making guidelines for identifying the range of concentrations that represent background are presented in Table 2-14.

Begin the combined plots analysis by inspecting the probability plot to identify a segment boundary or inflection point that may mark the upper limit of the background range. In the example shown on Figure 2-21, an inflection point is observed on the probability plot at approximately 3.10 mg/kg.

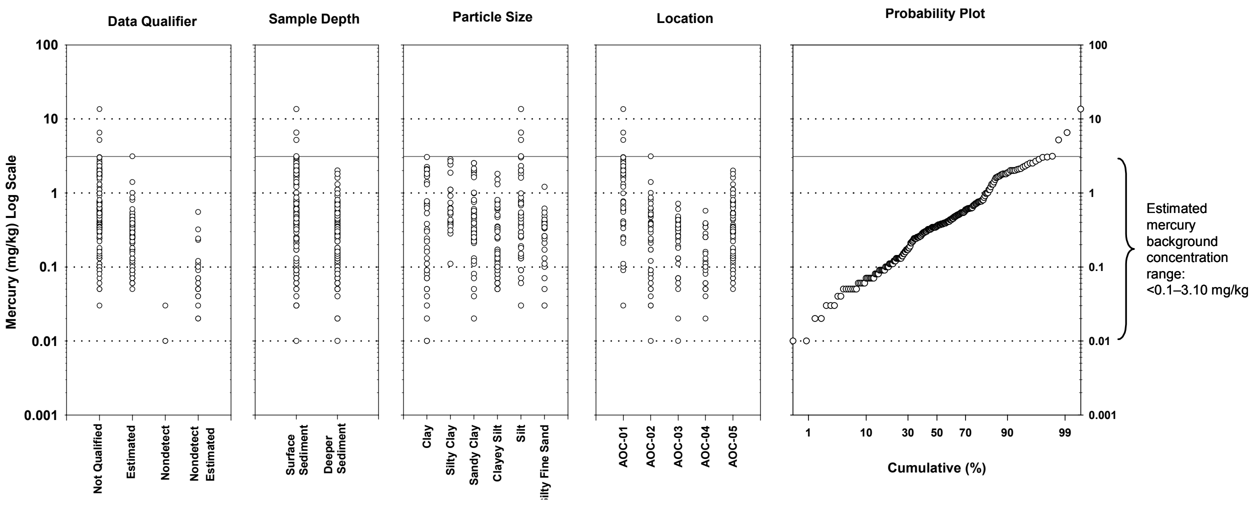


FIGURE 2-21. Combined mercury plots (univariate and cumulative probability)

The next step is to determine whether this initial estimate 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 univariate plots should be evaluated according to the decision questions listed in Table 2-14 to make this determination. On Figure 2-21, results of the combined probability-univariate plot analysis indicate that 3.10 mg/kg represents a reasonable estimate of the upper bound of the background range.

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

❑ Background chemicals such as metals may occur over wide concentration ranges in natural sediments (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 deeper sediments.

After completing the graphical analysis described above, the estimated background ranges should be compared to background ranges reported in the literature for sediment basins/watersheds with similar characteristics (if suitable literature data are available). This comparison will help to ensure that a reasonable conclusion has been reached. However, it is important to remember that sediment characteristics (e.g., grain size, mineralogy) must be similar if background ranges reported in

BOX 2-22. Procedures for creating the combined plots figure

1. After creating the univariate plots (data qualifier, sampling depth, particle size, 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 2-14. 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. On Figure 2-21, a majority of the mercury concentrations are either NQ or J values. The U and UJ value concentrations are all below the upper bound of the background concentration range. Therefore, all data 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 on Figure 2-21), then the sampling depth univariate plot provides evidence to support the conclusion that the inflection point at 3.10 mg/kg represents the upper bound of the background range. Although the time of the suspected chemical release and the rate of sediment accumulation must be considered, it is often reasonable to assume that near-surface sediments will be more contaminated than deeper sediments. On Figure 2-21, the maximum concentration detected in deeper interval sediment corresponds to the inflection point on the probability plot (3.10 mg/kg), and is likely to represent the upper bound of the background concentration range. Concentrations above 3.10 mg/kg are likely to represent contamination. If the answer is no (i.e., there are no significant differences between surface and subsurface concentrations), then the metal is likely to be naturally occurring, and the maximum concentration detected at the sediment site is likely to represent the upper bound of the background range.
3. Does the particle size univariate plot indicate that relatively high concentrations tend to occur only in certain sediments?	If the answer is yes (i.e., relatively high metals concentrations tend to occur only in certain types of sediment within the basin), then the maximum concentration detected in the fine-grained sediment may be useful to estimate the upper bound of the background range. If concentrations in coarse-grained sediments are higher than in fine-grained sediments, then contamination should be suspected. The particle size univariate plot shown on Figure 2-21 indicates that mercury concentrations above 3.10 mg/kg occur in relatively coarse-grained sediments (silt). As discussed in Section 2.1.3, the highest concentrations of naturally occurring metals usually occur in the finest-grained sediments (i.e., clays). Therefore, concentrations above 3.10 mg/kg are likely to be the result of a chemical release, and concentrations below 3.10 mg/kg are likely to be within the background range. If the answer is no (i.e., concentrations show little dependence on location), then contamination should be suspected (a chemical release can impact any of the different sediment types that occur in a sediment basin).
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. On Figure 2-21, mercury concentrations above the estimated background range occur at only one of the five locations (AOC-01). If AOC-01 is located near a suspected source of mercury contamination, the anomalously high concentrations most likely represent 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 sediment is not likely to be contaminated. The maximum detected concentration therefore would likely represent the upper bound of the background range.

the literature are to be used for comparison to estimated background ranges at an investigation site (see Section 2.1.3).

2.2.5 Determine Acceptability of Background Ranges

After completing the exploratory data analysis, the project team 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 sediment basin (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 ambient chemical concentrations (i.e., total concentrations of both naturally occurring chemicals and anthropogenic chemicals not related to specific point sources or site releases).

If the above conditions are met, then the background analysis is completed for the target chemical: 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 Methods for Further Analysis

If the procedures described above do not yield technically defensible and reliable estimates of the background concentration ranges for the suspected COPC metals, the project team should implement appropriate methods for further analysis. The methods recommended for further analysis 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 or parent soil, and the geochemical processes that occur during and after soil formation and sediment deposition. It should be noted that other physical and chemical parameters, such as grain size and TOC, also may be correlated with chemical concentrations in sediments (see Section 2.1.3). Therefore, although the geochemical background analysis techniques presented in this document focus on bivariate analysis of correlated metals, other physical and chemical parameters also may be used (Cal/EPA, 1998).

- ***Comparative Method:*** This method is specifically designed to compare data from potentially impacted sediment sites to data from reference (or background) areas. The statistical tests used in the Comparative Method are designed to assess whether metal concentrations at the potentially impacted portion of the sediment basin are statistically similar to reference area concentrations. Adequate background sampling data are required for this purpose.

As shown on Figure 1-4, for most sediment investigations it will be appropriate to implement the Geochemical Method first. If the Geochemical Method does not produce definitive results, the Comparative Method is usually the next step. Some of the key characteristics of these two classes of background analysis are listed in Table 2-15. In some situations, depending on the nature of the investigated chemicals, both methods may be appropriate.

The Geochemical Method is particularly likely to be successful if evidence suggests that geochemical processes (e.g., weathering, precipitation, or sorption) control the concentrations of background 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 relationships between chemicals, and can identify their corresponding background concentration ranges even when no off-site data are available.

TABLE 2-15. Features of background analysis methods

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

The Comparative Method is able to distinguish between COPCs and background chemicals by comparing data from a potentially impacted sediment site to reference area data. 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) 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 subsection. Table 2-16 provides some practical rules for identifying conditions that favor each method.

If these methods are not successful and the COPCs are organic chemicals, then ACF should be considered. As described in Appendix A, background analysis by ACF involves distinguishing the organic compounds associated with a site-related chemical release from those attributable to background sources. Although ACF can be expensive due to the high cost of generating the required analytical data, the method can be particularly useful when background concentration ranges estimated by Exploratory Data Analysis or the Comparative Method are not acceptable to all stakeholders.

In addition to the background analysis methods presented in this document, a technique known as "factor analysis" may be used to evaluate background conditions. Factor analysis is a statistical approach used to evaluate the interrelationships among multiple variables. If a group of chemicals exhibits a high degree of correlation, then the correlation is likely the result of one or more factors the chemicals have in common. For example, chemicals derived from a common source (e.g., a natural source) tend to be correlated. Therefore, the relationships identified by factor analysis can be used to differentiate between background and potentially impacted populations within a soil or

TABLE 2-16. Favorable conditions for background analysis by the Geochemical and Comparative Methods

Conditions Favorable for the Geochemical Method	Conditions Favorable for the Comparative Method
• The target chemicals are likely to be correlated with naturally occurring chemicals (e.g., metals).	• The targeted background chemicals are either naturally occurring or anthropogenic.
• The sediment investigation site dataset includes data from both potentially impacted and background areas.	• The sediment investigation site can be either completely or partially impacted.
• Suitable reference area data are not available due to physical, geographical, or budgetary constraints.	• A reference area can be identified that is geochemically, and anthropogenically similar to the potentially impacted sediment site.
• Regulatory acceptance of the method is expected due to precedent, guidance, or pre-approval.	• Regulatory acceptance of the method is expected due to precedent, guidance, or pre-approval.
	• The spatial extent of the reference area has been defined, and all stakeholders agree that the reference area is appropriate for the comparison.

sediment dataset. The details of factor analysis are beyond the scope of this guidance; further information is available in Reyment and Jöreskog (1993) and Rummel (1970). Statistical software programs for factor analysis are available (e.g., SPSS and StatView; see Appendix D.3). The application of factor analysis to the evaluation of background conditions for PAHs is described in Appendix A.

2.3.1 Feasibility and Applicability

The Geochemical Method can identify background concentration ranges by evaluating correlated background chemicals even when background or reference data are not available (i.e., a separate [reference area] dataset is not required). If portions of a sediment basin have been impacted, the Geochemical Method can identify background ranges by distinguishing between the two populations (background concentrations and above-background concentrations). If a sediment investigation 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. Geochemical association analysis is capable of identifying background ranges without the need for a predefined reference dataset. However, if the Geochemical Method does not yield a technically defensible estimate of the background range for a suspected COPC metal, it may be necessary to implement the Comparative Method.

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, biologically, and anthropogenically similar to the potentially impacted areas of the sediment basin. Chemical, biological, and ecological data should be used to investigate candidate areas away from the source area before they can be considered background areas. However, because of physical, geographical, and cost constraints, it will not always be possible to identify or adequately sample a suitable reference area. Furthermore, the Comparative Method requires that both the potentially impacted and background areas 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 2-17 provides practical rules for evaluating the applicability of each method to analysis of organic and inorganic background chemicals.

2.3.2 Cost/Benefit Analysis

Both methods require significant effort, cost, and time. For example, the Geochemical Method requires 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 background analysis method justify its associated costs. In general, the benefits of background analysis are most likely to justify the associated costs when a sediment basin is not impacted or only marginally impacted by site-related releases. In addition, background analysis

TABLE 2-17. Practical rules for applicability of background analysis methods

Class of Chemicals	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 reference area are available.
Inorganic	Applicable when the co-presence of inorganic chemicals is attributable to geochemical processes.	Applicable when adequate data from a reference area are available.

can be very beneficial for sites located within sediment basins that have been exposed to long-term anthropogenic (not site-related) sources.

2.3.3 Regulatory Acceptance

The methods used for background analysis at a particular site must be acceptable to all stakeholders. The most favorable conditions are: (a) the stakeholders promote the use of the method; (b) the stakeholders have accepted the 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

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.

Alloway, B.J. 1990. *Heavy Metals in Soils.* Glasgow: Blackie and Son.

American Public Health Association. 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). 2003a. *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). 2003b. *D2488-00 Standard Practice for Description and Identification of Soils (Visual-Manual Procedure)*. West Conshohocken, PA.

American Society for Testing and Materials (ASTM). 2003c. *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). 2003d. *D422-63(1998) Standard Test Method for Particle-Size Analysis of Soils*. West Conshohocken, PA.

American Society for Testing and Materials (ASTM). 2003e. D5549-94e1 Standard Guide for the Contents of Geostatistical Site Investigation Report. West Conshohocken, PA.

American Society for Testing and Materials (ASTM). 2003f. D5922-96e1 Standard Guide for Analysis of Spatial Variation in Geostatistical Site Investigations. West Conshohocken, PA.

American Society for Testing and Materials (ASTM). 2003g. *D5923-96e1 Standard Guide for Selection of Kriging Methods in Geostatistical Site Investigations.* West Conshohocken, PA.

American Society for Testing and Materials (ASTM). 2003h. *D5924-96e1 Standard Guide for Selection of Simulation Approaches in Geostatistical Site Investigations.* West Conshohocken, PA.

American Society for Testing and Materials (ASTM). 2003i. *E1527-00 Standard Practice for Environmental Site Assessments: Phase I Environmental Site Assessment Process*. West Conshohocken, PA.

American Society for Testing and Materials (ASTM). 2003j. *E1903-97 Standard Guide for Environmental Site Assessments: Phase II Environmental Site Assessment Process*. West Conshohocken, PA.

Bertine, K., and E. Goldberg. 1977. History of Heavy Metal Pollution in Southern California Coastal Zone—Reprise. *Environmental Science and Technology* 11:297-299.

Buchman, M.F. 1999. *NOAA Screening Quick Reference Tables*. NOAA HAZMAT Report 99-1. Seattle, WA: National Oceanic and Atmospheric Administration, Coastal Protection and Restoration Division.

California Environmental Protection Agency (Cal/EPA). 1998. *Ambient Concentrations of Toxic Chemicals in San Francisco Bay Sediments*. Staff report, Regional Water Quality Control Board, San Francisco Bay Region. May. http://www.swrcb.ca.gov/rwqcb2/download/sfbaysediment.pdf

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

Covay, K.J., and D.A. Beck. 2001. *Sediment-Deposition Rates and Organic Compounds in Bottom Sediment at Four Sites in Lake Mead, Nevada, May 1998.* United States Geological Survey (USGS). Open-File Report 01-282. http://water.usgs.gov/pubs/of/ofr01282/index.htm

CRC Press. 1992. *CRC Handbook of Chemistry and Physics*. 73rd ed. Boca Raton, FL.

Daskalakis, K, and T. O'Connor, 1995. Normalization and Elemental Sediment Contamination in the Coastal United States. *Environmental Science and Technology* 29:470-477.

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. July.

Department of the Navy (DON). 1999b. *Navy Installation Restoration Chemical Data Quality Manual (IR CDQM).* NFESC Special Report SP-2056-ENV. Washington, D.C.: Naval Facilities Engineering Command, Naval Facilities Engineering Service Center. September.

Department of the Navy (DON). 1999c. *Navy Policy for Conducting Ecological Risk Assessments.* Letter from A.A. Granuzzo, Chief of Naval Operations, to Commander, Naval Facilities Engineering Command. Ser. N453E/9U595355. Washington, D.C. 5 April.

Department of the Navy (DON). 2001. *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). 2003. *Final Implementation Guide for Assessing and Managing Contaminated Sediment at Navy Facilities.* Naval Facilities Engineering Command. February.

Di Toro, D.M., Mahony, J.H., D.J. Hansen, K.J. Scott, M.B. Hicks, S.M. Mayr, and M. Redmond. 1990. Toxicity of Cadmium in Sediments: The Role of Acid Volatile Sulfides. *Environmental Toxicology and Chemistry* 9:1,487-1,502.

Di Toro, D.M., Mahony, J.H., D.J. Hansen, K.J. Scott, A.R. Carlson, and G.T. Ankley. 1991. Acid Volatile Sulfide Predicts the Acute Toxicity of Cadmium and Nickel in Sediments. *Environmental Toxicology and Chemistry* 26:96-101.

Environment Canada. 1992. *Criteres Intermimaires pour l'Evaluation de la Quality des Sediments du Saint-Laurent*. Environment Canada et Ministere de l'Environnement du Quebec.

Environment Canada. 1994. *Interim Sediment Quality Assessment Values.* Manuscript Report No. ECD. Ottawa, Canada: Ecosystem Conservation Directorate, Environment Canada.

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.

Hanson, P., D. Evans, D. Colby, and V. Zdanowics. 1993. Assessment of Elemental Contamination in Estuarine and Coastal Environments Based on Geochemical and Statistical Modeling of Sediments. *Marine Environmental Research* 36:237-266.

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

Huvane, J.K., and S.R. Cooper. 1999. Diatoms as Indicators of Environmental Change in Florida Bay. Florida Bay Science Conference, Key Largo, FL.

Ingersoll, C.G., P.S. Haverland, E.L. Brunson, T.J. Canfield, F.J. Dwyer, C.E. Henke, N.E. Kemble, D.R. Mount, and R.G. Fox. 1995. Draft Calculation and Evaluation of Sediment Effect Concentrations for the Amphipod, *Hyalella Azteca* and the Midge, *Chrionomus Riparius,* submitted to *J. of Great Lakes Research*.

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.

Judson, S., M.E. Kauffman, and L.D. Leet. 1987. *Physical Geology.* 7th ed. New York: Prentice-Hall.

Kalkhoff, S.J., and P.C. Van Metre. 1997. *Organochlorine Compounds in a Sediment Core from Coralville Reservoir, Iowa.* United States Geological Survey (USGS). Fact Sheet FS-129-97. Iowa City, Iowa. http://wwwdiaiwc.cr.usgs.gov/nawqa/factsheets/fs-129/fs-129.html

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

Lee, H., B.L. Boese, J. Pelletier, M. Winsor, D. T. Specht, and R. C. Randall. 1989. *Guidance Manual: Bedded Sediment Bioaccumulation Tests*. EPA/600/x89/302, ERLN-N111. Newport, OR: U.S. Environmental Protection Agency, Bioaccumulation Team, Pacific Ecosystems Branch.

Long, E.R., D.D. MacDonald, S.L. Smith, and F.D. Calder. 1995. Incidence of adverse biological effects within ranges of chemical concentration in marine and estuarine sediments. *Environ. Manag*. 19:81-87.

Loring, D. 1990. Lithium—A New Approach for Granulometric Normalization of Trace Metal Data. *Marine Chemistry* 29:155-168.

Los Alamos National Laboratory (LANL). 2000. *Periodic Table of the Elements.* Los Alamos, NM: LANL, Chemistry Division.

Mangerud, J. 1972. *Radiocarbon Dating of Marine Shells, Including a Discussion of Apparent Ages of Recent Shells from Norway. Boreas* 1:143-172.

Martin, J.M. and M. Whitfield. 1983. *The Significance of the River Inputs to the Ocean, Trace Metals in Seawater*. Eds. C.S. Wong, E. Boyle, K.W. Bruland, J.D. Burton, and E.D. Goldberg. New York: Plenum Press.

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

Merck. 1989. *Merck Index—An Encyclopedia of Chemicals, Drugs and Biologicals.* 11th ed. Rahway, NJ: Merck and Company.

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.

Morse, J., B. Presley, R. Taylor, G. Benoit, and P. Santschi. 1993. Trace Metal Chemistry of Galveston Bay: Water, Sediments and Biota. *Marine Environmental Research* 36:1-37.

Olmez, I., E. Sholkovitz, D. Hermann, and R. Eganhouse. 1991. Rare Earth Elements in Sediments off Southern California: A New Anthropogenic Indicator. *Environmental Science and Technology* 25:310-316.

Persaud, D., R. Jaagumagi, and A. Hayton. 1989. *Development of Provincial Sediment Quality Guidelines*. Toronto: Ontario Ministry of the Environment, Water Resources Branch, Aquatic Biology Section.

Persaud, D., R. Jaagumagi, and A. Hayton. 1992. *Guidelines for the Protection and Management of Aquatic Sediment Quality in Ontario*. Toronto: Ontario Ministry of the Environment.

Reyment, R.A., and K.G. Jöreskog. 1993. *Applied Factor Analysis in the Natural Sciences.* Cambridge: Cambridge University Press.

Rummel, R.J. 1970. *Applied Factor Analysis.* Evanston, IL: Northwestern University Press.

Schropp, S., G. Lewis, H. Windom, J. Ryan, F. Caldner, and L. Burney. 1990. Interpretation of Metal Concentrations in Estuarine Sediments of Florida Using Aluminum as a Reference Element. *Estuaries* 13:227-235.

Shields, E.J. 1988. *Pollution Control Engineers Handbook.* Des Plaines, IL: Cahners Publishing.

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

Trefry, J., S. Metz, R. Trocine, and T. Nelsen. 1985. A Decline in Lead Transport by the Mississippi River. *Science* 230:439-441.

United States Army Corps of Engineers (USACE). 1977. *Ecological Evaluation of Proposed Discharge of Dredged or Fill Material into Navigable Water, Interim Guidance for Implementation of Section 404(b) (1) of Public Law 92-500 (Federal Water Pollution Control Act Amendments of 1972.* Misc. Paper D 76-17. Vicksburg, MS: U.S. Army Corps of Engineers, Waterways Experiment Station. 1-EZ.

United States Army Corps of Engineers (USACE). 1997. *Engineering and Design – Practical Aspects of Applying Geostatistics at Hazardous, Toxic, and Radioactive Waste Sites.* Technical Letter. ETL 1110-1-175. June 30.

United States Environmental Protection Agency (U.S. EPA). Region 10. 1987. *Recommended Protocols for Sampling and Analyzing Subtidal Benthic Macroinvertebrate Assemblages in Puget Sound.* Puget Sound Water Quality Authority. January.

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.

Washington, DC: 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). 1994a. *Methods for Assessing the Toxicity of Sediment-Associated Contaminants with Estuarine and Marine Amphipods.* EPA/600/R-94/025. Washington, DC: Office of Research and Development.

United States Environmental Protection Agency (U.S. EPA). 1994b. *Standard Test Methods for Measuring the Toxicity of Sediment-Associated Contaminants with Fresh Water Invertebrates*. EPA/600/R-94/024. Washington, DC: Office of Science and Technology.

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. Washington, DC: Office of Solid Waste and Emergency Response. June.

United States Environmental Protection Agency (U.S. EPA). 2000a. *Bioaccumulation Testing and Interpretation for the Purpose of Sediment Quality Assessment, Status and Needs*. EPA-823-R-00-001. Washington, DC: Office of Water; Office of Solid Waste. February.

United States Environmental Protection Agency (U.S. EPA). 2000b. *Coastal 2000 Northeast Component, Field Operations Manual*. EPA/620/R-00/002. Washington, DC: Office of Research and Development. April.

United States Environmental Protection Agency (U.S. EPA). 2000c. *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). 2000d. *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). 2002a. *Guidance on Choosing a Sampling Design for Environmental Data Collection: For Use in Developing a Quality Assurance Project Plan. EPA QA/G-5S.* EPA/240/R-02/005. Washington, DC: Office of Environmental Information. December.

United States Environmental Protection Agency (U.S. EPA). 2002b. *USEPA Contract Laboratory Program National Functional Guidelines for Inorganic Data Review.* Final. EPA 540-R-01-008. Washington, DC: Office of Emergency and Remedial Response. July.

United States Environmental Protection Agency (U.S. EPA). 2003. *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 Environmental Protection Agency (U.S. EPA) and United States Army Corps of Engineers (USACE). 1991. *Evaluation of Dredged Material Proposed for Ocean Disposal–Testing Manual*. EPA 503/8-91/001. Washington, DC: U.S. EPA Office of Water. February.

United States Environmental Protection Agency (U.S. EPA) and United States Army Corps of Engineers (USACE). 1998. *Evaluation of Dredged Material Proposed for Discharge in Waters of the U.S. – Testing Manual (Inland Testing Manual).* EPA-823-B-98-004 Washington, DC: U.S. EPA Office of Water and Department of the Army, U.S. Army Corps of Engineers.

United States Geological Survey (USGS). 1998. *Short-Lived Isotopic Chronometers: A Means of Measuring Decadal Sedimentary Dynamics.* Fact Sheet FS-73-98. St. Petersburg, FL.

United States Geological Survey (USGS). 2002. *Online Map Lists.* URL: <http://mapping.usgs.gov/mac/maplists.html>.

Washington State Department of Ecology (WDOE). 1992. *Statistical Guidance for Ecology Site Managers*. Olympia, WA: Toxics Cleanup Program. August.

3. GEOCHEMICAL METHOD

The Geochemical Method uses statistical techniques based on geochemical principles to analyze sediment basin 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 sediments are considered carefully during the analysis to ensure that the calculated background ranges accurately reflect natural background conditions.

The Geochemical Method usually requires only data from the sediment basin, which typically represent a combination of “impacted” and “background” areas. At most sediment basins, chemical concentrations detected in certain areas will represent background conditions; therefore, no reference area or additional off-basin sampling is necessary.

In most situations, a chemical release 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 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 most soils and sediments. Unless a release is suspected at a sediment basin, 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).
- ❑ Based on a wide range of observations, metal concentrations in rocks, soils, and sediments generally tend to be lognormally distributed. It should be noted, however, 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 observations, certain groups of metals tend to occur together in natural rocks, soils, and sediments, 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 where finding suitable reference areas is difficult or impossible.

3.1 Overview

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 a plot indicates strong correlation between the two metals, then the plots 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, 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 often are 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, soils, and sediments are discussed in Section 2.1.3.)

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).

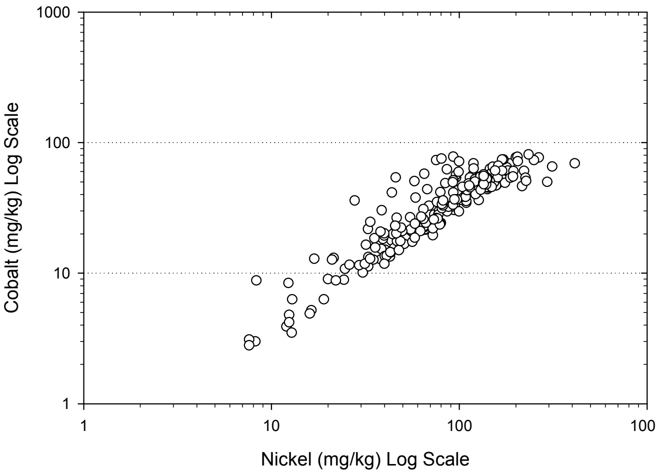


FIGURE 3-1. Scatter plot of cobalt vs. nickel concentrations (log scale) showing strong association

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

Metal pairs that show weak correlation should not be used for geochemical association analysis. For example, in the scatter plots presented in Figure 2-7, eight different metals in Florida sediments are plotted with respect to aluminum. The scatter plots indicate correlation ranging from very weak (e.g., mercury vs. aluminum) to strong (e.g., zinc vs. aluminum).

The distribution of metals in sediments depends on factors including source rock or soil types, weathering processes, surface adsorption phenomena, and characteristics of depositional environment. Therefore, metal/metal relationships may vary significantly, and must be evaluated on a site-specific basis. The most important factors that cause metal association in background sediments 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 sediments, including aluminum, iron, and calcium. Sediment is composed primarily of particles of chemically resistant minerals such as quartz and clay minerals, resulting from weathering of the parent rock. Adsorption of metals to fine-grained sediment particles—primarily clay and colloidal organic material—can result in elevated metal concentrations in subaqueous sediments. 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 sediments, 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 sediment basin should be evaluated first. Follow-on evaluation of the relationships between suspected COPC metals then can be conducted.

Iron may not be useful for identifying COPCs in anaerobic sediments by geochemical association analysis. Under anaerobic conditions, iron is reduced to a soluble form, and tends to migrate within the porewater between sediment particles. Aluminum is less reactive and has lower reduction potential than iron; therefore, if anaerobic conditions exist, aluminum is usually a better reference element. Note, however, that if a strong correlation

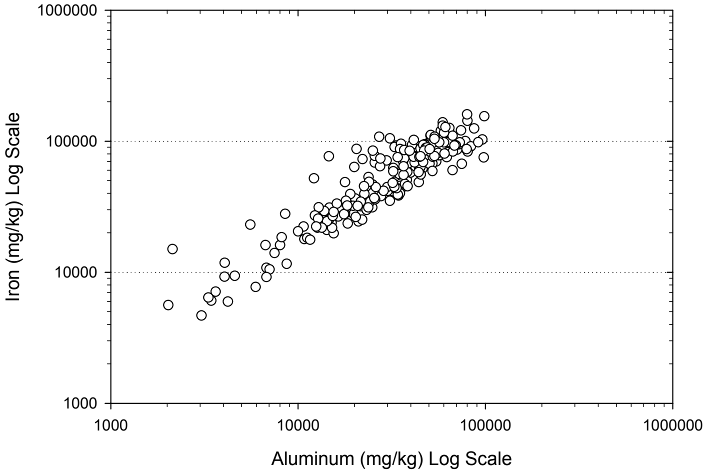


FIGURE 3-2. Scatter plot of iron vs. aluminum concentrations (log scale) showing strong association

between aluminum and iron is observed, iron can be considered a non-COPC, and may be a suitable reference metal.

As discussed in Section 2.1.3, TOC and grain size also can be used as reference constituents when evaluating background conditions with geochemical association analysis. Aluminum, however, is the most commonly used reference constituent.

3.2.1 Geochemical Regression Analysis

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 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 sediment basin. 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 should also 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 basin-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

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 basin sediment 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.

metals should be identified. It may be necessary to construct two or more scatter plots using different reference metals 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 sediment is required to select suitable reference metals (see Section 2.1.3). 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).

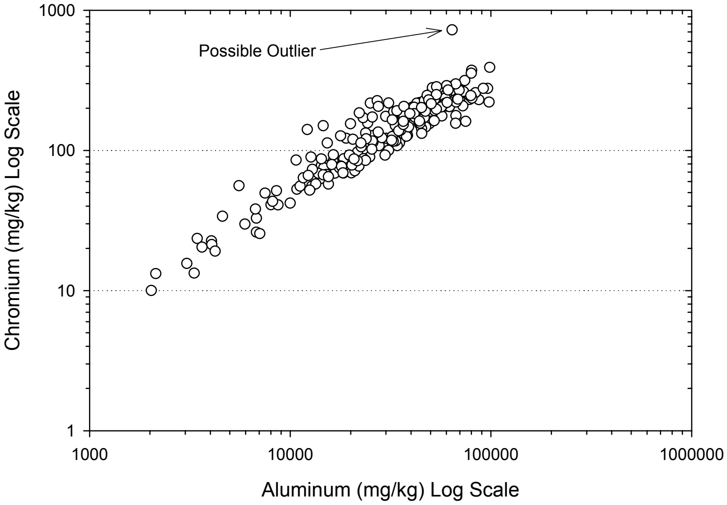


FIGURE 3-3. Scatter plot of chromium vs. aluminum concentrations (log scale) showing strong association

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. The overall pattern can be described in terms of *form* (e.g., a linear relationship), *direction* (e.g., the slope of 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 least-squares linear 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 a chromium vs. aluminum 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 sediment 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.

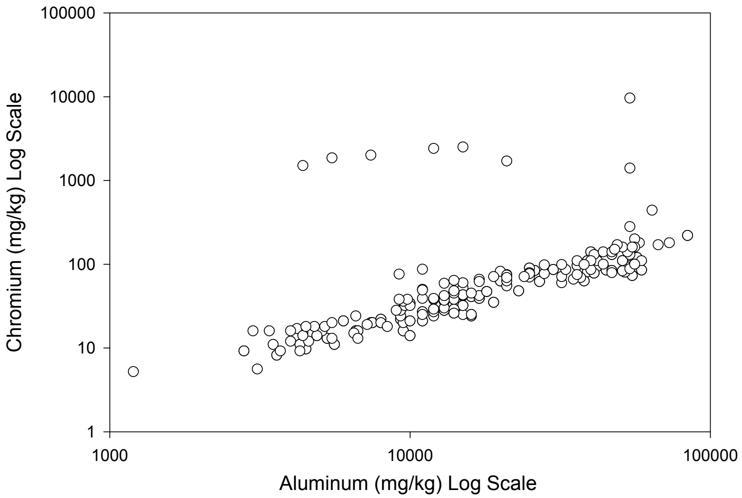


FIGURE 3-4. Scatter plot of chromium vs. aluminum concentrations (log scale) showing unclear association

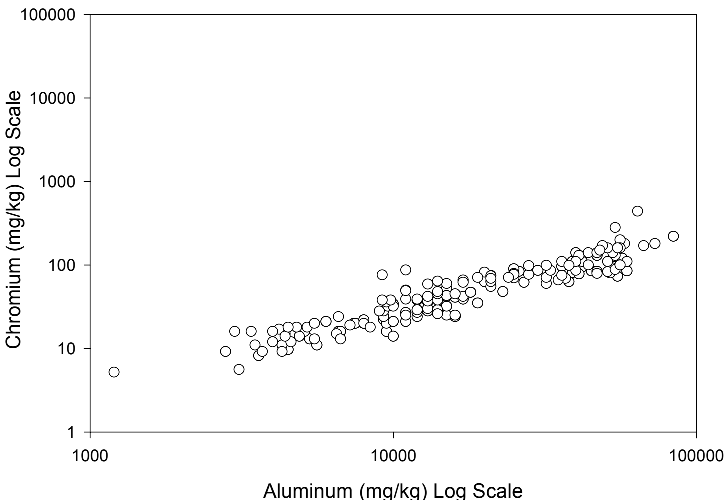


FIGURE 3-5. Scatter plot of chromium vs. aluminum concentrations (log scale) with outliers removed

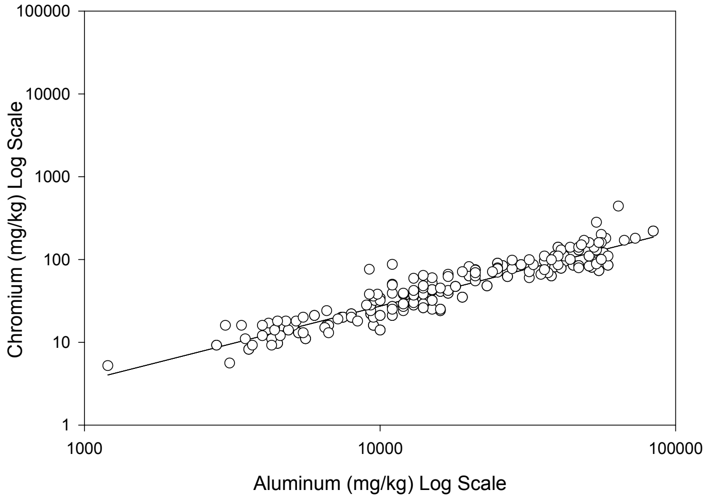


FIGURE 3-6. Scatter plot of chromium vs. aluminum concentrations showing least-squares regression line

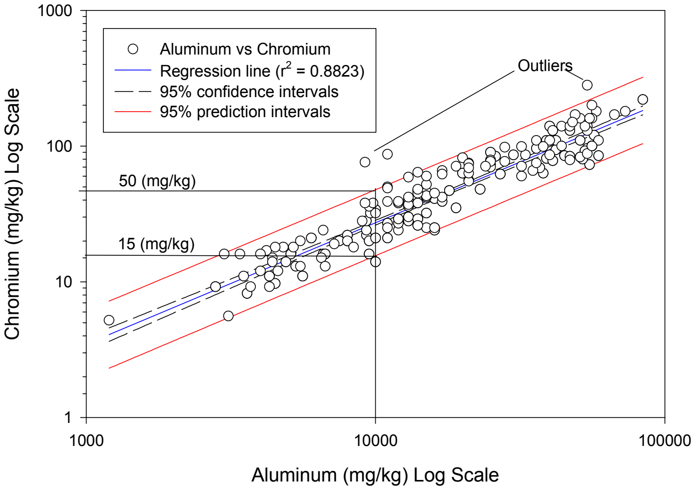


FIGURE 3-7. Geochemical regression: chromium vs. aluminum concentrations (log scale)

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:

$$\log(\mathrm{Cr}) = a + b \log(\mathrm{Al})$$

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.

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-squares 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-squares 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

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
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 done 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.

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.

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 established and should be documented for

TABLE 3-1. 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.

use in subsequent basin-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).

If exploratory data analysis, the Geochemical Method, and the Comparative Method fail to provide satisfactory results, investigators may decide to use factor analysis, as described in Section 2.3.

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 sediment relative to the 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.

3.3.1 Natural Enrichment

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, soils, and sediments (e.g., quartz vein and placer metal deposits). However, ore deposits are relatively rare; the enrichment that occurs when metals are redistributed by weathering, soil formation, and sedimentary processes are more important for background analysis. The most significant natural enrichment processes are summarized below:

- ***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. These secondary aluminosilicates may replace minerals with high trace metal concentrations, resulting in aluminum enrichment and trace metal depletion.
- ***Sorption.*** Fine-grained soils and sediments containing high concentrations of clay minerals or organic matter tend to contain high concentrations of naturally occurring metals due to sorption of metallic ions.
- ***Biological reactions***. Metal uptake by vegetation and adsorption by organic matter produced by plants and animals may result in high concentrations of certain metals.
- ***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).

3.3.2 Enrichment Factors

The enrichment factor (*E*) corresponding to each sediment 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, basin-specific metal concentrations and ratios can be established by whole rock analysis.

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

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

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 sediment relative to basalt. The enrichment factor will equal 1 if no enrichment has occurred. If natural enrichment has occurred through processes such as chemical precipitation or surface adsorption, enrichment factors will be >1. Metal contamination will also 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).

3.3.3 Enrichment Factor Calculation and Plotting

The first step in the analysis is to investigate the geology of the basin and the sediment characteristics to identify the parent rock from which the sediment 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 in the vicinity of the sediment basin are available, they should be used in place of literature values.) The suspected COPC metal and normalizing metal ratios in sediment samples then are calculated, and the enrichment factor 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 generate 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 sediment than in the basalt parent rock, the enrichment factors ($[C_{Cr}/C_{Al}]_{sample}/[C_{Cr}/C_{Al}]_{basalt}$) indicate that the sediment is enriched in chromium relative to the parent rock.

3.3.4 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 sediments by factors of up to approximately 286. The inflection point (increase in slope) on the probability plot marks the approximate

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 sediment samples (e.g., $[C_{Cr}/C_{Al}]_{sample1,}$ $[C_{Cr}/C_{Al}]_{sample\ 2,}$....$[C_{Cr}/C_{Al}]_{sample\ n}$)
6. Calculate the enrichment factor for each sample (e.g., $[C_{Cr}/C_{Al}]_{sample}/[C_{Cr}/C_{Al}]_{basalt}$).
7. Arrange the data in order of increasing enrichment factors.
8. Calculate the cumulative percentage that corresponds to each data point (see Section 2.2.3).
9. Construct a probability plot using the paired enrichment factor–cumulative percentage data.
10. 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

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.

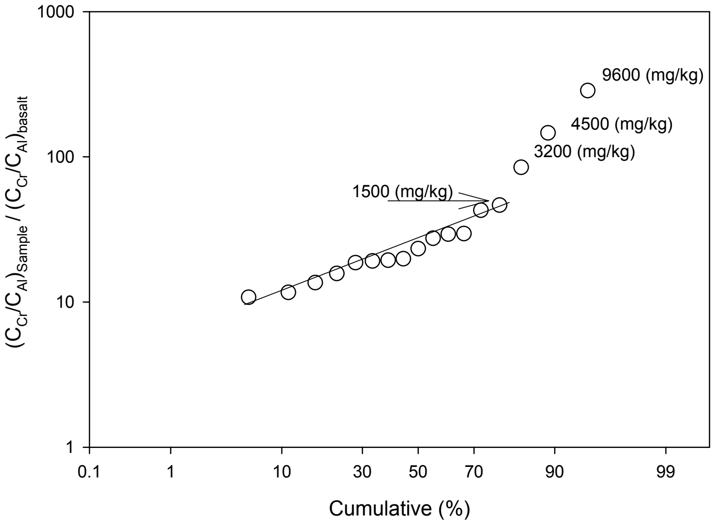


FIGURE 3-8. Probability plot of geochemical enrichment factors

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 basin-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 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 basin 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 re-analyze samples by a method that will yield lower RLs.

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

The Comparative Method uses statistical hypothesis tests to evaluate background conditions by comparing various properties of the potentially impacted and background sediment datasets. The bases of these tests are hypotheses concerning the presence of naturally occurring or anthropogenic chemicals within a sediment basin. These hypotheses must be supported by available information on the nature of potential releases within the sediment basin and the watersheds that drain into the basin. Examples of such hypotheses are provided in Section 2.1.4. The Comparative Method requires a dataset that represents the potentially impacted area, and one or more datasets that represent background or reference areas. Ideally, both potentially impacted and background datasets should be unbiased and representative of geochemically, hydrodynamically, ecologically and anthropogenically similar domains. Furthermore, the two datasets should be nearly the same size. In practice, potentially impacted datasets are usually large and biased or clustered toward areas of concern, such as the immediate vicinity of a discharge or outfall point, whereas background datasets are typically small and randomly collected. In such cases, the potentially impacted datasets may have to be declustered prior to statistical testing, as discussed in Section 4.1.2.

If background datasets do not exist, then appropriate reference areas must be identified for sediment sampling. This sampling process must consider the unique aspects of sediment background areas (Box 4-1) and must be conducted in accordance with DQO principles. The scope of the background sediment analyses also must be adequately comprehensive in order to demonstrate the geochemical, physical, ecological and anthropogenic similarity of the potentially impacted and background areas. Additional data, particularly biological and ecological information, may be necessary for this purpose (Section 2.1). Concurrence of various stakeholders on the appropriateness and representativeness of the background dataset is necessary for 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

BOX 4-1. Unique aspects of sediment background area selection

Sediment reference areas must be identified based on an understanding of the hydrodynamic regime of the basin. If the water flow is unidirectional, such as in stream sediment environments, background data can be collected at locations upstream of point sources. For example, locations upstream of an outfall that discharges COPCs to a stream could be used as reference areas (if the physical, chemical, and biological characteristics of the upstream and downstream locations are sufficiently similar). The reference samples should be collected far enough upstream of the release source to ensure that turbulent mixing of potentially impacted and nonimpacted sediments is minimal.

For estuarine sediments, it may not be possible to identify an "upstream" reference area. In this situation, candidate locations away from the source or contaminated areas should be thoroughly investigated before they can be viewed as suitable reference areas. For example, in the marine sediment investigation of Allen Harbor (located adjacent to the former U.S. Naval Construction Battalion Center, Davisville, North Kingstown, RI), ecological and biological indicators were characterized at a number of nearshore locations to identify areas where no observable adverse benthic effects existed (SAIC, 1998). The results of 10-day amphipod bulk sediment bioassays, and sea urchin fertilization porewater toxicity tests were evaluated to identify nonimpacted areas. Nontoxic conditions were defined as amphipod survival ≥80% or sea urchin fertilization ≥70%. Chemical concentrations from the no-observable-benthic-effect locations then were used as reference data.

graphic and statistical analyses to determine if chemical concentrations in sediment at a potentially impacted area are significantly elevated relative to concentrations in background areas. If chemical concentrations at the potentially impacted area are significantly higher than those detected in the background area, the chemicals are declared COPCs. Uncertainty in these decisions due to a limited number of sediment samples (as a result of inevitable resource constraints) is accounted for by statistical tests of hypotheses. The key questions 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 to reduce the probability of falsely concluding that a chemical is a COPC?
- ❑ How is the necessary number of samples (data points) for the selected statistical test determined?
- ❑ How is the selected statistical test performed?

This section focuses primarily on the application of the Comparative Method to sediment chemical data, but it also can be applied to certain types of ecological and biological data (Box 4-2). Such data include results of location-specific toxicity and bioaccumulation tests, as well as quantitative indicators of benthic conditions.

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: (1) the number of samples required for the various tests to achieve the specified DQO performance goals, (2) the particular distribution (normal or lognormal) expected of the data to be collected, (3) the likely spatial pattern of sediment contamination, and (4) information in published statistical papers that demonstrate the performance of the candidate tests for various data distributions and contamination scenarios. After all data have been collected and exploratory data analyses have been conducted as discussed in Section 2, a final selection of the statistical test(s) can be made. To aid the user in selecting the most appropriate statistical test(s), the assumptions, advantages and disadvantages of each test discussed in this chapter are provided in Table 4-1.

BOX 4-2. Comparative analysis of ecological and biological data

Sediment background analyses are usually conducted in conjunction with ecological investigations of the sediment basin. These investigations commonly generate ecological and biological data, such as toxicity and bioaccumulation test results. The Comparative Method is applicable to such data if: (1) the data can be expressed as quantitative, location-specific measurements, and (2) the data can be separated into distinct datasets for statistical comparison (i.e., potentially impacted data vs. background data).

For example, toxicity test results from potentially impacted and reference areas can be compared to determine whether the numeric difference between the two datasets is statistically significant. For this purpose, the test results (e.g., observed amphipod survival rates, or sea urchin fertilization rates) are segregated according to location as potentially impacted or background values. The resulting numeric datasets then can be analyzed by the Comparative Method. Other examples of quantitative biological data that could be analyzed by the Comparative Method include bioaccumulation data expressed as tissue concentrations of organisms such as benthic invertebrates that inhabit both the potentially impacted and reference areas.

If the biological data are not quantitative or cannot be separated into at least two distinct datasets (e.g., potentially impacted vs. nonimpacted), the Comparative Method cannot be used. Examples of such data are qualitative benthic community indicators, and tissue concentration data from mobile organisms, such as fish that inhabit or roam over the entire sediment basin (including both potentially impacted and nonimpacted [reference] areas).

TABLE 4-1. Assumptions and advantages/disadvantages of statistical tests to detect when potentially impacted 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 potentially impacted and background concentration distributions. • More nondetects are allowed than for other tests considered in this section. • At least one detected (quantified) background measurement is present and it is larger than the largest nondetect value. • No assumptions are required with regard to the shape of potentially impacted and background data concentration distributions.	• Very simple to conduct the test. • No distribution assumptions are necessary. • Many nondetects are permitted. • Can be used in conjunction (in tandem) with tests that focus on the detecting differences in the mean or median.	• May require a large number of measurements to have adequate power to detect differences in potentially impacted and background concentrations.
Quantile Test	• Objective is to test for differences in the right tail (largest values) of the potentially impacted and background concentration distributions. • Below-detection values are not among the largest *r* data values in the pooled set of potentially impacted and background data. • No assumptions are required with regard to the shape of the potentially impacted 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 potentially impacted 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 a large number of measurements to have adequate power to detect differences in potentially impacted and background concentrations. • Test may be inconclusive if nondetects are present among the largest *r* data values.
Wilcoxon Rank Sum Test	• Objective is to test for differences in the medians of the potentially impacted and background populations. • Only one RL (all nondetects have the same value), which is less than the smallest detected datum. • No more than 40% of both the potentially impacted and background datasets are nondetects. • The potentially impacted and background data concentration distributions have the same shape (variance).	• Nonparametric—i.e., no distribution assumptions necessary (however, the test is based on the assumption that the variance of the potentially impacted distribution is the same as the variance of the background distribution). • In general, the test has more power to detect shift in potentially impacted median than the two-sample t tests when the potentially impacted 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 potentially impacted and background distributions, as well as differences in medians, can be detected. • Although manual calculations are labor intensive, this test can easily be performed using a number of readily available statistical software packages.	• Relatively more complex to compute by hand. • Too many nondetects prevent use of the test.

TABLE 4-1. (cont'd)

Test Statistic	Objectives/Assumptions	Advantages	Disadvantages
Gehan Test	• Objective is to test for differences in the medians of the potentially impacted and background populations. • Nondetects do *not* have the same value (multiple RLs exist). • The censoring mechanism that generated the nondetects is the same for the potentially impacted and background populations. • No assumptions are required with regard to the shape of the potentially impacted and background data concentration distributions.	• Can be used when multiple RLs 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 potentially impacted and background data.
Two-Sample t Test	• Objective is to test for differences in the means of the potentially impacted and background populations. • Both potentially impacted and background mean concentrations are normally distributed. • Below-detection-values have no significant impact on computed means (e.g., less than 15% of measurements are below detection). • The potentially impacted and background data distributions have the same shape (variance).	• Most powerful test for detecting a shift in the potentially impacted mean from the background mean, if the potentially impacted and background data are normally distributed. • Certain transformations (e.g., Box-Cox) may be able to normalize the data. These transformations can be readily performed using statistical software.	• The test requires a statistical evaluation of the assumption of equal total variances for the potentially impacted 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 often is violated. • Outliers can affect the test results. • Not well suited for datasets that contain nondetects.
Satterthwaite Two-Sample t Test	• Objective is to test for differences in the means of the potentially impacted and background populations. • Both potentially impacted and background data have a normal distribution. • No nondetects are present. • Potentially impacted and background data distributions are expected or known to have different shapes (i.e., unequal variances).	• Test can be used when the potentially impacted and background distributions have unequal variances.	• The test is relatively complicated to compute by hand. • Same disadvantages as for the two-sample t test.
Two-Sample Test of Proportions	• Objective is to test for differences in the proportions of the potentially impacted and background data above a given cutoff level. • Test may be used when more than 50% of the potentially impacted or background datasets are nondetects. • No assumptions are required with regard to the shape of the potentially impacted and background data concentration distributions.	• No distribution assumptions are necessary. • Relatively simple test to perform. • Can be used when many nondetects 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).

In this regard, selection of the optimal test depends in part on the spatial pattern of the contaminated sediment, which is expressed as a hypothesis, such as:

- ***Basin-wide impact:*** The entire distribution of measurements from the potentially impacted area is likely to be simply shifted to higher values than the distribution of background measurements. In this case, the difference between the two datasets can be assessed by comparing their mean or median concentrations.

- ***Localized or Hotspot Impact:*** Only a small portion of the distribution of potentially impacted measurements is expected to be higher than the distribution of background measurements. In this case, the difference between the two datasets can be assessed by comparing their highest concentrations.

For the case of a basin-wide impact, the Wilcoxon Rank Sum (WRS) test, the Gehan test, the two-sample t test, and the Satterthwaite two-sample t test are preferred. The Slippage test, Quantile test, and two-sample test of proportions are best suited to identify chemicals that have elevated concentrations in only a small portion of the potentially impacted area (i.e., localized or hotspot impact). If portions of a sediment basin can be segregated into distinct areas or strata, the WRS, Gehan, or two-sample t test can be used to compare data from each area or stratum to the background dataset. The main features of these tests are discussed in Table 4-1 and subsequent parts of this section.

Basic Assumptions of the Recommended Tests

All tests discussed in this section require that the potentially impacted and background measurements are independent (spatially or temporally uncorrelated) and representative of their respective populations. As discussed in Section 2.1.6, this assumption requires:

- ***Unbiased Coverage of the Investigated Areas:*** Biased and clustered measurements may lead investigators to over- or under-estimate background concentrations; therefore, an unbiased (e.g., probability-based) sampling design should be used to select sampling locations. If the analysis relies on previously existing sediment data, the basis of the sampling design must be evaluated. Because judgmental sampling designs often are biased toward potentially contaminated areas, sampling locations often are clustered in impacted areas. In this case, background concentrations may be over-estimated. Other judgmental sampling designs can lead investigators to under-estimate background concentrations. For example, background concentrations are likely to be underestimated if the sampling design is biased toward areas suspected to contain the lowest levels of non-site related contamination, or toward areas with large grain-size sediments. Isaaks and Srivastava (1989, Chapter 10) provide a thorough introduction to the effects of biased and clustered data, and describe various declustering techniques (see Box 4-3).

- ***Independent Data:*** The sediment samples should be collected far enough apart in space and time to minimize any spatial or temporal correlations among measured concentrations. In many instances, however, the sediment data will exhibit spatial correlations. This is particularly true for small investigation areas, where sampling locations tend to be located close together. In these situations, spatial procedures, including geostatistics (Isaaks and Srivastava, 1989), should be used to analyze the data and assess differences between measured concentrations at potentially impacted and background locations.

4.1.3 Statistical Testing Approaches Not Recommended

This section describes two methods for comparing potentially impacted and background sediment data that are not recommended to identify COPCs. These methods are not acceptable, because the probability of error can be very high.

Comparing Maximum Potentially Impacted and Maximum Background Measurements

One approach to test whether a chemical is a COPC is to compare the maximum measurement

BOX 4-3. Declustering of sediment data

In many investigations, the data available for background analysis include sampling results from locations that are clustered and biased toward zones of concern. Typical sediment zones of concern are drainage channels at impacted sites, outfall discharge points, and areas downstream of impacted terrestrial areas. Such biased data tend to overrepresent highly impacted sediments. If excessive overestimation is anticipated, the existing clustered dataset should be declustered prior to analysis. A variety of declustering techniques exist (Isaaks and Srivastava, 1989, Chapter 10). These techniques include cell declustering, in which the investigated sediment area is divided into equal cells by a regular grid. The data within each cell are then reduced or replaced according to one of the following alternatives:

- Randomly selected data points: Only a fixed number of randomly selected sediment data points, e.g., 2 or 3 points, within each cell are retained.
- Average data point: All sediment data points within the cell are replaced by a single data point with a value equal to the average of all measured concentrations within the cell. The location of this point is defined as the average easting and northing of the data points within the cell. This approach is appropriate only if the potentially impacted dataset and the background dataset are both represented by cell-averaged values. For example, comparison of cell-averaged values from the potentially impacted area to background concentrations representing individual sampling locations could yield misleading results.
- Maximum data point: Only the sediment data point with the highest measured value within each cell is retained. This approach is appropriate only if the potentially impacted dataset and the background dataset are both represented by maximum cell values. For example, comparison of maximum cell values from the potentially impacted area to background concentrations representing individual sampling locations could yield misleading results.

Any of the above declustering alternatives will result in smaller datasets. For all the statistical tests, the number of samples (data values) in each dataset (the background and potentially impacted datasets) should be at least 10 and, best case, more than 20. This will help to ensure that the tests have sufficient power to reliably detect COPCs.

among potentially impacted data with the maximum background measurement, using the following decision rule:

❑ If the maximum potentially impacted measurement exceeds the maximum background measurement, then declare the chemical a COPC; otherwise declare the chemical not a COPC.

As discussed in O'Brien and Gilbert (1997), if the potentially impacted and background datasets have the same concentration distribution and the same number of data points, the probability is 50% that the maximum measurement occurs in the potentially impacted dataset and 50% that it occurs in the background dataset. Thus, the chance is 50% that the chemical will be declared to be a COPC, when in fact the chemical occurs at background levels in the potentially impacted area. The probability of erroneously declaring a chemical to be a COPC increases if the potentially impacted dataset has more data points that the background dataset. In fact, if the potentially impacted 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 $p = 20/30 = 2/3$. In this case, the probability that this testing approach would erroneously identify a chemical as a COPC is 67%.

Clearly, this decision rule is not acceptable because: (1) the probability of incorrectly identifying a chemical as a COPC can be very high, and (2) correct determination of whether or not a chemical is a COPC is critically dependent on which area—the potentially impacted area or the background area—is represented by the most measurements. (Procedures for background analysis should not be confused with U.S. EPA or Navy risk assessment protocols: although maximum concentrations typically are compared to risk-based screening criteria in the initial phase of the risk assessment process, they should not be used to test whether a chemical is a COPC with respect to background.)

Comparing the Maximum Potentially Impacted Measurement to a Background Threshold

Another decision rule that might be used to decide if a chemical at the potentially impacted area is a COPC is:

- If one or more potentially impacted measurements exceed the 95th percentile of the background area data distribution, declare the chemical of interest to be a COPC.

Suppose the potentially impacted 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 that the probability that one or more of n potentially impacted measurements will exceed the 95th percentile is equal to $1 - (0.95)^n$, where 0.95 is the probability that any randomly drawn (representative) potentially impacted 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 potentially impacted distributions are identical and n = 21, the probability that one or more of the potentially impacted measurements will exceed the 95th percentile of the background distribution is 0.67. In other words, there is a 67% chance of falsely identifying a chemical as a COPC. If more extensive sampling is conducted at the potentially impacted area, for example, if n = 64, the probability of falsely concluding that the chemical is a COPC is 96%!

Threshold values such as the 90th or 99th percentiles could also be used. Other threshold values that might be suggested as appropriate decision criteria include the background mean, two times the background mean, or an upper confidence limit on the background mean. 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, if potentially impacted measurements are individually compared to the threshold value, the basic problem of excessive decision errors remains—only the specific probability of making an erroneous decision changes.

TABLE 4-2. Probabilities that one or more of *n* potentially impacted measurements will exceed the 95th percentile of the background distribution if the potentially impacted 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

Based on the above discussion, it is recommended any background threshold comparison:

- Should only be considered as a means to suggest the need for additional investigation of whether or not a chemical is a COPC; and
- Should never be the only test applied to determine if a chemical is a COPC.

4.2 Recommended Comparative Statistical Tests

Unlike the above background threshold comparisons, comparative statistical tests provide tools that do not have the problems of elevated false decision error rates. These tests are discussed in this section.

Following are some general words of advice about using comparative statistical methods to decide which chemicals are COPCs:

- ***Datasets*:** Background datasets should be comparable to the potentially impacted data. Ideally, both datasets will be unbiased and representative of geochemically, hydrodynamically, ecologically, and anthropogenically similar sediment domains. Furthermore, the two datasets should be nearly the same size. In practice, potentially impacted datasets are usually large and biased or clustered toward areas of concern, whereas background datasets are typically small and randomly collected. In these situations, the potentially impacted datasets may have to be declustered prior to testing (Section 4.1.2).

- ❑ ***Support Analyses*:** Results of the exploratory data analyses (Section 2.2), such as descriptive statistics and probability plots, should be used as the foundation for formal hypothesis testing by the Comparative Method.
- ❑ ***Assumption Verification*:** The assumptions that underlie a formal statistical hypothesis test should always be reviewed. For example, some statistical tests require normally distributed data, or that the variances of the potentially impacted and background datasets are equal.
- ❑ Preferred Tests:
  - ♦ Use the nonparametric Slippage test (Section 4.2.1) as a quick way to decide which chemicals are COPCs by comparing extreme measurements from the potentially impacted area to the maximum background measurements. This test assumes that if an impacted sediment zone exists, it is limited to only a portion of the potentially impacted area, i.e., the localized or hotspot impact hypothesis.
  - ♦ Use the nonparametric Quantile test (Section 4.2.2) if an important criterion for deciding which chemicals are COPCs is whether the extreme concentrations in the potentially impacted area are higher than the extreme background concentrations. Similar to the Slippage test, the Quantile test assumes that if an impacted sediment zone exists, it is limited to only a portion of the potentially impacted area.
  - ♦ Consider using the nonparametric WRS test (Section 4.2.3) to compare median potentially impacted measurements to median background measurements and thus decide which chemicals are COPCs. The WRS test assumes that if an impact has occurred, it covers the potentially impacted area more or less uniformly, i.e., the basin-wide impact hypothesis.
  - ♦ Use the nonparametric Gehan test (Section 4.2.4) instead of the WRS test if the background or potentially impacted datasets contain multiple nondetects. Similar to the WRS test, the Gehan test assumes that if an impact has occurred, it covers the potentially impacted area more or less uniformly.
  - ♦ Use the two-sample t test (Section 4.2.5) if the mean concentrations of the background and potentially impacted datasets are assumed to be normally distributed with about the same variance, and if very few or no nondetects are present. This test assumes that if an impact has occurred, it covers the potentially impacted area more or less uniformly.
  - ♦ Use the Satterthwaite two-sample t test (Section 4.2.6) if the mean concentrations of the background and potentially impacted datasets are assumed to be normally distributed with different variances, and if very few nondetects are present. Similar to the two-sample t test, this test assumes that if an impact has occurred it covers the potentially impacted area more or less uniformly.
  - ♦ Use the nonparametric two-sample test of proportions (Section 4.2.7) if more than 50% of the background or potentially impacted measurements are nondetects. This test focuses primarily on the portion of measurements in excess of a given cutoff value.
- ❑ ***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 nondetects and do not require any specific statistical distribution assumptions.

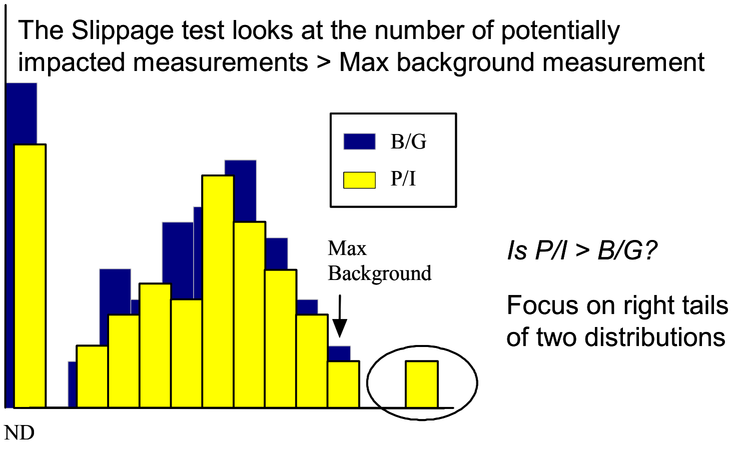
- ***Comparing Multiple Datasets:*** In many sediment investigations, particularly estuarine sediment investigations, measurements from multiple areas or stations should be compared simultaneously to identify zones that are statistically elevated with respect to other areas. In these instances, the Kruskal-Wallis or similar tests (Box 2-7) should be used.

- ***Direct Statistician Involvement:*** Consult an experienced environmental statistician whenever disputes regarding the most appropriate graphic or statistical testing methods arise. Analysis of spatially correlated sediment data also requires the involvement of an experienced geostatistician. The application of statistics and geostatistics requires a thorough knowledge of statistical methods for environmental applications, and the appropriate conditions for each statistical test.

4.2.1 Slippage Test

Sediment Contamination Scenario

Site history information indicates that operations may have released a chemical into the nearby sediment basin. The particular chemical of interest also is known to be present in sediments due to naturally occurring sources. The decision question is: are concentrations of this chemical within the potentially impacted area greater than those in the background area? If so, the chemical will be considered a COPC. Knowledge of site operations suggests that if releases of the chemical did occur, the chemical is likely to have impacted only a portion of the potentially impacted area of the sediment basin (i.e., localized impact).


The Slippage test looks at the number of potentially impacted measurements > Max background measurement

Role of the Data Quality Objectives Process

The stakeholders and regulators used the DQO planning process to agree:

- On the methods that should be used to collect, handle, prepare, and measure the sediment samples.

- That nondetects may occur frequently.

- That larger measurements among the potentially impacted and background data should be the primary criteria used to decide whether a chemical is a COPC.

- On the values of the design parameters used to determine the required number of potentially impacted and background sediment measurements (see the following subsection entitled “Guidance on Implementing the Slippage Test”).

The Slippage test is appropriate for this scenario because it uses only the largest few data values and does not require any assumptions about the underlying distributions of the potentially impacted and background measurements. The assumptions behind the Slippage test are summarized in Table 4-1.

Advantages and Disadvantages

- The Slippage test involves counting the number of potentially impacted measurements that exceed the largest background datum and comparing that count with a critical value from a special table (see Box 4-4). Therefore, the Slippage test is extremely easy to conduct.

BOX 4-4. Procedure for conducting the Slippage test

1. Specify the tolerable probability, α, that the Slippage test will incorrectly declare that the potentially impacted concentrations tend to be higher than the background concentrations, i.e., the probability that the chemical will be incorrectly identified as a COPC. α can be set only at 0.01 or 0.05, because critical values for the test are available only for these two α values (Step 7 below). 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 levels selected for each test.
2. Specify the values of ε (the proportion of a site within which chemical concentrations are substantially greater than background levels) and of the power (1 – β) the stakeholders and regulators have decided are important for the Slippage test.
3. Determine the approximate 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 analyze each sample for the chemical of interest. Some of the measurements may be nondetects.
5. Determine the value of the largest *detected* background measurement. In making this determination, ignore all nondetects that may be present in the background dataset.
6. Count the number, *K,* of detected *potentially impacted* measurements that are larger than the largest detected background measurement. In making this determination, ignore all nondetects in the potentially impacted dataset.
7. If α was set at approximately 0.01, determine the critical value K_c from Table C-2. If α was set at 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 that the potentially impacted concentrations for the chemical of interest tend to be larger than the background concentrations for that chemical, i.e., the chemical is a COPC.

- The Slippage test considers only the largest background measurement and the largest measurements among the potentially impacted data. Therefore, it is important to verify that these elevated values do not represent mistakes or errors made during sample collection, handling, measurement, or data processing. Statistical tests for outliers (Appendix B.3) can be used for this purpose. If the extreme values prove to be outliers, they should be scrutinized and eliminated if they are the result of 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.

- In general, the Slippage test is designed to address the localized or hotspot impact hypothesis. If the exact nature of sediment contamination is unknown, other tests, such as the WRS test, should be used to assess the basin-wide impact hypothesis.

- The Slippage test can be viewed as a quick test to assess the COPC-status of a chemical.

- 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 potentially impacted and background measurements, *n* and *m*, respectively, required for the test to have adequate power to correctly declare that 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 potentially impacted data that has concentrations greater than background.

- The magnitude of the difference between potentially impacted area concentrations and background area concentrations.
- The tolerable probability, α, that the Slippage test will declare the chemical to be a COPC when in fact it is not a COPC.
- The underlying distributions (for example, normal or lognormal) of the potentially impacted 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 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. These results are for the case where the 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:

- If potentially impacted and background measurements have already been collected and the budget does not allow for additional sampling, the information in Table 4-3 can be used to approximately determine if a power of 0.80 or 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 site-basin hydraulic connections, sediment basin hydrodynamics, and site operations, should be used to the maximum extent in making COPC decisions.
- If a value of $\alpha < 0.025$ is selected, the number of samples in Table 4-3 would have to be increased for the Slippage test to retain a power of 0.80 or 0.90. If a value of an $\alpha > 0.05$ 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.

Box 4-4 gives the Slippage test procedure. Examples are provided in Boxes 4-5 and 4-6.

4.2.2 Quantile Test

Sediment Contamination Scenario

The sediment contamination scenario described for the Slippage test also applies to the Quantile test. The decision question is: are concentrations of this chemical within the potentially impacted area greater than those in the background area? If

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 potentially impacted area has concentrations substantially larger than background concentrations

	Number of Required Measurements (*n* and *m*)	
	Power	
Selected Proportion	**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).

BOX 4-5. 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
   Potentially Impacted 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 potentially impacted measurements are larger than 79.
7. Using Table C-2 with $n = m = 10$, we find the critical value K_c is 6.
8. Therefore, the Slippage test declares that the evidence is insufficient to declare that 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.

BOX 4-6. 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
   Potentially Impacted 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 potentially impacted measurements are larger than 25.
7. Using Table C-3 with $n = m = 15$, we find the critical value K_c is 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 also would be performed to complement the results of the Slippage test. However, the datasets contain so many nondetects 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.

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 impact is most likely limited to a few hotspots. This situation suggests that the Quantile test is appropriate.

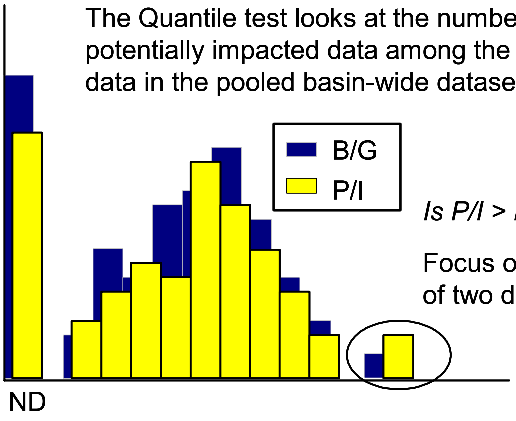
Role of the Data Quality Objectives Process

The stakeholders and regulators used the DQO planning process to agree:

- ❑ On the methods that should be used to collect, handle, prepare, and measure the sediment samples.

- That nondetect measurements may occur frequently.
- On the values of the design parameters used to determine the necessary number of potentially impacted and background sediment measurements (see the following subsection entitled "Guidance on Implementing the Quantile Test").

The Quantile test is appropriate for this scenario because (1) it is a valid test regardless of the underlying distribution of the potentially impacted and background data, and (2) the test looks for differences in the right tails of the potentially impacted and background concentration distributions.



The Quantile test looks at the number of potentially impacted data among the largest data in the pooled basin-wide dataset.

Is P/I > B/G?

Focus on right tails of two distributions.

In many instances, the spatial distribution of impacted sediments is not accurately known, i.e., the contaminated sediments may be either uniformly distributed throughout the basin, or localized. In these cases, the stakeholders and regulators may consider using other tests, such as the WRS test, in addition to the Quantile test. The assumptions behind using the Quantile test are summarized in Table 4-1.

Advantages and Disadvantages

- The Quantile test is closely related to the Slippage test. It consists of looking at the largest *r* measurements in the pooled potentially impacted and background datasets, and counting the number of those measurements that are from the potentially impacted area. If *k* or more of the *r* measurements are potentially impacted measurements, the Quantile test declares the chemical to be a COPC.
- Any number of nondetects is permitted in the potentially impacted and background datasets, as long as all nondetects are smaller than the smallest of the *r* largest detected measurements in the pooled dataset.
- 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.
- The Quantile test focuses on comparing the right tails of the potentially impacted and background distributions rather than comparing the medians or means of the two distributions. This implies that the Quantile test is appropriate to address the localized impact hypothesis. If the nature of the impact is unknown, the Quantile test should be used in tandem with the WRS test. The WRS test addresses the basin-wide impact hypothesis by looking for differences in the medians.
- Table 4-1 provides a summary of the advantages and disadvantages of the Quantile test. The Quantile test procedure is shown in Box 4-7. Boxes 4-8 and 4-9 provide two examples of its use.

Guidance on Implementing the Quantile Test

As with other tests discussed in this document, the first step in implementing the Quantile test is to determine the required number of potentially impacted and background measurements, *n* and *m*, respectively. The required number of measurements depends on the desired power of the test to declare (when it is true) that the chemical of interest is a COPC. As with the Slippage test, the required values of *n* and *m* also depend on the following:

BOX 4-7. Procedure for conducting the Quantile test

1. Select the tolerable probability, α, that the Quantile test will incorrectly declare that the potentially impacted area concentrations tend to be larger than the background concentrations. The probability α may be set at 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 values of ε and power specified in Table 4-4 to approximate the required number of potentially impacted and background measurements. Table 4-5 may be used if it is important to detect potentially impacted concentrations that are only slightly larger than background.
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 analyze each sample for the chemical of interest. Some of the measurements may be nondetects. If samples have already been collected and analyzed, verify that the number of samples is in agreement with Table 4-4 or Table 4-5. Collect and analyze additional samples if necessary.
5. List the pooled potentially impacted and background measurements from smallest to largest. The total number of pooled measurements is n + m.
6. Using the values of *n* and *m*, use 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 potentially impacted and background measurements whether or not *k* or more of the largest *detected r* measurements are potentially impacted measurements. (Note: ignore any nondetects when determining the largest detected *r* measurements). If so, the Quantile test indicates that the chemical is a COPC. If not, the test indicates that the data are insufficient for the Quantile test to conclude that the chemical is a COPC, and the WRS test should be conducted.

BOX 4-8. Example 1 of the Quantile test

1. Suppose α = 0.05 is selected.
2. Suppose ε = 0.50 is selected, and a power of 0.80 is needed to determine whether potentially impacted concentrations are distinctly higher 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 were used to illustrate the Slippage test in Box 4-2):

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

5. The 20 pooled and ordered background and potentially impacted data are (*P* and *B* indicate Potentially Impacted and Background, respectively):

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

6. As α = 0.05 was selected in Step 1, Table C-6 indicates that 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 potentially impacted area. Therefore, because 3 < k= 4, the Quantile test indicates the data are insufficient to conclude the chemical is a COPC.

BOX 4-9. Example 2 of the Quantile test

1. Suppose α = 0.01 is selected.
2. Suppose ε = 0.50 and a power of 0.80 is needed to determine whether potentially impacted concentrations are distinctly higher 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
   Potentially Impacted Data: <5, <10, 11, 13, <22, 23, <24, <36, <40, 70, 89, 100, 115, 200, 300

5. The 30 pooled and ordered background and potentially impacted data are:

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

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

6. As α = 0.01 was selected in Step 1, Table C-4 indicates that 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 potentially impacted dataset. Therefore, because *k* (i.e., 6) of the highest *r* (i.e., 6) measurements are from the potentially impacted area, the Quantile test indicates that the chemical is a COPC.

1. The proportion, ε, of the potentially impacted area that has concentrations greater than the background area.
2. The magnitude of the difference between potentially impacted area concentrations and background area concentrations.
3. The tolerable probability, α, that the Quantile test will declare that the chemical is a COPC when in fact it is not a COPC.
4. The underlying distribution (e.g., normal or lognormal) of the potentially impacted and background concentration measurements.

U.S. EPA (1994, 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 a prescribed power to correctly declare that a chemical is a COPC. A portion of this information is summarized in Tables 4-4 and 4-5. These tables list the approximate number of measurements needed for the Quantile test to have a power of approximately 0.80 or 0.90 to correctly declare that a chemical is a COPC, for cases where the tolerable probability, α, of incorrectly declaring a chemical to be a COPC is 0.01, 0.025, 0.05, or 0.10. Stakeholders and regulators should agree on an acceptable value of α. The numbers of measurements listed in Tables 4-4 and 4-5 were obtained by assuming that the data 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 probably be increased somewhat to achieve the 0.80 and 0.90 power levels.

The measurement numbers listed in Table 4-4 are for cases in which approximately 85% of actual concentrations in the potentially impacted area are higher than background concentrations, i.e., distinctly higher concentrations. The measurement numbers listed in Table 4-5 are for cases in which only approximately 5% of actual concentrations in the potentially impacted area are higher than background concentrations, i.e., somewhat higher concentrations. The numbers of measurements listed in Table 4-5 are larger than in Table 4-4 because more measurements are necessary to achieve the same power when the concentration differences are small.

The Quantile test can be computed using the software Environmental Stat for S-PLUS (see Appendix D.3).

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 potentially impacted area has concentrations distinctly higher than background concentrations[(a)]

α	0.01		0.025		0.05		0.10	
Power	**0.80**	**0.90**	**0.80**	**0.90**	**0.80**	**0.90**	**0.80**	**0.90**
$\varepsilon = 0.10$	>100	>100	100	>100	80	100	55	70
$\varepsilon = 0.20$	55	60	40	40	35	40	25	35
$\varepsilon = 0.30$	25	30	20	25	20	20	15	15
$\varepsilon = 0.40$	20	25	15	20	15	15	10	15
$\varepsilon = 0.50$	15	20	15	15	10	10	10	10
$\varepsilon = 0.60$	10	15	10	10	10	10	10	10
$\varepsilon = 0.70$	10	10	10	10	10	10	10	10
$\varepsilon = 0.80$	10	10	10	10	10	10	10	10
$\varepsilon = 0.90$	10	10	10	10	10	10	10	10
$\varepsilon = 1.0$	10	10	10	10	10	10	10	10

(a) n = m were obtained for the case where the normal potentially impacted concentration distribution is shifted to the right of the normal background concentration distribution by the amount $\Delta/s = 4$ (U.S. EPA, 1994, 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 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 potentially impacted area has concentrations somewhat higher than background concentrations[(a)]

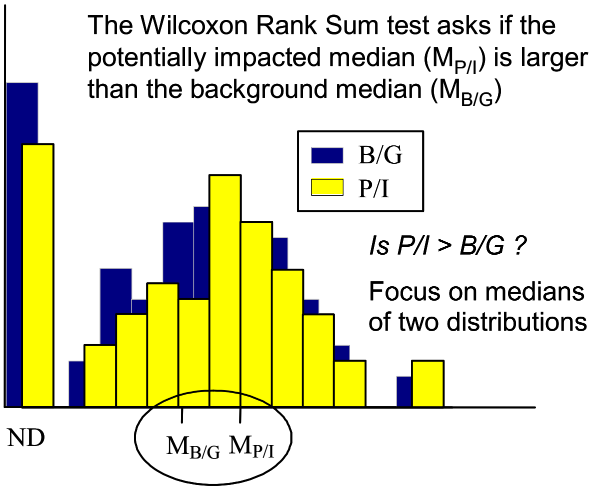
α	0.01		0.025		0.05		0.10	
Power	**0.80**	**0.90**	**0.80**	**0.90**	**0.80**	**0.90**	**0.80**	**0.90**
$\varepsilon = 0.10$	>100	>100	>100	>100	>100	>100	>100	>100
$\varepsilon = 0.20$	>100	>100	>100	>100	>100	>100	>100	>100
$\varepsilon = 0.30$	>100	>100	>100	>100	>100	>100	>100	>100
$\varepsilon = 0.40$	>100	>100	>100	>100	>100	>100	>100	>100
$\varepsilon = 0.50$	>100	>100	>100	>100	>100	>100	>100	>100
$\varepsilon = 0.60$	>100	>100	>100	>100	>100	>100	>100	>100
$\varepsilon = 0.70$	>100	>100	100	>100	75	>100	70	>100
$\varepsilon = 0.80$	>100	>100	75	>100	60	>100	50	>100
$\varepsilon = 0.90$	>100	>100	60	100	50	100	40	100
$\varepsilon = 1.0$	>100	>100	50	75	50	75	30	75

(a) n = m were obtained for the case where the normal potentially impacted concentration distribution is shifted to the right of the normal background concentration distribution by the amount $\Delta/s = 1$ (U.S. EPA, 1994, 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 that the chemical is a COPC.

4.2.3 Wilcoxon Rank Sum Test

Sediment Contamination Scenario

The sediment contamination scenario developed by the stakeholders and regulators during Steps 1 and 2 of the DQO process is based on historical data and expert knowledge which indicate that if contamination exists it is probably more-or-less uniformly distributed throughout the potentially impacted area (rather than occurring in hotspots within that area).


The Wilcoxon Rank Sum test asks if the potentially impacted median ($M_{P/I}$) is larger than the background median ($M_{B/G}$)

Is P/I > B/G ?

Focus on medians of two distributions

Role of the Data Quality Objectives Process

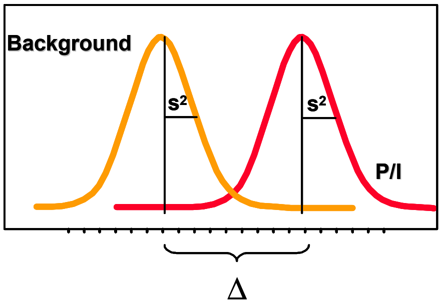
- The stakeholders and regulators used the DQO planning process to agree:
- On the methods that should be used to collect, handle, prepare, and measure the sediment samples.
- That it is unlikely that more than 40% of the measurements will be nondetects.
- On the values of the design parameters used to determine the necessary number of potentially impacted and background sediment measurements (see the subsection entitled "Guidance on Implementing the WRS Test" that follows).

The WRS test is appropriate for this scenario because (1) it is a valid and reliable test regardless of the underlying distribution of the investigated sediment data, and (2) the test determines if the difference between median concentrations of potentially impacted and background data is statistically significant.

The exact nature of the spatial extent of sediment contamination often is unknown. In this case, the Slippage or Quantile test should be conducted in conjunction with the WRS test to address the scenario of hotspots. As noted above, the WRS test is appropriate to assess uniform, basin-wide impacts, whereas the Slippage and Quantile tests are appropriate for hotspot scenarios. Use of the Quantile test along with the WRS test improves the probability of detecting either basin-wide or localized sediment contamination patterns. The assumptions behind the WRS test are summarized in Table 4-1.

Advantages and Disadvantages

- The WRS test is nonparametric—i.e., it does not require any specific assumptions about the exact form of the underlying data distributions.
- The performance (power) of the WRS test (for detecting when the median potentially impacted area concentration is greater than the median background concentration) is known from theory and practice to be as high or higher than other statistical tests that evaluate differences in averages.
- The WRS test is based on the assumption that the underlying distribution of the potentially impacted dataset has the same shape (variance) as the distribution of the background dataset. This assumption implies that the two distributions are the same, except that the potentially impacted data distribution may be shifted to higher concentrations than the background data distribution. The assumption of equal variances should be evaluated using descriptive statistics and graphic plots of the potentially impacted and background data (see Box 2-7).
- If nondetects occur, all of them must have the same RL, and that RL must be less than the smallest reported concentration. If multiple nondetects are present 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 potentially impacted or background datasets are nondetects.
- The WRS test does not place great importance (weight) on the larger potentially impacted and background measurements. It uses and considers all measurements, rather than focusing on the largest measurements (as do the Slippage and Quantile tests).
- The WRS test should be used in conjunction with the Quantile Test so that either basin-wide or localized contamination can be detected with adequate probability.
- The software Environmental Stats for S-PLUS (see Appendix D.3) can be used to compute the WRS test and the Quantile test.



The WRS test asks: Is the potentially impacted data distribution shifted to the right of the background data distribution by a significant amount Δ?

Guidance on Implementing the WRS Test

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

- The tolerable 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 correctly 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 potentially impacted median concentration exceeds the background median concentration that must be detected with the required power.
- The proportion of the total number of potentially impacted and background sediment samples that will be collected in the background area. If this proportion is 0.50, then n = m.

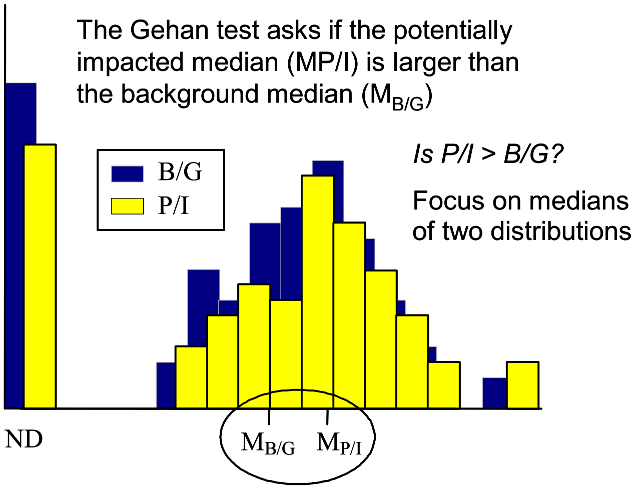
When n = m is desired (the ideal case), a formula for determining the number of potentially impacted and background measurements is given in MARSSIM (NRC et al., 1997, Equation 5-1, p. 5-28). However, rather than use the formulas in U.S. EPA (1994) or NRC et al. (1997), it is simpler to select *n* and *m* from Table 4-6 if it is desired to have n = m. The values of n = m listed in Table 4-6 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).

Table 4-1 summarizes the advantages and disadvantages of the WRS test. Box 4-10 describes the steps necessary to perform the WRS test when n <20 and m <20, and Box 4-11 provides an example of that procedure. Box 4-12 describes how to conduct the WRS test when n ≥20 and m ≥20, and Box 4-13 provides an example of that procedure.

4.2.4 Gehan Test

Sediment Contamination Scenario

The sediment contamination scenario developed by the stakeholders and regulators during Steps 1 and 2 of the DQO process is the same as for the WRS test in Section 4.2.3: If contamination has occurred, it is probably more-or-less uniformly distributed throughout the potentially impacted area of the sediment basin. However, in this case, the datasets are expected to contain multiple non-detects with different RLs.



The Gehan test asks if the potentially impacted median (MP/I) is larger than the background median ($M_{B/G}$)

TABLE 4-6. Number of potentially impacted and background measurements (n and m, $n = m$) required by the WRS test to achieve a desired power[(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	5452	4627	3972	3278	2268	4827	3870	3273	2846	1748	3972	3273	2726	2157	1355	3278	2846	2157	1655	964	2268	1748	1355	964	459
0.2	1370	1163	998	824	570	1163	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 4-10. Procedure for conducting the WRS test when the number of potentially impacted 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 potentially impacted concentrations tend to be larger than the background concentrations, i.e., the probability that the chemical will be incorrectly identified as 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 potentially impacted 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 potentially impacted 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 potentially impacted and background measurements needed when it is desired to have n equal to m. If equal n and m values are not desired, use Equation 6.3 in U.S. EPA (1994) and increase that value by 20% to guard against missing or unusable measurements.
4. Collect the n and m samples and analyze for the chemical of interest; some measurements may be nondetects. 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 and rank the pooled set of $n + m$ potentially impacted and background measurements from smallest to largest, keeping track of which measurements came from the potentially impacted 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 are tied, average the ranks separately for each of those measurement values.

   If a few nondetects occur (say, <10%), and if all such values are less than the smallest detected measurement in the pooled dataset, handle the nondetects 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 nondetects (the same procedure as for tied detected measurements).

   If between 10% and 40% of the pooled dataset are nondetects, and all are less than the smallest detected measurement, use the WRS test procedure in Box 4-12, even if n and m are less than 20. Note: The procedure in Box 4-12 is for the case where m and n are both 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, the test should not be used to decide whether the chemical is a COPC until additional information is obtained by collecting more samples and analyzing by a method with a lower RL.
6. Calculate the sum of the ranks of the potentially impacted 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 potentially impacted concentration distribution is shifted to the right of the background concentration distribution, i.e., that the chemical is a COPC.
8. If the WRS test declares that the chemical is *not* a COPC, this conclusion may indicate (1) the chemical is indeed not a COPC, or (2) the assumptions that underlie the WRS test are not valid for the potentially impacted 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 potentially impacted and background concentration distributions.

BOX 4-10. (cont'd)

The possibility that the causes in items (2) or (3) may have resulted in the WRS test declaring that the chemical is not a COPC should be evaluated. Review the DQO planning process records to make sure the actual number of samples (*n* and *m*) corresponds to the number of samples determined necessary to detect a difference (shift) between the potentially impacted and background medians of magnitude Δ/s. 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 potentially impacted and background measurements needed by using a software package such as DataQUEST (U.S. EPA, 1997). Collect additional samples if needed.

BOX 4-11. Example of the WRS test when the number of potentially impacted and background measurements is small ($n<20$ and $m<20$)

Suppose it is necessary to determine if a chemical in surface sediment in Region A at a potentially impacted 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 potentially impacted concentration distribution is greater than the median of the background distribution by the amount $\Delta/s = 1.5$, 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, we find that $n = m = 18$ measurements are needed for the WRS test.
4. Then, 18 samples from both the potentially impacted and the background areas 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 each sample was analyzed for the chemical of interest. Suppose the measurements were:

Background Data:	22,	32,	9,	12,	3,	7,	11,	2,	9,	11,	13,	16,	20,	25,	<1,	<1,	17,	21
Potentially Impacted 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 potentially impacted data are determined (the potentially impacted and background data and ranks are denoted by P and B, respectively):

	B	B	P	B	P	B	P	P	B	P	P	P	B	B	P	P	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	P	P	B	B	P	P	B	B	B	P	B	P	B	P	P	P
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 potentially impacted 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 $w_{0.05} = 110$.

   Computing $nm - w_{\alpha} = 18 \times 18 - 110 = 214$. Therefore, $W < nm - w_{\alpha}$, i.e., $179.5 < 214$. The WRS has indicated the evidence 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 potentially impacted 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 potentially impacted and background samples.

BOX 4-12. Procedure for conducting the WRS test when the number of potentially impacted 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 potentially impacted concentrations tend to be larger than the background concentrations, i.e., the probability that the chemical will be incorrectly identified as 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 median potentially impacted 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 potentially impacted 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 potentially impacted and background measurements needed when it is desired to have n equal m. If equal n and m values are not desired, use Equation 6.3 in U.S. EPA (1994) and increase that value by 20% to guard against missing or unusable measurements.
4. Collect the n and m samples and analyze for the chemical of interest; some measurements may be nondetects. 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 and rank the pooled set of $n + m$ potentially impacted and background measurements from smallest to largest, keeping track of which measurements came from the potentially impacted 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 are tied, average the ranks separately for each of those measurement values.

   If <40% of the pooled dataset are nondetects, and if all such values are less than the smallest detected measurement in the pooled dataset, handle those nondetects 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 nondetects should not exceed 40% of the total number of measurements.

   If more than 40% of the pooled data are nondetects, 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 potentially impacted 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 $100(1-\alpha)$ 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 $R > w_{1-\alpha}$.

Role of the Data Quality Objectives Process

- The stakeholders and regulators used the DQO planning process to agree:
- On the methods that should be used to collect, handle, prepare, and measure the sediment samples.
- That it is unlikely that more than 40% of the measurements will be nondetects.
- On the values of the design parameters used to determine the necessary number of potentially impacted and background sediment measurements (see the subsection

BOX 4-13. Example of the WRS test when the number of potentially impacted 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 potentially impacted concentration distribution is $\Delta/s = 1.8$ units greater than the median of the background distribution, then 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 potentially impacted and background measurements, respectively.
4. Then 20 samples from both the potentially impacted 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

   Potentially Impacted 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 potentially impacted data are determined (the potentially impacted and background data and ranks are denoted by *P* and *B,* respectively).

	B	B	B	B	P	P	P	B	B	B	B	B	P	B	B	P	P	B	B	B	P
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

	P	B	B	B	B	P	B	P	P	P	P	B	P	P	P	P	P	P	P
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 potentially impacted data is $R = 4 + 4 + 4 + 13 + 16 + \ldots + 39 + 40 = 507.5$.
7. Also,

$$\begin{aligned} w_{0.99} &= n(n+1)/4 + z_{0.99}\,[n(n+1)(2n+1)/24]^{1/2} \\ &= 20(21)/4 + 2.33[20(21)(41)/24]^{1/2} \\ &= 167.4 \end{aligned}$$

   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.

entitled “Guidance on Implementing the Gehan Test” that follows).

The Gehan test (Palachek et al., 1994) is appropriate for this scenario because the potentially impacted and background datasets were likely to contain multiple nondetects with different RLs. This test is appropriate for the basin-wide impact hypothesis. In cases where the exact nature of the spatial extent of sediment contamination is not known, the Quantile or Slippage tests should be used in conjunction with the Gehan test to assess possible localized impact (i.e., hotspots). The assumptions behind the Gehan test are summarized in Table 4-1.

Advantages and Disadvantages

- The Gehan test can be used when the background or potentially impacted datasets contain multiple nondetects with different RLs.
- If the RLs are different for the potentially impacted and background datasets, then the test results may be an indication of this analytical difference rather than an indication that the chemical is a COPC.

Table 4-1 summarizes the advantages and disadvantages of the Gehan test.

Guidance on Implementing the Gehan Test

The number of potentially impacted and background measurements required to conduct the Gehan test may be approximated using the method described for the WRS test in Section 4.2.3.

The Gehan test procedure for n ≥10 and m ≥10 is presented in Box 4-14. An example of the test is presented in Box 4-15. If n <10 or m <10, the procedure in Box 4-16 may be used to conduct the Gehan test.

BOX 4-14. Procedure for conducting the Gehan test when n≥10 and m≥10

1. Specify the tolerable probability, α, that the Gehan test will incorrectly declare that the potentially impacted median is larger than the background median, i.e., the probability that the chemical will be incorrectly identified as a COPC.
2. Specify the value of Δ/s and the power, where Δ/s is the magnitude of the difference in median potentially impacted 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 potentially impacted concentration distributions for the chemical of interest.
3. Use the specified values of α, Δ/s, and the power in Table 4-6 to determine the number of potentially impacted and background measurements needed when it is desired to have *n* equal to *m*. If it is not desired to have *n* equal to *m*, use Equation 6.3 in U.S. EPA (1994) and increase that value by 20% to guard against missing or unusable measurements.
4. Collect the *n* and *m* samples and analyze for the chemical of interest; some measurements are likely to be nondetects. 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* potentially impacted measurements, including the nondetects, from smallest to largest, where the total number of combined samples is N = m + n. The below-detection symbol (<) is ignored when listing the *N* data from smallest to largest.
6. Determine the *N* ranks, $R_1, R_2, \ldots, R_N$, for the *N* ordered data values using the method described in the example given in Box 4-15.
7. Compute the *N* scores, $a(R_1), a(R_2),\ldots,a(R_N)$ 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 potentially impacted 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 declares that the chemical is a COPC if $G \geq Z_{1-\alpha}$, where $Z_{1-\alpha}$ is the 100(1 – α)th percentile of the standard normal distribution, which is obtained from Table C-1. Otherwise, the test declares that the evidence is not strong enough to conclude that the chemical is a COPC.

BOX 4-15. 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 potentially impacted concentration distribution is greater than the median background distribution by the amount Δ/s = 2.0, 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, we find that n = m = 10 measurements are needed to conduct the Gehan test.

4. The 10 samples from the potentially impacted 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 each sample was analyzed for the chemical of interest. Suppose the measurements are:

Background:	1	<4	5	7	<12	15	18	<21	<25	27
Potentially Impacted:	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 and the bullet list below as you go 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

Data	h_i	Index_i	d_i	e_i	R_i	$a(R_i)$
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* potentially impacted measurements, including the nondetects, from smallest to largest, as illustrated in column 1 of Table 1. Ignore the below-detection 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 the background dataset
  = 1 if the *i*th measurement is from the potentially impacted dataset

- 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 nondetect 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, i.e., 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 nondetect value, i.e., 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, i.e., whenever Index = 0.
  - For each successive row, increase *e* by 1 whenever the datum in column 1 in that row is a nondetect value, i.e., when Index = 1.

BOX 4-15. (cont'd)

- Let T denote the total number of nondetects in the pooled background and potentially impacted datasets. For the previous data there are T = 6 nondetects. 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, i.e., if $h_i = 0$ for the *i*th row.
  - $R_i = (T + 1 + d_I)/2$ if the datum in column 1 of *i*th row is a nondetect value, i.e., if $h_i = 1$ for the *i*th row.
- Compute the n = 20 scores, $a(R_1), a(R_2), \ldots, a(R_{20})$, using the formula

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

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

8. Compute the Gehan statistic, G:

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

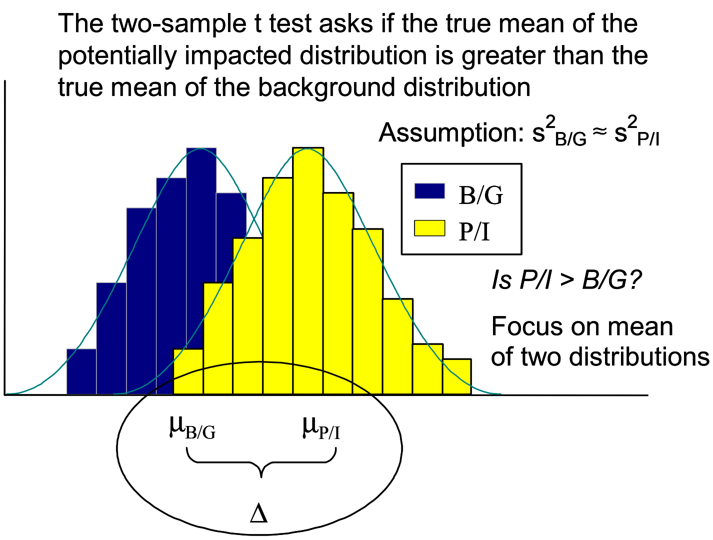
$$= 40 / [(100*1942) / (20*19)]^{1/2}$$
$$= 40 / 22.606$$
$$= 1.77$$

9. In Step 1 above we specified that $\alpha = 0.05$. When $\alpha = 0.05$, Table C-1 yields $Z_{1-\alpha} = Z_{0.95} = 1.645$. As G > 1.645, i.e., 1.77 > 1.645, the Gehan test declares that the chemical is a COPC.

BOX 4-16. Procedure for conducting the Gehan test when n<10 and m<10

1. Generate on a computer all possible orderings of the combined *n* potentially impacted and background measurements. Denote the *number* of possible orderings by *M.*
2. Compute the G statistic (Equation 1 in Box 4-14) for each of these orderings to generate an empirical distribution (histogram) of the 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 *k*th 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 potentially impacted data,* equals or exceeds the 100(1−α)th percentile obtained above, conclude that the chemical is a COPC.

4.2.5 Two-Sample t Test



Sediment Contamination Scenario

The sediment contamination scenario developed by the stakeholders and regulators during Steps 1 and 2 of the DQO process is the same as for the WRS and Gehan tests: If contamination has occurred, it is probably more-or-less uniformly distributed throughout the potentially impacted area of the sediment basin.

Role of the Data Quality Objectives Process

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 sediment samples.
- ❑ That it is likely that very few nondetects will be reported by the laboratory.
- ❑ On the values of the design parameters used to determine the necessary number of potentially impacted and background sediment measurements (see the subsection entitled “Guidance on Implementing the Two-Sample t Test”).
- ❑ That normally distributed mean concentrations are expected (for large datasets, n and $m > 30$, 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 if the datasets are small (n and $m < 30$), tests for normality indicate the measurements are not normally distributed, and stakeholders do not agree that normal transformation is appropriate, then non-parametric methods such as the WRS test or the Quantile test should be used in place of the two-sample t test because the means cannot be assumed to be normally distributed.
- ❑ That measurements from the potentially impacted area 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) indicates 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 is appropriate for this scenario because the assumptions of normality of mean, equal variances for background and potentially impacted data, and the absence of nondetects were expected to be valid. However, once the measurements are obtained, these assumptions should be evaluated by observation and statistical testing. If the site and background variances appear to be approximately equal, but normal distributions of the mean concentrations are not expected, 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 the two-sample t test are summarized in Table 4-1.

Advantages and Disadvantages

- To implement the two-sample t test, all nondetects must be replaced with a surrogate value, such as one-half the DL. However, if a large number of nondetects exist (e.g., more than 15% of measurements) the test results cannot be viewed as reliable.

- If a large number of nondetects exist, other tests should be used to address the basin-wide impact hypothesis. The WRS test should be used if there is only one RL and no more than 40% of measurements are nondetects. The Gehan test should be used if multiple RLs are present. The two-sample test of proportions (Section 4.2.7) is appropriate if more than 40% of the data are nondetects.

- If the exact nature of the spatial extent of the sediment contamination is not known, the two-sample t test should be accompanied by the Quantile or Slippage tests. These tests address the localized (hotspot) impact hypothesis.

- If the datasets are small (*n* and *m* less than 30), and the tests for normality indicate the measurements are not normally distributed, nonparametric tests such as the WRS and Gehan tests should be used in place of the two-sample t test.

- If the mean concentrations of potentially impacted and background measurements are normally distributed, but their variances are different, then the Satterthwaite two-sample t test should be used instead of the two-sample t test. The difference between variances can be assessed by statistical tests, such as the F test described in Conover (1998, p. 300) and U.S. EPA (2000). (Note, however, that real data are never completely normal, and the power of these common parametric tests is reduced when the data distributions depart significantly from normality.)

- The two-sample t test is the most common statistical test for assessing the differences between mean concentrations of two datasets. Most statistical software packages can compute the two-sample t test.

Table 4-1 summarizes the advantages and disadvantages of the two-sample t test.

Guidance on Implementing the Two-Sample t Test

The number of potentially impacted (*n*) and background (*m*) measurements required for the two-sample t test should be approximated using the procedure outlined in Box 4-17. An example of the evaluation of Equation 1 in Box 4-17 is given in Box 4-18. After *n* and *m* have been determined, the samples collected, and measurements reported by the laboratory, summary statistics should be computed for both the potentially impacted and background datasets. In particular, the computed sample variance of the potentially impacted 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 whether the two sample variances are equal is provided in Conover (1998) and U.S. EPA (2000, p. 4-33). This procedure is commonly available 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). Outliers should be identified, investigated to determine if they represent mistakes or errors, and, if necessary, discarded. If the potentially impacted and background datasets are small (n and $m < 30$) they should then 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 test procedure is presented in Box 4-19; an example of the procedure is presented in Box 4-20.

BOX 4-17. Procedure for calculating the number of potentially impacted and background measurements required to conduct the two-sample t test

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

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

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

α = the tolerable probability that the two-sample t test will incorrectly declare that 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 required power (probability) that the two-sample t test will declare that the chemical is a COPC when that is indeed the case (β is usually specified to be ≥ 0.80)

$\mu_s - \mu_b$ = true potentially impacted mean (μ_s) minus the true background mean (μ_b); i.e., the difference in the *true* (unknown) means of the potentially impacted and background areas 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-1 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 $Z_{0.80} = 0.84$)

Stakeholders and regulators should determine the appropriate values of the parameters in Equation 1 in this box during the DQO planning process.

BOX 4-18. Example of the procedure for calculating the number of potentially impacted and background measurements required to conduct the two-sample t test

Suppose the values of the parameters in Equation 1 in Box 4-17 were specified by the stakeholders and regulators as follows:

s^2 = 7.5
α = 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 potentially impacted 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 potentially impacted 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.

BOX 4-19. Procedure for conducting the two-sample t test

1. Stakeholders and regulators used the DQO process to select values of s^2, α, $1 - \beta$, and $\mu_s - \mu_b$, and the procedure in Box 4-17 (as illustrated in Box 4-18) to determine the number of potentially impacted (*n*) and background (*m*) measurements.
2. Collect the samples and obtain the *n* and *m* potentially impacted and background measurements.
3. Suppose
   - the *n* potentially impacted 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* potentially impacted measurements
$\bar{y}$ = the arithmetic mean of the *m* background measurements
s_x^2 = the sample variance of the *n* potentially impacted measurements (the formula for computing this variable is given in Box 2-9)
s_y^2 = the sample variance of the *m* background measurements (see Box 2-9)

5. The two-sample t test declares:
   - that the chemical is a COPC if $T \geq t_{1-\alpha, n+m-2}$
   - that 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. 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 that the value of α was specified in Step 1, as part of the process for determining the required number of potentially impacted and background measurements.

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

The possibility that the test did not declare the chemical to be a COPC due to items (2) or (3) should be evaluated.

- First, review the DQO planning process records to make sure the number of samples (*n* and *m*) agrees with what was determined necessary to detect the difference between potentially impacted and background means that was considered important.
- Second, review the computations that were 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 potentially impacted 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 potentially impacted and background measurements needed to detect the critical (important) shift in the potentially impacted mean by calculating the variance of the potentially impacted and background measurements (s_x^2 and s_y^2, respectively), and use the larger of these two estimated variances in Equation 1 of Box 4-17. If this new value, denoted by n', is larger than either the number of potentially impacted or background measurements obtained and used in the t test, collect additional samples so n' potentially impacted and n' background measurements are available. Then redo the t test.

BOX 4-20. Example of the two-sample t test

1. Suppose the values of the parameters in Equation 1 of Box 4-17 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 4-18 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:

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

   No potential outliers are apparent in either dataset. Therefore, tests for outliers do not appear necessary. Each dataset should tested 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:

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

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

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

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

$$= 3.84$$

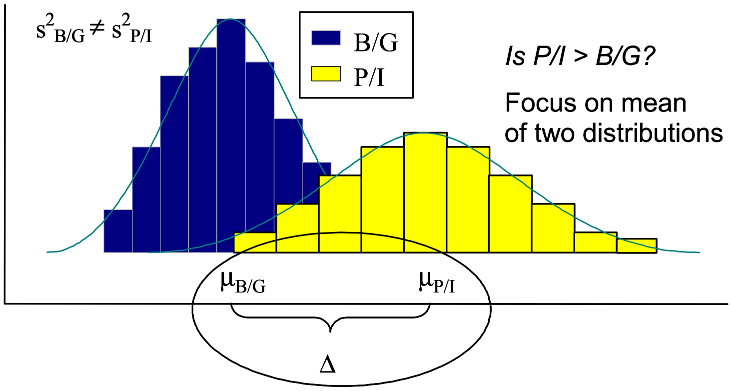
5. The value of $t_{1-\alpha,\ n+m-2}$, i.e., of $t_{0.975,\ 18}$ is found from Table C-9 to be 2.101. Therefore, as T >2.101, i.e., 3.84 >2.101, the two-sample t test declares that the chemical is a COPC.

4.2.6 Satterthwaite Two-Sample t Test

Sediment Contamination Scenario

The sediment contamination scenario for the Satterthwaite two-sample t test is the same as for the two-sample t test (Section 4.2.5): If contamination exists, it is probably uniformly distributed throughout the impacted portion of the sediment basin. However, the Satterthwaite two-sample t test was selected (instead of the two-sample t test) because it can be used when measurements from the potentially

The Satterthwaite two-sample t test asks if the true mean of the potentially impacted distribution is greater than the true mean of the background distribution



impacted area are not expected to have approximately the same variance as the background measurements.

Role of the Data Quality Objectives Process

The stakeholders and regulators concluded that the potentially impacted and background datasets meet all the conditions for use of the two-sample t test (Section 4.2.5) with one exception: based on prior data and statistical tests, or expert knowledge, the potentially impacted and background datasets are expected to have different variances.

Advantages and Disadvantages

- The Satterthwaite two-sample t test can be applied to datasets with unequal variances. A statistical procedure for testing whether two sample variances are equal is provided in Conover (1998) and U.S. EPA (2000, p. 4-33).

- Like the two-sample t test, the Satterthwaite two-sample t test assumes that the mean concentrations are normally distributed. This assumption is valid for large datasets (n and $m > 30$); however, for small datasets, tests for normality of the measurements should be conducted to assure the validity of this assumption.

- To implement the Satterthwaite two-sample t test, all nondetects must be replaced with a surrogate value, such as one-half the DL. However, if a large number of nondetects exist (e.g., more than 15% of measurements) the test results cannot be considered reliable.

- If a large number of nondetects exist, other tests should be used to address the basin-wide impact hypothesis. For example, the Gehan test should be used if multiple RLs are present. The two-sample test of proportions (Section 4.2.7) is appropriate if more than 40% of the data are nondetects.

- If the exact nature of the spatial extent of the sediment contamination is unknown, the Satterthwaite two-sample t test should be accompanied by the Quantile or Slippage tests. These tests address the localized (i.e., hotspot) impact hypothesis.

- If the datasets are small (n and $m < 30$), and the tests for normality indicate the measurements are not normally distributed, a nonparametric test such as the Gehan test should be used in place of the two-sample t test.

Table 4-1 summarizes the advantages and disadvantages of the Satterthwaite two-sample t test.

Guidance on Implementing the Satterthwaite Two-Sample t Test

Ideally, the same number of measurements should be obtained for both the potentially impacted and background areas. The required number of measurements (n) should be approximated by the same procedure used for the two-sample t test (see Box 4-17). The larger of the potentially impacted or background variance should be used for s^2 in Equation 1 of Box 4-17.

When the n measurements have been obtained, the normality of mean concentrations can be assumed to be reasonable if $n > 20$ to 30. For smaller datasets, the assumption of normality should be evaluated (see Section 2.2.3). The Satterthwaite two-sample t test procedure is described in Box 4-21. An example of this procedure is presented in Box 4-22.

BOX 4-21. 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 potentially impacted and background variances (s^2). Then use the procedure in Box 4-17, as illustrated in Box 4-18 to determine the number of measurements for both the potentially impacted and the background areas.
2. Collect the samples and obtain the *n* potentially impacted and *n* background measurements.
3. Suppose
   - the *n* potentially impacted 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* potentially impacted measurements
$\bar{y}$ = the arithmetic mean of the *n* background measurements
s_x^2 = the sample variance of the *n* potentially impacted measurements (the formula for computing this variable is given in Box 2-9)
s_y^2 = the sample variance of the *n* background measurements (see Box 2-9).

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 t-test can be computed when the number of potentially impacted and background measurements are not equal. In that case, n_x and n_y would replace *n* in these equations, 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 exists 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 entering $1-\alpha$ and f. If *f* is not an integer, linear interpolation may be used to determine $t_{1-a,\,f}$ from Table C-9.

If the two-sample t test does not declare that the chemical is a COPC, it may indicate (1) the chemical is indeed not a COPC, or (2) the assumptions that underlie the t test are not valid for the potentially impacted and background measurements, or (3) an insufficient number of measurements were obtained for the Satterthwaite t test to detect the difference in potentially impacted and background concentration distributions that actually exists.

The possibility that the test did not declare the chemical to be a COPC due to items (2) or (3) should be evaluated.

1. First, review the DQO planning process records to make sure the number of samples agrees with what was determined necessary to detect the difference between potentially impacted and background means that was considered important.
2. Second, review the computations that were 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.
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 potentially impacted and background measurements needed by calculating the variance of the potentially impacted and background measurements (s_x^2 and s_y^2, respectively), and using the larger of these two estimated variances in Equation 1 of Box 4-17. If this new value, denoted by *n′*, is larger than the number of potentially impacted and background measurements obtained and used in the t test, then collect additional samples to obtain *n′* potentially impacted and *n′* background measurements. Then redo the Satterthwaite t test.

BOX 4-22. Example of the Satterthwaite two-sample t test

1. Suppose a preliminary study was conducted to estimate the variance of the background and potentially impacted measurements, and the variance of the potentially impacted data was found to be significantly larger than the background data variance. Suppose the larger of the two estimated variances was 15. Therefore, 15 was selected as the s^2 value to use in Equation 1 in Box 4-17. (If very few potentially impacted 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 in Equation 1 in Box 4-17 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, the reader may verify that Equation 1 in Box 4-17 gives the value n = 14.1, rounded down to n = 14.
2. Therefore, n = 14 potentially impacted and n = 14 background samples were collected and measured using the methods specified during the DQO process and as documented in the Quality Assurance Project Plan (QAPP).
3. Suppose the measurements are as follows:

Potentially Impacted Data (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 Data (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 necessary. Each dataset should be evaluated graphically (Section 2.2.3) and with a formal statistical test (Appendix B.1) to verify that the normality assumption is reasonable for each dataset. The reader may verify that the normality assumption appears to be reasonable for both datasets.
5. Next, the following calculations are conducted:

$$\bar{x} = 5.41$$
$$\bar{y} = 6.84$$
$$s_x^2 = 18.708$$
$$s_y^2 = 14.316$$

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

$$= \frac{5.41 - 6.84}{(18.708/14 + 14.316/14)^{1/2}}$$

$$= \frac{-1.43}{1.536}$$

$$= -0.931$$

$$f = \frac{(18.708/14 + 14.316/14)^2}{(18.709/14)^2/13 + (14.316/14)^2/13}$$

$$= \frac{5.564}{0.1374 + 0.08043}$$

$$= 25.54 \text{ degrees of freedom}$$

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 $T_s < 1.3155$, i.e., as $-0.931 < 1.3155$, the Satterthwaite two-sample t test does *not* declare that the chemical is a COPC. Indeed, the estimated mean of the potentially impacted measurements is less than the estimated mean of the background measurements.

BOX 4-22. (cont'd)

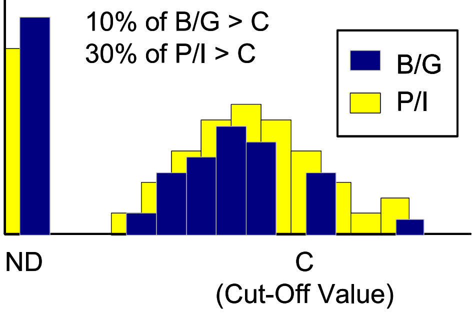
7. As the test did not declare that the chemical is a COPC, the DQO process records and QAPP should be reviewed to confirm that all requirements for collecting the type, quantity, and quality of measurements were correctly followed. Next, evaluate whether the number of measurements used in the test ($n = 14$) 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, compute Equation 1 in Box 4-17 using the larger of the estimated potentially impacted and background variances, i.e., using $s^2 = 18.7$, as computed in Step 5. From Equation 1, $n = 16.1$, 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 each area (the background and potentially impacted areas). Simple random sampling should be used to determine the new sampling locations in the field. Also, the collection and measurement protocols for obtaining the new data should be exactly the same as for the original data (specified in the QAPP). Then the Satterthwaite two-sample t test should be recomputed using the new background and potentially impacted datasets, each of which consists of 14 old and 3 new measurements. Before conducting the Satterthwaite t test, graphic and statistical tests for normality should be conducted on the new datasets ($n = 17$) to verify that the normality assumption is still reasonable.

4.2.7 Two-Sample Test of Proportions

Sediment Contamination Scenario

Consider the case where the potentially impacted portion of the sediment basin is expected to contain hotspots with no distinct (gradient-like) contamination pattern. In this case, the localized impact hypothesis is appropriate. In addition, a large number of measurements from both the potentially impacted and background areas are expected to be nondetects. Due to the large number of nondetects, no specific assumptions can be made about the statistical distribution of the sampling data.

The two-sample test of proportions asks if a larger proportion of the potentially impacted data than of the background data exceeds a concentration *C*.



Is P/I > B/G?

Focus on proportions of potentially impacted and background data greater than *C*.

Role of the Data Quality Objectives Process

Given the anticipated large number of nondetects, it is difficult to conduct a valid statistical test on differences between mean or median concentrations of the potentially impacted and background datasets. Therefore, the stakeholders and regulators decided to test the potentially impacted and background data to identify the dataset that has the largest proportion of concentrations greater that a specified cut-off concentration, *C*. After determining the magnitude of the cut-off concentration limit, *C*, the DQO planning team agreed:

- ❑ That the null and alternative hypotheses to be tested are

$$H_o: \quad P_p \le P_b$$
$$H_a: \quad P_p > P_b$$

  where Pp and Pb are the true proportions of the potentially impacted and background measurements, respectively, that exceed C (the predefined cut-off concentration level).

- ❑ On the methods that will be used to collect, handle, prepare, and measure the sediment samples.

- ❑ On the values of the design parameters used to determine the necessary number of potentially impacted and background sediment measurements (see the subsection entitled "Guidance on Implementing the Two-sample Test of Proportions").

The assumptions behind the two-sample test of proportions are summarized in Table 4-1.

Advantages and Disadvantages

- ❑ The two-sample test of proportions may be conducted regardless of the underlying distribution of measurements, i.e., the test is nonparametric.
- ❑ The two-sample test of proportions is most suitable in cases where the potentially impacted and background datasets contain a large number of nondetects.
- ❑ The two-sample test of proportions does not evaluate whether the potentially impacted mean (or median) exceeds the background mean (or median). The test simply focuses on the proportions of measurements that exceed a given cut-off value.

Table 4-1 summarizes the advantages and disadvantages of the two-sample test of proportions.

Guidance on Implementing the Two-Sample Test of Proportions

The number of potentially impacted (*n*) and background (*m*) measurements required to conduct the two-sample test of proportions should be approximated by the procedure outlined in Box 4-23. An example of the procedure is given in Box 4-24.

After the data evaluation process (Section 2.1.7) has been completed (i.e., once it has been determined that the data contain no errors, and have been collected, handled, and measured according to the DQO specifications), and the assumptions required for the two-sample test of proportions have been shown to be reasonable, then the

BOX 4-23. Procedure for calculating the number of potentially impacted and background measurements required for the two-sample test of proportions

The formula for calculating the number of potentially impacted (*n*) and background (*m*) measurements required for the two-sample test of proportions is as follows (from U.S. EPA, 2000):

$$n = m = \frac{2(Z_{1-\alpha} + Z_{1-\beta})^2 \, \overline{P}\,(1 - \overline{P})}{D^2} \qquad (1)$$

where: $\overline{P}$ = $(P_p + P_b) / 2$

P_p = the proportion of the true potentially impacted 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, i.e., the probability that the chemical will be incorrectly identified as 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 potentially impacted and background distributions 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 a 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-5 (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 $1 - \beta = 0.80$, we find from Table C-5 that $Z_{0.80} = 0.84$)

Stakeholders and regulators should determine the appropriate values of the parameters in Equation 1 during the DQO planning process.

BOX 4-24. Example of the procedure for calculating the number of potentially impacted and background measurements required for the two-sample test of proportions

Suppose the values of the parameters in Equation (1) of Box 4-23 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_p and P_b are true values and are therefore unknown, estimates of these true proportions must be supplied by a preliminary sampling study at the background and potentially impacted areas. This study must be conducted according to the same sampling and analysis protocol that will be used in the main study. Suppose a preliminary study based on 20 background samples and 20 potentially impacted samples yields P_s and P_b estimates of 0.30 and 0.15, respectively. Therefore, $\overline{P} = (0.30 + 0.15) / 2 = 0.225$. Therefore, equation (1) in Box 4-23 is:

$$n = m = 2(1.96 + 0.84)^2\ 0.225\ (1 - 0.225) / 0.20^2 = 68.35$$

68.35 is rounded up to 69. Therefore, 69 background area samples and 69 potentially impacted area samples are needed. Because the 20 potentially impacted and 20 background samples have already been collected, handled, and measured by the methods required for the full study, only 49 new potentially impacted and 49 new background measurements need to be collected.

two-sample test of proportions may be conducted. The test procedure is presented in Box 4-25. An example is provided in Box 4-26.

4.3 Determination of Background Ranges for COPCs

After sediment COPCs have been identified by the tests described in Section 4.2, it may be necessary to estimate the upper limits of their background ranges. This is usually required when (a) delineating impacted sediments with respect to background or (b) evaluating incremental (residual) risks associated with above-background concentrations. For these purposes, the available basin-wide data—both the background dataset and the impacted area dataset—should be combined for exploratory data analysis as described in Section 2.2. The univariate, post plot, and probability plot analyses presented in Section 2.2.4 provide methods for estimating the upper limits of the background range for each COPC. If the basin-wide dataset has already been subjected to exploratory data analysis, there is no need to repeat the process. The resulting background ranges should be used to determine the upper limit of the background range.

As noted in Section 2.2.3, the upper limit of the background range can be identified as the higher of the maximum concentration value or the upper 95th percentile of the concentration values within the background subpopulation. Note, however, that estimating the upper limit of the background range based solely on the background dataset is not recommended. The primary reason for this recommendation is the small size of most background datasets: a small dataset may not yield a reliable estimate of the upper limit of the background range for a particular COPC. This limitation can be corrected by evaluating the basin-wide dataset as described above. In this approach, the background data are combined with data from the potentially impacted area. This allows data that represent background conditions within the potentially impacted area to supplement the background dataset. The larger dataset provides a more reliable basis for estimating the upper limits of the background ranges.

BOX 4-25. Procedure for conducting the two-sample test of proportions (from U.S. EPA, 2000)

1. Stakeholders and regulators used 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 analysis study at the background and potentially impacted areas to obtain estimates of the true proportions P_p and P_b of the potentially impacted and background populations that exceed *C*. Then use the procedure in Box 4-23 to determine *n* and *m*, the necessary number of potentially impacted and background measurements.
3. Collect, handle, and analyze the *n* and *m* samples, as specified in the sampling and analysis plan and the QAPP.
4. Suppose

   - *n* potentially impacted 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 we recommend that n = m. However, the following formulas are for the more general case where the number of potentially impacted measurements, *n*, and the number of background measurements, *m*, are not equal.
5. Let k_p and k_b be the number of potentially impacted and background measurements, respectively, that exceed *C.*
6. Compute $p_p = k_p/n$, which is the estimated proportion of the true distribution of potentially impacted measurements that exceed *C.*
7. Compute $p_b = k_b/m$, which is the estimated proportion of the true distribution of background measurements that exceed *C.*
8. Compute

   $$p = (k_s + k_b) / (n + m)$$
9. Compute k_p, k_b, $n(1 - p_p)$, and $m(1 - p_b)$. If all of these quantities are greater than or equal to 5, continue with step 10. If not, seek assistance from a statistician, because the computations for the test are more complicated when these quantities are less than 5.
10. Compute the test statistic:

    $$Z_p = (p_p - p_b) / [p(1 - p)(1/n + 1/m)]^{1/2}$$
11. Use Table C-1 to find $Z_{1-\alpha}$
12. If $Z_p \geq Z_{1-\alpha}$ the test has declared that $P_p > P_b$, i.e., that the true proportion of the potentially impacted measurements greater than the concentration value *C* is greater than the true proportion of the background measurements greater than *C.*

    If $Z_p < Z_{1-\alpha}$, then the data do not provide enough evidence to conclude that $P_p > P_b$. In that case, go to step 13.
13. Suppose the test declares that the data do not provide enough evidence to conclude that $P_p > P_b$. This conclusion may indicate (1) the chemical is not a COPC, or (2) the assumptions that underlie the test are not valid for the potentially impacted and background measurements, or (3) an insufficient number of measurements (*n* and *m*) were obtained for the test to detect the difference *D* that actually exists. The possibility that the test did not declare that $P_p > P_b$ due to items (2) or (3) should be evaluated. Review the DQO planning process records to make sure the number of measurements (*n* and *m*) agrees with what was determined necessary to detect the specified difference *D.* Use Equation 1 in Box 4-23 to recalculate the number of measurements required for the test. Those computations should be done using the estimates p_p and p_b in place of P_p and P_b, respectively. If the new *n* value is greater than what was used to compute the test statistic, collect and analyze the additional samples necessary to achieve the new *n* value and redo the test.

BOX 4-26. 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 parts per billion (ppb) for the chemical of interest.
2. Also suppose that a preliminary study was conducted at the potentially impacted and background area to obtain estimates of the true proportions P_p and P_b. Suppose these estimates were 0.30 and 0.15, respectively. Then, as illustrated in Box 4-24, 69 measurements are needed from the potentially impacted area and 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,* i.e., greater than 1 ppb. Furthermore, suppose that $k_p = 24$ of the potentially impacted measurements were greater than *C*. Therefore,

   $p_b = 19/69 = 0.275$
   $p_p = 24/69 = 0.347$
   $p = (k_p + k_b) / (n + m) = (19 + 24) / (69 + 69) = 0.3116$
4. Also,

   $mp_b = 69(0.275) = 19$
   $np_p = 69(0.347) = 24$
   $m(1 - p_b) = 69(1 - 0.275) = 50$
   $n(1 - p_p) = 69(1 - 0.347) = 45$

   All the above numbers are greater than 5. Therefore, we continue on with the test as described in Box 4-25.
5. The test statistic is computed as follows:

$$\begin{aligned} Z_p &= (p_p - p_b) / [p(1 - p)(1/n + 1/m)]^{1/2} \\ &= (0.347 - 0.275) / [0.3116(1 - 0.3116)(1/69 + 1/69)]^{1/2} \\ &= 0.072 / [0.2145*(0.014493 + 0.014493)]^{1/2} \\ &= 0.072 / 0.0789 \\ &= 0.913 \end{aligned}$$

6. From Table C-1 we find that $Z_{1-\alpha} = Z_{0.975} = 1.96$
7. As $Z_p < 1.96$, i.e., $0.913 < 1.96$, the data do not provide sufficient information for the test to reject H_o and declare that the chemical is a COPC.
8. Equation 1 in Box 4-23 is then re-evaluated to determine whether the test did not reach a statistically significant conclusion because the datasets have fewer measurements than are required to achieve the power of $1 - \beta = 0.80$ when D = 0.20 (at the $1 - \alpha = 0.975$ confidence level). We obtain:

$$n = m = 2(1.96 + 0.84)^2\ 0.3116(1 - 0.3116) / 0.2^2 = 84.09$$

   This indicates that 85 potentially impacted measurements (*n*) and 85 background measurements (*m*) are needed; therefore, the number of measurements (n = m = 69) is insufficient.

In conclusion, based on the data, the true difference *D* is estimated to be 0.347 – 0.275 = 0.072. However, the two-sample test of proportions was not able to declare that this difference is large enough to conclude that $P_p > P_b$.

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.

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.

Science Applications International Corporation (SAIC). 1998. *Preliminary Remediation Goals (PRGs) for NCBC Allen Harbor Landfill (Site 09)*. Final Report. Narragansett, RI. March

United States Environmental Protection Agency (U.S. EPA). 1994. *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). 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 STUDY

For the purposes of this document, the hypothetical case study area is called "Big Spruce Bay" (the Bay), and is assumed to be a significant port area on the west coast of the United States.

The Bay case study area is a four-lobed maritime harbor in a tidal estuarine setting, with a primary navigation channel and several upland freshwater tributaries. A sediment basin map is presented as Figure 5-1. (Although not included here due to space considerations, note that concentration data should be posted on a sediment basin map to evaluate the spatial characteristics of the datasets [as recommended in Section 2.2.4].)

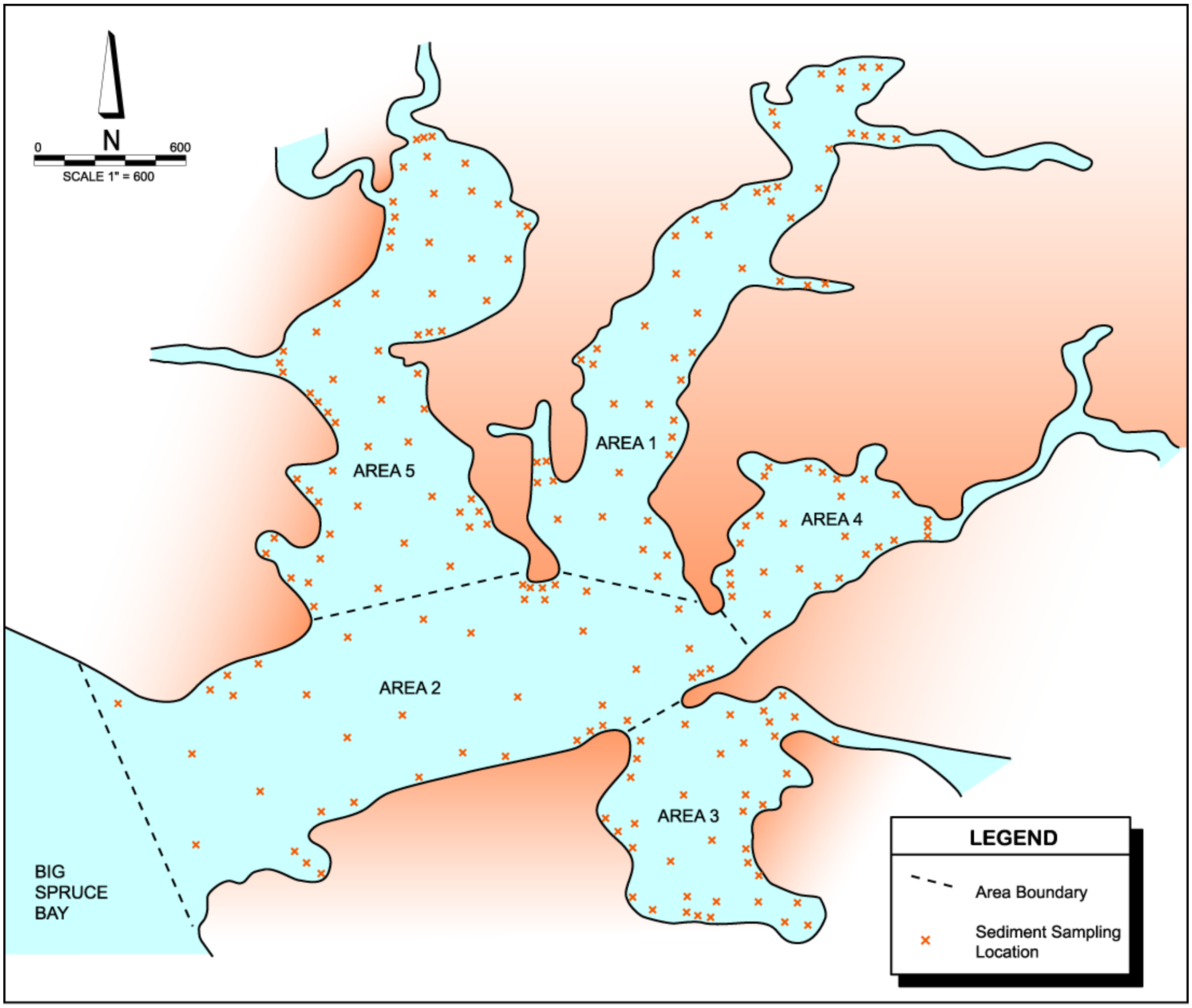


FIGURE 5-1. Case study sampling locations, Big Spruce Bay

The Bay is used for both commercial (large and small) and military (Navy) activities, including ship anchoring, mooring at piers and wharves, ship maintenance, and shipbuilding. The upland watershed area is used for residential, commercial, industrial, and agricultural purposes. Sediments in the basin are derived primarily from basaltic rocks and soils that occur in the surrounding mountains. Areas 1 and 5, the most heavily used lobes of the Bay, were first developed approximately 50 years ago as primary freight staging areas for overseas trade. The Navy subsequently became the primary operator within Areas 1 and 5. Area 4 is predominantly a commercial port, and Area 3 receives minimal commercial/industrial activity due to its poor navigability. Area 2 is the primary navigation channel.

A Remedial Investigation (RI) was performed to characterize the nature and extent of suspected contamination within the various lobes of the Bay. Objectives of the RI were as follows:

- Assess the nature and extent of potential metals contamination resulting from past activities in the harbor and surrounding area.
- Evaluate potential ecological risks associated with metals detected within the sediment basin (the Bay).
- Assess the need for further action to address potential risks associated with subaqueous sediments in the basin.

The datasets evaluated for this case study represent an unusually large number of samples and a wide range of analytes. However, datasets of this magnitude are not required; the same methods and procedures can be used to evaluate background conditions at sites with smaller datasets.

5.1 Field Investigation

Sediment samples were collected within the Bay during a single assessment event. Only “surface” sediment samples were collected; these extended from the sediment-water interface to a maximum depth of approximately 7 inches below the interface. The samples were collected with a Van Veen grab sampling device, which collects sediments beneath a surface area of approximately 1 square foot (144 square inches).

A total of 219 locations were sampled using a stratified random sampling design to fully characterize harbor sediments, with emphasis on areas of suspected contaminant impact (e.g., near areas of known industrial activities, stream mouths). Sampling locations were selected based on available bathymetric and sedimentological harbor characteristics, information regarding watersheds discharging to the Bay, and known or suspected point sources of contamination within the various harbor areas. The sampling locations were focused on the nearshore environment due to its proximity to upland sources of impact and the wide range of chemicals that may have been released near the shoreline. The secondary focus areas, the primary navigation channel and channels within each lobe, were evaluated to assess the effects of contaminant mixing and the lateral extent of contamination. As indicated above, all samples were subaqueous surface sediment samples; however, overlying water column depths varied as follows: Shallow Subtidal (<6 feet); Deep Subtidal (6 to 35 feet); Navigation Channel (>35 feet); and Mooring Areas (>100 feet). Sediment sampling locations are indicated on Figure 5-1.

The sediment samples were analyzed for target metals. Based on data evaluated in the baseline ecological risk assessment (BERA), four metals—arsenic, cobalt, copper, and nickel—were detected at concentrations above their respective risk-based screening criteria. These suspected COPC metals were evaluated further to determine whether concentrations that exceed screening criteria could be within the background concentration ranges.

5.2 Exploratory Data Analysis

The geological setting of the Big Spruce Bay sediment basin is quite uniform, and only minor variations in particle size were observed among the sediment samples; therefore, no particle size univariate plots were constructed. Similarly, because all sediment samples were "surface" samples, no sampling depth univariate plots were constructed. Descriptive statistics of site data for aluminum, arsenic, cobalt, copper, and nickel are provided in Table 5-1.

To illustrate the spatial (location) dependence of each metal's concentration distribution, aluminum, arsenic, cobalt, copper, and nickel concentrations were plotted on univariate plots as functions of data qualifier and sampling location. A probability plot was also constructed for each metal. Concentrations were plotted in mg/kg with respect to a $\log_{10}$ scale (as described in Section 2.2.4).

The probability plots were prepared by plotting the metal concentration associated with each data point against 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 background concentration ranges.

The following primary decision questions were addressed to evaluate the combined plots:

- Does the probability plot exhibit a distinct increase in slope (inflection point) that may separate two different populations?
- Does the estimate of the upper bound of the background concentration range depend on any U or UJ values?
- Does the sampling location univariate plot indicate that relatively high concentrations occur only in certain areas?

Probability, spatial distribution, and data qualifier plots were constructed for aluminum and each suspected COPC metal. As discussed above, because all samples were "surface" sediment samples and particle sizes are relatively uniform, no sampling depth or particle size univariate plots were constructed.

TABLE 5-1. Descriptive statistics

Statistic	Aluminum (mg/kg)	Arsenic (mg/kg)	Cobalt (mg/kg)	Copper (mg/kg)	Nickel (mg/kg)
Mean	36,462.41	14.68	36.58	114	95.75
Std.Dev	21,868.23	7.49	19.84	233	62.50
Count	219	219	219	219	219
Min.	2,030	4.20	28.17	8.40	7.60
Max.	98,700	66.90	81.2	2,020	345
Median	34,438	12.70	38.85	114	82

Aluminum

Univariate plots and a probability plot were constructed to evaluate the spatial distribution of aluminum and confirm that it can be used as a non-COPC reference metal (see Figure 5-2). Because the probability plot shows no inflection point and no outliers in the upper concentration range, the initial conclusion was that all observed aluminum concentrations are within the background concentration range. The combined plots then were evaluated to address the remaining decision questions and confirm this conclusion (see Table 5-2).

Analysis of the combined plots indicates that the maximum detected aluminum concentration (98,700 mg/kg) is also the upper bound of the estimated background concentration range. Therefore, the analysis indicates that aluminum is not a COPC.

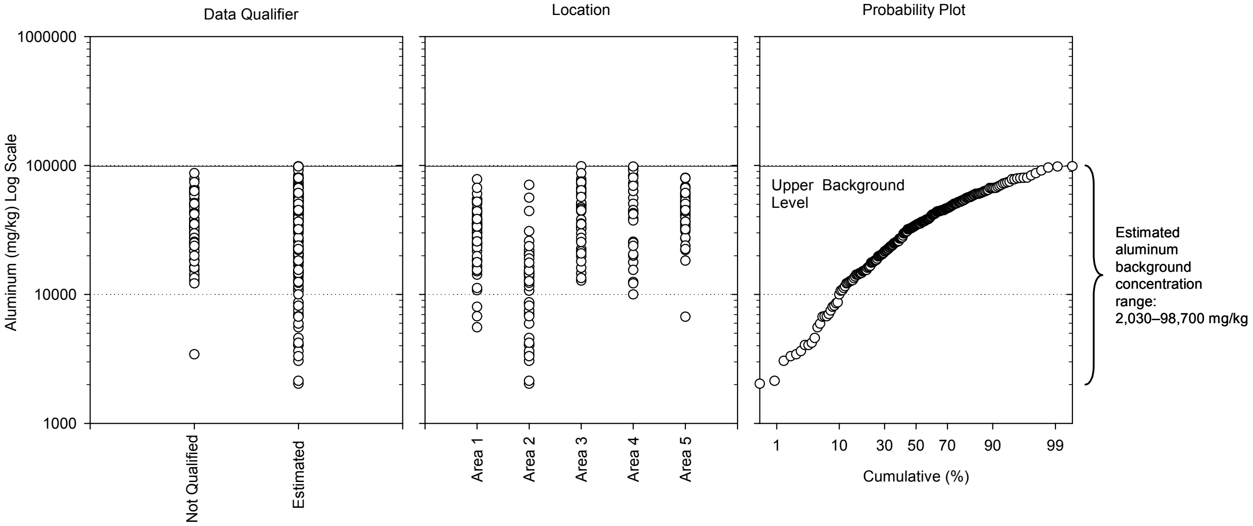


FIGURE 5-2. Combined univariate and cumulative probability plots for aluminum

TABLE 5-2. Combined plot analysis for aluminum

Decision Question	Conclusion
• Is there an inflection point in the probability plot?	No, there is no inflection point, and no data outliers. Therefore, the aluminum data indicate the concentrations are all background.
• Does the estimate of the upper bound of the background concentration range depend on any U or UJ values?	No, none of the aluminum concentration values were qualified as UJ or U. Therefore, all conclusions are based on reliable and high-quality data.
• Does the sampling location univariate plot indicate that relatively high concentrations occur only in certain areas of the site?	No, the maximum concentrations of aluminum are very similar for each of the five areas, and they are all within the background range identified on the probability plot.

Arsenic

The probability plot for arsenic (Figure 5-3) shows an inflection point at 30.6 mg/kg; therefore, the initial conclusion was that 30.6 mg/kg represents the upper bound of the background concentration range. The combined plots were then evaluated to address the remaining decision questions and confirm this conclusion (Table 5-3).

Analysis of the combined plots confirms that 30.6 mg/kg is the upper bound of the background range for arsenic. The four outliers that represent contamination are associated with samples collected in Area 5, which exhibited concentrations of 36.1, 37.7, 38.0, and 66.9 mg/kg. Therefore, the analysis indicates that arsenic is a COPC.

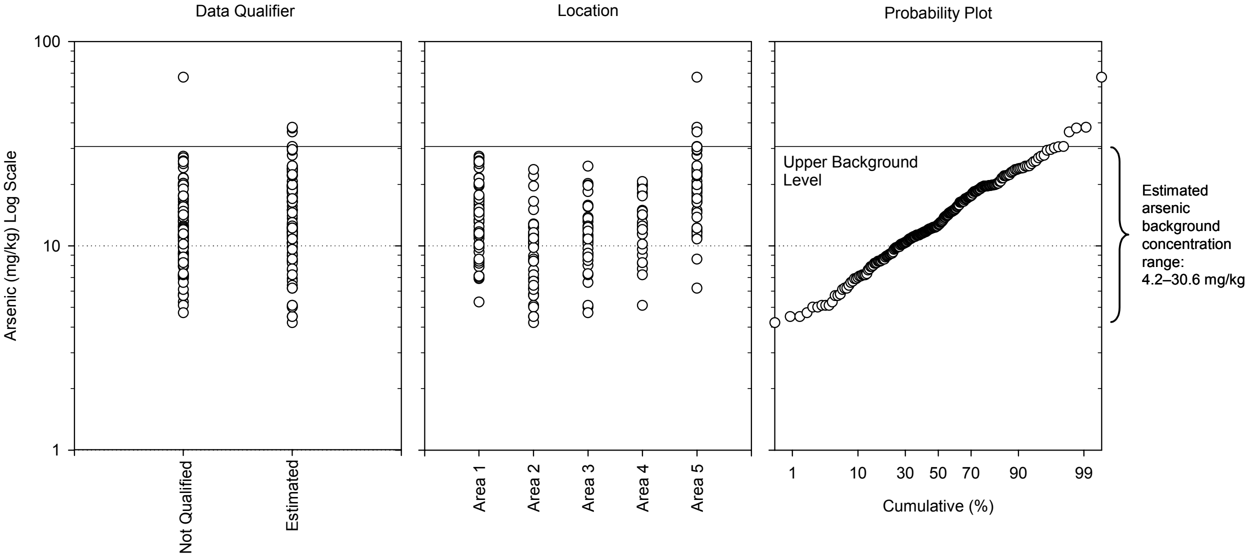


FIGURE 5-3. Combined univariate and cumulative probability plots for arsenic

TABLE 5-3. Combined plot analysis for arsenic

Decision Question	Conclusion
• Is there an inflection point in the probability plot?	Yes, there is a distinct inflection point, with four data outliers. Therefore, the arsenic probability plot indicates that the outliers beyond the inflection point are likely to represent arsenic contamination.
• Does the estimate of the upper bound of the background concentration range depend on any U or UJ values?	No, none of the arsenic concentrations were qualified as U or UJ. Therefore, the inflection point is based on reliable and high-quality data.
• 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 5. This indicates that concentrations above 30.6 mg/kg are likely to represent arsenic contamination.

Cobalt

The probability plot for cobalt (Figure 5-4) shows no inflection point and no outliers in the upper concentration range; therefore, the initial conclusion was that all observed cobalt concentrations are within the background concentration range. The combined plots were then evaluated to address the remaining decision questions and confirm this conclusion (Table 5-4).

Analysis of the combined plots indicates that the maximum detected cobalt concentration (81.2 mg/kg) is also the upper bound of the estimated background concentration range. Therefore, the analysis indicates that cobalt is not a COPC.

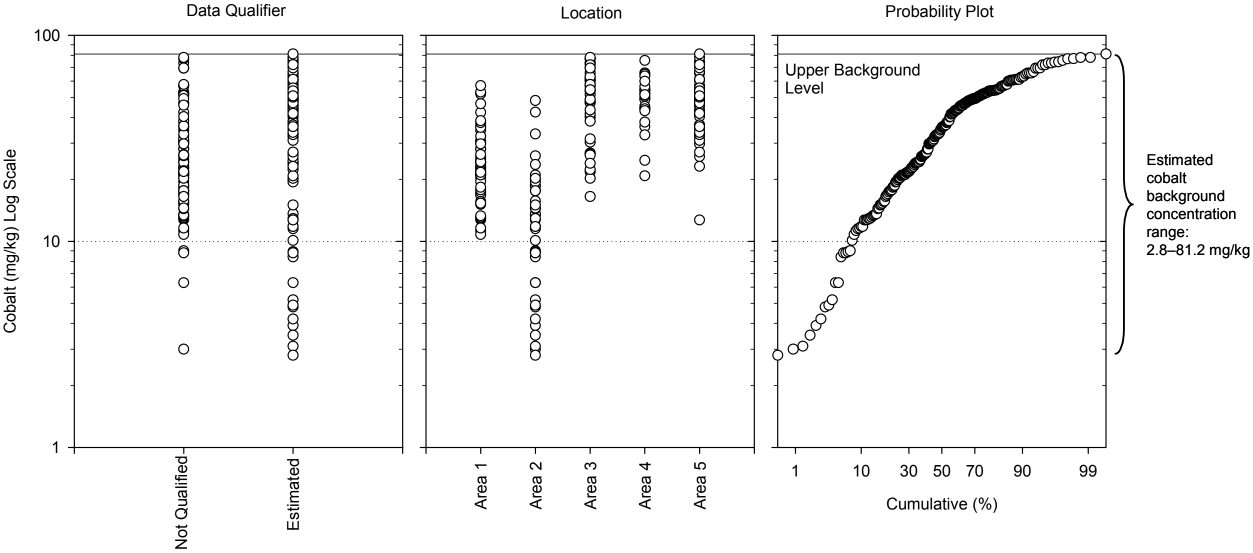


FIGURE 5-4. Combined univariate and cumulative probability plots for cobalt

TABLE 5-4. Combined plot analysis for cobalt

Decision Question	Conclusion
• Is there an inflection point in the probability plot?	No, there is no inflection point, and no data outliers. Therefore, the cobalt data indicate that the concentrations are all background.
• Does the estimate of the upper bound of the background concentration range depend on any U or UJ values?	No, all cobalt concentration data were qualified as either NQ or J. Therefore, all conclusions are based on reliable and high-quality data.
• Does the sampling location univariate plot indicate that relatively high concentrations occur only in certain areas of the site?	Yes, the highest cobalt concentrations were detected in samples from Areas 3, 4, and 5. However, the probability plot indicates that the high concentrations are within the background population distribution, and do not represent contamination. (This conclusion was confirmed by geochemical association analysis.)

Copper

The probability plot for copper (Figure 5-5) shows an inflection point at 819 mg/kg; therefore, the initial conclusion was that 819 mg/kg represents the upper bound of the background concentration range. The combined plots then were evaluated to address the remaining decision questions and confirm the initial conclusion (Table 5-5).

Analysis of the combined plots confirms that 819 mg/kg is the upper bound of the background range for copper. Three outliers that appear to represent contamination (1,300, 1,890, and 2,020 mg/kg) are associated with sediment samples from Area 1. Therefore, the analysis indicates that copper is a COPC.

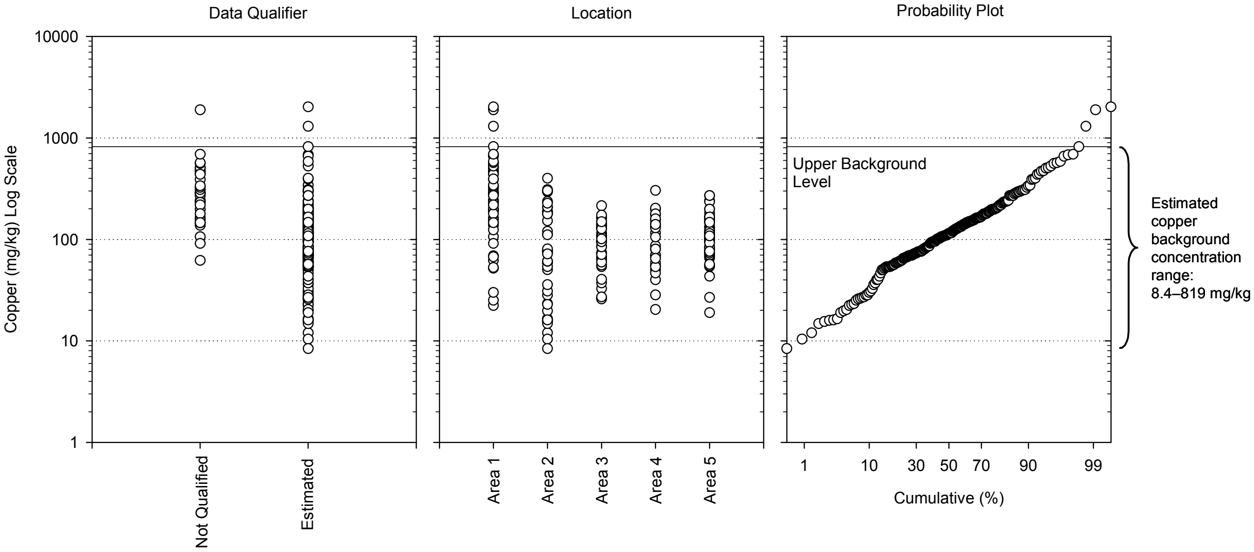


FIGURE 5-5. Combined univariate and cumulative probability plots for copper

TABLE 5-5. Combined plot analysis for copper

Decision Question	Conclusion
• Is there an inflection point in the probability plot?	Yes, there is a distinct inflection point, with three data outliers. Therefore, the copper probability plot indicates the outliers beyond the inflection point are likely to represent copper contamination.
• Does the estimate of the upper bound of the background concentration range depend on any U or UJ values?	No, none of the copper data were qualified as either U or UJ data. Therefore, the inflection point was based on reliable and high-quality data.
• Does the sampling location univariate plot indicate that relatively high concentrations occur only in certain areas of the site?	Yes, copper concentrations above the inflection point were detected only in samples from Area 1. This indicates that concentrations above 819 mg/kg are likely to represent copper contamination.

Nickel

The probability plot for nickel (Figure 5-6) shows no inflection point and no outliers in the upper concentration range; therefore, the initial conclusion was that all observed nickel concentrations are within the background concentration range. The combined plots were then evaluated to address the remaining decision questions and confirm this conclusion (Table 5-6).

Analysis of the combined plots indicates that the maximum detected nickel concentration (345 mg/kg) is also the upper bound of the estimated background concentration range. Therefore, the analysis indicates that nickel is not a COPC.

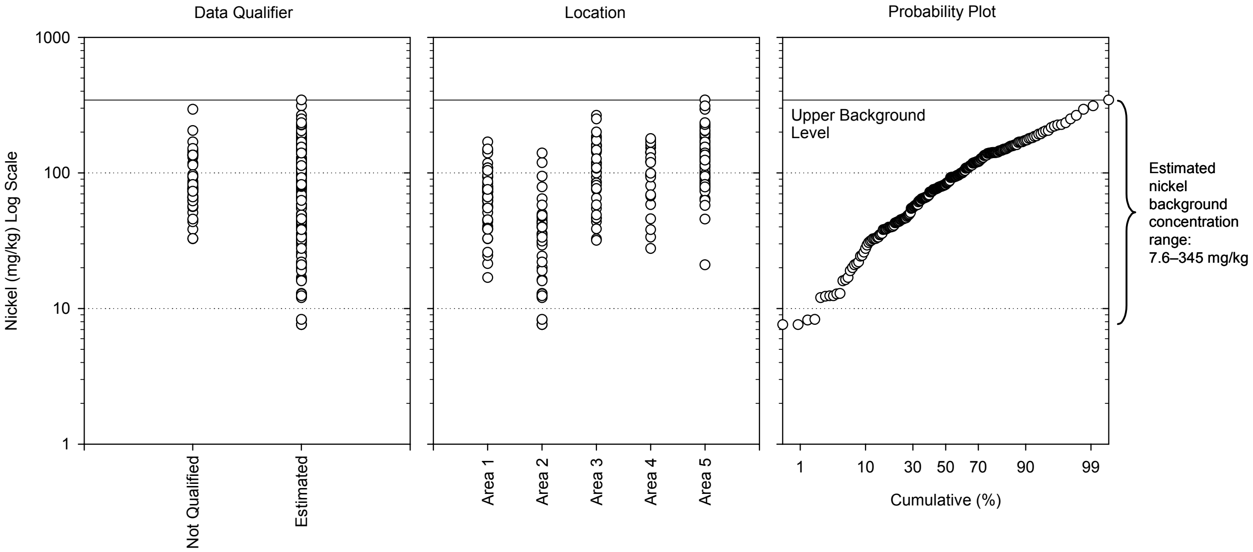


FIGURE 5-6. Combined univariate and cumulative probability plots for nickel

TABLE 5-6. Combined plot analysis for nickel

Decision Question	Conclusion
• Is there an inflection point in the probability plot?	No, there is no inflection point, and no data outliers. Therefore, the nickel data indicate the concentrations are all background.
• Does the estimate of the upper bound of the background concentration range depend on any U or UJ values?	No, all nickel concentration data were qualified as either NQ or J. Therefore, all conclusions are based on reliable and high-quality data.
• Does the sampling location univariate plot indicate that relatively high concentrations occur only in certain areas of the site?	Yes, the highest nickel concentrations were detected in samples from Areas 3 and 5. However, the probability plot indicates that the high concentrations are within the background population distribution, and do not represent contamination. (This conclusion was confirmed by comparative analysis.)

Summary of Results

- All detected aluminum concentrations are within the estimated background range—aluminum is not a COPC.
- Four arsenic outliers were identified—arsenic is a COPC.
- All detected cobalt concentrations are within the estimated background range—cobalt is not a COPC.
- Three copper outliers were identified—copper is a COPC.
- All detected nickel concentrations are within the estimated background range—nickel is not a COPC.

The exploratory data analysis yielded technically defensible estimates of the background concentration ranges for aluminum and all four suspected COPC metals. However, to illustrate application of the Geochemical and Comparative Methods, it was assumed that some stakeholders did not accept the estimated arsenic, cobalt, and nickel background ranges.

5.3 Geochemical Method Analysis

Step 1 of the Geochemical Method, geochemical association analysis (Section 3.2), was used to estimate the cobalt and nickel background ranges. After evaluating the geochemical characteristics of the Bay sediments, and completing the exploratory data analysis described above, aluminum was identified as a non-COPC metal and was selected for use as a reference metal. Although not presented in this case study, iron and aluminum concentrations exhibited a high degree of correlation; therefore, iron also was considered a potential reference metal. However, because aluminum is more commonly selected as a reference metal for the Geochemical Method, it was used as the reference metal for the case study.

TABLE 5-7. Matrix of correlation coefficients

	Aluminum	Arsenic	Cobalt	Copper	Nickel
Aluminum	1.00	0.591	0.856	0.500	0.923
Arsenic		1.00	0.525	0.513	0.569
Cobalt			1.00	0.267	0.919
Copper				1.00	0.423
Nickel					1.00

As shown in Table 5-7, arsenic, cobalt, copper, and nickel are all correlated with aluminum. The cobalt-aluminum correlation coefficient (0.856) and nickel-aluminum correlation coefficient (0.923) indicate strong geochemical association. Although arsenic was not evaluated by the Geochemical Method, the arsenic-aluminum correlation coefficient (0.591) indicates the two metals are associated, and the background range could have been estimated by geochemical association analysis. However, for the purposes of this case study, it is assumed that the Geochemical Method did not yield acceptable estimates of the arsenic background range. Arsenic was therefore carried forward for evaluation by the Comparative Method.

The log-log scatter plots shown in Figures 5-7 and 5-8 were constructed to estimate the background ranges for cobalt and nickel. Least-squares linear regression was applied to model the best-fit straight line through the data, and evaluate the

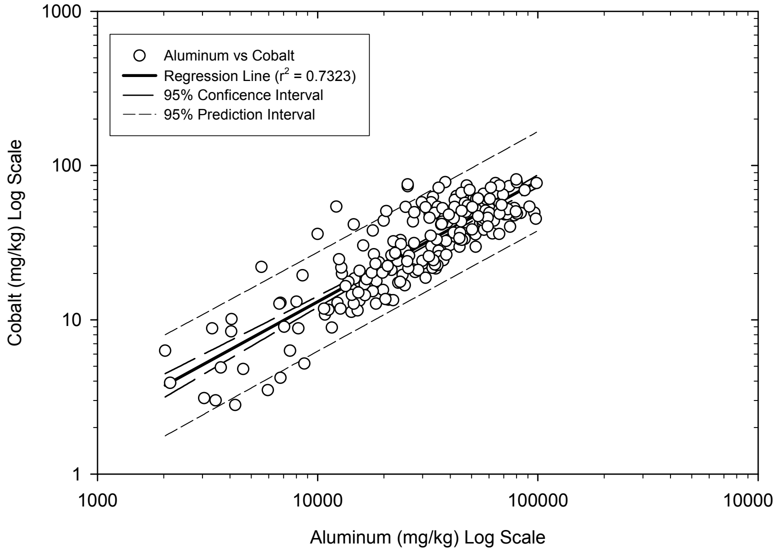


FIGURE 5-7. Geochemical regression: cobalt vs. aluminum

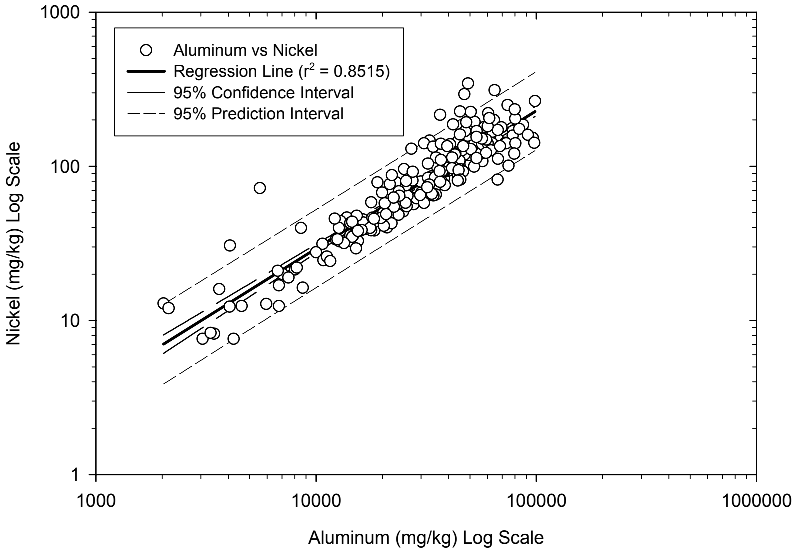


FIGURE 5-8. Geochemical regression: nickel vs. aluminum

strength of geochemical association between aluminum and each suspected COPC metal.

Cobalt

The correlation coefficient associated with the cobalt-aluminum least-squares regression line shown in Figure 5-7 (0.856) indicates a strong geochemical association relationship. All but ten of the observed cobalt concentrations lie within the 95% prediction interval; this is further evidence of the strength of the cobalt-aluminum relationship. Points outside the 95% prediction interval are either equal to or less than the maximum concentrations inside the 95% prediction interval plot. Therefore, these data are not considered outliers, and are considered to be within the estimated background range. Results of the geochemical regression analysis are consistent with results of the combined plot analysis; therefore, the maximum detected cobalt concentration (81.2 mg/kg) represents the approximate upper bound of the natural background range.

Nickel

The correlation coefficient associated with the nickel-aluminum least-squares regression line shown in Figure 5-8 (0.923) indicates a very strong geochemical association relationship. Although the nickel-aluminum scatter plot exhibits a higher degree of correlation than the cobalt-aluminum plot, there are three outliers with concentrations higher than the maximum values within the 95% prediction interval. If geochemical association analysis were the only method used to evaluate nickel, these three outliers would be assumed to be above the background range. In this case, the estimated upper bound of the background range would be the concentration value associated with the highest data point within the 95% prediction interval (265 mg/kg).

Summary of Results

Results of the geochemical association analysis indicate that cobalt is not a COPC in the case study area, and thus confirm conclusions of the combined plots analysis. Although the combined plots analysis indicated that all detected nickel concentrations are most likely within the background range, the geochemical association analysis indicated that three concentration values may be above the background range. Therefore, nickel was carried forward for evaluation by the Comparative Method. For illustration purposes, it was assumed that nickel could not be evaluated by geochemical enrichment analysis due to the lack of chemical data representing the (sediment source) parent rock.

5.4 Comparative Method Analysis

The Comparative Method was used to confirm findings of the exploratory data analysis for arsenic, and to resolve the conflicting results for nickel. For the Comparative Method analysis, Area 3 arsenic and nickel concentrations were considered background or "reference data," whereas Area 5 arsenic and nickel concentrations were considered "contaminated site data." (Area 3 was the least industrialized area of the Bay. Area 5 was one of the most industrialized areas.) Figure 5-1 shows the relative locations of Areas 3 and 5. Table 5-8 lists the arsenic and nickel data associated with each area.

TABLE 5-8. Area 5 and Area 3 (reference) data

Area 5 Data			Area 3 (Reference) Data		
Sample	**Arsenic (ppm)**	**Nickel (ppm)**	**Sample**	**Arsenic (ppm)**	**Nickel (ppm)**
1	18.80	84.90	1	10.46	161.23
2	19.70	93.00	2	6.60	157.79
3	12.10	145.00	3	11.30	150.00
4	23.80	172.00	4	9.20	118.00
5	22.80	139.00	5	9.11	85.95
6	19.60	141.00	6	12.10	158.00
7	19.30	127.00	7	13.10	157.00
8	24.00	149.00	8	24.50	265.00
9	18.40	122.00	9	13.80	186.00
10	17.30	99.60	10	13.60	142.00
11	24.50	112.00	11	12.30	101.00
12	27.70	121.00	12	7.20	250.00
13	11.70	45.60	13	8.01	180.14
14	23.30	182.00	14	8.46	173.96
15	18.50	216.00	15	15.80	200.00
16	11.10	87.60	16	12.48	145.60
17	6.20	141.00	17	20.17	146.97
18	11.70	130.00	18	11.02	75.24
19	29.30	194.00	19	11.10	92.80
20	22.20	139.00	20	9.08	84.30
21	20.70	110.00	21	7.30	65.10
22	17.40	151.00	22	12.32	99.69
23	23.90	169.00	23	10.60	119.00
24	23.00	161.00	24	5.00	43.70
25	16.50	187.00	25	5.10	38.60
26	16.30	221.00	26	4.70	32.60
27	12.20	135.00	27	18.40	108.00
28	11.50	193.00	28	11.80	54.70
29	17.20	204.00	29	10.20	46.20
30	10.80	134.00	30	8.93	46.43
31	18.50	114.00	31	19.82	113.91
32	23.60	169.00	32	15.50	80.90
33	19.60	155.00	33	19.70	127.00
34	30.10	104.00	34	9.70	76.70
35	37.70	92.30	35	15.60	109.00
36	17.00	140.00	36	9.20	49.00
37	19.00	64.40	37	8.80	31.80
38	18.10	92.60	38	10.80	58.00
39	16.50	97.10			
40	30.60	94.20			
41	38.00	135.00			
42	22.00	113.00			
43	14.90	76.00			
44	19.40	79.30			
45	29.60	135.00			
46	16.30	81.90			
47	19.40	80.80			
48	19.80	82.00			
49	36.10	345.00			
50	18.70	294.00			
51	14.50	225.00			
52	19.90	205.00			
53	24.60	227.00			
54	17.00	234.00			
55	22.00	21.00			
56	8.60	73.10			
57	12.20	62.70			
58	22.30	78.70			
59	66.90	124.00			
60	30.50	312.06			
61	13.80	57.60			

5.4.1 Comparative Method Application

Most of the background issues were conclusively addressed through exploratory data analysis, and geochemical association analysis. The remaining background questions were related to the arsenic and nickel concentrations detected in Area 5 sediments: Are arsenic and nickel concentrations in Area 5 statistically different from those in the "Reference Area" (Area 3)?

To answer this question, the Area 5 arsenic and nickel data (i.e., site data) were statistically compared to the Area 3 arsenic and nickel data (i.e., reference [or background] area data). Histograms of the observed concentrations and their corresponding fitted normal curves are displayed on Figure 5-9.

The observed Area 5 contamination could be spread over the entire area (i.e., basin-wide impact), or could be limited to small zones within the area (i.e., localized impact). Consequently, two types of tests were selected:

- ❑ The Slippage test, to assess the differences among elevated concentrations in both Area 5 and the reference area (Area 3); and
- ❑ The WRS test, to assess the difference between median concentrations.

Both these tests are nonparametric and do not require any specific distributional assumptions. However, given the large size of the datasets (i.e., >30), the data were also subjected to two parametric tests of the equality of mean concentrations: the two-sample t and Satterthwaite tests.

5.4.2 Slippage Test Results

Results of the Slippage test, shown in Table 5-9, indicate an unusually large number of elevated arsenic concentrations in Area 5 compared to the maximum reference area concentrations. Elevated nickel concentrations in Area 5, however, do not differ significantly from the maximum reference area nickel concentrations. These results confirm the exploratory data analysis findings, which indicated that the elevated Area 5 arsenic concentrations represent above-background contamination, whereas nickel occurs only as a background chemical.

5.4.3 WRS Test Results

As shown in Table 5-10, the WRS test results indicate that the median arsenic concentration in Area 5 is significantly different from the median arsenic concentration in the reference area. Conversely, the WRS test results indicate that the median nickel concentration in Area 5 is not significantly different from the median nickel concentration in the reference area. These results also confirm the exploratory data analysis findings, which indicated that the elevated Area 5 arsenic concentrations represent above-background contamination, whereas nickel occurs only as a background chemical.

5.4.4 Two-Sample t and Satterthwaite Two-Sample t Test Results

Given the large number of Area 5 and reference area measurements (more than 30 measurements from each area), the arsenic and nickel data also were subjected to the two-sample t test and the Satterthwaite two-sample t test. As indicated in Table 5-11, the test results indicate that the mean Area 5 arsenic concentration is significantly different from the mean reference area arsenic concentration. The test results also confirm that the mean Area 5 nickel concentration is not significantly different from the mean reference area nickel concentration. The results thus support earlier findings, which indicated that the elevated Area 5 arsenic concentrations represent contamination, whereas nickel occurs only as a background chemical.

5.4.5 Summary of Comparative Method Results

Results of the Comparative Method analysis are summarized as follows:

- ❑ All nickel concentrations in Bay sediments are within the background range.
- ❑ The nickel background range was accurately estimated by the combined plots analysis presented in Section 5.2.

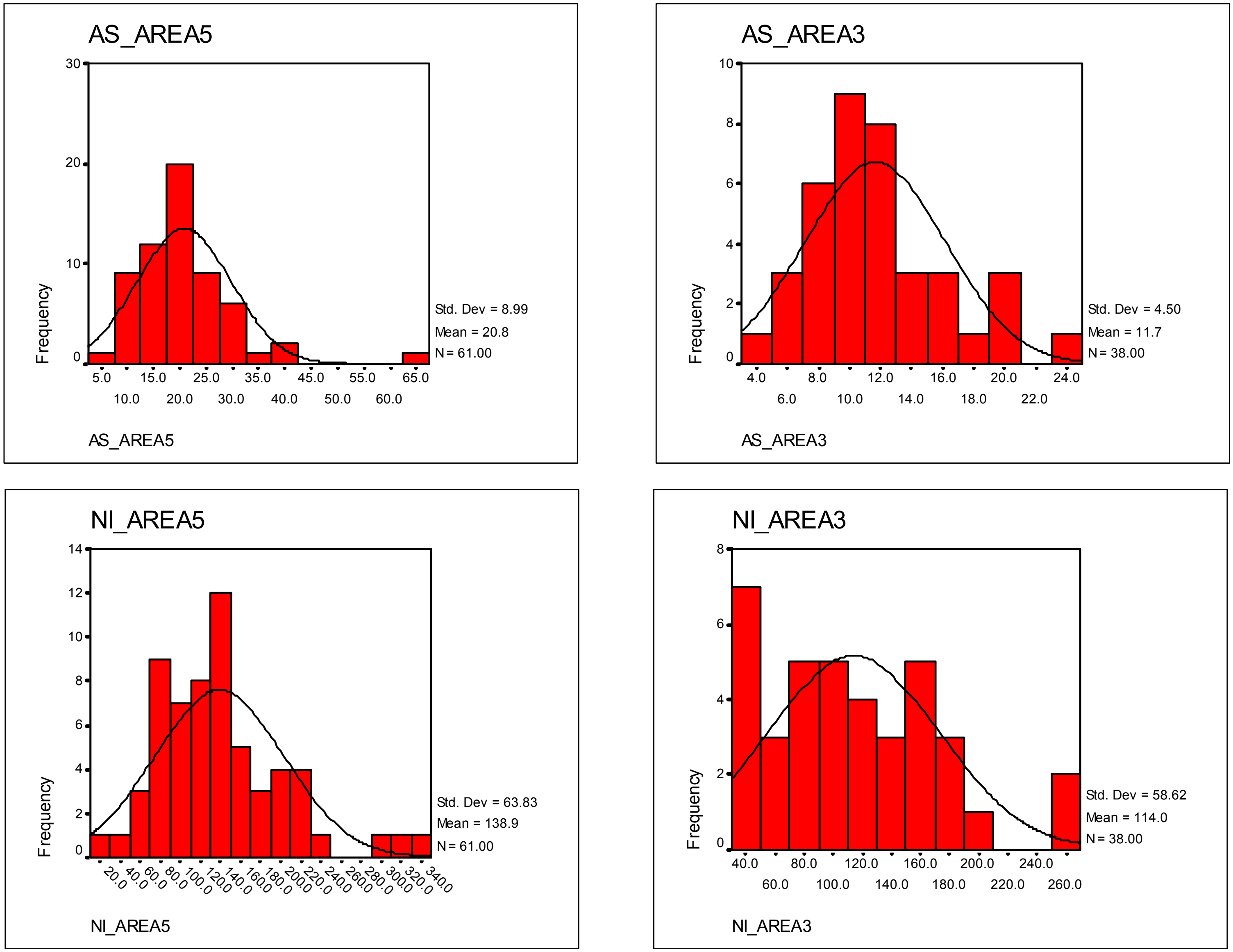


FIGURE 5-9. Histograms of arsenic and nickel data in Areas 5 and 3

TABLE 5-9. Results of Slippage test

Arsenic (Area 5)	Arsenic (Area 3)		Nickel (Area 5)	Nickel (Area 3)	
66.90	24.50	K = 11	345.00	265.00	K = 3
38.00	20.17	$K_c \sim 7$	312.06	250.00	$K_c \sim 7$
37.70	19.82	$K > K_c$	294.00	200.00	$K < K_c$
36.10	19.70	**Arsenic is a COPC**	234.00	186.00	**Nickel is not a COPC**
30.60	18.40		227.00	180.14	
30.50	15.80		225.00	173.96	
30.10	15.60		221.00	161.23	
29.60	15.50		216.00	158.00	
29.30	13.80		205.00	157.79	
27.70	13.60		204.00	157.00	
24.60	13.10		194.00	150.00	
24.50	12.48		193.00	146.97	
24.00	12.32		187.00	145.60	
23.90	12.30		182.00	142.00	
23.80	12.10		172.00	127.00	
23.60	11.80		169.00	119.00	
23.30	11.30		169.00	118.00	
23.00	11.10		161.00	113.91	
22.80	11.02		155.00	109.00	
22.30	10.80		151.00	108.00	
22.20	10.60		149.00	101.00	
22.00	10.46		145.00	99.69	
22.00	10.20		141.00	92.80	
20.70	9.70		141.00	85.95	
19.90	9.20		140.00	84.30	
19.80	9.20		139.00	80.90	
19.70	9.11		139.00	76.70	
19.60	9.08		135.00	75.24	
19.60	8.93		135.00	65.10	
19.40	8.80		135.00	58.00	
19.40	8.46		134.00	54.70	
19.30	8.01		130.00	49.00	
19.00	7.30		127.00	46.43	
18.80	7.20		124.00	46.20	
18.70	6.60		122.00	43.70	
18.50	5.10		121.00	38.60	
18.50	5.00		114.00	32.60	
18.40	4.70		113.00	31.80	
18.10			112.00		
17.40			110.00		
17.30			104.00		
17.20			99.60		
17.00			97.10		
17.00			94.20		
16.50			93.00		
16.50			92.60		
16.30			92.30		
16.30			87.60		
14.90			84.90		
14.50			82.00		
13.80			81.90		
12.20			80.80		
12.20			79.30		
12.10			78.70		
11.70			76.00		
11.70			73.10		
11.50			64.40		
11.10			62.70		
10.80			57.60		
8.60			45.60		
6.20			21.00		

TABLE 5-10. Results of WRS test

Arsenic

Area	Number of Measurements	Mean Rank	Sum of Ranks
5	61	63.5	3,876.5
3 (reference area)	38	28.3	1,073.5
Test Statistics:			
Mann-Whitney U	332.5		
Wilcoxon W	1,073.5		
Z	−5.947022915		
Asymp. Sig. (2-tailed)	2.73063E−09	<0.05	95% significance
Conclusion: Arsenic is a COPC			

Nickel

Area	Number of Measurements	Mean Rank	Sum of Ranks
5	61	54.1	3,298.5
3 (reference area)	38	43.5	1,651.5
Test Statistics:			
Mann-Whitney U	910.5		
Wilcoxon W	1,651.5		
Z	−1.788014531		
Asymp. Sig. (2-tailed)	0.07377366	>0.05	95% significance
Conclusion: Nickel is not a COPC			

- Arsenic is a COPC, and the elevated Area 5 concentrations represent above-background contamination.

- The background range for arsenic was accurately estimated by the combined plots analysis presented in Section 5.2.

5.5 Case Study Summary

- Exploratory data analysis conclusively demonstrated that aluminum is not a COPC; the estimated background range is 2,030-98,700 mg/kg.

- Exploratory data analysis indicated that arsenic is a COPC; the estimated background range is 4.2-30.6 mg/kg. These results were confirmed by the comparative analysis.

- Exploratory data analysis indicated that cobalt is not a COPC; the estimated background range is 2.1-81.2 mg/kg. These results were confirmed by geochemical association analysis.

- Exploratory data analysis conclusively demonstrated that copper is a COPC; the estimated background range is 8.4-819 mg/kg.

- Exploratory data analysis indicated that nickel is not a COPC, and that the estimated background range is 7.6-345 mg/kg. The geochemical method did not confirm these results—three elevated concentration outliers were identified and the upper bound of the background range was estimated at only 265 mg/kg. The comparative analysis supported the conclusions of the combined plot analysis; therefore, the final estimated upper bound of the nickel background range is 345 mg/kg.

The arsenic outliers appear to represent contamination within Area 5. The copper outliers appear to represent contamination within Area 1. As previously discussed, Areas 1 and 5 were the most heavily used lobes of the Bay; therefore, these results are supported by the site history.

TABLE 5-11. Results of two-sample t and Satterthwaite tests

Arsenic

Area	Number of Measurements	Mean	Std. Deviation
5	61	20.798	8.988
3 (reference area)	38	11.654	4.497

Test Statistics:	Levene's Test for Equality of Variances			Test for Equality of Means			
	F	Sig.	Result	t	f	Sig. (2-tailed)	
t-Test (Equal variances assumed)	4.723	0.032	<0.05 (Equal variances rejected)	Not Applicable			
Satterthwaite Test (Equal variances not assumed)				6.711	93.444	0.000	<0.05 (Equality of mean rejected)
Conclusion: Arsenic is a COPC.							

Nickel

Area	Number of Measurements	Mean	Std. Deviation
5	61	138.860	63.830
3 (reference area)	38	113.982	58.619

Test Statistics:	Levene's Test for Equality of Variances			Test for Equality of Means			
	F	Sig.	Result	t	f	Sig. (2-tailed)	
t-Test (Equal variances assumed)	0.004	0.948	>0.05 (Equal variances not rejected)	1.945	97	0.055	>0.05 (Equality of mean not rejected)
Satterthwaite Test (Equal variances not assumed)				Not Applicable			
Conclusion: Nickel is not a COPC.							

Levene, H. 1960. Robust Tests for Equality of Variances. pp. 278-292 in *Contributions to Probability and Statistics,* I. Olkin et al., eds. Stanford, CA: Stanford University Press.

APPENDIX A: CHEMICAL FINGERPRINTING OF PAHS IN SEDIMENTS – RECOGNIZING THE CONTRIBUTION OF URBAN BACKGROUND

A.1 Introduction

Polycyclic aromatic hydrocarbons (PAHs), also called polynuclear aromatic hydrocarbons, are nearly ubiquitous contaminants of freshwater and marine sediments worldwide. They are acutely toxic and have carcinogenic properties, and are being recognized with increasing frequency as major contaminants in sediments, particularly in urban environments. As a result, the occurrence of PAHs in sediments is regulated under numerous U.S. laws, which often require remedial activities or, at a minimum, long-term monitoring.

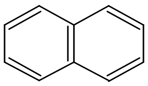
Because of the significant costs associated with either of these actions, it is prudent to identify the sources of PAHs at contaminated sites before accepting liability for their presence and designing a remedial strategy. It is particularly important to understand the sources of PAHs in waterways and coastal areas, because multiple PAH point sources (such as spills/seeps of petroleum or coal-derived liquids) co-exist at these sites alongside persistent nonpoint sources (such as atmospheric particles and general surface runoff from the surrounding urban or industrial communities). This situation leads to the formation of complex mixtures of PAHs in the nearby sediments.

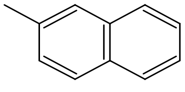
The term "urban background" denotes the variety of nonpoint sources of PAHs within urban environments. The PAHs in these background materials have been long recognized to impart an "urban background" signature to urban sediments that can be confused with or (at least) contribute to PAHs derived from point sources in the area. This appendix provides an overview of the PAHs that occur in urban sediments; analytical methods used to characterize PAHs in sediments; and the data analysis methods used to recognize and allocate different sources of PAHs in urban sediments, including those attributable to urban background.

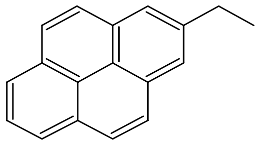
PAH sources, including urban background sources, often can be identified by using "advanced chemical fingerprinting" (ACF) methods. However, the high cost of generating analytical data for background analysis by ACF can be prohibitive for large sediment investigations. Fortunately, costs can be reduced if rapid sediment characterization (RSC) methods are used to screen a large number of sediment samples and select a subset of PAHs for further evaluation by ACF. The combined approach (RSC and ACF) provides a cost-effective tool for evaluating PAH sources and estimating background PAH concentrations in sediments.

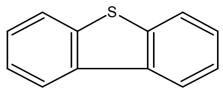
A.1.1 General Chemistry of PAHs

Recognizing the sources of PAHs in sediments requires a basic understanding of their chemistry and nomenclature. As their name implies, ***p***olycyclic ***a***romatic ***h***ydrocarbons literally: (1) contain multiple ring structures, (2) which are aromatic in nature, and (3) comprised of hydrogen and carbon. Naphthalene, consisting of two fused benzene rings, is the simplest PAH (see right). The arrangement and number of fused rings is used to distinguish different PAHs.


naphthalene


2-methyl naphthalene


2-ethyl pyrene


dibenzithiophene

Some PAHs contain carbon side chains of varying numbers, lengths, and locations. These carbon chains are termed "alkyl groups." PAHs that do not contain alkyl groups are termed "nonalkylated" or "parent" PAHs, such as naphthalene. PAHs with one or more of these groups are said to be "alkylated," such as 2-methyl naphthalene and 2-ethyl pyrene (see right). Also, compounds called heterocyclics co-occur with PAHs and are similar in structure. These compounds contain either a nitrogen, oxygen, or sulfur atom, such as dibenzothiophene (see right).

A.1.2 Source Categories of PAHs in Urban Sediments

Sources of PAHs in urban sediments can be separated into several categories (Table A-1). **Petrogenic PAHs** are hydrocarbons formed by the geochemical alteration of organic matter at moderate temperature (50-150°C) and pressure over very long (i.e., geologic) timescales. These PAHs enter urban environments from anthropogenic sources such as petroleum (crude oil or fuels) spills/leaks, coal-fired power plants, and municipal sewage treatment plants. **Pyrogenic PAHs** form when fuels and other organic matter are incompletely or inefficiently combusted or pyrolyzed at moderate to high temperatures (>400°C) over very short time intervals. (It is important to note that not all petrogenic and pyrogenic PAHs found as sediment contaminants are derived from anthropogenic sources; in some areas, the presence of petrogenic PAHs results from eroded coal/shales or naturally seeped oil; also, pyrogenic PAHs can be generated from naturally occurring forest and brush fires.)

Biogenic PAHs include those derived from oxidation of microbial- or plant-derived compounds. Biogenic and other nonanthropogenic sources of PAHs typically occur in older, deeper sediments that were deposited prior to industrialization/urbanization or in sediments in remote locations. As such, most nonanthropogenic PAH sources are generally considered to be minor contributors to the PAHs found in near-surface sediments in modern urbanized areas.

TABLE A-1. Source Categories of PAHs in Sediments

PAH Category	Origin
Petrogenic	Generated from organic matter in ancient sediments/rocks by geologic conditions (temperature and pressure) over geologic time
Pyrogenic	Generated from combustion or pyrolysis of organic matter (wood, coal, petroleum, wastes)
Biogenic	Generated by modern biological processes or by diagenetic processes (e.g., oxidation of organic matter) in Recent sediments

A.1.3 Sources of Background PAHs in Sediments

PAHs are introduced into sediment environments through a variety of anthropogenic activities from both point and nonpoint sources that may exist along urban waterways. Common point sources of PAHs in many urban sediments include direct or indirect discharges from petroleum terminals, shipyards, aluminum smelting, manufactured gas production facilities, tar distillation plants, rail yards, loading/unloading facilities (e.g., creosote pilings), marinas, discharge canals, and stormwater outfalls; and spilled or seeped petroleum or coal- or oil-derived tars and their associated distillation products (e.g., creosote). Common nonpoint sources include atmospheric (soot) particulates and dripped/leaked petroleum washed from the surrounding urban roadways, parking lots, vegetation, and structures during rainfall events. Other nonpoint sources of PAHs in urban waterways include recreational boat traffic, ship traffic, general surface and stormwater runoff (i.e., not entering from a specific outfall location), and direct atmospheric particulate deposition (soot from petroleum combustion, forest fires, wood stoves, coal-fired power plants, smelters, etc.) to the waterway.

On a global basis and in areas remote from urban influence, background PAHs generally are limited to pyrogenic PAHs derived from particles transported over large distances (Ohkouchi et al., 1999). In selected geologically active environments, oil seeps and erosion from petroleum source rocks and coal can result in elevated concentrations from natural sources of petrogenic PAHs (Boehm et al., 2000a and 2000b). The concentrations of background PAHs in these remote areas are generally much lower than background PAH concentrations in urban waterways, where direct deposition of combustion-related PAHs from proximal sources and urban runoff have occurred for much of the last century.

In urban sediments, background PAHs associated with pyrogenic sources usually are more abundant than those associated with petrogenic sources, due to the high volume of fossil fuels combusted in urban areas. Concentrations of pyrogenic PAHs normally are highest in upper sediment layers, and decrease to a relatively constant "natural background" concentration at depths corresponding to deposition prior to urbanization. In most settings, the most highly PAH-contaminated sediments were deposited between the start of heavy industrial use of fossil fuels and the present, indicating that the fraction of the total "urban background" PAHs above natural background levels is derived from the combustion of fossil fuels (see example in Section A.4.3). Forest and brush fires probably are major sources of the "natural background" pyrogenic PAHs deposited before the industrial revolution.

Background petrogenic PAHs also are present in most urban sediments, although they usually are less abundant than background pyrogenic PAHs. Their occurrence is largely attributable to uncombusted petroleum spilled or dripped onto roadways and parking lots (e.g., crankcase oil) that enters a waterbody via runoff following storm events. Although oil spills receive considerable attention from industry, regulators, and the media, they typically do not account for a large fraction of PAHs entering most settings. This is because oil spills account for only a fraction of the oil released into the environment, and because the concentrations of PAHs in the petroleum are much less than those of other sources of PAHs (Stout et al., 2001a).

A.2 Quantifying PAHs in Sediments

The first step in characterizing the contribution of urban background PAHs to contaminated sediments is to recognize and quantify all PAHs present. This effort requires good spatial (and perhaps depth) coverage of the impacted sediments, as well as a method for identifying the signatures of PAHs in the sediments. These two requirements can be cost-effectively achieved through the combination of rapid sediment characterization (RSC) and advanced chemical fingerprinting (ACF). RSC methods are used to cover large sediment areas in order to identify PAH trends, hotspots, and key samples (Section A.2.2). ACF then is used to analyze a selected subset of sediments in order to recognize distinct PAH (and other hydrocarbon) source "fingerprints" and concentrations, including those of urban background (Section A.2.3). Overall, this approach is designed to determine if distinct PAH sources can be recognized and, if so, to determine the nature and extent of those contaminants in sediments near Navy facilities.

This section provides a summary of RSC and ACF methods. Section 3.0 provides a more detailed description of using ACF to identify PAH signatures. Also, a detailed discussion of the combined RSC-ACF approach to contaminant fingerprinting is provided in a new Navy Guidance Document currently under preparation.[1]

A.2.1 Developing a Conceptual Site Model

Before RSC or ACF are applied at a site, a conceptual site model (CSM) of the site should be developed in order to gain a solid understanding of the area's history. The primary objectives of developing a CSM for a PAH source study are to:

- ☐ Identify all of the known or suspected sources of PAHs within the study area; and,
- ☐ Develop specific objectives (hypotheses) to be evaluated by the PAH source study that address the greatest environmental risks, and provide the greatest potential benefit for the Navy.

These objectives can be achieved by reviewing pre-existing data for the study area and nearby areas, which might provide additional insight to

---

[1] *A User's Guide for Optimizing the Determination of Contaminant Sources in Sediments Near Naval Facilities*. Contact: Stacey Curtis, SPAWAR Systems Center, San Diego, CA.

regional background issues, and by reviewing the anthropogenic use history of an area. An important consideration in the historical assessment of an area is the occurrence and extent of past dredging activities that may have occurred within the study area. This is important because dredging activities influence the distribution of PAHs, and therefore impact decisions on where to collect sediment samples. Another consideration is the magnitude of stormwater drainage, particularly in large urban areas with sizable watersheds which are collectively drained and which may enter the waterway at specific outfalls.

A.2.2 Overview of Rapid Sediment Characterization

RSC serves as an initial, low-cost method for screening a large number of sediment samples in order to select a subset of PAHs for further evaluation by ACF. RSC can cost-effectively provide the spatial (and depth) coverage necessary to:

❑ Recognize concentration gradients (which helps identify the fate and transport processes at work in an area and changes in PAH input over time);

❑ Recognize PAH hotspots (which often indicate point source areas); and

❑ Help define background PAH concentrations (which will impact cleanup goals and strategies).

Data from the RSC screening process then are synthesized to identify samples of sediments thought to represent background conditions in the study area, and determine which samples are most appropriate for ACF analysis.

Numerous RSC methods are described in existing Navy guidance (http://web.ead.anl.gov/ecorisk/issue/pdf/rsc.pdf). Techniques for the RSC of PAHs in sediments have been adapted from methods developed for use in soils. They rely upon either immunoassay or ultraviolet fluorescence techniques that require dewatering of the sediment followed by solvent (e.g., methanol) extraction from the sediment matrix. RSC also can be enhanced significantly if additional data on sediment physical properties are collected (for example, XRF screening for metals, sediment grain size, or total organic carbon; see Section A.4.2).

A.2.3 Overview of Advanced Chemical Fingerprinting

ACF is an umbrella term for identifying and distinguishing PAH sources in sediments using several different and complementary analytical techniques.

The U.S. EPA lists 16 nonalkylated PAHs of principal environmental concern on its Priority Pollutant List, and has developed a standard method (SW-846 Method 8270, *Semi-Volatile Organic Compounds by Gas Chromatography/Mass Spectrometry*) to analyze for these PAHs in various media. However, hundreds of PAHs exist, and many of these are essential to PAH source investigations as *diagnostic* tools to differentiate target PAH analytes from complex mixtures. Table A-2 lists common PAH analytes targeted during PAH source investigations; the 16 U.S. EPA priority pollutants are indicated by asterisks.

One common ACF technique is to modify U.S. EPA Method 8270, which targets only the 16 priority pollutants in its standard form, and ignores all alkylated PAHs and related heterocyclics. In the modified method, a GC is operated with a very slow oven temperature program to optimize separation of target PAH compounds, and a mass spectrometer is operated in the selected ion monitoring (SIM) mode to minimize interferences from nontarget compounds and lower the detection limits. These simple modifications reduce PAH detection limits from 660 μg/kg (wet) for the standard method to 1 μg/kg (dry) for the modified method. These lower detection limits for PAHs are particularly important when attempting to define background concentration levels. In addition, modified Method 8270 measures the ions characteristic of the PAHs and heterocyclics useful in PAH source investigations.

Internal surrogate and recovery standards are used to measure performance and concentrations, relative to an external calibration solution containing the parent PAHs on the target analyte list. The response factors for the parent PAHs are applied to

TABLE A-2. Inventory of PAHs commonly used to distinguish PAH sources

Analyte/Analyte Group	Abbr.	Ring #	Analyte/Analyte Group	Abbr.	Ring #
Naphthalene*	N0	2	C3-dibenzothiophenes	D3	3
C1-naphthalenes	N1	2	C4-dibenzothiophenes	D4	3
C2-naphthalenes	N2	2	Fluoranthene*	FL	4
C3-naphthalenes	N3	2	Pyrene*	PY	4
C4-naphthalenes	N4	2	C1-fluoranthenes/pyrenes	FP1	4
Biphenyl	Bph	2	C2-fluoranthenes/pyrenes	FP2	4
Acenaphthylene*	Acl	3	C3-fluoranthenes/pyrenes	FP3	4
Acenaphthene*	Ace	3	Benz(a)anthracene*	BaA	4
Dibenzofuran	DdF	3	Chrysene*	C0	4
Fluorene*	F0	3	C1-chrysenes	C1	4
C1-fluorenes	F1	3	C2-chrysenes	C2	4
C2-fluorenes	F2	3	C3-chrysenes	C3	4
C3-fluorenes	F3	3	C4-chrysenes	C4	4
Anthracene*	AN	3	Benzo(b)fluoranthene*	BbF	5
Phenanthrene*	P0	3	Benzo(k)fluoranthene*	BkF	5
C1-phenanthrenes/anthracenes	P1	3	Benzo(e)pyrene	BeP	5
C2-phenanthrenes/anthracenes	P2	3	Benzo(a)pyrene*	BaP	5
C3-phenanthrenes/anthracenes	P3	3	Perylene	Per	5
C4-phenanthrenes/anthracenes	P4	3	Indeno(1,2,3-c,d)pyrene*	ID	6
Dibenzothiophene	D0	3	Dibenz(a,h)anthracene*	DA	5
C1-dibenzothiophenes	D1	3	Benzo(g,h,i)perylene*	BgP	6
C2-dibenzothiophenes	D2	3			

* Priority pollutant PAH.

the appropriate alkylated PAHs. This method is fully described in the NOAA Status and Trends methodology and numerous peer-review publications (e.g., Sauer and Uhler, 1994-5; Page et al., 1995; Boehm et al., 1997; Stout et al., 2002b).

Another common ACF technique is to modify U.S. EPA Method 8015B, which employs gas chromatography in combination with flame ionization detection (GC/FID). The primary modification to Method 8015B requires use of a slow GC heating rate, which provides a more detailed chromatographic "fingerprint" of the concentration and character of the total extractable hydrocarbons in sediments. These data then are compared to total PAH concentrations in order to differentiate target PAH analytes from the often complex mixture of hydrocarbons present in sediment samples.

Figure A-1 is an example of the results of using U.S. EPA Method 8270, both modified and unmodified, to characterize PAHs in the same petroleum source (fuel oil #6). The bottom histogram is an analysis for only the 16 priority pollutants, whereas the top histogram illustrates an analysis for the fuller suite of PAHs listed on Table A-2 using modified Method 8270. The additional "fingerprinting" information obtained in the full PAH analytes is of significant benefit in PAH source studies.

A.3 Using ACF to Distinguish PAH Sources

Because of their nature of formation and similar physical/chemical properties, groups of petrogenic or pyrogenic PAHs tend to co-occur in sediments. This knowledge allows the investigator to recognize specific PAH assemblages, or "fingerprints," as being derived from a certain source. PAH source recognition studies require identification of both nonalkylated and alkylated PAHs, as well as selected heterocyclics, in order to distinguish various PAH sources (including urban background).

Pyrogenic and petrogenic PAHs can be readily distinguished on the basis of their alkyl group distributions. Figure A-2 shows the basic relative distributions of variously alkylated PAHs, depending on the temperature (and rate) of formation. Petrogenic PAH profiles form a "bell-shaped curve" due to the relative abundance of alkylated PAHs (C_1, C_2, C_3, C_4). Pyrogenic PAH profiles form a decreasing, or "sloped" curve, due to the domination of the nonalkylated, parent PAH (C_0), over the alkylated PAHs; this dominance increases

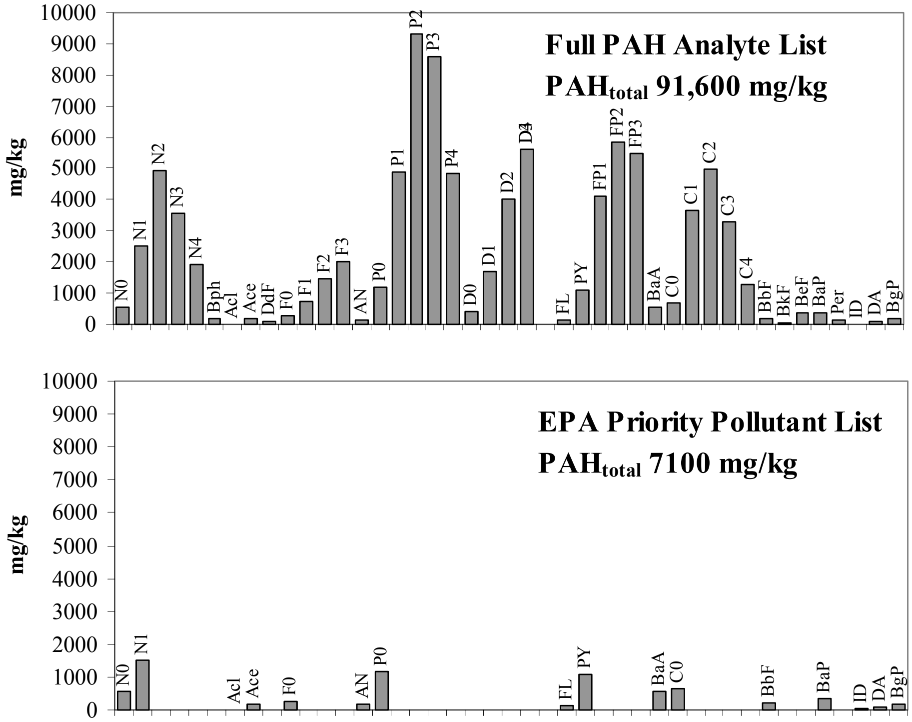


FIGURE A-1. Comparison of PAH histograms for the same fuel oil #6 using two different PAH analyte lists. For PAH compound abbreviations, see Table A-2.

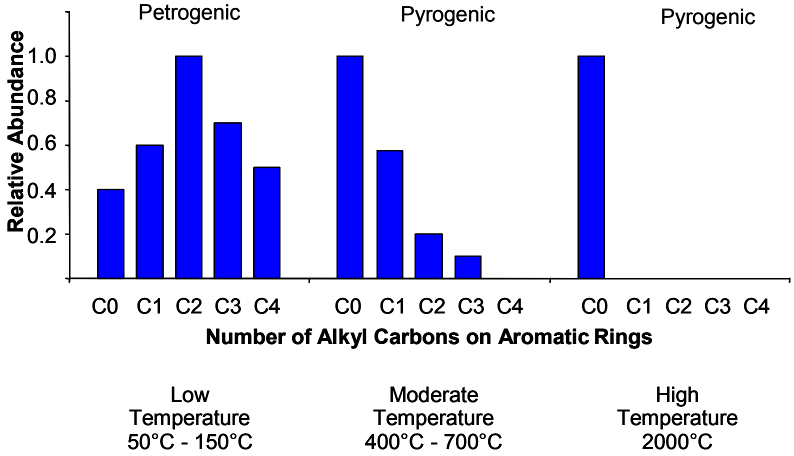


FIGURE A-2. Representative distribution of alkylated PAHs formed at different temperatures (after Blumer, 1976)

with increasing temperature of formation. (Note that these identifying curves are not exhibited in PAH histograms if only the 16 priority pollutant PAHs are measured.) More detailed descriptions of PAH sources "fingerprinting" features is available in Stout et al. (2002a).

A.3.1 Typical Petrogenic PAH Source Signatures

Petrogenic PAH source signatures are comprised of lower molecular weight (2- and 3-ring) PAHs; higher molecular weight PAHs usually are present only at low concentrations (<100 mg/kg; Kerr at al., 1999). Petrogenic PAHs also are characterized as having higher concentrations of alkyl groups than pyrogenic PAHs, and exhibit "bell curve" profiles of PAH histograms (Figure A-2). These PAHs commonly enter urban and coastal waterways from anthropogenic sources such as spills or leaks of crude oil and other fuels, but also are comprised of effluents from oil terminals and refineries, discharges of ballast and bilge water from ships, coal-fired power plants, as a component within urban runoff, and from municipal sewage treatment plants.

As crude oil is refined, it is subjected to heating under mild conditions (<550°C) and separates into various distillates, including light distillates (gasolines, kerosene, jet fuel), middle distillates (diesel fuel #2, fuel oil #2 and #4), and residuals (lube oils, fuel oil #6, bunker C fuel). Because distillation heating is relatively mild, there is no significant formation of new (more highly condensed) PAHs. Thus, the resulting refined products only contain PAHs that were present in the parent crude oil.

Different crude oils, and the petroleum products refined from them (such as gasoline), exhibit different PAH distributions, which can be useful in differentiating among individual petrogenic sources. Figure A-3 shows the PAH profiles determined for two different crude oils (one fresh and one weathered) and for diesel fuel #2 (weathered). Some obvious differences in the distributions of PAHs are evident. For example, the unweathered crude oil is enriched in lower molecular weight, 2-ring PAHs (C_0-C_4 naphthalenes), which are depleted in the weathered crude oil. However, overall, the C_0-C_4 alkyl series in each oil exhibits a bell-shaped profile that is characteristic of petrogenic (i.e., petroleum-derived) PAHs. This profile is predictably altered in the weathered crude oil due to the effects of evaporation, biodegradation, and solubilization following a release into the environment (see Section A.3.4).

It is worth noting that PAHs comprise only a small fraction of most crude oils and petroleum products (a typical crude oil may contain from 0.2% to more than 7% total PAHs). For example, total PAH concentrations (i.e., the sum of the 43 PAH analytes listed in Table A-2) in the petroleum products shown in Figure A-3 range from only 1.3 to 2.4 wt% of total petroleum. This range is markedly lower than the level observed in the creosote profile shown in Figure A-4, at 14.2 wt%. Thus, it requires a significantly higher mass of petroleum to introduce the same PAH mass into the environment than a creosote release does.

A.3.2 Typical Pyrogenic PAH Source Signatures

Pyrogenic PAH source signatures are complex, and, unlike the signatures in petroleum, are dominated by higher molecular weight (4-, 5-, and 6-ring) PAHs. Pyrogenic PAH assemblages are characterized by a dominance of the unalkylated (parent) PAHs, and a decreasing abundance of PAHs with increasing degree of alkylation, thereby exhibiting a sloped profile on PAH histograms (Figure A-2). These PAHs may be released to the environment in vapor phases, as airborne particles, or in the solid byproducts of the heating process. Fossil fuel combustion, particularly the combustion of petroleum in gasoline and diesel engines, is an important and prevalent source of vapor and particulate pyrogenic PAHs. Vapor-phase sources of PAHs are not a particular concern to urban sediments; particulate pyrogenic PAHs are discussed in Section A.3.3.

Pyrogenic PAHs are produced by the incomplete combustion (O_2 is present) or pyrolysis (O_2 is absent) of organic matter. A commonly encountered anthropogenic source of pyrogenic PAHs in sediments are the byproducts of the carbonization (coking) processes associated with historic manufactured gas production (MGP). These processes yielded coal- and petroleum-derived liquid tar residues (coal tar and petroleum tar) that were produced in the course of heating coal or crude oil for gas production (Gas Research Institute, 1987). In the Pacific Northwest, gas production from crude oil was very common (more so than from coal) due to the ready supply of west coast (mostly California) crude oil. These tarry byproducts of the carbonization process once were discarded, but later often were distilled into useful products. For example, creosote is a tar distillation product used to preserve wood (e.g., the wooden pilings found in many urban waterways). The residues from tar distillation (e.g., pitch) were enriched in pyrogenic PAHs and were useful in the smelting of aluminum ore (e.g., Naes and Oug, 1998).

Figure A-4 shows typical PAH profiles for three pyrogenic materials: a typical unweathered coal tar, creosote, and coal tar pitch. As expected, each of these pyrogenic PAH source materials are enriched in higher molecular weight PAHs, include several 5- and 6-ring PAHs. Pyrogenic materials contain higher concentrations of PAHs than do petrogenic materials (i.e., petroleum products). The coal tar, creosote, and coal tar pitch shown in Figure A-4 contain 103,000, 142,000 and 141,000 mg/kg of total PAHs (i.e., 10.3-14.2 wt%), all of which are much higher than in most petrogenic source materials, as shown in Figure A-3. This indicates that even small quantities of pyrogenic materials entering an urban waterway could contribute significant quantities of PAHs to sediments.

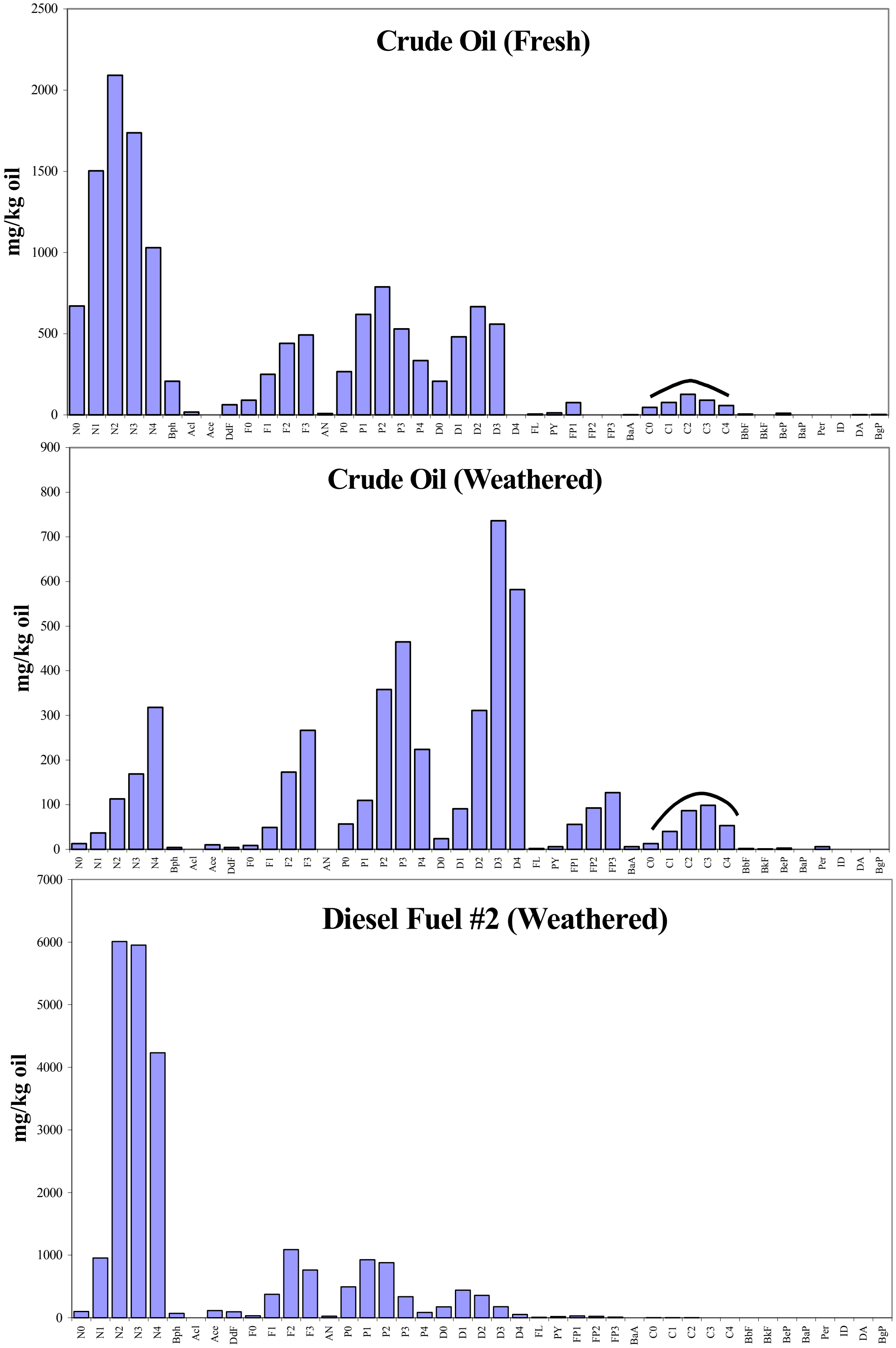
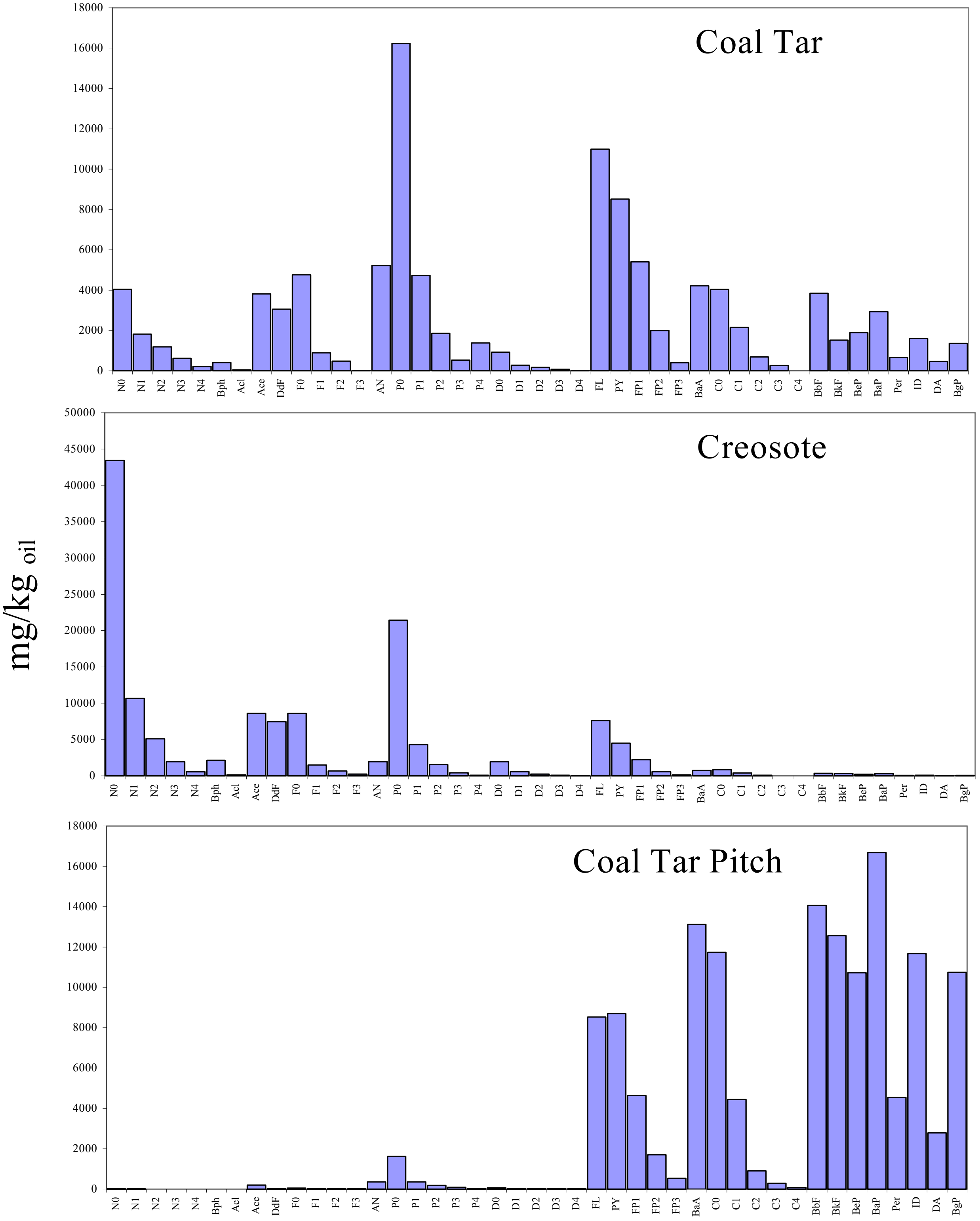


FIGURE A-3. Histograms of selected petrogenic PAH sources. For PAH compound abbreviations, see Table A-2.



Coal Tar

Creosote

Coal Tar Pitch

Figure A-4. Histograms of selected pyrogenic sources. For PAH compound abbreviations, see Table A-2.

A.3.3 Typical Urban Background PAH Signatures

General Characteristics

Stormwater runoff is probably the largest chronic contributor of background PAHs to urban sediments. During storm events, rainfall washes countless small nonpoint PAH (and other pollutant) sources from the entire catchment area and discharges this pollution at point sources (i.e., stormwater outfalls) along urban waterways. These sources have impacted waterways for decades, and are often significant sources of the background PAHs detected in urban sediments.

Discharged stormwater often contains the following types of PAHs:

- ❑ Urban dust/soot particles containing combustion-related (i.e., pyrogenic) PAHs, principally derived from incomplete combustion within automobile and truck engines, especially diesel-based engines,
- ❑ Used lubricating oils (i.e., petrogenic PAHs), principally from oil drippings from automobiles and trucks onto roadways and parking lots, and
- ❑ Waste oil and petroleum products (i.e., petrogenic PAHs) that are illegally or unintentionally discharged into a city's storm drain systems.

Many of these particles settle out of the water column and enter the sediment column. Because of the variable effects of dilution and transport processes, total PAH concentrations in urban sediments near stormwater outfalls cover very broad ranges and are highly site-dependent; however, they are typically in the 1-50 µg/g (dry) range. (Note this total concentration is the sum of all 43 PAH analytes listed in Table A-2).

Although the character of PAHs within urban runoff varies between different catchment areas and at different times, the overall PAH signatures of urban runoff and the receiving urban sediments typically are dominated by pyrogenic PAH assemblages (see Figure A-5). This probably results from a greater mass of pyrogenic PAH-laden combustion-derived particles that actually enter the sediments (as compared to petrogenic PAH-laden oils that disperse as sheens).

Also, 2- and 3-ring PAHs (i.e., those PAHs more likely associated with petrogenic sources) are more water soluble and degradable than higher-ring PAHs. Thus, although there is some degree of mixing of petrogenic and pyrogenic source materials in urban runoff (e.g., oils and combustion particles, respectively), urban runoff is predominantly a source of pyrogenic PAHs.

Characteristics of Urban Background PAHs

A typical PAH profile for sediments impacted by urban background sources is shown in Figure A-5. (The total concentration of the PAHs shown in this figure is 49.8 µg/g dry). The most abundant PAHs are high molecular weight (4- to 6-ring) compounds, particularly the fluoranthene and pyrene isomers. The fluoranthene and chrysene homologue series each exhibit the sloped pattern characteristic of pyrogenic sources (see Figure A-2).

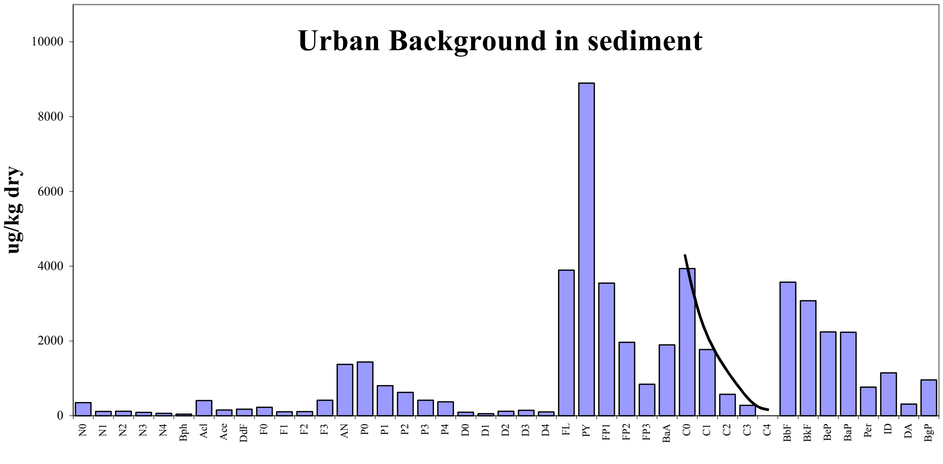


FIGURE A-5. Histogram of typical urban background PAHs in sediment. For PAH compound abbreviations, see Table A-2.

Very few lower molecular weight (2- and 3-ring) PAHs are present in the PAH fingerprint; the most abundant of these are the anthracene and phenanthrene isomers. This homologue series also exhibits the characteristic sloped pattern. These features generally are distinct from those of other pyrogenic PAH sources (e.g., Figure A-4), thereby allowing for the recognition of urban background. However, weathering can cause some pyrogenic PAH source signatures to resemble that of urban background (see Section A.3.4).

Characteristics of Urban Background Total Extractable Hydrocarbons

Sediments containing urban runoff generally exhibit a characteristic GC/FID fingerprint, one dominated by discrete resolved 3- to 5-ring PAH peaks and an unresolved complex mixture (or UCM) hump in the baseline. An example GC/FID fingerprint from PAHs in a Pacific Northwest urban sediment is shown in Figure A-6. The UCM hump is characteristic of a (mostly) residual range petroleum, such as lubricating or hydraulic oil(s), which are expected to occur in urban runoff (Gogou et al., 2000). The hump results from the thousands of compounds that are present but cannot be separated by conventional chromatographic methods. The presence of this residual petroleum in sediments is recognized by a UCM hump that ranges from about C_{17} to C_{35}, reaching a peak around C_{31}.

GC/FID fingerprints of urban runoff-impacted sediments also exhibit many discrete peaks, which are recognized as various nonalkylated 3-, 4-, and 5-ring PAHs, including phenanthrene (P0), anthracene (AN), fluoranthene (FL), pyrene (PY), chrysene (C0), benzo[*b,j,k*] fluoranthene (BbjkF), and indeno(1,2,3-c,d)pyrene (ID) (Figure A-6). The prominence of 3-, 4- and 5-ring, nonalkylated PAHs indicates the presence of combustion-derived particles typical of motor exhaust or wood smoke.

Also present in this example urban sediment are numerous odd-dominated normal hydrocarbons (n-C_{27}, n-C_{29}, n-C_{31}) associated with modern plant materials (e.g., plant waxes from leaf debris). These hydrocarbons are not always present due to the scarcity of vegetation in some urban areas, but are common in the Pacific Northwest (where this example sediment is from). The presence of these compounds suggests that some naturally occurring, biogenic PAHs may be present (e.g., retene).

A.3.4 Effects of Weathering on PAH Source Signatures

Weathering is the process of change that can occur after a material is released into the environment. PAH weathering occurs primarily through a combination of evaporation/volatilization, degradation by microorganisms, and dissolution into water. Two key effects on PAH signatures from weathering are common:

- ❑ Levels of low-molecular-weight (2- and 3-ring) PAHs are reduced, thereby increasing the proportion of 4- to 6-ring PAHs; and
- ❑ Levels of nonalkylated PAHs are reduced, thereby increasing the proportion of more alkylated PAHs.

The implication of these effects is that the PAH signature of a contaminant dominated by 2- and 3-ring PAHs ultimately may be converted through weathering to a PAH signature dominated by 4- to

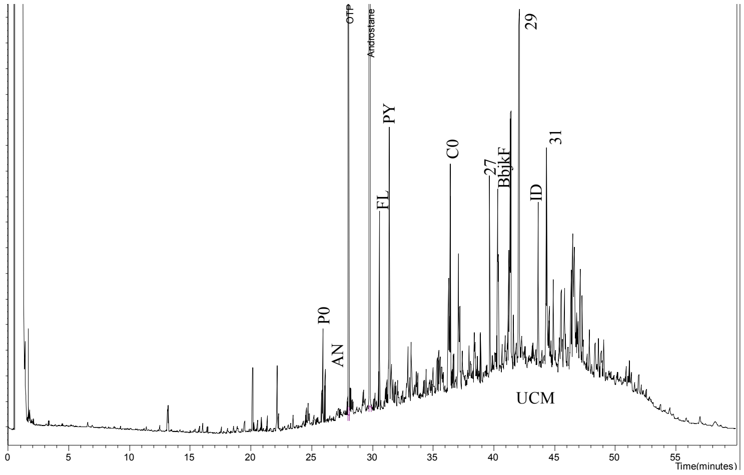


FIGURE A-6. GC/FID chromatogram of total extractable hydrocarbons for typical urban background PAHs in sediment

6-ring PAHs. And, because urban background PAHs are already dominated by 4- to 6-ring PAH, it can be difficult to distinguish between urban background and some other highly weathered pyrogenic source (e.g., severely weathered coal tar). Fortunately, the use of other methods of data analysis (e.g., GC/FID fingerprints) can help reduce these complications.

A.4 Methods for Recognizing and Allocating PAH Sources in Urban Sediments

More often than not in urban environments, no single anthropogenic source accounts for all of the PAHs contributed to the local sediment throughout its history. Instead, petrogenic and pyrogenic PAHs from discrete point sources are mixed with PAHs derived from nonpoint urban background sources. Defining the contribution of PAHs from urban background sources is the objective of most PAH source studies.

There is no single method for unraveling PAH mixtures, as no two datasets or study areas are the same. Instead, recognizing and unmixing PAH source signatures in sediments is best achieved through careful analysis. This section describes the following four analytical approaches:

1. Qualitative "pattern recognition" that includes visual inspection of the available GC/FID "fingerprints" of the total extractable hydrocarbons and PAH histograms, including comparison to known standards in a sample library.
2. Semi-quantitative graphical analysis of PAH concentration data using source-specific diagnostic ratios or indices (e.g., cross-plots or ternary diagrams), again, sometimes including comparisons to known standards.
3. Spatial and temporal analysis of PAHs (and other) signatures and concentrations, including a comparison to historical information compiled for the area (as part of the conceptual site model; see Section A.2.1 above).
4. Quantitative chemometric analysis involving numerical analysis methods such as principle component analysis (PCA).

A.4.1 Qualitative "Pattern Recognition" Approach

The first step in evaluating data from ACF of sediments is to review the available data. Comparing samples by visual inspection of concentrations in data tables is difficult. Therefore, this initial review of data is best achieved by (1) qualitative comparison of the available GC/FID "fingerprints" of the total extractable hydrocarbon in the sediments and (2) qualitative comparison of the PAH histograms generated from the concentration data. A longer discussion of histogram construction is provided in Section 2.2.3 and in Boxes 2-11 and 2-12 of the main document.

There is great value in simply visually examining the GC/FID "fingerprints" and PAH histograms acquired in the course of ACF (Section A.2.3). A qualitative assessment of the "fingerprints" of the total extractable fraction of sediments can reveal the presence of various petroleum fractions or other contaminants, or simply the characteristic urban background features (as described in Section A.3.3). Figure A-7 shows GC/FID fingerprints of several PAH source materials commonly found in urban sediments. The visual differences among these chromatograms (and for other PAH sources not shown) provide a good indication of the variety of PAH sources that may be present at a given sediment site. (Notably, GC/FID chromatographic data is usually not provided by the laboratory and must be requested.) An assessment of the character of the total extractable hydrocarbons can be extremely useful in any subsequent interpretation of the PAH concentration data associated with a particular sample.

Also, PAH histograms that include the alkylated PAHs are useful for identifying the presence of petrogenic (bell-shaped) or pyrogenic (sloped) homologue patterns (see Sections A.3.1 and A.3.2). These graphs can provide a quick assessment of PAH source category, which in some instances may be sufficient to rule out the presence or influence of urban background: if only petrogenic PAH profiles are evident. Examples of the visual differences between different PAH source materials are shown in Figures A-3, A-4, and A-5.

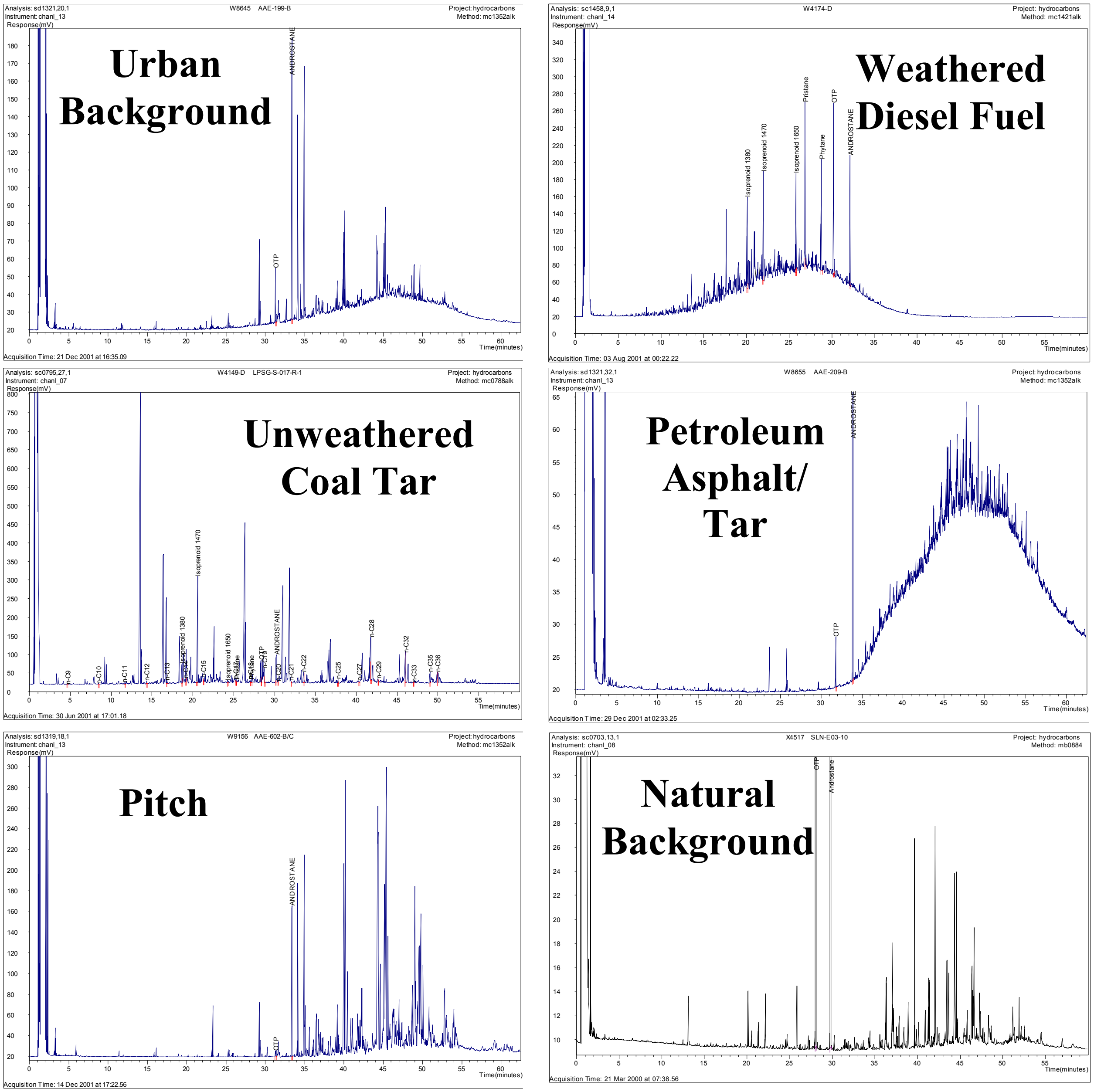


FIGURE A-7. GC/FID "fingerprints" for several PAH sources in sediments

Although experience in viewing chromatograms or histograms can provide significant detail, even an inexperienced interpreter can visually compare the GC/FID "fingerprints" or PAH histograms of different samples and assess any likely relationships between the contaminants in sediments with similar appearances. Also, the influences of weathering on the distributions must always be considered (Section A.3.4). "Odd-looking" samples will be quickly identified based upon their unique "fingerprints" or PAH histograms. And, hopefully, there will be some agreement between the classifications based on both the total extractable hydrocarbon signatures and on PAH distribution graphs.

Although a full review of the qualitative "pattern recognition" approach to PAH source identification is beyond the scope of this appendix, the reader is directed to Stout et al. (2002a) and references therein for a review of the features of different PAHs and other hydrocarbon sources.

A.4.2 Quantitative Approaches

Use of RSC Screening Data

RSC screening data can be very useful when evaluating the general condition of sediments at a given site. Spatial analysis of RSC PAH concentration data (described in Section A.4.3) can reveal gradients or hotspots, which can point toward or indicate potential PAH sources. RSC data also can help identify urban background PAH signatures can prove valuable through an evaluation of the population distribution. In a large enough population of sediments from a particular site, most samples containing PAHs derived only from urban background sources likely would correspond to the lower end of the distribution, whereas samples that represent point sources likely would correspond to the higher end of the distribution (Figure A-8).

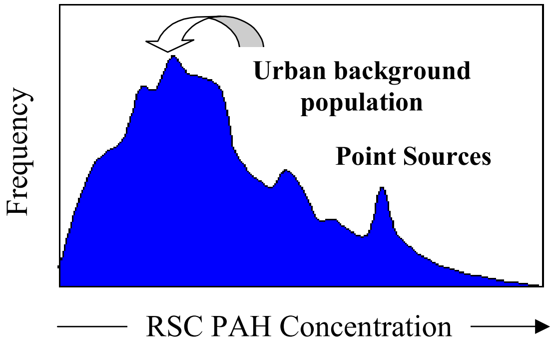


FIGURE A-8. Stylized distribution of RSC PAH concentrations revealing likely urban background population and point sources

Another use of the RSC data is to construct various cross-plots (or scatter plots) of PAH concentrations plotted against other types of screening data. For example, chemical variations often follow changes in sediment texture, so cross-plotting physical property data (e.g., grain size or total organic carbon) can be useful in evaluating background conditions. In many sediment studies, RSC screening for metals (e.g., using XRF) is conducted simultaneously with RSC of PAHs. The metals data act as a proxy for grain size because iron (and aluminum) concentrations are expected to be highest in fine-grained sediments where clays are more abundant.

Figure A-9 shows RSC PAH concentrations (y-axis) plotted against RSC %Fe (x-axis) for nearly 200 sediments from the Elizabeth River, Norfolk, VA. Many of the sediments fall along a single ambient trend line that is believed to represent the background conditions for sediments in the study area. The PAH concentrations in these sediments are generally low, but increase with the %Fe (a proxy for the percentage of fine-grained sediments). Samples that plot above the ambient trend line suggest the presence of hotspots (potential point sources of PAH). This approach allows for a quick assessment of the likely influence of urban background using only the RSC data. See Section 3.2 of the main document for detailed discussions of the use of scatter plots for background analysis.

Use of PAH Concentration Data

PAH histograms are only one of several methods for visually evaluating PAH concentration data. Numerous standard graphing techniques (e.g., population boxplots and probability plots) can aid in PAH data analysis, and should be performed during initial data exploration (see Section 2.2 of the main document). For example, boxplots of various parameters (concentrations or PAH ratios) can be used to identify potential outliers and population quartiles. "High" outliers are likely to represent point sources of PAHs. If a sufficient number of sediments containing only background PAHs is included in the sample set, then the "low" outliers or quartile are likely to represent urban background. Detailed discussions of graphic techniques are provided in Section 2.2.3 of the main document.

Numerous ratios among PAH analytes can be cross-plotted to reveal outliers or clusters of samples that could include an urban background population. For example, because of the fundamental differences in petrogenic and pyrogenic PAHs, various ratios between a nonalkylated (i.e., parent) PAH and its alkylated equivalent(s) can be used to quickly distinguish these source categories from one another, and also reveal variability within a category (Douglas et al., 1996; Bence et al., 1996; Zeng and Vista, 1997; Stout et al., 2001b; Stout et al., in press). Although some of these ratios may be affected by preferential weathering of the parent PAH, these ratios usually are less than 1 in petrogenic PAH assemblages and greater than 1 in pyrogenic assemblages.

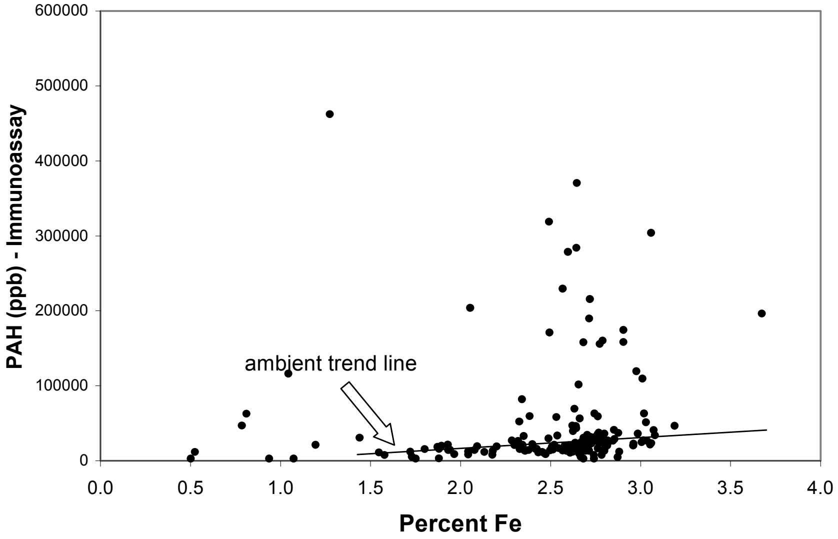


FIGURE A-9. Cross-plot of RSC PAH immunoassay data and percent iron for Elizabeth River sediments. Many sediments plot along an ambient trend line.

For example, the ratio of chrysene to the C2-chrysenes (C_0/C_2) was used in an assessment of PAHs in sediments from Thea Foss Waterway, Tacoma, WA. This ratio was selected because these 4-ring PAHs are only minimally affected by all but the most severe weathering. Figure A-10 shows that (with one exception) the sediments studied were uniformly enriched in the parent PAH ($C_0/C_2>1$), indicating the presence of overwhelmingly pyrogenic sources. (It should be noted that other chemical features also were examined before reaching this conclusion; pyrogenic PAHs should not be identified on the basis of this ratio alone.)

Other ratios that depend upon the thermal stability of related isomers also can be useful in classifying PAH sources (e.g., Mitra et al., 1999; Walker and Dickhut, 2001). For example, ratios of phenanthrene to anthracene (P0/AN) and fluoranthene to pyrene (FL/PY) are useful for differentiating between sediment PAH assemblages containing primarily pyrogenic or petrogenic PAHs (Table A-3). Anthracene and fluoranthene are thermodynamically less stable than their respective isomers, phenanthrene and pyrene (Baumard et al., 1998). Anthracene and fluoranthene are produced during rapid, high temperature pyrosynthesis, but are less favored to persist during the slow organic diagenesis leading to the generation of fossil fuels. Thus, as shown in Table A-3, the P0/AN ratios of pyrogenic PAH assemblages usually are less than 5, whereas petrogenic ratios usually are greater than 5. The FL/PY ratios usually approach or exceed 1 in pyrogenic assemblages and usually are substantially less than 1 in petrogenic PAH assemblages (Table A-3). Because of the extreme variability in PAH ratios among the different sources, and because most of the references used to assemble Table A-3 lack alkyl PAH and other chemical “fingerprinting” data, both ratios (at a minimum) should be used together when differentiating between PAH sources.

These differences were used to evaluate variations in the nature of pyrogenic-dominated PAH sources in Thea Foss Waterway sediments, as shown in Figure A-10. A cross-plot of the AN/P0 and FL/PY ratios for these sediments is shown as Figure A-11. The samples containing a greater influence of urban runoff exhibited lower FL/PY ratios and higher AN/P0 ratios than sediments containing manufactured gas plant tars.

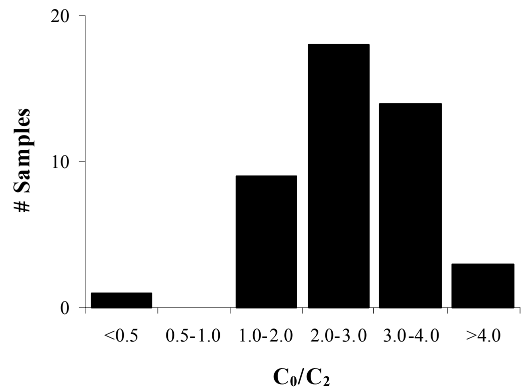


FIGURE A-10. Histogram showing the C_0/C_2 ratio for 43 Thea Foss Waterway sediments

TABLE A-3. Compilation of selected PAH isomer ratios for different PAH source materials. (Source: Neff et al., in press.)

Source	P0/AN	FL/PY	Reference
	Primarily Pyrogenic Sources		
Coke oven emissions	**1.27 – 3.57**	0.76 – 1.31	Maher and Aislabie, 1992
Iron/Steel plant (soot)	0.24	0.62	Yang et al., 2002
Iron/Steel plant (flue gas)	0.06	1.43	Yang et al., 2002
Wood-burning emissions	6.41	1.26	Page et al., 1999
Auto exhaust soot (gasoline)	1.79	0.90	O'Malley et al., 1996
Diesel engine soot	0.06	1.26	Bence et al., 1996
Highway dust	4.7	1.4	Christensen et al., 1999
Urban runoff	0.68 – 1.78	0.23 – 1.07	Stout et al., 2001a
Creosote	0.11 – 4.01	1.52 – 1.70	Neff, 2002
Coal tar	3.11	1.29	Neff, 2002
Coke	0.24	1.49	Stout (unpublished data)
	Primarily Petrogenic Sources		
Used crankcase oil (gas engine)	6.2	0.79	Pruell and Quinn, 1988
60 Crude oils (mean)	52.0	0.25	Kerr et al., 1999
Weathered bunker fuel tar	12.9 – 17.4	0.05 – 0.18	Wang et al., 1998
Diesel fuel (#2 fuel oil)	>800[(a)]	0.38	Bence et al., 1996
No 4 fuel oil	11.8	0.16	Stout (unpublished data)
Bunker C residual fuel oil	14.8	0.14	Stout (unpublished data)
Coal	4.86	0.91	Neff, 2002

(a) Anthracene concentration was below detection limit.

Similar quantitative methods can be used to distinguish among multiple petrogenic sources in sediments (Boehm et al., 1997; Stout et al., 2001c). For example, the alkyl homologue groups of the phenanthrenes and dibenzothiophenes have been demonstrated to weather at the same rates, and are fairly persistent in contaminated sediments (Douglas et al., 1996; Boehm et al., 1997). The concentrations and relative abundances of alkyl-dibenzothiophenes is a function of the amount of sulfur in the petroleum, a feature that can vary widely among different petroleum sources. Thus, alkyl-PAH ratios can be very useful for identifying PAH assemblages from different petrogenic sources in sediments. The ratios of total C$_2$-dibenzothiophenes to total C$_2$-phenanthrenes (D2/P2) and of the C$_3$-homologues (D3/P3) were particularly useful for distinguishing among sediment PAHs from North Slope crude oil (the oil released from the *Exxon Valdez*) and from other petrogenic sources (seep oil, weathered petroleum tar, diesel fuel) in spill path areas of Prince William Sound (Figure A-12). If the sediment PAH data are graphed in double-ratio plots (e.g., D2/P2 vs. D3/P3), the PAH assemblages from different petrogenic sources cluster separately, often allowing clear differentiation among multiple sources.

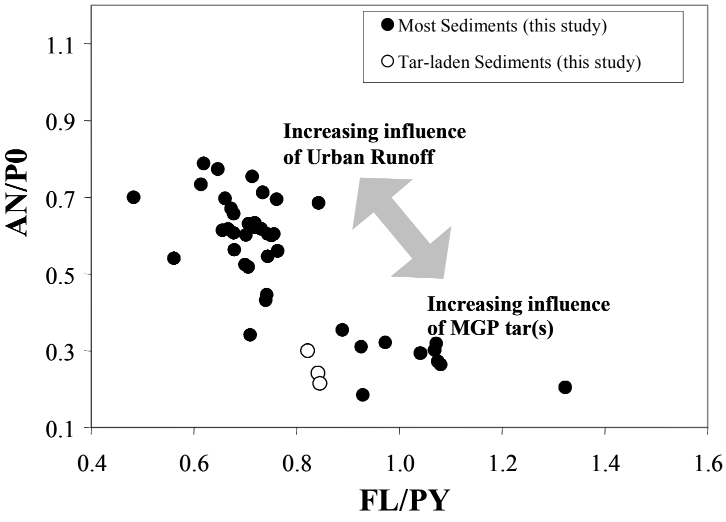


FIGURE A-11. Cross-plot of PAH isomer ratios for sediments containing multiple and mixed pyrogenic sources

In some studies it is useful to grossly quantify the proportion of PAHs derived from petrogenic and pyrogenic sources. The first attempt to quantify

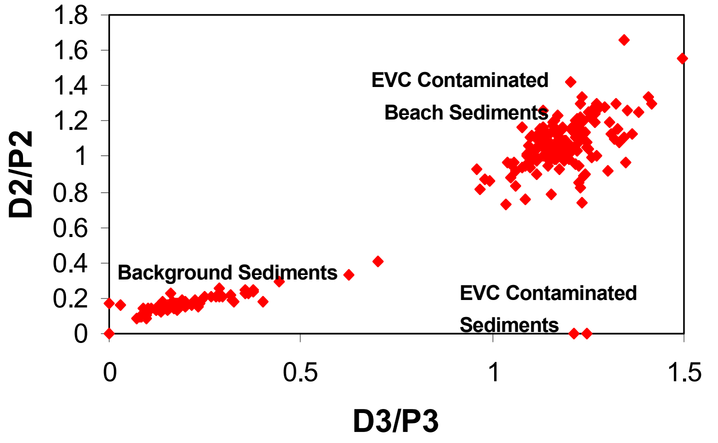


FIGURE A-12. Cross-plot of dibenzothiophenes and phenanthrene concentrations in Prince William Sound sediments. EVC – ExxonValdez cargo. (Source: Page et al., 1996.)

the proportions of pyrogenic and petrogenic PAHs in sediments relied upon a simple expression using PAH homologue concentrations (Boehm and Farrington, 1984). In this approach, each PAH analyte (Table A-2) is assigned to either a petrogenic, pyrogenic, or mixed category. These classifications are based upon the fundamental features of petrogenic and pyrogenic source materials and the expected weathering they endure upon release into the environment. For instance, most low molecular weight (2- and 3-ring) and alkylated PAHs are assigned to petrogenic source materials, whereas most high molecular weight (4- to 6-ring) and nonalkylated parent PAHs are assigned to pyrogenic sources. Clearly, some coal-derived liquids, such as coal tar and creosote, are exceptions to this generalization; Figure A-4). Some PAHs (e.g., P0 and P1) originate from both petrogenic and pyrogenic sources, and therefore are considered to have mixed sources.

An example of this simple PAH classification and allocation technique is demonstrated with a sample of sediments impacted by urban background sources (Figure A-13). The calculations reveal that approximately 26% of the PAHs detected in this sediment sample are derived from petrogenic sources and 74% from pyrogenic sources (i.e., 4.8 and 13.6 mg/kg, respectively). This proportion is typical of urban background conditions in many areas.

The utility of this simple approach for background analysis is that when a large population of sediments is examined, sediments containing only PAHs from urban background sources can be distinguished from sediments impacted by a site-related release. Deviations from the study area's

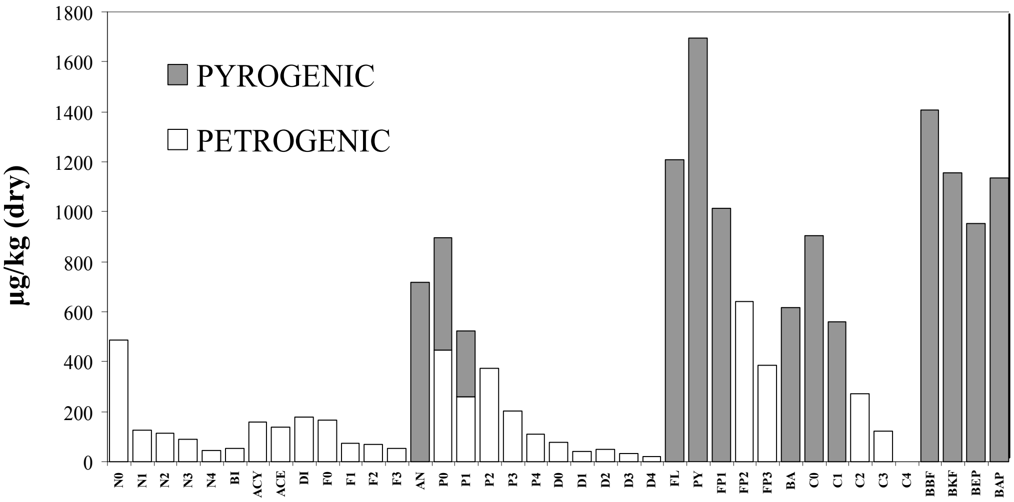


FIGURE A-13. Simplified classification of PAHs in a sediment impacted with urban runoff

established proportion of pyrogenic and petrogenic PAHs could be used to recognize a contribution from any additional pyrogenic or petrogenic PAH sources. A large number of sediment samples analyzed in the same manner can be used to map the distribution of petrogenic and pyrogenic PAHs in sediments. An “excess” of the petrogenic or pyrogenic fraction at a particular location may indicate the presence of a point source of PAHs superimposed on the urban background.

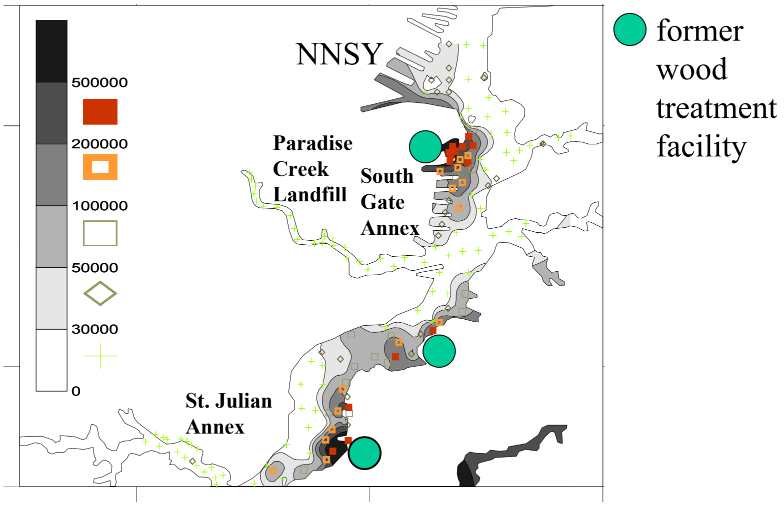


FIGURE A-14. Map of RSC PAH data for Elizabeth River sediments. Three hotspots correspond to locations of former wood treatment facilities.

A.4.3 Spatial and Temporal Trend Analysis

Spatial Analysis of PAH Data

Plotting and mapping PAH data (from both RSC and SCF) can reveal the general concentrations expected for sediments containing PAHs from urban background. More importantly, these techniques can help identify PAH concentration gradients or hotspots, either of which could suggest the location of possible point sources of PAHs and establish the overall background conditions. The identification of possible point source locations can be bolstered if the locations happen to correspond to existing or former facilities at which PAH-laden materials were handled or produced (see examples below). Conversely, if spatial displays of PAH data do not reveal hotspots or concentration gradients associated with a suspected point source, then the contribution of the suspected source should be re-evaluated.

Spatial analysis of RSC PAH data can provide a technical basis for selecting representative sediment samples for further ACF analysis that occur along a given trend or within a given hotspot for ACF. If data density and structure allow, the RSC data may even be contoured to interpolate concentrations across the study area. Figure A-14 shows the RSC immunoassay screening results for nearly 200 sediment samples collected from the Elizabeth River, VA. (These data were generated as part of the PAH source study surrounding the Norfolk Naval Shipyard).[2] The screening data showed the existence of three PAH hotspots which, based on the CSM of the study area, were determined to be associated with former wood treatment facilities. PAH concentrations were shown to decrease quickly to background levels with distance away from these facilities (Figure A-9).

Spatial analysis of ACF PAH also can be useful, especially for examining total PAH concentrations or identifying gross percentages of petrogenic and or pyrogenic PAHs within an urban waterway (see Figure A-13). Specifically, various PAH ratios can be plotted to evaluate spatial trends (Section A.4.2), sometimes without requiring GIS-based maps. Figure A-15 shows a plot of the PAH ratios AN/P0 and FL/PY for Thea Foss Waterway sediments (same data as shown in Figure A-11). The FL/PY and AN/P0 ratios for each sediment sample are plotted against distance from the head of the waterway (as determined by latitude in this north-south oriented waterway). The plot reveals an overall shift in these ratios with distance from the head of the waterway, which corresponds to an increasing influence of urban background toward the mouth of the waterway (Stout et al., in press).

---

[2] *A User’s Guide for Optimizing the Determination of Contaminant Sources in Sediments Near Naval Facilities*. Contact: Stacey Curtis, SPAWAR Systems Center, San Diego, CA.

The greater influence of MGP tars near the head of the waterway is consistent with the suspected source of PAH-laden tars, a MGP facility formerly located at the head of the waterway.

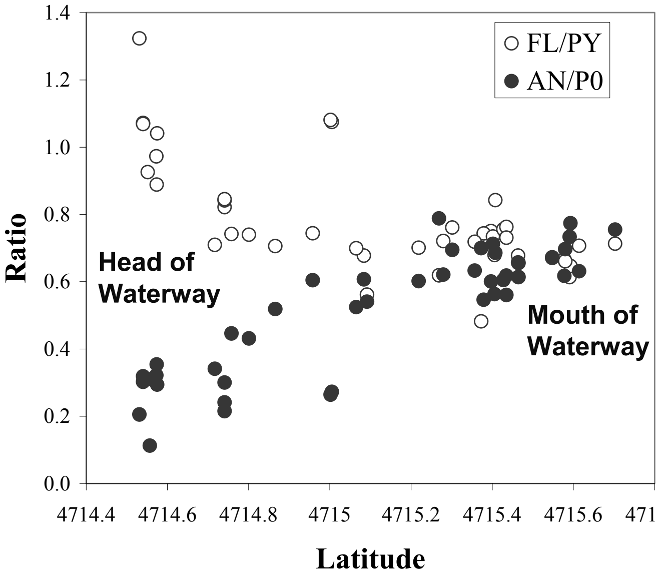


FIGURE A-15. Plot showing the spatial trends in selected PAH ratios for Thea Foss Waterway sediments (Source: Stout et al., in press.)

Temporal Analysis of Sediment Cores

In waterways where sediments are deposited regularly, an understanding of the temporal changes in PAH sources can be obtained through the study of PAH profiles in age-dated sediment cores (i.e., the geochronology of the sediments). When the ACF and geochronology data are correlated with the CSM of a given sediment site, it is possible to characterize the urban background-derived PAH component.

Radiochemical techniques, using ^{210}Pb or ^{137}Cs isotopes, are the most common method for age-dating sediment cores over years or decades (e.g., Lavelle et al., 1986; Bloom and Crecelius, 1987). The isotope ^{210}Pb is a decay product of the uranium (^{238}U) decay series, and has a half-life of 22.3 years. The isotope ^{137}Cs arises primarily from atmospheric fallout from nuclear weapons tested between 1957 and 1965. ^{137}Cs can be used to calibrate ^{210}Pb measurements to the historical onset of nuclear weapons testing (c. 1955) because ^{137}Cs does not have a constant historical or natural flux into sediments. This type of radiochemical dating relies on two simplifying assumptions to support the requirement of a constant ^{210}Pb flux into the sediments: (1) that the sediments have a relatively uniform grain distribution with depth, and (2) that the sediments have a relatively constant historical deposition rate.

Sediment geochronology recently was used in a study of sediments in Eagle Harbor, WA (Brenner et al., 2002). When geochronology data were combined with ACF data, it became clear that urban background PAHs were pervasive in the sediments from the central part of the harbor deposited over the last 50 years (Stout et al., 2001b). Figure A-16 shows the PAH concentration profile for an age-dated core from the central Harbor. Analysis of the GC-FID fingerprints for sediments within the core revealed a transition from naturally occurring background PAHs in the sediments deposited prior to the 1930s to increasing concentrations of PAHs attributable to urban background in the sediments deposited after urbanization in the 1940s. Understanding the time period when these different sediments had been deposited was helpful in recognizing the sources of the PAHs in the sediments and distinguishing the urban background component.

A.4.4 Numerical Analysis Approach Using Factor Analysis

Factor analysis, also known as principal component analysis (PCA), is one of several types of ordination techniques by which multivariate data sets are explored, reduced, interpreted and/or studied further (see Section 2.3 of the main document for additional discussion). Factor analysis generates new independent variables (i.e., factors) that are linear combinations of the original input variables (e.g., PAH concentrations). The method reduces the dimensionality of the data to a few important "principal components" (PCs) or axes that best describe variations in the data. The first axis (1st PC) demonstrates the most prominent trend, and successive axes (2nd PC, 3rd PC, etc.) demonstrate additional trends in decreasing order of importance.

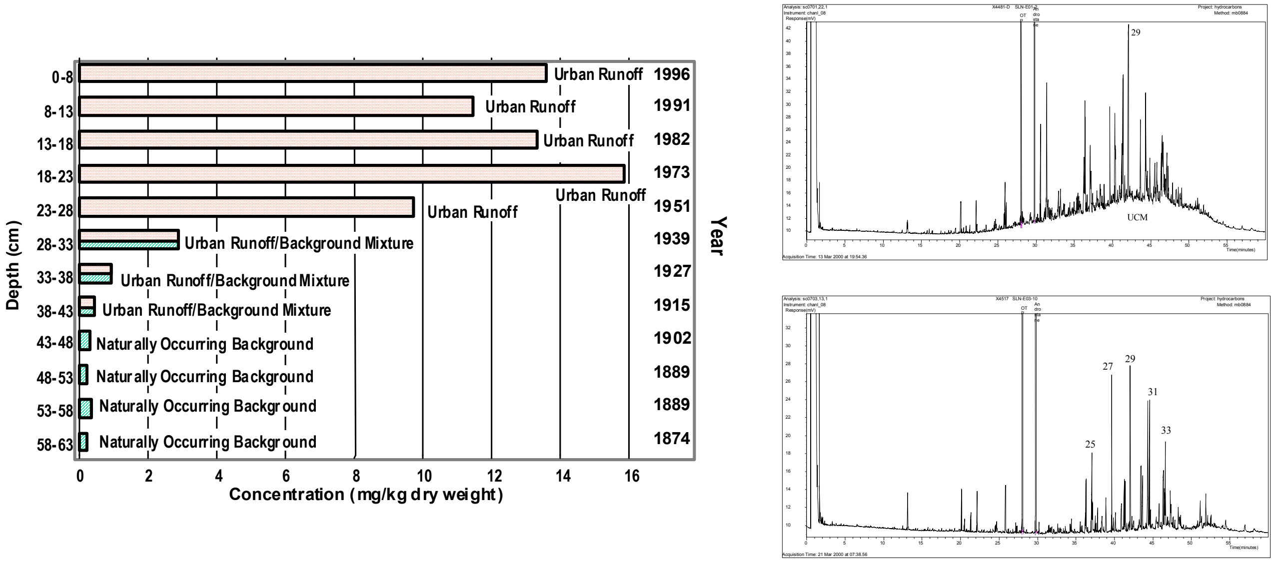


FIGURE A-16. PAH concentration profile for an age-dated core from Eagle Harbor. Corresponding GC-FID fingerprints for urban background (top) and naturally-occurring background (bottom) are shown. (Source: Stout et al., 2001b.)

One significant advantage of factor analysis methods is that they provide a strictly mathematical means of analyzing PAH concentration data, thus removing any biases of the interpreter, which could creep into the other techniques described in Sections A.4.1 to A.4.3. Factor analysis has the additional advantage of being able to convey the complex chemical differences among many samples with many individual PAH measurements in a clear, visual manner.

Factor analysis has been used in many types of studies and has been applied to PAH fingerprinting and allocation studies (Bence and Burns, 1995; Boehm et al., 1997; Naes and Oug, 1997; Pena-Mendez et al., 1999; Maxon et al., 1997; Burns et al., 1997; Mitra and Wilson, 1992). Excellent summaries of factor analysis and other numerical methods commonly employed can be found in Johnson and Ehrlich (2002). This section outlines a basic factor analysis approach, from required data input to analysis of data output.

Input Data Requirements and Transformations

The raw input to factor analysis is most often the PAH concentration data for a set of sediment samples. Factor analysis has proven to be an especially effective means of comparing the PAH concentration data from a large number of sediment samples. If only a small number of samples are available (<10), most comparisons can be made without the use of factor analysis. Factor analysis can be performed on the sediment samples alone, or it can include various training data sets that include the PAH concentrations in known source materials.

Most PAH datasets for sediments have concentrations that vary widely between a highly contaminated samples (e.g., those nearer a source) and diluted, but still contaminated samples (e.g., those removed from the source or representing background). Because of this variability, the raw concentration data are normally pre-processed in some manner prior to conducting the analysis. Preprocessing can be achieved through a variety of transformations; however, normalizing the raw PAH concentration data (in order to remove the effect of widely varying concentrations between samples and between individual analytes) is most common. Normalization is performed for each sample (i.e., constant row-sum) and for each PAH analyte (i.e., variable normalization). Because of these normalization steps, the effects of absolute PAH concentration are removed and all PAH analytes are given equal weight in the factor analysis. As a result of the normalization steps, factor analysis conducts the comparison between the samples based upon the relative proportions of the PAH analytes, rather than their absolute concentrations, and prevents high concentration samples or variables from dominating the analysis (Johnson and Ehrlich, 2002).

Factor Analysis Output and Interpretation

Factor analysis yields a distribution of data in n-dimensional space, where n is the number of variables (e.g., PAH analytes). The 1st PC is a line through this space upon which each sample point can be projected. The line's orientation is such that the variance of these projections is maximized. The 2nd PC is another line defining the next highest variance. These first two lines (i.e., the 1st and 2nd PC) define a plane. These planes are called **factor score plots**, which are one end product of factor analysis. The distances between sample points on these plots represent the variances captured in each PC; in other words, samples that cluster together are chemically similar and outliers are chemically dissimilar.

Figure A-17 shows an example of a factor score plot for nearly 100 sediments from Eagle Harbor, WA. (Data from this same site were shown in Figure A-16). The factor score plot indicates that three sources of PAHs were recognized: natural background (arising from pre-industrial, natural forest fires), urban runoff, and creosote (from a former coal tar distillation facility on the waterway). Many sediment samples from this urban waterway contained only (or primarily only) one of these three end-members. These "single-source" samples tend to plot as clusters at or near the apices of the trends revealed by the factor score plot (Figure A-17; top). Also, although not performed as part of the study, additional calculations involving spatial distributions, concentrations, and volumes of impacted sediments of each sediment sample in Eagle Harbor then could be used to allocate responsibility among the three end-member sources.

Many other sediments plot in locations intermediate between the three end-members, suggesting they contain PAH mixtures. The spatial relationships among samples on a factor score plot can be used to estimate the proportions of each end-member in each sediment sample. Of course, the effects of total PAH concentration must be accounted for. For instance, the normalized distributions of the PAHs suggest a mixture of creosote and urban background in many samples, whereas the significantly higher concentrations of PAHs in creosote (over urban background; see Sections A.3.2 and A.3.3) must be accounted for in any allocation model.

Another output of factor analysis is a **factor loading plot**. Factor loadings are measures of the relative contribution of every variable (e.g., PAH analyte) to a factor axis. To construct a factor loading plot, one factor loading is calculated for each variable (e.g., PAH analyte) contributing to each PC. A cross-plot of the factor loadings for the first few PCs reveals the individual PAHs responsible for the variance in each PC. The factor loading plots can be used to interpret the variables responsible for the clustering/separations observed on factor score plots. The corresponding factor loading plot for the first two PCs in the Eagle Harbor dataset is shown in Figure A-17 (bottom). This shows that, as expected, the urban runoff-dominated sediments were enriched in 4- to 6-ring PAHs (see Section A.3.3), and that the creosote-dominated sediments were enriched in 2- and 3-ring PAH, which are characteristic of this material (see Figure A-4).

Results of the factor analysis output were interpreted in conjunction with all available data, including qualitative interpretations of the GC/FID "fingerprints" and the PAH histograms (as described in Section A.4.1). Figure A-18 shows some representative GC/FID "fingerprints," as well as PAH histograms constructed for the three end-members identified in the PCA (Figure A-17, top). These figures indicate that the end-members exhibit qualitative differences in their fingerprints (GC/FID and PAH), which were quantitatively substantiated by the factor analysis. In this study, the combination of these two ACF techniques, interpreted in light of known product types and weathering, were able to identify the sources that were represented (i.e., natural background, urban runoff, and creosote; Stout et al., 2001b; Brenner et al., 2002).

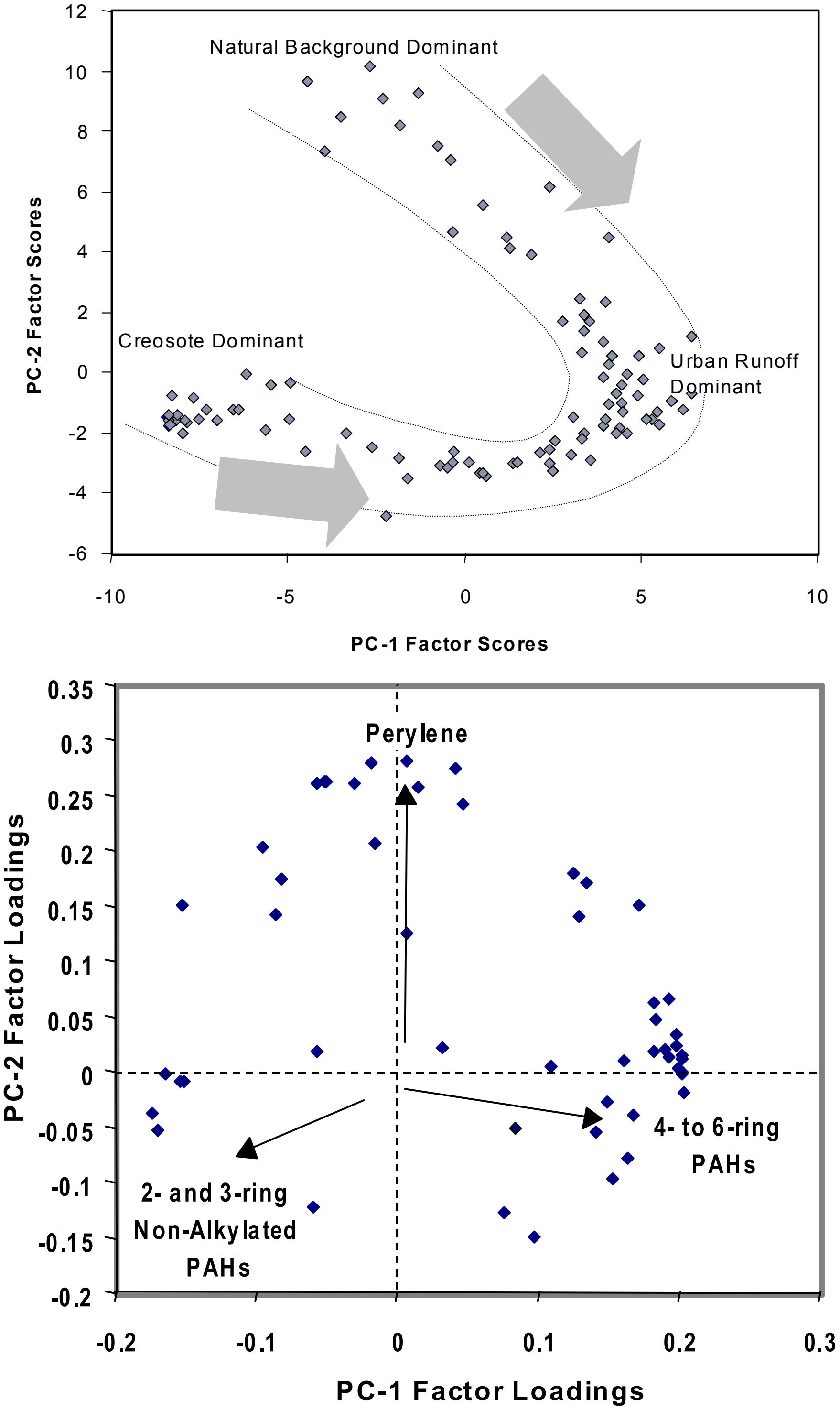


FIGURE A-17. PCA factor score plot (top) and corresponding factor loading plot (bottom) for sediment PAH data from Eagle Harbor. Gray arrows indicate increasing influence of urban background. (Source: Stout et al., 2001b.)

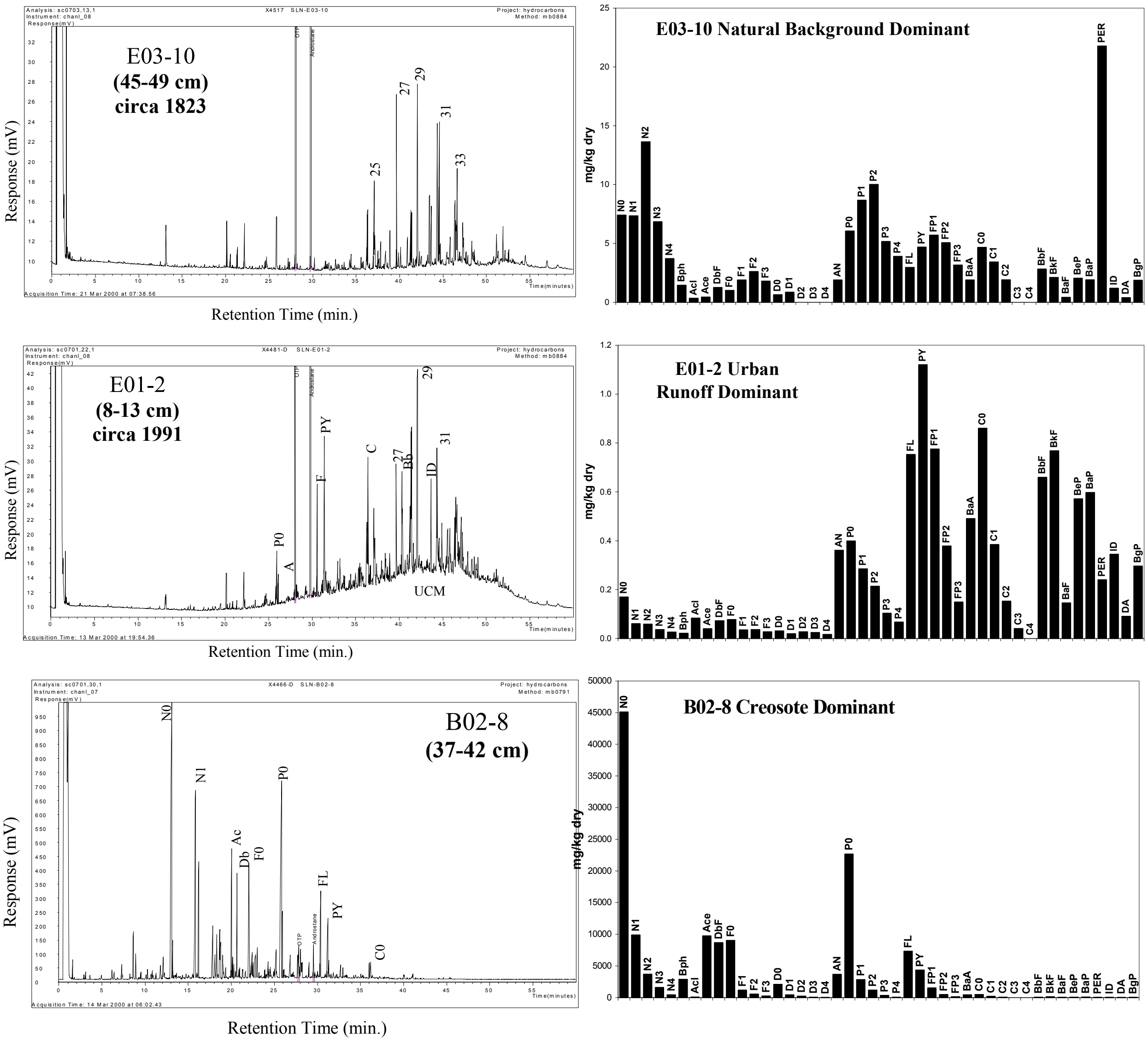


FIGURE A-18. GC/FID "fingerprints" (left) and PAH histograms (right) produced from ACF of sediments identified as "end-members" in the PCA analysis shown in Figure A-17. Top: Sediment containing naturally occurring PAH; middle: sediment containing PAHs derived from urban runoff; and bottom: sediment containing unweathered creosote. (Source: Stout et al., 2001b.)

Factor analysis sometimes can benefit from including candidate source (e.g., upland) materials along with the sediment samples. At some sediment sites where property access is not problematic, it may be reasonable to include candidate source soil samples for suspected source properties for comparison to the sediments. If candidate source materials are available, their PAH distributions can be directly compared to the PAH distributions in target sediments, particularly those proximal to the suspect property(s) in question.

In the absence of upland samples from the study area, the comparable data from "library" standards can be included along with sediment data in order to help interpret the PCA output. Figure A-19 shows an example of this approach, for PAH data from Thea Foss Waterway. In this analysis, PAH

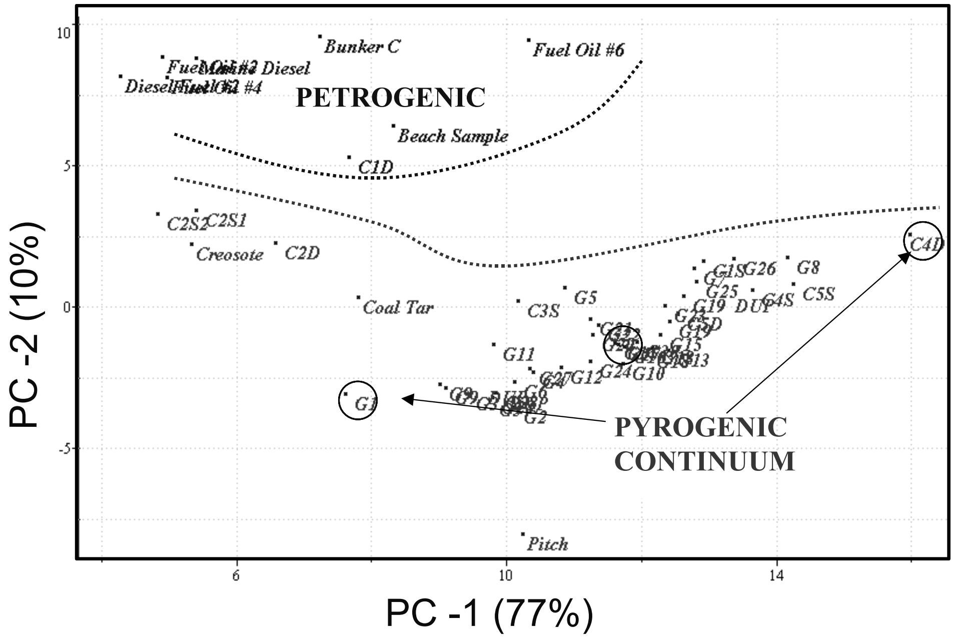


FIGURE A-19. PCA factor score plot for Thea Foss Waterway sediments and relevant library standards (i.e., selected petroleum fuels, creosote, coal tar, and pitch)

data from Battelle's sample library for various petroleum fuels (fuel oil #6, bunker C, and various distillate fuels), creosote, coal tar, and pitch were included along with 43 sediment samples. Factor analysis indicated that only two sediments plotted near the petroleum fuel standards (top left), indicating that these two sediments contain a predominantly petrogenic PAH component. Similarly, a few samples plotted near the creosote standard, indicating that they contain a predominant pyrogenic PAH component consistent with creosote. However, most of the sediments plotted along a continuum that was revealed by the factor loadings plot (not shown) to be dominated by pyrogenic PAHs. The PAH distributions in the sediments along this pyrogenic continuum was captured by investigating the PAH distributions in three circled samples shown in Figure A-19. The PAH histograms for these three samples are shown in Figure A-20.

These histograms indicate that all 43 sediment samples demonstrate an overall pyrogenic PAH character. For example, each has (1) a greater relative abundance of parent PAHs than alkylated PAHs (giving rise to the skewed patterns for each PAH homologue series); (2) an enrichment of fluoranthene (FL), pyrene (PY) and anthracene (AN); and (3) a predominance of 5- and 6-ring PAHs. Some of the differences between these end-members are consistent with changes due to weathering; in particular, the relative loss of phenanthrene (P0) and chrysene (C0) compared to their alkylated equivalents (P1-P4 and C1-C4) in the Core 4 deep sample could be caused by preferential weathering of the nonalkylated parent PAHs (see Section A.3.4). As a result, the slope of the degree of skewness of the Core 4 deep sample's C_0-C_4 phenanthrene and chrysene distributions is reduced (Figure A-20). Alternatively, the continuum also could indicate the presence of a mixture of more than one type of pyrogenic source material. In this study, the presence of mixed PAH signatures arising from coal tar and urban background was identified as the most likely cause for the observed continuum (see Figure A-11; Stout et al., in press).

The two examples provided above demonstrate that factor analysis output should be interpreted with caution, and that many of the techniques discussed in Sections A.4.1 to A.4.3 should be incorporated during the interpretation.

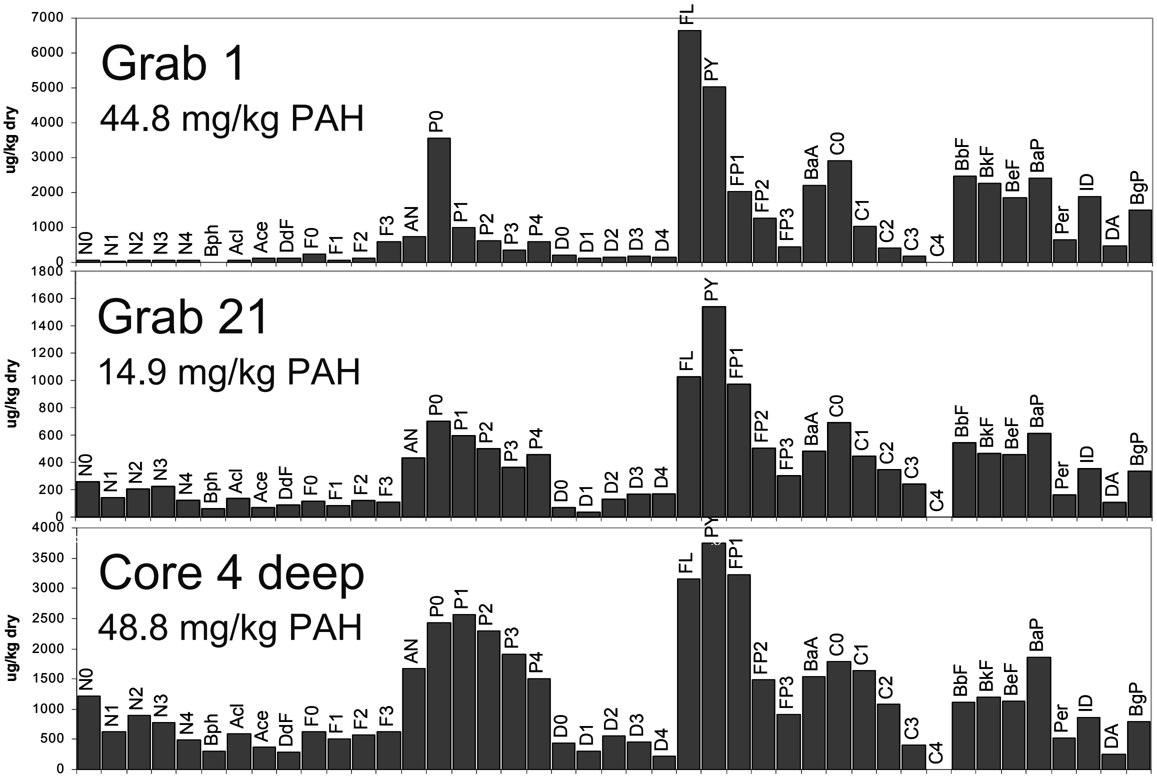


FIGURE A-20. Histograms of selected Thea Foss Waterway sediment samples that occur along the pyrogenic continuum shown in Figure A-19

A.5 References

Baumard, P., H. Budzinski, and P. Garrigues. 1998. “Polycyclic Aromatic Hydrocarbons in Sediments and Mussels of the Western Mediterranean Sea.” *Environ. Toxicol. Chem., 17*(5): 765-776.

Bence, A.E., and W.A. Burns. 1995. “Fingerprinting Hydrocarbons in the Biological Resources of the *Exxon Valdez* Spill Area.” In: P.G. Wells, J.N. Butler, and J.S. Hughes (eds.), *Exxon Valdez Oil Spill: Fate and Effects in Alaskan Waters*. ASTM STP 1219. American Society for Testing and Materials, Philadelphia, PA. pp. 84-140.

Bence A.E., K.A. Kvenvolden, and M.C. Kennicutt. 1996. “Organic Geochemistry Applied to Environmental Assessments of Prince William Sound, Alaska, after the Exxon Valdez Oil Spill – a Review.” *Org. Geochem., 24*(1): 7-42.

Bloom, N., and E.A. Crecelius. 1987. “Distribution of Silver, Mercury, Lead, Copper and Cadmium in Central Puget Sound Sediments.” *Mar. Chem., 21*: 377-90.

Blumer, M. 1976. “Polycyclic Aromatic Hydrocarbons in Nature.” *Scientific American, 234*(3): 34-45.

Boehm, P.D., and J.W. Farrington. 1984. “Aspects of the Polycyclic Aromatic Hydrocarbon Geochemistry of Recent Sediments in the Georges Bank Region.” *Environ. Sci. Technol., 18*(11): 840-845.

Boehm, P.D., G.S. Douglas, W.A. Burns, P.J. Mankiewicz, D.S. Page, and A.E. Bence. 1997. “Application of Petroleum Hydrocarbon Chemical Fingerprinting and Allocation Techniques after the Exxon Valdez Oil Spill.” *Mar. Pollut. Bull., 34*(8): 599-613.

Boehm, P.D., D.S. Page, W.A. Burns, A.E. Bence, P.J. Mankiewicz, and J.S. Brown. 2000a. “Resolving the Origin of the Petrogenic Hydrocarbon Background in Prince William Sound, Alaska.” *Environ. Sci. Technol., 35*: 471-479.

Boehm, P.D., G.S. Douglas, J.S. Brown, D.S. Page, A.E. Bence, W.A. Burns, and P.J. Mankiewicz. 2000b. “Comment on ‘Natural Hydrocarbon

Background in Benthic Sediments of Prince William Sound, Alaska: Oil versus Coal'." *Environ. Sci. Technol., 34*(10): 2064-2065.

Brenner, R.C., V.S. Magar, J.A. Ickes, J.E. Abbott, S.A. Stout, E.A. Crecelius, and L.S. Bingler. 2002. "Characterization and Fate of PAH-Contaminated Sediments at the Wyckoff/Eagle Harbor Superfund Site." *Environ. Sci. Technol., 36*(12): 2605-2613.

Burns, W.A., P.J. Mankiewicz, A.E. Bence, D.S. Page, and K.R. Parker. 1997. "A Principal Component and Least Squares Method for Allocating Polycyclic Aromatic Hydrocarbons in Sediment to Multiple Sources." *Environ. Toxicol. Chem., 16*(6): 1119-1131.

Douglas, G.S., A.E. Bence, R.C. Prince, S.J. McMillen, and E.L. Butler. 1996. "Environmental Stability of Select Petroleum Hydrocarbon Source and Weathering Ratios." *Environ. Sci. Technol., 30*: 2332-9.

Gas Research Institute. 1987. *Management of Manufactured Gas Plant Sites*. GRI-87/0260.1. October.

Gogou, A., I. Bouloubassi, and E.G. Stephanou. 2000. "Marine Organic Geochemistry of the Eastern Mediterranean: 1. Aliphatic and Polyaromatic Hydrocarbons in Cretan Sea Surficial Sediments." *Mar. Chem., 68*: 265-282.

Johnson, G.W., and R. Ehrlich. 2002. "State of the Art Report on Multivariate Chemometric Methods in Environmental Forensics." *Environmental Forensics, 3*(1): 59-79.

Kerr, J.M., S.J. Melton, R.I. Magaw, G. Naughton. 1999. "Polyaromatic Hydrocarbon Content in Crude Oils Around the World." 1999 SPE/EPA Exploration and Production Environmental Conference, Austin, TX. Soc. Petrol. Engin., Richardson, TX (SPE Paper No. 52724).

Lavelle J.W., G.J. Massoth, and E.A. Crecelius. 1986. "Accumulation Rates of Recent Sediments in Puget Sound, Washington." *Marine Geology, 72*: 59-70.

Maxon, C.L., A.M. Barnett, and D.R. Diener. 1997. "Sediment Contaminants and Biological Effects in Southern California: Use of a Multivariate Statistical Approach to Assess Biological Impact." *Environ. Toxicol. Chem., 16*(4): 775-784.

Mitra, S., T.M. Dellapenna, and R.M. Dickhut. 1999. "Polycyclic Aromatic Hydrocarbon (PAH) Source, Sediment Deposition Patterns, and Particle Geochemistry as Factors Influencing PAH Distribution Coefficients in Sediments of the Elizabeth River, VA, USA." *Mar. Chem., 66*: 113-127.

Mitra, S., and N.K. Wilson. 1992. "Pattern of Polynuclear Aromatic Hydrocarbons in Indoor Air: Exploratory Principal Component Analysis." *Environment International, 18*: 477-487.

Naes, K., and E. Oug. 1997. "Multivariate Approach to Distribution Patterns and Fate of Polycyclic Aromatic Hydrocarbons in Sediments from Smelter-Affected Norwegian Fjords and Coastal Waters." *Environ. Sci. Technol., 31*(5): 1253-1258.

Naes, K., and E. Oug. 1998. "The Distribution and Environmental Relationships of Polycyclic Aromatic Hydrocarbons (PAHs) in Sediments from Norwegian Smelter-Affected Fjords." *Chemosphere, 36*(3): 561-576.

Neff, J.M., S.A. Stout, and D. Gunster. In press. "Ecological Risk Assessment of PAHs in Sediments: Identifying Sources and Toxicity." *Environmental International*.

Ohkouchi, N., Kawamura, K., Kawahata, H. 1999. "Distributions of Three- to Seven-Ring PAH on the Deep Sea Floor in the Central Pacific." *Environ. Sci. Technol., 33*: 3083-3090.

Page, D.S., P.D. Boehm, G.S. Douglas, and A.E. Bence. 1995. "Identification of Hydrocarbon Sources in the Benthic Sediments of Prince William Sound and the Gulf of Alaska Following the *Exxon Valdez* Oil Spill." In: P.G. Wells, J.N. Butler, and J.S. Hughes (eds.), *Exxon Valdez Oil Spill: Fate and Effects in Alaskan Waters*. ASTM STP 1219. American Society for Testing and Materials, Philadelphia, PA. pp. 41-83.

Page, D.S., P.D. Boehm, G.S. Douglas, A.E. Bence, W.A. Burns, and P.J. Mankiewicz. 1996. "The Natural Petroleum Hydrocarbon Background in Subtidal Sediments of Prince William Sound, Alaska, USA." *Environ. Toxicol. Chem., 15*(8): 1266-1281.

Pena-Mendez, E.M., M.S. Astorga-Espana, and F.J. Garcia-Montelongo. 1999. "Interpretation of Analytical Data on n-Alkanes and Polynuclear Aromatic Hydrocarbons in *Arbacia lixula* from the Coasts of Tenerife (Canary Islands, Spain) by Multivariate Data Analysis." *Chemosphere, 39*(13): 2259-2270.

Sauer, T.C., and A.D. Uhler. 1994/95. "Pollutant Source Identification and Allocation: Advances in Hydrocarbon Fingerprinting." *Remediation, Winter, 1994/95*: 25-50.

Stout, S.A., A.D. Uhler, and P.D. Boehm. 2001a. "Recognition of and Allocation among Multiple

Sources of PAH in Urban Sediments." *Environ. Claims J., 13*(4): 141-158.

Stout, S.A., V.S. Magar, R.M. Uhler, J. Ickes, J. Abbott, and R. Brenner. 2001b. "Characterization of Naturally-Occurring and Anthropogenic PAHs in Urban Sediments: Wycoff/Eagle Harbor Superfund Site." *Journal of Environmental Forensics,* 2(4).

Stout, S.A., A.D. Uhler, and K.J. McCarthy. 2001c. "A Strategy and Methodology for Defensibly Correlating Spilled Oil to Source Candidates." *Environmental Forensics, 2*: 87-98.

Stout, S.A., A.D. Uhler, K.J. McCarthy, and S. Emsbo-Mattingly. 2002a. "Chemical Fingerprinting of Hydrocarbons." In: B.L. Murphy and R.D. Morrison (eds.), *Introduction to Environmental Forensics*. Academic Press, NY. pp. 137-260.

Stout S.A., A.D. Uhler, V.S. Magar, K.J. McCarthy, S.J. Emsbo-Mattingly, and E.A. Crecelius. 2002b. "Sediment Geochronology Reveals Temporal Changes in Contaminant Sources." *Contaminated Soil Sediment and Water, July/Aug*: 104-6.

Stout S.A., A.D. Uhler, and S.J. Emsbo-Mattingly. In press. "Characterization of PAH Sources in Sediments of the Thea Foss/Wheeler Osgood Waterways, Tacoma, Washington." *Soil and Sediment Contamination.*

Walker, S.E., and R.M. Dickhut. 2001. "Sources of PAHs to Sediments of the Elizabeth River, VA." *Soil and Sediment Contamination, 10*(6): 611-632.

Zeng, E.Y., and C.L. Vista. 1997. "Organic Pollutants in the Coastal Environment of San Diego, California: 1. Source Identification and Assessment by Compositional Indices of Polycyclic Aromatic Hydrocarbons." *Environ. Toxicol. Chem., 16*(2): 179-188.

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, 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 easily conducted using appropriate statistical software, such as DataQUEST
- Is not designed to process datasets with nondetects
- May not have sufficient power to detect nonnormality if the underlying distribution is only slightly different than the normal distribution or if the number of data in the dataset is too small.

The computations needed to conduct the W test are provided in Box B-1 along with an example.

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 the performance of the test compares favorably with other tests. The same assumptions and their verification 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 not designed to process datasets with nondetects
- May not have large power to detect nonnormality 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 necessary to conduct the test are provided in Box B-2 along with an example.

B.1.3 Other Tests

In addition to the W and D'Agostino tests, other statistical procedures 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 is provided in D'Agostino and Stephens (1986). 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 is recommended. The Filliben test (Filliben, 1975; U.S. EPA, 1997) is closely related to the W test. The studentized range test also is recommended 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 using the Cohen method or calculating a rimmed mean or a Winsorized mean and standard deviation. These methods are defined and their assumptions, advantages, and disadvantages are listed in Box B-3. Examples of computing the median, trimmed mean, the Winsorized mean and standard deviation are illustrated in Box B-4. The Cohen method for computing the mean and standard deviation of a normally distributed set of data that contains nondetects (i.e., a censored dataset) is explained and illustrated in Box B-5.

Cautionary Note

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 underlying distribution of measurements. The only descriptive statistics that might be possible to compute are *p*th 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 2-9 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-1. 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-6, B-7, and B-8, respectively. The Rosner test is described in Box B-9 and illustrated in Box B-10. It should be noted that in background analysis, outlier measurements are deleted, only if they are determined to be the results of field, laboratory or processing errors.

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.

BOX B-1. Shapiro-Wilk W test procedure

Select the significance level, α, desired for the test, where $0 < \alpha < 0.5$. That is, select the probability, α, that can be tolerated of the W test declaring 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 $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$.
- 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 $\alpha = 0.05$
- 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, 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, 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 $D = \{[1 - 0.5(n+1)]x_{(1)} + [2 - 0.5(n+1)]x_{(2)} + \ldots + [n - 0.5(n+1)]x_{(n)}\,\} / n^2s$
- 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 $\alpha = 0.05$
- 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^2s, 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, $D = 1833.3 / 6583 = 0.2785$
- Therefore, $Y = (0.2785 - 0.282094) / (0.02998798 / 115^{1/2}) = -1.29$
- Using Table C-12, linear interpolation indicates 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, it cannot be concluded that the measurements are not normally distributed.

BOX B-3. 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 in the usual way as illustrated in Box 2-9 of main document.	• The largest nondetect is less than the median of the entire dataset (detects + nondetects); i.e., 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.

BOX B-3. (cont'd)

Method	Assumptions	Advantages	Disadvantages
Winsorized Mean ($\overline{x}_w$): 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, $\overline{x}_w$, 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) Suppose 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 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.
*p*th Sample Percentile The *p*th sample percentile is computed as described in Box 2-9 of main document.	- All nondetects have the same DL. - All detects are greater than the DL. - The computed value of k (see Box 2-9) 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).

BOX B-4. 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 RL, 0.15. If multiple RLs 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, 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, i.e., discard the three nondetects and the measurements 0.62, 0.63, 0.79. Compute the arithmetic mean on the remaining six 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:

$$\begin{aligned} s = [&(0.18 - 0.32)^2 + (0.18 - 0.32)^2 + (0.18 - 0.32)^2 + (0.18 - 0.32)^2 + (0.25 - 0.32)^2 + (0.26 - 0.32)^2 + \\ &(0.27 - 0.32)^2 + (0.36 - 0.32)^2 + (0.50 - 0.32)^2 + (0.50 - 0.32)^2 + (0.50 - 0.32)^2 + (0.50 - 0.32)^2] / 11 \\ = {}& 0.1416 \end{aligned}$$

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-5. Cohen method for computing the mean and variance of a censored dataset (U.S. EPA, 2000; Gilbert, 1987, p. 182)

- Let the single nondetect limit be denoted by ND. Let $x_1, x_2, \ldots, x_n$ denote the *n* measurements in the dataset, including those that are less than ND. Let *k* be the number out of *n* that are greater than the ND.
- Compute $h = (n - k)/n$, which is the fraction of the *n* measurements that are below the ND.
- Compute the arithmetic mean of the *k* measurements that exceed the ND 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 >ND.

- Compute the following statistic using the *k* measurements that exceed the ND:

$$s_c^2 = [(x_1 - \bar{x}_c)^2 + (x_2 - \bar{x}_c)^2 + \ldots + (x_k - \bar{x}_c)^2] / k$$

- Compute $G = s_c^2 / (\bar{x}_c - \mathrm{ND})^2$
- 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:

$$\text{Cohen mean} = \bar{x}_c - \lambda(\bar{x}_c - \mathrm{ND})$$

$$\text{Cohen variance} = s_c^2 + \lambda(\bar{x}_c - \mathrm{ND})^2$$

- Cohen standard deviation is the square root of Cohen variance.

Example:

- n = 25 measurements of a chemical in sediment samples were obtained. The nondetect limit was equal to 36. Five measurements were reported as <36. 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 ND.
- Compute the arithmetic mean of the 20 measurements that exceed the ND:

$$\bar{x}_c = (49 + 49 + 59 + \ldots + 142 + 144) = 83.85$$

- 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 $G = 882.63 / (83.85 - 36)^2 = 0.385$.
- Using Table C-13, linear interpolation between $\gamma = 0.35$ and $\gamma = 0.40$ for $h = 0.20$ indicates that $\lambda = 0.291$.
- 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 = $(1{,}548.9)^{1/2} = 39.4$

TABLE B-1. Assumptions, advantages, and disadvantages of outlier tests

Statistical Test	Assumptions	Advantages	Disadvantages
Dixon Test	• $n \le 25$ • Measurements are representative of the underlying population. • The measurements without the suspect outlier are normally distributed; otherwise, see 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 identifying 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	• $3 < n \le 50$ • Measurements are representative of underlying population. • The measurements without the suspected outlier are normally distributed; otherwise, see 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 identifying 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.
Rosner's Test	• $n \ge 25$ • Measurements are representative of underlying population. • The measurements without the suspected outliers are normally distributed; otherwise, see 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's test. • Computations are more complex than for Dixon's test or the Discordance test.
Walsh's 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 identifying 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 $60 < n \le 220$. • 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.

BOX B-6. Procedure for conducting the Dixon extreme value 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. 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 $= [x_{(n)} - x_{(n-1)}] / [x_{(n)} - x_{(1)}]$ if $3 \le n \le 7$

  $= [x_{(n)} - x_{(n-1)}] / [x_{(n)} - x_{(2)}]$ if $8 \le n \le 10$

  $= [x_{(n)} - x_{(n-2)}] / [x_{(n)} - x_{(2)}]$ if $11 \le n \le 13$

  $= [x_{(n)} - x_{(n-2)}] / [x_{(n)} - x_{(3)}]$ if $14 \le n \le 25$

  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. 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. Any test for normality will have little ability to detect nonnormality on the basis of only four data values. Suppose α is selected to be 0.05, i.e., there should be no more than a 5% chance that the test will incorrectly declare the largest observed measurement to be 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-7. 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, $\overline{x}$, and the sample standard deviation, *s*.
- Compute $D = [x_{(n)} - \overline{x}] / s$
- 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. We wish 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 $\overline{x} = 51.6$ 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 normally distribution.

BOX B-8. 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 $60 < n \leq 220$, 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 (i.e., 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 $60 < n \leq 220$. 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 three largest measurements are not outliers.

BOX B-9. Procedure for conducting the Rosner outlier test (U.S. EPA, 2000)

STEP 1:

- Select the desired significance level α, i.e., 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 $\overline{x}^{(0)}$, and $s^{(0)}$ using all *n* measurements. Determine the measurement that is farthest from $\overline{x}^{(0)}$ and label it $y^{(0)}$.

$$\overline{x}^{(i)} = (x_1 + x_2 + \ldots + x_{n-i}) / (n - i)$$

$$s^{(i)} = \{[(x_1 - \overline{x}^{(i)})^2 + (x_2 - \overline{x}^{(i)})^2 + \ldots + (x_{n-i} - \overline{x}^{(i)})^2] / (n - i)\}^{1/2}$$

- Delete $y^{(0)}$ from the dataset of *n* measurements and compute (using i = 1 in the above formulas) the sample arithmetic mean, labeled $\overline{x}^{(1)}$, and $s^{(1)}$ on the remaining n – 1 measurements. Determine the measurement that is farthest from $\overline{x}^{(1)}$ 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 $\overline{x}^{(2)}$, 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 $\overline{x}^{(0)}$, $\overline{x}^{(1)}$, …, $s^{(0)}$, $s^{(1)}$, … are computed using the above formulas.

STEP 3:

- To test if there are *r* outliers in the dataset, compute

$$R_r = [\,|\,y^{(r-1)} - \overline{x}^{(r-1)}\,|\,] / s^{(r-1)}$$

- 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

$$R_{r-1} = [\,|\,y^{(r-2)} - \overline{x}^{(r-2)}\,|\,] / s^{(r-2)}$$

- 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 there are a certain number of outliers are present or that no outliers exist at all.

BOX B-10. 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 four 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, i.e., we can tolerate a 5% chance of the Rosner test falsely declaring four outliers. In Table C-16, $\lambda_4 = 2.89$ when $n = 32$, $r = 4$ and $\alpha = 0.05$. As $R_4 = 1.97$ is less than 2.89, it is concluded that four outliers are not present. Therefore, test if three 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 $\lambda_3 = 2.91$ when $n = 32$, $r = 3$ and $\alpha = 0.05$. Because $R_4 = 2.26$ is less than 2.91, it is concluded that three outliers are not present. Therefore, test if two 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, $\lambda_2 = 2.92$ for $n = 32$, $r = 2$ and $\alpha = 0.05$. As $R_2 = 2.56$ is less than 2.92, it is concluded that two outliers are not present in the dataset. Therefore, test if one 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 $\lambda_1 = 2.94$ for $n = 32$, $r = 1$ and $\alpha = 0.05$. Because $R_1 = 3.05$ is greater than 2.94, then it is concluded at the $\alpha = 0.05$ significance level that one outlier is present in the dataset. Therefore, the measurement 395.67 is considered to be a statistical outlier. It will be further investigated to determine if the measurement 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)

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_c) for the Slippage test for $\alpha = 0.01$

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_c) for the Slippage test for $\alpha = 0.01$ (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_c) 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_c) for the Slippage test for $\alpha = 0.05$ (continued)

Number of Site Measurements, n

Number of Background Measurements, m

SITE	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 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		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 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		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 (ω_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

$_k\backslash^n$	**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

$_k\backslash^n$	**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

$_k\backslash^n$	**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**
0.00	.010100	.020400	.030902	.041583	.052507	.063625	.074953	.08649	.09824	.11020	.17342	.24268
0.05	.010551	.021294	.032225	.043350	.054670	.066159	.077909	.08983	.10197	.11431	.17925	.25033
0.10	.010950	.022082	.033398	.044902	.056596	.068483	.080563	.09285	.10534	.11804	.18479	.25741
0.15	.011310	.022798	.034466	.046318	.058356	.070586	.083009	.09563	.10845	.12148	.18985	.26405
0.20	.011642	.023459	.035453	.047829	.059990	.072539	.085280	.09822	.11135	.12469	.19460	.27031
0.25	.011952	.024076	.036377	.048858	.061522	.074372	.087413	.10065	.11408	.12772	.19910	.2762
0.30	.012243	.024658	.037249	.050018	.062969	.076106	.089433	.10295	.11667	.13059	.20338	.2819
0.35	.012520	.025211	.038077	.051120	.064345	.077736	.091355	.10515	.11914	.13333	.20747	.2873
0.40	.012784	.025738	.038866	.052173	.065660	.079332	.093193	.10725	.12150	.13595	.21129	.2925
0.45	.013036	.026243	.039624	.053182	.066921	.080845	.094958	.10926	.12377	.13847	.21517	.2976
0.50	.013279	.026728	.040352	.054153	.068135	.082301	.096657	.11121	.12595	.14090	.21882	.3025
0.55	.013513	.027196	.041054	.055089	.069306	.083708	.098298	.11208	.12806	.14325	.22225	.3072
0.60	.013739	.027849	.041733	.055995	.070439	.085068	.099887	.11490	.13011	.14552	.22578	.3118
0.65	.013958	.028087	.042391	.056874	.071538	.086388	.10143	.11666	.13209	.14773	.22910	.3163
0.70	.014171	.028513	.043030	.057726	.072505	.087670	.10292	.11837	.13402	.14987	.23234	.3206
0.75	.014378	.029927	.043652	.058556	.073643	.088917	.10438	.12004	.13590	.15196	.23550	.32489
0.80	.014579	.029330	.044258	.059364	.074655	.090133	.10580	.12167	.13775	.15400	.23858	.32903
0.85	.014773	.029723	.044848	.060153	.075642	.091319	.10719	.12225	.13952	.15599	.24158	.33307
0.90	.014967	.030107	.045425	.060923	.075606	.092477	.10854	.12480	.14126	.15793	.24452	.33703
0.95	.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**
0.00	.31862	.4021	.4941	.5961	.7096	.8388	.9808	1.145	1.336	1.561	2.176	3.283
0.05	.32793	.4130	.5066	.6101	.7252	.8540	.9994	1.166	1.358	1.585	2.203	3.314
0.10	.33662	.4233	.5184	.6234	.7400	.8703	1.017	1.185	1.379	1.608	2.229	3.345
0.15	.34480	.4330	.5296	.6361	.7542	.8860	1.035	1.204	1.400	1.630	2.255	3.376
0.20	.35255	.4422	.5403	.6483	.7673	.9012	1.051	1.222	1.419	1.651	2.280	3.405
0.25	.35993	.4510	.5506	.6600	.7810	.9158	1.067	1.240	1.439	1.672	2.305	3.435
0.30	.36700	.4595	.5604	.6713	.7937	.9300	1.083	1.257	1.457	1.693	2.329	3.464
0.35	.37379	.4676	.5699	.6821	.8060	.9437	1.098	1.274	1.475	1.713	2.353	3.492
0.40	.38033	.4735	.5791	.6927	.8179	.9570	1.113	1.290	1.494	1.732	2.376	3.520
0.45	.38665	.4831	.5880	.7029	.8295	.9700	1.127	1.306	1.511	1.751	2.399	3.547
0.50	.39276	.4904	.5967	.7129	.8408	.9826	1.141	1.321	1.528	1.770	2.421	3.575
0.55	.39679	.4976	.6061	.7225	.8517	.9950	1.155	1.337	1.545	1.788	2.443	3.601
0.60	.40447	.5045	.6133	.7320	.8625	1.007	1.169	1.351	1.561	1.806	2.465	3.628
0.65	.41008	.5114	.6213	.7412	.8729	1.019	1.182	1.368	1.577	1.824	2.486	3.654
0.70	.41555	.5180	.6291	.7502	.8832	1.030	1.195	1.380	1.593	1.841	2.507	3.679
0.75	.42090	.5245	.6367	.7590	.8932	1.042	1.207	1.394	1.608	1.851	2.528	3.705
0.80	.42612	.5308	.6441	.7676	.9031	1.053	1.220	1.408	1.624	1.875	2.548	3.730
0.85	.43122	.5370	.6515	.7781	.9127	1.064	1.232	1.422	1.639	1.892	2.568	3.754
0.90	.43622	.5430	.6586	.7844	.9222	1.074	1.244	1.435	1.653	1.908	2.588	3.779
0.95	.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

n	Level of Significance	
	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. (continued)

n	r	α	
		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
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
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. Standard Guide for the Contents of Geostatistical Site Investigation Report. 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, WA: 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.

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.

Thompson, S.K., and G.A. Seber. 1996. *Adaptive Sampling*. New York: Wiley.

Sinclair, A.J. 1974. Selection of Threshold Values in Geochemical Data Using Probability Graphs. *J. Geochem. Explor.* 3:129-149.

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, Xiao-Hua, Sujuan Gao, and Siu 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)

Microsoft TerraServer – USGS topographic maps (digital raster graphics) and orthorectified aerial photographs (downsampled digital orthophoto quadrangles) available free in tiled format: URL: http://www.terraserver.microsoft.com

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://enviro.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

Data Quality Objectives Decision Error Feasibility Trials (DQO/DEFT). United States Environmental Protection Agency. 1994. EPA QA/G-4D. EPA 600/R-96/056. Washington, DC: Office of Research and Development. URL: <http://www.epa.gov/superfund/programs/dfa/decsupp.htm>

Data Quality Assessment Statistical Toolbox (DataQUEST) (G-9D). User's Guide and Software. United States Environmental Protection Agency. 1997. EPA/600/R-96/085. Washington, DC: Office of Research and Development. URL: <http://www.epa.gov/quality/dqa.html>

ELIPGRID-PC (Hot Spot Probability Calculations): Upgraded Version. Davidson, J. R. 1995. ORNL/TM-13103. Oak Ridge, TN.: Oak Ridge National Laboratory. December. URL: <http://dqo.pnl.gov/software/elipgrid.htm>

Environmental Stats for S-PLUS. Seattle: Insightful Corporation. (Millard, S.P., and N.K. Neerchal. 2001. *Environmental Statistics with S-PLUS*. Boca Raton, FL: CRC Press.) URL: <www.insightful.com>

GEOPACK (geostatistical software). United States Environmental Protection Agency. 1990. EPA/600/8-90/004. Washington, DC: Environmental Sciences Division. URL: <http://www.epa.gov/ada/csmos/models/geopack.html>

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/superfund/programs/dfa/decsupp.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>

Visual Sample Plan (sampling design software) from Department of Energy: URL: <http://dqo.pnl.gov/VSP/Index.htm>