PROCEDURAL GUIDANCE FOR STATISTICALLY ANALYZING ENVIRONMENTAL BACKGROUND DATA

Prepared by:

SWDIV and EFA WEST
of
Naval Facilities Engineering Command

September 1998

ACKNOWLEDGMENT

This background analysis procedures document was prepared for the Naval Facilities Engineering Command, Southwest Division through collaboration of individuals with expertise in statistics, toxicology, risk assessment, and geochemistry, with special thanks and acknowledgements for the contributions of the following individuals and firms. Richard DeGrandchamp, Ph.D. of Scientia Vertias, L.L.P. for his significant contribution in preparing the guidance document, and providing expertise in toxicology and risk assessment. Richard O. Gilbert, Ph.D. of Pacific Northwest Laboratory, and John Warren, Ph.D. of the U.S. Environmental Protection Agency's (USEPA), Statistical Policy Branch for contributing vital recommendations and insightful comments regarding the use and validation of statistical procedures recommended in this document. Dan Quo, Ph.D. of Bechtel National, Inc. for providing useful comments and written materials dealing with geochemical analyses that improve the understanding of environmental background. Tom Grieb, M.S. of Tetra Tech Environmental Management Inc. for providing review and data modeling of the statistical methods used for background analyses.

Additionally, numerous toxicologists and managers from California's Department of Toxic Substances Control (DTSC), and USEPA Region IX participated in a joint workshop with US Navy representatives to discuss establishment and use of background in cleanup processes, and risk assessments. In particular, John Christopher, Ph.D. of DTSC contributed valuable information regarding background analyses in support of risk assessment. Representatives from PRC Environmental Management Inc., Tetra Tech EM Inc., Battelle, HSI Geotrans, MK Environmental, and Bechtel National Inc. also participated in the joint workshop and provided information that supported preparation of this procedure document.

The procedures document was prepared under the supervision and contributions of Southwest Division representatives Camille Garibaldi, Gilbert Rivera, and Nick Bollo of Engineering Field Activity West, and Dennis Askvig of Engineering Field Division, Southwest.

Questions concerning this document may be directed to:

Dennis Askvig, Code 03
Southwest Division, Naval Facilities Engineering Command
1220 Pacific Highway
San Diego, CA 92132
(619) 532-2510

or

Gilbert Rivera, Code 7023
Engineering Field Activity, West, Naval Facilities Engineering Command
900 Commodore Drive
San Bruno, CA 94066
(650) 244-2565

CONTENTS

Section **Page**

CONTENTS (CONTINUED)

Section **Page**

FIGURES

Figure **Page**

TABLES

Table **Page**

EXECUTIVE SUMMARY

This document is procedural guidance for conducting background statistical analyses in the state of California. Background chemicals are defined as either naturally occurring (nonanthropogenic) or anthropogenic (ambient), which are unrelated to Navy activities or operations, and should not be considered chemical releases. Many background chemicals are detected during routine sampling and analyses, and if incorrectly identified as chemicals of concern (COCs) can obscure investigations and confound cleanup strategies. Failing to distinguish between Navy releases and background conditions can unknowingly lead to targeting background chemicals for remediation or setting remediation target goals below background levels. These situations can result in costly and protracted site cleanup, and delay property transfer and reuse.

Establishing background conditions is a preliminary step in identifying COCs for human health and ecological risk assessments. It is vital in selecting cost-effective remedial alternatives and setting reasonable and attainable remediation target goals. The analytical approach developed in this document is based on widely used statistical methods that will yield verifiable results.

When a background analytical approach is carefully planned and implemented according to this procedures document, uncertainty in all related Navy investigations and cleanup actions will be reduced. The Navy will implement this procedural guidance at all California installations to promote consistency throughout the Navy's Installation Restoration Program (IRP) and other compliance related programs which will increase public and regulatory confidence in Navy cleanup activities.

As part of Base Realignment and Closure (BRAC) and the IRP, the Navy is conducting various environmental investigations and cleanups. These include: Environmental Baseline Surveys (EBS), Preliminary Assessments (PA), Site Inspections (SI), Remedial Investigations (RI) and Feasibility Studies (FS), Underground Storage Tank Investigations (USTI), and air and groundwater monitoring. In order for these investigations to yield meaningful results and conclusions, it is necessary to establish background conditions to estimate site-related human health risks, identify threats to the environment, analyze compliance with regulatory standards, and evaluate remedial alternatives to cost-effectively remediate Navy releases.

This procedural guidance document is consistent with environmental laws, regulations, and U.S. Environmental Protection Agency (EPA) technical guidance. The data analysis and statistical testing methods closely follow *Guidance for the Data Quality Objectives Process* (DQO) (EPA 1994a) and *Guidance for Data Quality Assessment* (DQA) (EPA 1996b) developed by EPA’s Quality Assurance Division. Additionally, the overall technical approach for determining background conditions is generally consistent with *Determination of Background Concentrations of Inorganics in Soils and Sediments at Hazardous Waste Sites* (EPA 1995) and with the background analytical approach recently developed by the California Department of Toxic Substances Control (DTSC) which is presented in the *Interim Final Policy Determining Ambient Concentrations of Metals for Risk Assessments at Hazardous Waste Sites and Permitted Facilities* (DTSC 1996).

DOCUMENT ORGANIZATION

This procedural guidance document briefly introduces the importance of background analyses, provides guidance for developing a representative background data set, and conducting statistical testing. It is divided into four chapters, which are as follows:

- Chapter 1: Introduction
- Chapter 2: Developing Background Data sets
- Chapter 3: Analyzing Data and Statistical Testing
- Chapter 4: Summary

Chapter 1 introduces background analyses and states the purpose and intent of this document. Statutory requirements, regulations, and risk assessment guidance are briefly reviewed to provide the structural framework for the procedural guidance document. California and EPA guidance is referenced in defining background analytes and conditions.

Chapter 2 suggests approaches for developing background data sets. An approach is presented for installations where investigations are getting underway and where few samples have been collected. Another approach is described for extracting background data sets from an existing installation-wide database where prior investigations have been conducted. An overview of EPA's DQO and DQA processes are presented as the decision-making framework that project teams should employ to develop sampling and analysis plans, to conduct data analysis to assess, and to verify whether data collected meet data quality and quantity requirements.

Chapter 3 presents methods for analyzing data and statistical testing. A decision-making flow chart has been developed to guide project teams from the initial step involving data quality evaluations to the last step in which a determination is made as to whether a chemical release has occurred.

Chapter 4 summarizes the Navy's background statistical procedural guidance.

1.0 INTRODUCTION

1.1 PURPOSE

This document provides Navy project teams with consistent background procedural guidance that is protective of human health and the environment, cost-effective, and scientifically defensible. This guidance will enable project teams to determine site-specific background conditions in order to make correct and appropriate remedial decisions.

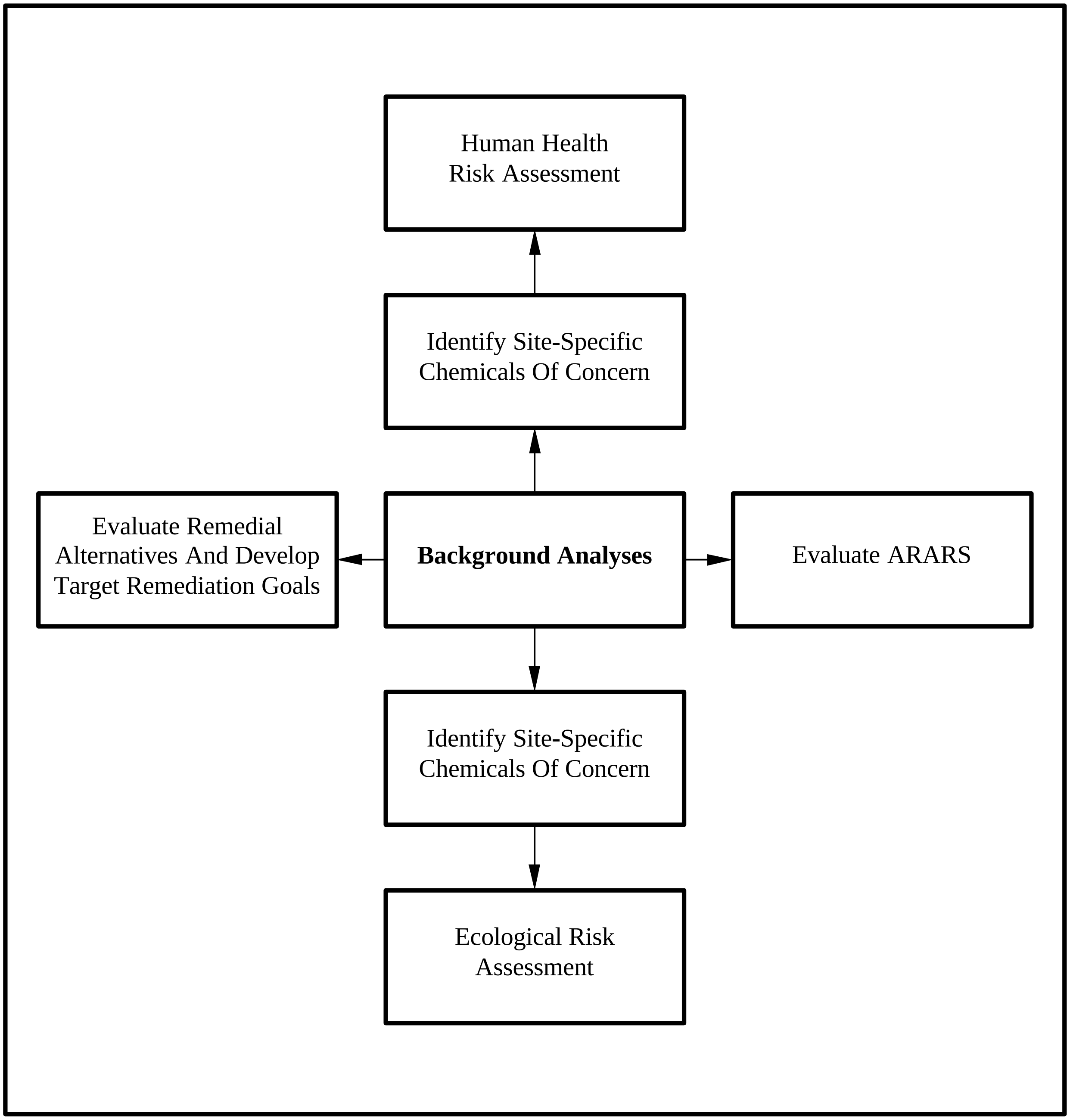
To successfully implement the Navy's environmental programs, project teams must accurately establish background conditions to prevent (1) background chemicals from unknowingly being targeted for remediation, (2) setting remediation goals below naturally occurring and anthropogenic background levels, or (3) overlooking a Navy release by assuming it is background. Remediating background chemicals can be costly and delay property transfer and reuse. The Navy seeks to ensure that all site releases are identified and that they do not pose risk to human health and the environment. By implementing the planning and statistical strategy in this document, investigations will be streamlined and significantly reduce the uncertainty in background analyses. This will increase stakeholder confidence not only in background determinations, but in the Navy's environmental cleanups.

1.2 IMPORTANCE OF BACKGROUND ANALYSES

The Navy is currently conducting environmental restoration activities on a massive scale in the state of California. Determining the nature and extent of chemical releases to define the Navy's cleanup responsibility is central to these activities. The Navy is committed to remediating each chemical release to protect human health and the environment, and to meeting statutory requirements. In order for the Navy to fulfill this responsibility and remediate each release at all installations, chemical releases must be characterized. To accomplish this goal, the Navy must distinguish between chemical releases associated with ongoing Navy activities or operations or prior site releases, and naturally occurring or anthropogenic background conditions. In order to make cost-effective remedial decisions, project teams must conduct detailed statistically based background studies to avoid pitfalls that are inherent in qualitative or semi-quantitative background investigations where uncertainty of unknown magnitude may be introduced.

As part of Base Realignment and Closure (BRAC) and the Installation Restoration Program (IRP), the Navy is conducting diverse environmental investigations and cleanups. These include: Environmental Baseline Surveys (EBS), Preliminary Assessments (PA), Site Inspection (SI), Remedial Investigations (RI) and Feasibility Studies (FS), Underground Storage Tank Investigations (USTI), and air and groundwater monitoring. In order for all of these investigations to yield meaningful results and conclusions, it is necessary to establish background conditions to estimate site-related human health risks, identify threats to the environment, analyze compliance with regulatory standards, and evaluate effective remedial alternatives to clean up Navy releases. Figure 1 shows the central role background analyses play in the RI/FS [or environmental cleanup] process. In this process, establishing background conditions is

FIGURE 1
THE CENTRAL ROLE OF BACKGROUND ANALYSES
IN THE RI/FS PROCESS



essential to identify release-related COCs, estimate human health and ecological risks, analyze compliance with regulatory standards (applicable or relevant and appropriate requirements [ARARs]), evaluate remedial alternatives, and develop appropriate cleanup levels.

Project teams should expect the background analysis to be a challenging activity requiring diverse technical expertise. Figure 2 presents an overview of the primary steps involved in the Navy's background approach.

The background analytical approach for many installations may be multifaceted and iterative. For example, analyses at installations located in areas where land mass was constructed with dredged sediments will likely be more complex than those located in native soils. This is because there is a high probability that dredged sediments were contaminated - by diverse industrial sources - *prior to* the initiation of Navy dredging activities. For installations built with dredged fill, anthropogenic background conditions corresponding to chemical concentrations at the time of dredging need to be established in order to distinguish those conditions from Navy releases.

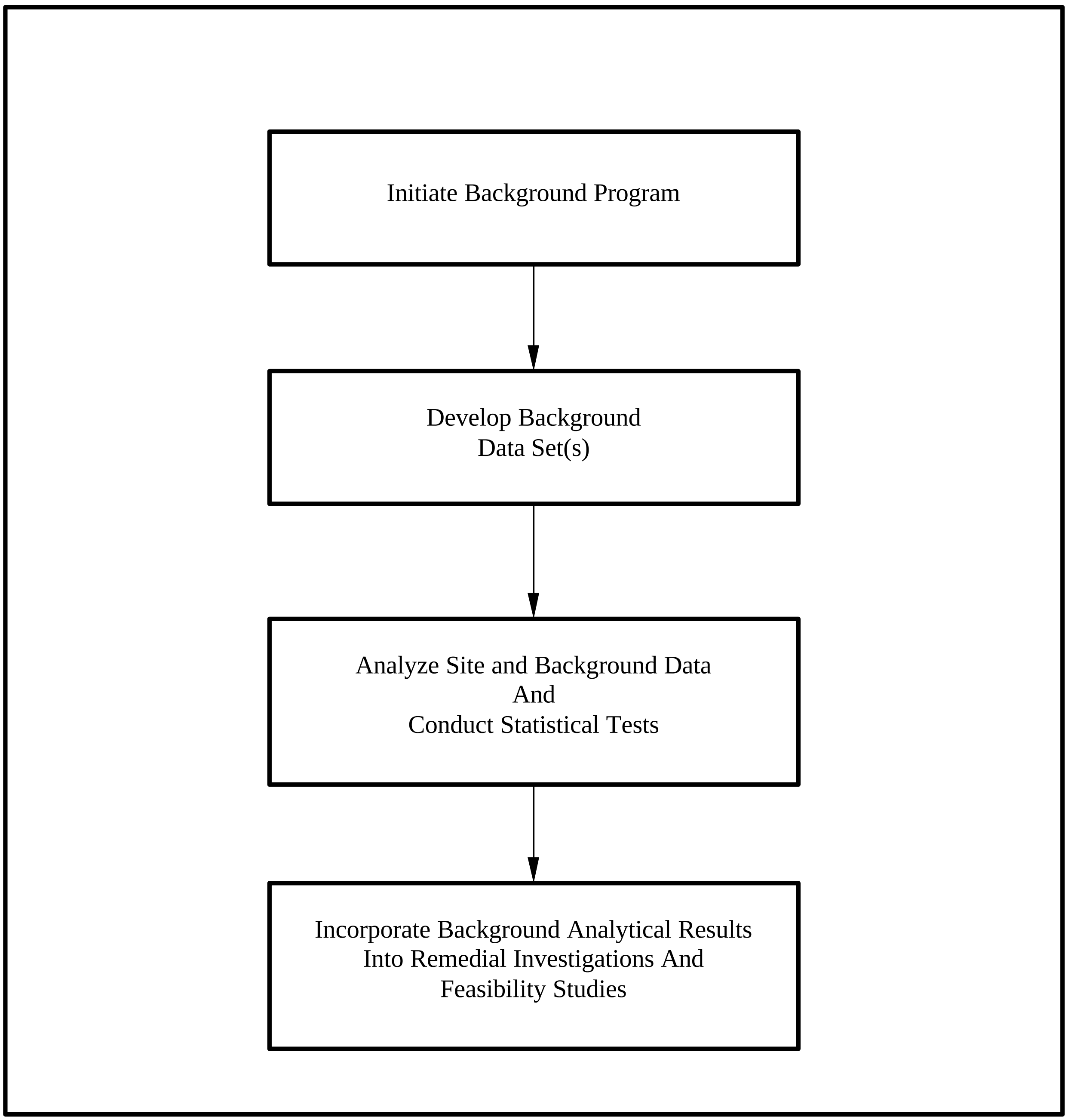
One of the principal objectives of background analyses is to eliminate background chemicals from subsequent phases of the investigation and remedial process, thereby focusing risk assessments and remedial alternative selection on those constituents associated with site releases. To confidently eliminate chemicals from further consideration requires that background analyses be conducted in a scientifically defensible and verifiable manner. The most effective time to conduct background analyses is in the early stages of the cleanup program. Eliminating background chemicals early in the process will reduce the number of chemicals that require further analysis, and decrease the complexity, time, and cost associated with future investigations and remedial activities. Waiting until investigations are completed to begin background analyses can be costly, and cause unnecessary public concern. This is because when background chemicals are mistakenly assumed to be chemicals of potential concern (COPCs) and later (COCs), stakeholders can be unnecessarily alarmed. Although statistical procedures may ultimately determine these COPCs (or COCs) to be background, public anxiety may be irrevocably heightened in the interim.

Background analyses using the DQO planning process should be conducted with collaboration among the Navy, regulatory agencies, and the public to foster strong working relationships and communication. The Navy is committed to using state of the art science in environmental investigations to remediate its chemical releases.

1.3 STATUTORY REQUIREMENTS, REGULATIONS, AND REGULATORY GUIDELINES

According to the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), samples to delineate the nature and extent of contamination should be collected and analyzed in phases. Initial information on site background characteristics is routinely gathered during the PA/SI phase.

FIGURE 2
OVERVIEW OF THE NAVY BACKGROUND ANALYSIS PROGRAM



According to EPA guidance (EPA 1988a, 1988b), the goals of the PA/SI are as follows:

- Gain an understanding of the nature and degree of threat posed by a release;
- Identify sites that may require immediate response (for example, a removal action).

According to *Code of Federal Regulations* 40 (CFR) Part 300.425, the PA is used for the following:

- To eliminate sites where CERCLA action is not necessary;
- To identify sites that require emergency response;
- To set priorities for the SI.

Confirming that a release has occurred requires that information on background conditions be available. For example, Section G, "Observed Releases," of the preamble to the Hazard Ranking System (HRS) final rule (55 *Federal Register* 51546) (EPA 1985) states that an observed release is established when a sample measurement equals or exceeds the sample quantitation limit (SQL) *and* is at least three times above the background level. Therefore, background conditions must be defined to establish that a release has occurred.

Also, during the RI scoping phase, additional detailed information on background conditions is required. Section 300.430(b)(8) of the National Oil and Hazardous Substances Pollution Contingency Plan (NCP) (EPA 1995) requires that a sampling and analysis plan (SAP) be developed during the scoping phase of the RI to provide a process for obtaining data of adequate quality and quantity to satisfy data objectives. Formulating the SAP should be based, in part, on *Guidance for Data Usability in Risk Assessment* (EPA 1990). According to this guidance, a primary objective is to determine “whether site concentrations are sufficiently different from background.” *Guidance for Conducting Remedial Investigations and Feasibility Studies Under CERCLA* (EPA 1988a) likewise states that background sampling should be conducted to differentiate between chemical releases resulting from site operations and background conditions.

Characterizing background conditions is an integral part of baseline human health and ecological risk assessments. These are conducted as part of the RI to ensure that remedy selection is protective of human health and the environment, which is one of the two threshold criteria of the NCP. *Risk Assessment Guidance for Superfund, Volume I, Human Health Evaluation Manual*, ([HHEM], EPA 1989b) is the principal guidance document developed by EPA for conducting baseline human health risk assessments (HHRAs) which are used to evaluate human health risks under baseline conditions and to determine whether remedial alternatives are protective of human health. A general discussion of statistical hypothesis testing and prespecified acceptance statistical criteria that should be used to eliminate background chemicals during the COC identification phase is presented in the HHEM.

Background analyses also play an important role in Resource Conservation and Recovery Act (RCRA) investigations. For example, the determination as to whether a chemical release has occurred ultimately depends on site-specific background levels for a solid waste management unit (SWMU). 40 CFR Part 264, Subpart F, requires groundwater monitoring at permitted hazardous waste land disposal facilities to detect groundwater contamination. It is recommended that statistical tests be applied to groundwater monitoring data to determine whether there is a significant exceedance of background or other allowable concentration level. Five recommended methods are 1) Parametric analysis-of variance, 2) Analysis-of-variance based on ranks, 3) Tolerance intervals, 4) Prediction intervals, and 5) Control charts.”

The Navy's procedures document is consistent with these methods to ensure that the statistical approach will satisfy RCRA background investigation requirements.

RCRA guidance (1989a) also stresses the importance of establishing soil background levels stating:

> "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…Selection and sampling of appropriate background areas may be important because verification of a release in a contaminated area may involve a comparison of study and background concentrations."

Regarding background related issues for RCRA cleanup, Section 3004(u) gives EPA the authority to require cleanup of contaminants from only the on-site waste management units (EPA 1992). Therefore, background levels need not be established for chemicals other than those chemicals associated with management units. Furthermore, proposed section 264.525(d)(1)(v) indicates that if the Potentially Responsible Party (PRP) can demonstrate that a particular portion of the contamination is attributable to another source, such as a naturally occurring substance, the cleanup levels for the facility related contaminants may be set at that background level.

More recent guidance has been developed to assist project teams determine whether cleanup has been successfully completed. *Statistical Methods for Evaluating the Attainment of Cleanup Standards, Volume 3: Reference-Based Standards for Soils and Solid Media* (EPA 1994b) and *A Nonparametric Statistical Methodology for the Design and Analysis of Final Status Decommissioning Surveys* (Gogolak 1997) present statistical testing procedures to evaluate whether remediated soils and solid media are statistically different from reference data sets. The analytical concepts and statistical methods promulgated by EPA and U. S. Nuclear Regulatory Commission (NRC) agree with those developed in this document.

EPA states (EPA 1994b) "The objective is to detect when the distribution of measurements for the remediated cleanup unit is 'shifted' in part or in whole to the right (to higher values) of the reference distribution." Although EPA 1994b does not explicitly state what a reference distribution is, it will usually be the background distribution for Navy installations. *Multi-Agency Radiation Survey and Site Investigation Manual* (Marssim 1997) contains non-parametric statistical tests for demonstrating compliance with dose or risk-based radionuclide regulations or standards, including comparisons with background areas. The common goal between EPA and NRC guidance, and the Navy's procedural guidance is to determine whether a statistically significant difference exists between site and background (reference) data sets.

Evaluating state or federal ARARs and selecting the appropriate remedial alternative for a site depends on site-specific background conditions. According to the NCP, compliance with ARARs is one of two threshold criteria that must be satisfied in remedy selection. However, when site-specific background information is lacking, ARAR evaluations can be compromised. For example, if naturally occurring background levels are higher than maximum contaminant levels (MCLs), it may be impracticable or impossible to remediate the site. This issue is addressed in CERCLA [Section 104(3)(A)], which states:

> "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."

It is clear from this provision that Congress recognized that remediating naturally occurring background chemicals to levels below background levels is not practical. This holds true even when background chemical concentrations exceed state or federally regulated levels.

1.4 IMPLEMENTING BACKGROUND ANALYSES

The Navy's procedural guidance for conducting statistically based background analyses utilizes a statistical test "tool box approach." In this approach, selecting the appropriate statistical test to optimize performance is based on specific data set characteristics. A statistical test that is appropriate for one data set may be inappropriate for another.

The tool box approach provides for superior background analyses when compared with a prescribed "one size fits all" approach that is frequently applied without regard to the type of data set being investigated or the underlying geochemical conditions. With a prescribed approach, data sets may be inappropriately "force-fit" into a single prescribed method rather than the best "fit" (highest statistical power) where several statistical tests are made available.

Prescribed methods are typically developed for ease of application, rather than scientific veracity, and unacceptable decision error rates. Decision errors can result in either unnecessary or can lead to insufficient remediation which can either be costly or pose unacceptable health risks, respectively. The statistical tool box approach formalizes statistical decision-making processes, allowing statistical test decision error rates to be defined by project teams. Ultimately, this approach will result in objective and scientifically defensible results that project teams can reproduce and verify.

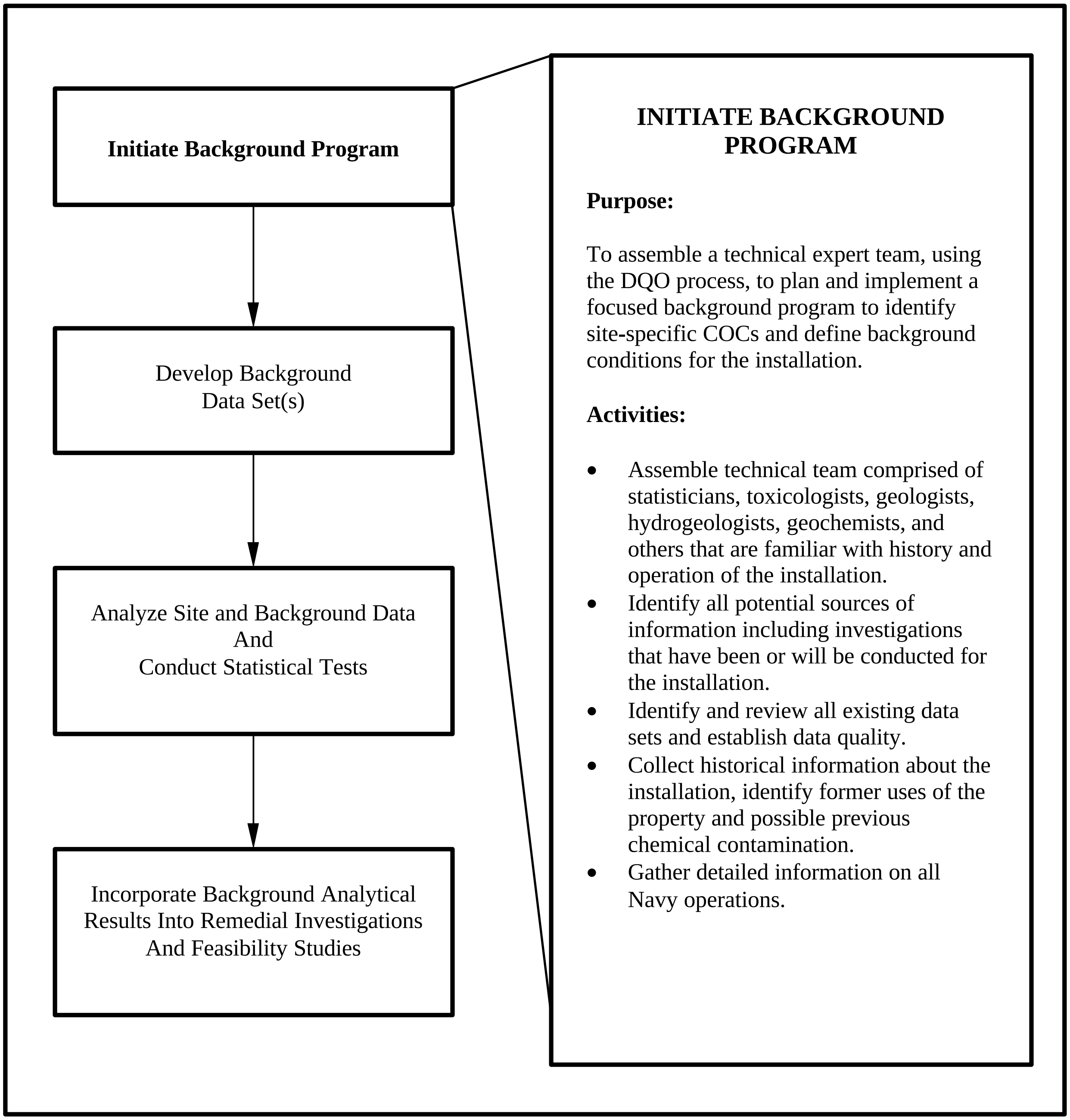
The Navy intends to implement this statistically based background procedural guidance at all California installations. In applying this guidance, project teams using the DQO planning process will conduct background analyses in two phases. In the first phase, a site-specific background data set(s) is developed. In the second phase, statistical tests are performed to compare site and background data sets to determine whether analytes are present above background levels. A statistically significant difference may indicate a release occurred. This difference may also be a decision error which would require the investigation of laboratory data to ensure that the quality required for decision making as specified in the DQO process was followed. Further samples and measurements may be needed to confirm that the significant differences are real and of a magnitude requiring remediation.

The specific analytical approach used to develop background data sets will depend on the installation site characteristics and the stage of investigation. For example, some installations are at an advanced stage of investigation where sampling is nearly complete and a data set already exists. For these installations, it may be considered impractical, inefficient, and costly to design and implement an entirely new background sampling program. An alternative is to extract background data sets from the existing data which may be scientifically defensible. It may be necessary to supplement the existing data set with additional samples to attain sufficient data to reach a scientifically defensible decision.

At other installations where investigations are just under way and few samples have been collected, it will be cost-effective to plan and implement a background sampling program to develop a separate stand-alone background data set at the earliest possible stage.

Figure 3 presents the purpose and activities involved in initiating a background program.

FIGURE 3
STEP 1-INITIATE BACKGROUND PROGRAM



1.5 DEFINING BACKGROUND CONDITIONS

Background conditions are defined as being unrelated to past or current site activities or operations. To determine whether a release has occurred, project teams must be able to distinguish them background conditions.

According to EPA (EPA 1989b), background chemicals fall into two categories:

- **Naturally Occurring or Nonanthropogenic Chemicals:** Minerals or inorganic elements that represent underlying geochemical conditions that have not been influenced by human activity;
- **Anthropogenic Chemicals:** Natural and man-made substances ubiquitously present in the environment as a result of human activities, but not related to site activities.

Naturally occurring organic or inorganic background chemicals are present in soil or water as part of the geological or hydrogeological conditions of the area. These chemicals have not been altered in form by any human activity. In contrast, anthropogenic background chemicals are generated through human activity, but are unrelated to Navy site operations or activities. Examples of anthropogenic background chemicals include pesticides (which are commonly detected in agricultural areas), polycyclic aromatic hydrocarbons (PAHs), and lead. PAHs and lead are commonly detected at relatively high concentrations throughout urban areas. PAHs are generated from diverse sources as a result of incomplete combustion of organic materials and lead has been widely distributed as vehicle and aircraft exhaust.

It is likely that anthropogenic background chemicals will be detected at many installations during routine sampling and analysis. Thus, it is imperative that project teams conduct carefully designed background analyses to distinguish background conditions from site releases in order to avoid unnecessary remediation. Failing to make these distinctions will confound the investigation and remedial decisions because some background chemicals pose risks *even* at background levels. Unless background conditions are defined, project teams may unnecessarily remediate sites where no releases have occurred, because the naturally occurring or anthropogenic background levels indicate unacceptable risks. Table 1 presents the background levels (range of concentrations detected throughout the U.S.) and corresponding background risk levels for a few chemicals. As this table indicates, background levels of arsenic and chromium are commonly detected at levels that pose unacceptable risks.

A technical challenge for the project teams, will be at those installations where a particular anthropogenic chemical (not associated with Navy activities) is ubiquitously distributed in a region but a Navy release involving the same chemical has also occurred. In this instance, it is necessary for project teams to differentiate the two chemical sources to determine the Navy's cleanup responsibility, should remediation become necessary. For example, if the total chemical concentration poses unacceptable risks or exceeds the target remediation goal, it may not be practical to achieve cleanup to pristine conditions due to the ambient or on-site background level of the chemical. This concept is depicted in Figure 4 (adapted from EPA 1992) which shows the concentration groupings for on-site and off-site

FIGURE 4
CONCENTRATION GROUPING ASSOCIATED WITH SAMPLES TAKEN FROM A HAZARDOUS WASTE SITE

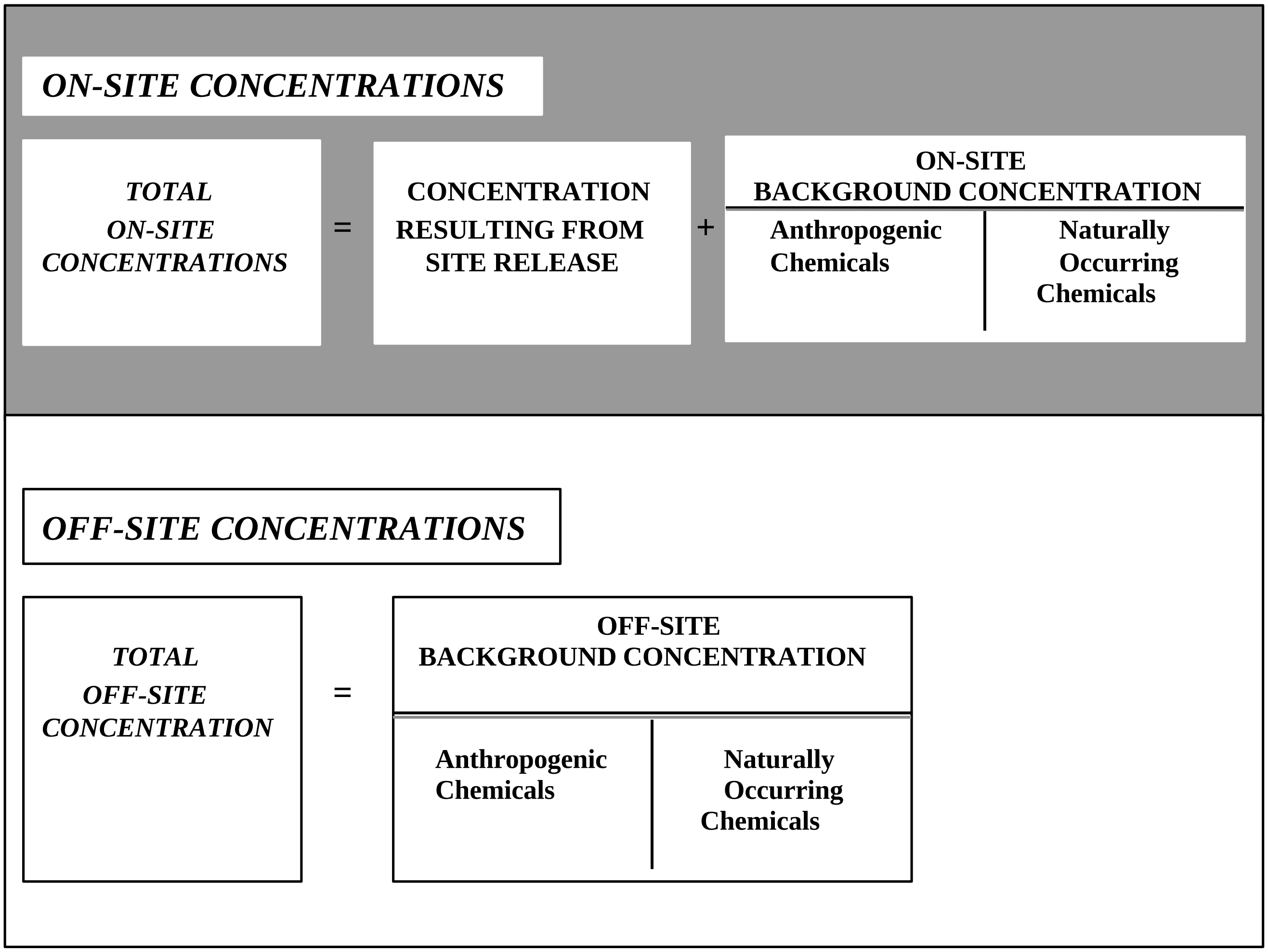


TABLE 1

NATURALLY OCCURRING AND ANTHROPOGENIC BACKGROUND CONCENTRATIONS AND CORRESPONDING RISK LEVELS

ANALYTE	BACKGROUND CONCENTRATION (range-ppm in soils)	BACKGROUND RISK (cancer risk range)	PRELIMINARY REMEDAITON GOALS[(4)] (ppm)
Naturally Occurring Background			
Arsenic	**0.1 to 97** [(1)]	**3E-7 to 2E-4** [(3)]	**0.38** [(5)]
Chromium (hexavalent form)	**1.0 to 2,000** [(1)]	**3E-8 to 7E-5** [(3)] **5E-6 to 1E-2** [(4)]	**30.0** [(5)] **0.2** [(6)]
Anthropogenic Background			
Polycyclic Aromatic Hydrocarbons (Benzo[a]pyrene)	**0.4 to 1.3 (rural soils)** [(2)] **0.4 to 650 (urban soils)** [(2)]	**7E-6 to 2E-5** [(3)] **7E-6 to 1E-2** [(3)]	**0.061** [(5)]

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

(2) ATSDR, 1987b. *Draft Toxicological Profile for Benzo(a)pyrene.* Agency for Toxic Substances and Disease Registry, 1987b.

(3) Background risks correspond to background concentrations and are based on residential exposure assumptions used to derive EPA Region IX preliminary remediation goals (PRGs).

(4) Background risks correspond to background concentrations and are based on residential exposure assumptions used to derive California Modified PRGs.

(5) EPA Region IX PRG. Concentrations downloaded August 1998 from: http://www.epa.gov/region09/waste/sfund/prg

(6) California Modified PRG. Concentrations downloaded August 1998 from: http://www.epa.gov/region09/waste/sfund/prg

concentrations containing anthropogenic and nonanthropogenic sources. As this figure shows, each sample taken from an operations area may contain three different components (natural, anthropogenic, and release) of a particular chemical. The goal for project teams is to determine what component of the total should be apportioned to each category to ensure that only that component associated with the release is targeted for remediation. It should be noted that anthropogenic and nonanthropogenic background conditions can be localized (from a point source) or generalized (from non-point sources) (EPA 1994b).

It is essential that naturally occurring and anthropogenic background levels be established to accurately identify chemicals of concern and to estimate site risks specifically associated with Navy releases. For example, anthropogenic (background) levels of PAHs exist at relatively high concentrations in many parts of the U.S. These compounds are generated from automobile exhaust; industrial operations; forest and urban fires An example is the great San Francisco fire of 1906 which released tons of PAHs into the air which subsequently accumulated in sediments, soils, and water. Because PAHs are resistant to degradation, they are still present at high concentrations. Thus, PAHs will confound many investigations in the San Francisco Bay area because many installations have been constructed—either in part or in whole—with PAH contaminated dredged sediments or surface soils, which must be distinguished from site releases.

1.6 IDENTIFYING CHEMICALS OF CONCERN

One of the most important uses of background analyses is for identifying chemicals of concern associated with Navy releases. EPA (EPA 1989b) risk assessment guidance provides procedures for dealing with naturally occurring background chemicals in the human health risk assessment, stating:

> "If inorganic chemicals are present at the site at naturally occurring levels, they may be eliminated from the quantitative risk assessment."

It also states that anthropogenic background chemicals can also "be omitted entirely from the risk assessment," but their exclusion requires additional site-specific information and analysis to characterize concentrations in the general geographic region. However, as previously discussed, establishing anthropogenic conditions to estimate potential site-related risks is important in making informed risk management decisions.

Unlike naturally occurring background chemicals, which should be completely eliminated from further consideration in all Navy investigations, anthropogenic chemicals can—but should not automatically—be eliminated from Navy investigations. However, all evaluations involving anthropogenic chemicals should be made separately from analyses conducted for site releases. This is particularly important when risks are estimated to determine whether remediation is necessary. When risks associated with anthropogenic conditions are added to risks associated with site releases, stakeholders may conclude that Navy activities included the release of anthropogenic background chemicals. Stakeholders would then expect anthropogenic chemicals to be targeted for remediation. While the cumulative risk associated with background *and* site release may exceed an acceptable risk level (triggering remediation), when evaluated separately, the site release may pose insignificant risks. In this case, cleanup would be unwarranted.

While background levels may ultimately play a role in risk management and/or remedial decisions regarding Navy releases, the Navy's response should not extend to remediating below background chemical levels. How the background information is documented and presented to stakeholders is equally important as the background investigation itself. Stakeholders should be able, while reading the documents or listening to a presentation, to immediately differentiate between background and site releases and their respective risk levels. This can be most easily accomplished by clearly communicating background information in a discrete chapter or appendix.

According to EPA (EPA 1989b), establishing background conditions is an integral part of the HHRA. EPA emphasizes that how background information and risks are presented in a risk assessment is a key component of describing overall site-related risk:

> "At a *minimum*, the discussion should include confidence that the key site-related contaminants were identified and discussion of contaminant concentrations relative to background concentration ranges."

Project teams will be most effective when background conditions are established as early in the investigative (characterization) or reuse categorization process as possible. Background analyses should *always* be conducted prior to identifying COCs associated with site releases. When the background analyses indicate that site-specific and naturally occurring background chemical concentration populations are not statistically different, the chemical should be eliminated as a COC and not evaluated in the risk assessment.

Statistical significance can occur when there is only a small difference between site and background means or medians if the number of samples is sufficiently large. Also, a large difference in site and background may exist and not be detected by a statistical test if only a few data are used in the test. Hence, it is important to specify during the DQO process the difference in the site and background means or medians that is *important* to detect with *specified probability* using the selected statistical test. Then if the test is conducted with a number of representative samples sufficient to achieve those performance goals, the statistical significance or non-significance is a suitable criterion for deciding whether a chemical is a COC.

In summary, project teams should proceed cautiously before designating a particular analyte as either a COPC or COC because these designations impart special significance for some stakeholders. Furthermore, many organic and inorganic anthropogenic and naturally occurring background chemicals pose unacceptable risk even at background levels. Lacking confirmation that these chemicals are present at background levels can lead to inappropriate designation as a COC and automatically trigger an unnecessary and impractical remedial response.

2.0 DEVELOPING BACKGROUND DATA SETS

Developing background data sets that accurately represent anthropogenic and nonanthropogenic conditions is one of the two primary steps in any background program. Although significant effort is typically focused on data evaluation and statistical analyses, the analytical process cannot begin until representative data sets are developed. No statistical approach can overcome the inherent problems posed by inaccurate or nonrepresentative background data sets. Therefore, project teams should devote as much time and effort to developing representative background data sets as will be focused on subsequent data analysis and statistical tests. Figure 5 presents the purpose and activities associated with developing background data sets. The following sections provide additional instructions.

2.1 OVERVIEW OF EXISTING DATA SETS AND INITIATING BACKGROUND PROGRAMS

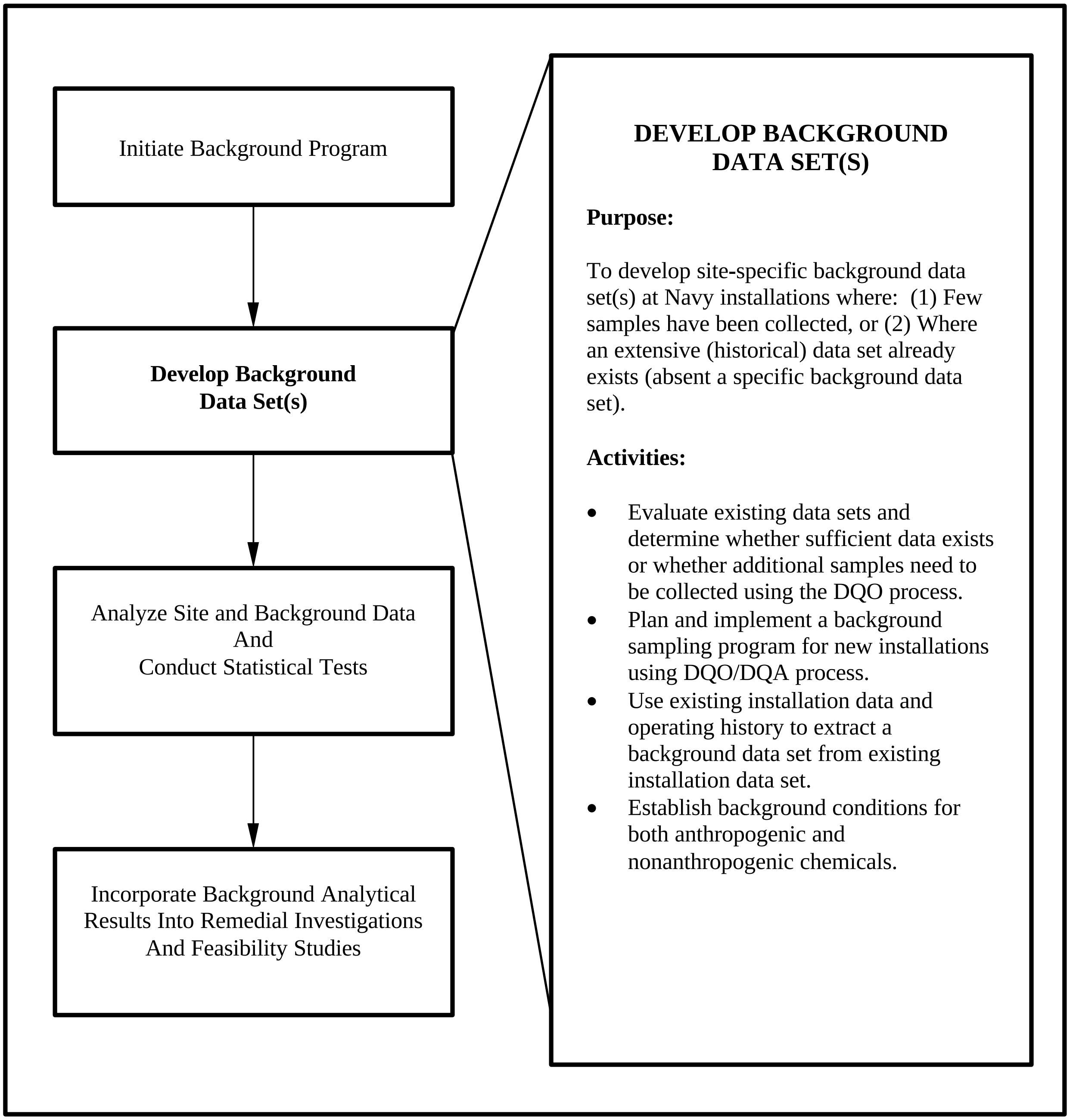
Navy project teams are involved in many different types and stages of environmental investigations. Some new investigations will have limited data sets, while more mature investigations have developed data sets with hundreds of samples. Regardless of the stage of data set development all installations should start a background program by formally assembling a project team to focus all activities using the DQO process.

Data with varying quality have been collected for diverse investigative purposes. Because a major component of background analyses involves data analysis, data quality should first be ascertained. While some data sets meet EPA's contract laboratory program (CLP) requirements, other data sets contain field screening data. This requires that data be carefully evaluated before combining data sets to ensure that data of comparable quality and type is appropriately combined or pooled. See EPA's DQA (QA/G-9) process and DataQUEST (QA/G-9D) for further details.

It is important to determine the underlying sampling design that was used during sample collection. Three acceptable sampling strategies are described in EPA guidance (EPA 1989b): authoritative sampling, random sampling, and systematic sampling. For the most part, authoritative sampling has been routine at most installations because the goal has been to define the nature and extent of contamination. With this approach, the most contaminated areas, which are expected to have the highest chemical concentrations, have been the focus of sampling activities. One of the principal advantages of this approach is that areas of suspected elevated concentrations can be specifically targeted.
According to EPA (EPA 1992):

"The definition of 'background' establishes the baseline from which site-related concentrations are measured. Usually, instead of a single background concentration, a distribution of concentration levels exists. Therefore, it may be necessary to characterize background statistically by estimating both the average background concentration (e.g. geometric or arithmetic mean) and the distribution of background concentrations around the mean (e.g. variance)."

FIGURE 5
STEP 2-DEVELOP BACKGROUND DATA SET(S)



This concept is important and should be emphasized because background levels should not be represented by a single concentration. This is particularly true for sites in which the geochemical conditions are heterogeneous. Background conditions should be represented as a population. Consequently, differentiating background and chemical releases should be based on the difference in populations. Taking many samples in a relatively small area can bias the data set and result in an abnormally low variability, incorrectly indicating the geochemical conditions are relatively homogeneous and not accurately representing the true geochemical variability of the entire site. Also, this type of data set cannot be used to assess spatial trends. Random and systematic sampling designs typically yield more information about operation areas and better characterize sites.

There are a number of problems that can arise when analyzing environmental data.

Most existing data sets have a significant number of non-detect or censored data, which can create difficulties during data analysis and statistical testing. Additionally, detection limits have changed amongst the laboratories over time, which increases the difficulty in combining data sets.

Project teams should make it a high priority to use all existing data despite data quality problems and limited information whenever possible. The time, effort, and cost associated with each sample taken should be considered before the datum is discarded. Even when the data quality is questionable, the results can sometimes indicate whether additional sampling is required or where additional samples should be collected. Data should be eliminated from data sets only when they are clearly shown to be of such poor relative quality that they confound the analysis or are outliers (e.g. sample representing contamination).

The remaining sections of this chapter provide guidance on how Navy project teams should develop background data sets. Due to the diversity of installations, it is necessary to develop an overall strategy that can be applied at any installations in California. For those installations that are beginning an investigation, the strategy involves assembling the background project team with diverse technical expertise. A typical team may be comprised of: a statistician, a hydro-geologist or geologist, toxicologist, and chemist; to develop and implement a carefully constructed sampling and analysis plan to collect background data. This plan may be initiated with other sampling events, such as sampling to determine the nature and extent of contamination.

For those installations where extensive data sets already exist, it may be impractical, costly, and unnecessary to initiate a new sampling program to collect background data. For many installations, background data are likely to be present within the existing installation data set. Rather than initiating an expensive and time consuming background sampling program to collect new data, the project team should focus its efforts on extracting background data from the existing database whenever possible and appropriate. This is a cost-effective approach that eliminates sample duplication. A careful evaluation of the existing database with regard to quality and quantity for the intended purposes will provide Navy project teams with the rationale to justify collecting additional samples to fill data gaps or augment the existing data set. EPA’s GEO-EAS (EPA 1991) software package may be useful for selecting the best new sampling locations when additional sampling is warranted.

It should be emphasized that extracting a background data set from an existing database is consistent with regulatory guidance and current practices. As an example, the analytical approach used by EPA Region IX, the California EPA Department of Toxic Substances Control (DTSC), and the Navy to evaluate background conditions and identify chemical releases is based on the presumption that background data sets already exist in many installation databases, making further sampling unnecessary.

To extract background data, DTSC recommends using graphical techniques (cumulative frequency distribution plots, CFDPs) to visualize the entire installation-wide database in one graph. Background data are subsequently extracted and chemical releases identified during visual inspection of the apparent inflection or break points along the curve. Although this technique may provide general information about the installation, it oversimplifies background analyses and does not account for differences in background geochemistry across the installation, anthropogenic background conditions, multiple data set subpopulations (or heterogeneity) and other complicating factors. While graphical techniques (including CFDPs) can provide some insight to project teams during the initial data analysis, CFDPs (or other graphical representations) should not be solely relied upon to distinguish releases from background conditions. From a statistical standpoint, the CFDP approach results in unacceptable Type I (or false positive) error rates. Examples of some of the complications involved with the CFDP approach are presented in the companion Handbook for Application of Statistical Techniques. Background conditions are more accurately established by using the battery of analytical and statistical methods which are described in Chapter 3.

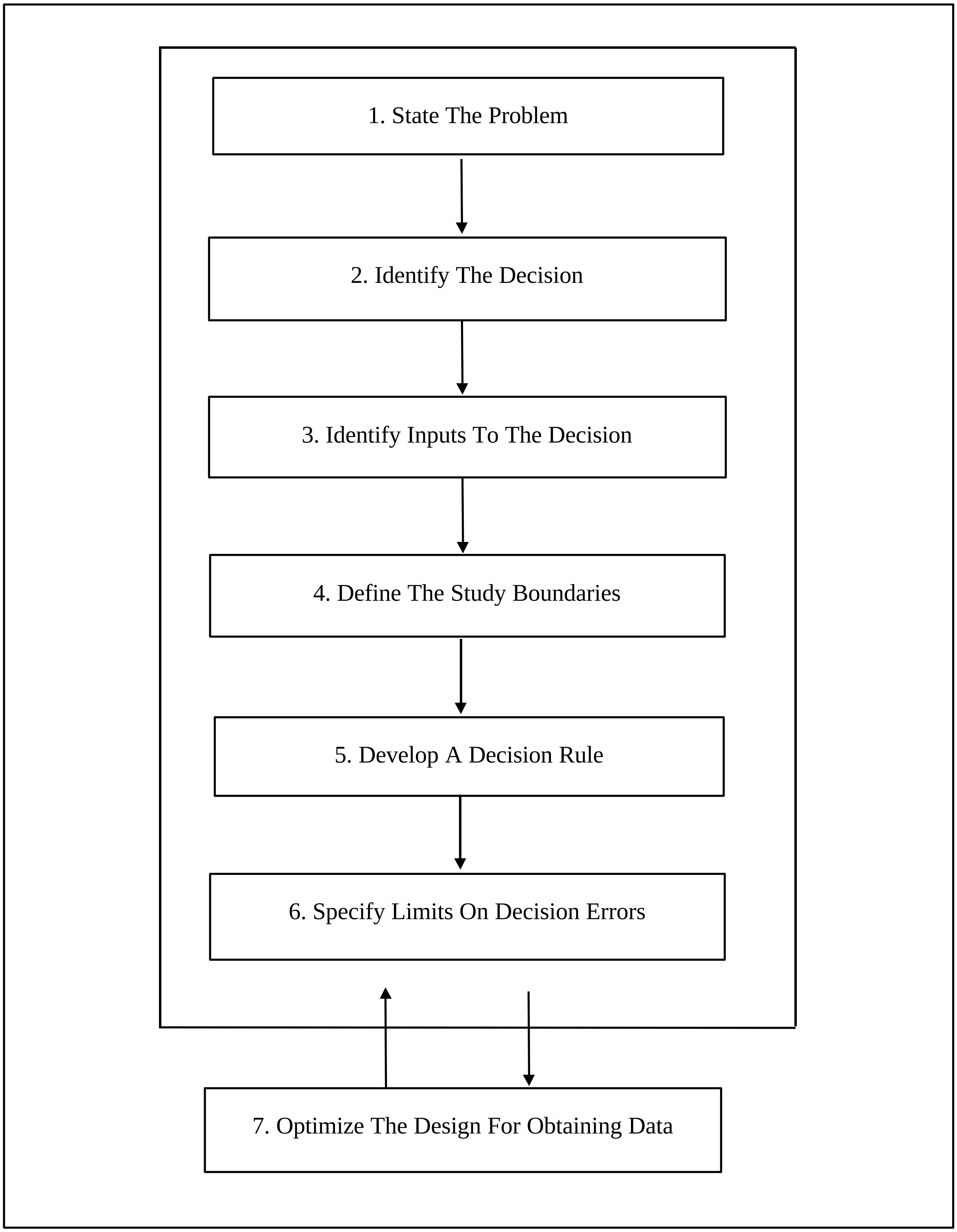
Background data can be successfully “extracted” from existing installation data when a focused and carefully planned data evaluation strategy is implemented. For example, samples collected to determine the nature and extent of organic contamination (at sites where solvents were used exclusively during Navy operations) are routinely analyzed for many inorganic elements in addition to solvent chemicals. When it can be confirmed that inorganic chemicals, such as arsenic or chromium, were not used or released during operations, the inorganic chemical data likely represent background conditions for the area and may be included in the installation-wide background data set.

2.2 DATA QUALITY OBJECTIVES AND DATA QUALITY ASSESSMENT

All project teams should follow a logic-based decision-making framework during all stages of the background analysis to first determine *if* additional samples need to be collected and if so, subsequently determine the quality and quantity of samples that need to be collected. EPA DQO/DQA guidance (EPA 1994a, 1996b) provides an excellent framework for Navy project teams. All project teams should use the DQO/DQA framework to guide decision-making processes, regardless of whether background sampling is being planned for newly investigated installations, or when teams are extracting background data sets from an existing installation database.

DQOs are defined in EPA guidance (EPA 1994a) as qualitative and quantitative statements to ensure that data of known and documented quality are obtained during the investigations to support decisions. Figure 6 shows the principal seven steps involved in the DQO process. The results from each step influence subsequent decisions and the process is flexible in that previous steps may be revisited as necessary. The process is designed to ensure that the type, quantity, and quality of environmental data used to make decisions are appropriate and adequate for the intended application. The DQO process was developed to aid data collection and to allow decision makers to define appropriate data requirements and acceptable levels of decision errors during the planning process.

FIGURE 6
THE DATA QUALITY OBJECTIVE PROCESS



EPA has recently developed DQA guidance that project teams should also use during all phases of the background analysis (EPA 1996b). The DQA process is defined as "the scientific and statistical evaluation of data to determine if the background and site data are of the right type, quality, and quantity to support their intended use" (EPA 1996b). The DQA process is a five-step evaluation process depicted in Figure 7.

A detailed discussion of the DQO/DQA process is provided in EPA (1994, 1996b). The importance of following the DQO/DQA process cannot be overstated. It can assist project teams in conducting scientifically rigorous data collection and analysis which underlies the Navy's background analytical approach.

2.3 DEVELOPING BACKGROUND DATA SETS FOR NEWLY INVESTIGATED INSTALLATIONS

For newly investigated Navy installations where few samples have been collected, the first step involves developing a background sampling and analysis plan. The steps involved are presented in Figure 8. It is advantageous to collect background data to establish background conditions for the Navy installations as early as possible to identify COPCs and COCs to focus all subsequent investigations on Navy chemical releases. Developing a representative background data set and following the Navy's statistical approach will result in the appropriate identification of COCs.

Eliminating background chemicals during the initial stage of all investigations is a practical approach that will streamline every phase of the Navy's investigations, from identifying COPCs or COCs to selecting the appropriate remedial alternative. An additional advantage of employing the Navy's background analytical approach early is that project teams will avoid the appearance of "deselecting" chemicals during the final stages of the investigation or at remedy selection. Deselection of chemicals late in the investigation or at the remedy selection stage should be avoided as it may cause confusion, concern, or anxiety among stakeholders when it should and could have been done early in the process. Project teams dispel the perception that the Navy is ignoring a chemical release by conducting background analyses *before* developing a site conceptual model (that may be incorrect). Every effort should be made to carefully identify site-specific COCs related to releases as early as possible, which will more accurately describe site conditions for the stakeholders and facilitate remedy selection.

2.3.1 Identifying Nonimpacted Background Areas

The first step in establishing background conditions involves identifying non-impacted areas that represent the underlying geologic and hydrogeologic conditions. Anthropogenic background conditions must also be defined, for example those installations that have been constructed with dredged sediments (likely contaminated prior to installation construction), or are located in agricultural regions where pesticides have been historically used, or may have been impacted by off-site sources of contamination.

FIGURE 7
THE DATA QUALITY ASSESSMENT PROCESS

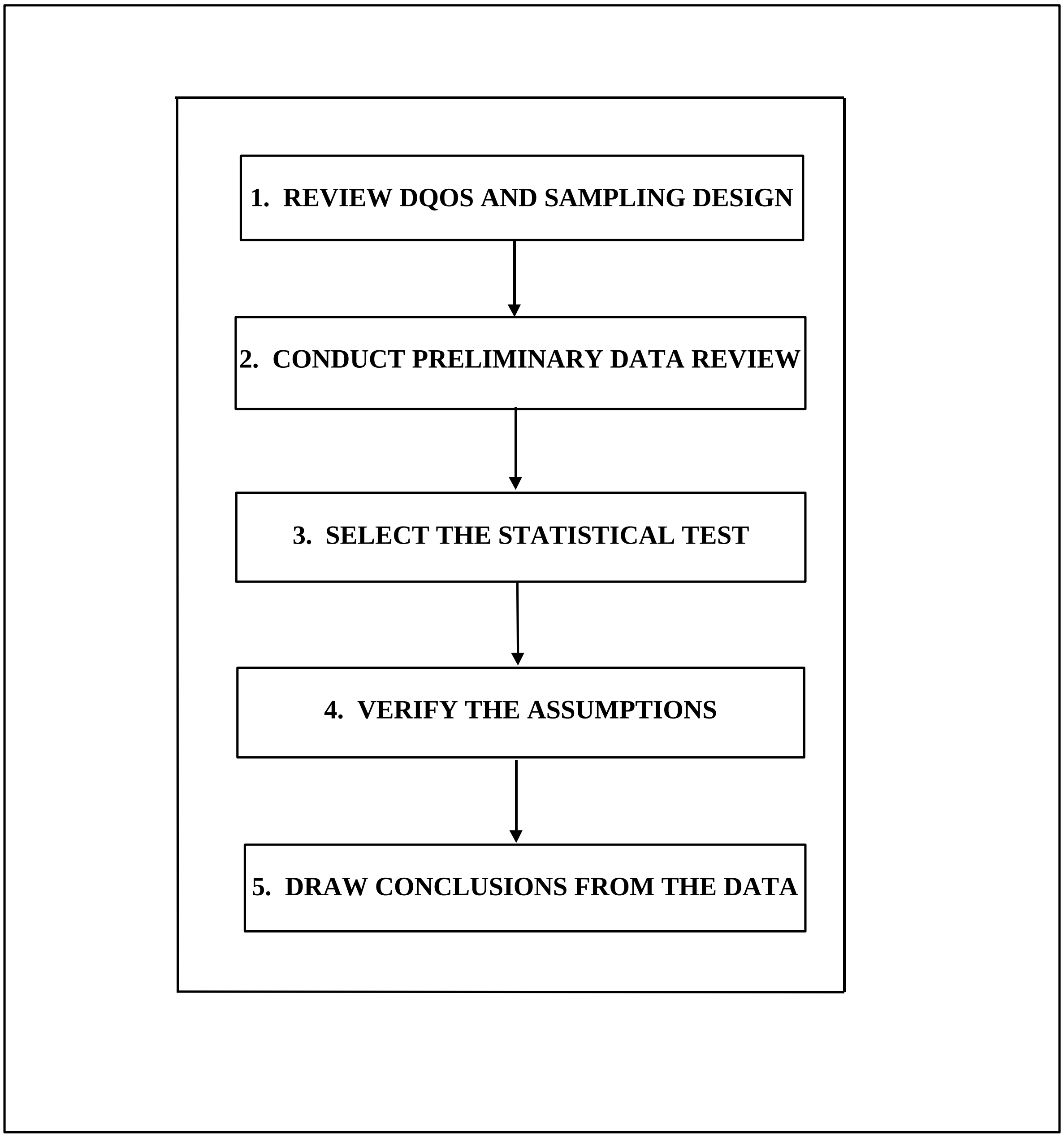
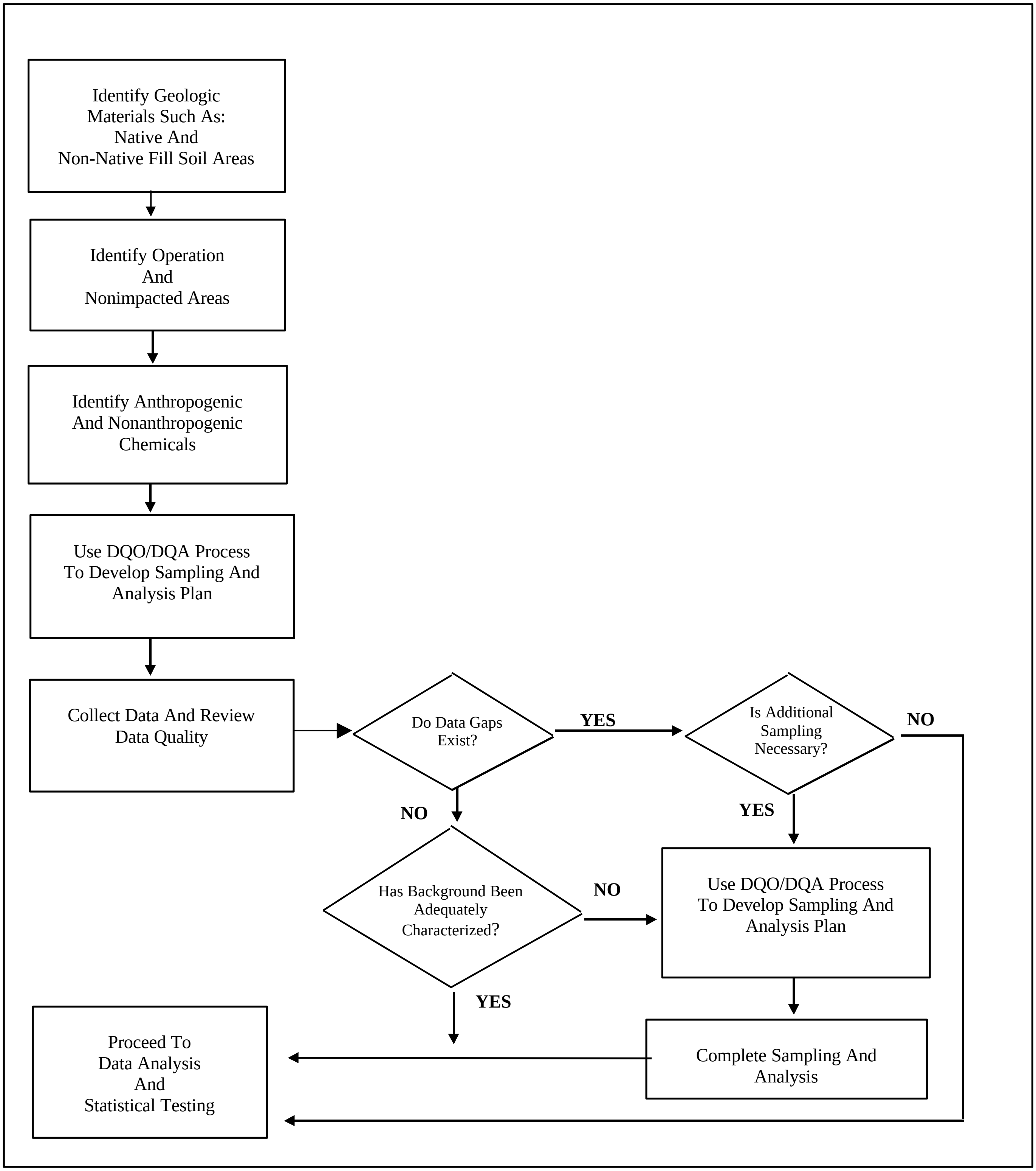


FIGURE 8
DEVELOPING A BACKGROUND DATA SET(S) FOR NEWLY INVESTIGATED INSTALLATIONS WITH LITTLE OR NO PRIOR DATA



EPA (EPA 1989b) defines non-impacted background areas as, "those that have not been influenced by site activities involving a chemical release." The range of background levels needs to be established for each environmental medium of concern. Water and soil background samples may be collected from different areas. DTSC (DTSC 1992, 1994) also recommends that samples be collected at or near the site, "but not in areas likely to be influenced by the contamination and/or facility operations (past or present)."

Several factors must be considered when potential background sampling areas are identified. The most important include geological and hydrogeological features, upgradient sources of contamination, and anthropogenic impacts on background areas or background sampling locations on site.

Care should be taken to ensure that background and site samples are of similar soil composition and groundwater type. Correctly matching the geologic and hydrogeologic underlying conditions will generally ensure that background and site data have similar chemical properties so that any differences in chemical concentrations may be attributable to site activity and releases.

The effect of potential upgradient sources must also be evaluated. Upgradient sources of groundwater contamination or airborne emissions that may increase soil concentrations should be carefully evaluated. If a potential background area is affected by an upgradient, but nonsite-related chemical source, background samples may contain organic chemicals that may invalidate the area as background. The sample area may be validated as representing inorganic chemicals by demonstrating that the organic chemical source was upgradient and did not affect the concentration of inorganic chemicals.

When both the operation area and the background locations are equally affected by the same upgradient source (for example exhaust stack emmisions), then this nonsite-related anthropogenic contamination source may complicate identifying those chemicals that are related to operation activities. Although it is best to select areas that have not been impacted by any anthropogenic chemicals, this is not always possible. When anthropogenic chemicals are present, the area selected to represent background should be affected by the same level of chemicals as the operation area under investigation. With this approach, differences between the site and background data sets may be attributable to site activity and site releases.

Project teams should carefully evaluate the presence of organic contamination in background samples and neither discard nor accept the sample as representing background without supporting rationale. Although DTSC guidance (DTSC 1992) states "with few exceptions, one may assume that background levels for man-made chemicals are zero," it should be noted that low levels of organic chemicals are often detected in areas where no operations or releases have occurred. Accordingly, project teams should not automatically assume samples containing anthropogenic chemicals are evidence of contamination which will typically nullify the sample as a potential background datum. Project teams should follow EPA guidance (EPA 1989b), which suggests that the presence of ubiquitous, anthropogenic chemicals be considered as background and be carefully analyzed. Even in the event that project teams cannot confirm that the chemical represents anthropogenic background conditions, the sample may still yield valuable data for other chemicals.

In general, selecting background sampling areas requires the use of professional judgment and a thorough understanding of site operations, historical uses of the property, and the surrounding industrial sources unrelated to installation activities. The following approaches should be used to identify non-impacted background sampling locations (EPA 1992).

- Use historical records and aerial photographs to determine where and how contaminated wastes were handled and may have been released at the site
- Use tax maps and historical information to determine uses of the property and surrounding properties that may have affected the Navy installation
- Sample from under stationary objects, such as building foundations, cemetery headstones (if permitted), and trees, or sample in natural areas not currently disturbed by human activity (woodlands)
- Avoid collecting background samples in locations where abnormally high nonsite-related anthropogenic concentrations of contaminants are likely to be found, such as along railroad track beds and highways

2.4 EXTRACTING BACKGROUND DATA SETS FROM EXISTING DATABASES

For installations that have been sufficiently sampled, it may not be necessary to collect additional samples specifically for background analyses. Samples already in the data set, collected to investigate nature and extent of contamination, may be used to provide background information needed to establish the naturally occurring levels of most inorganic analytes, provided the data are representative of the background population.

Where petroleum or organic chemicals were exclusively used in site operations, the concentrations of inorganic chemicals detected in samples may have been affected, and further investigation should be conducted to ensure that these sample data accurately represent background levels. For these types of sites, it may be necessary to collect new background samples. However, project teams should review the location of samples already collected to ensure samples completely represent the background population.

Extracting background data from an existing installation data set should be considered since, developing new and specific sample plans, collecting samples, laboratory analysis, and data validation can be costly and delay investigations and clean-up. Resampling an entire installation can delay cleanup, property transfer, and reuse; therefore, resampling should be judiciously conducted and based on a cost-benefit basis. The DQO process should be used to determine if additional sampling is necessary, including clearly stated objectives and goals to guide all activities.

The analytical approach used to extract background data from an existing database will depend on the installation characteristics and how the installation was constructed. For example, extracting background data at installations constructed primarily with native soils should be a straightforward process. In comparison, developing background data for installations constructed on sites comprised of fill material (non-native soils or sediments) that were historically contaminated with anthropogenic chemicals prior to construction of the installation is highly complicated. For installations built with sediments, historical information should be obtained such as dredging, construction history, and maps to identify whether different background conditions exist. These conditions can arise when the installation construction occurred over many years or sediments from different locations were used to construct the base. It is likely that the sediments had various anthropogenic levels of chemical composition.

The first few steps of extracting background data should be similar for most installations. For example, one of the initial steps (after assembling the project team with diverse expertise) involves compiling all installation-wide environmental data. Next, this data set should be filtered by eliminating poor-quality or inappropriate data. These data could represent spurious laboratory errors or abnormally high detection limits. The data set is then winnowed by carefully reviewing historical site information, sample collection processes and analytical methods, data validation procedures, detection limits, matrix interferences, installation construction activities, and detailed site operations. Eliminating data anomalies before, rather than after, data analysis and statistical testing will prevent confounding the analyses.

Graphical techniques can and should be used for visualizing data. They can be particularly useful when examining large databases and can be used to identify data set subpopulations. However, overextending any single analytical technique can lead to erroneous conclusions. For example, CFD plots should often be used in conjunction with posting plots and spatial contour plots because important spatial information is lost when CFD plots are used alone. (Posting plots are a graph that shows the data values of the sampling locations.) A group of data may appear as a separate population cluster at the high-concentration portion of a CFD plot, but further analysis may reveal that there is no spatial relationship among the data. That is, the data do not represent a discreet location and each datum in the cluster may represent disparate and widely separated areas. Other errors can occur when different underlying geochemical characteristics are not taken into account. Examples of these graphical techniques are presented in the companion handbook.

2.4.1 Identifying Areas Containing Non-Native Fill

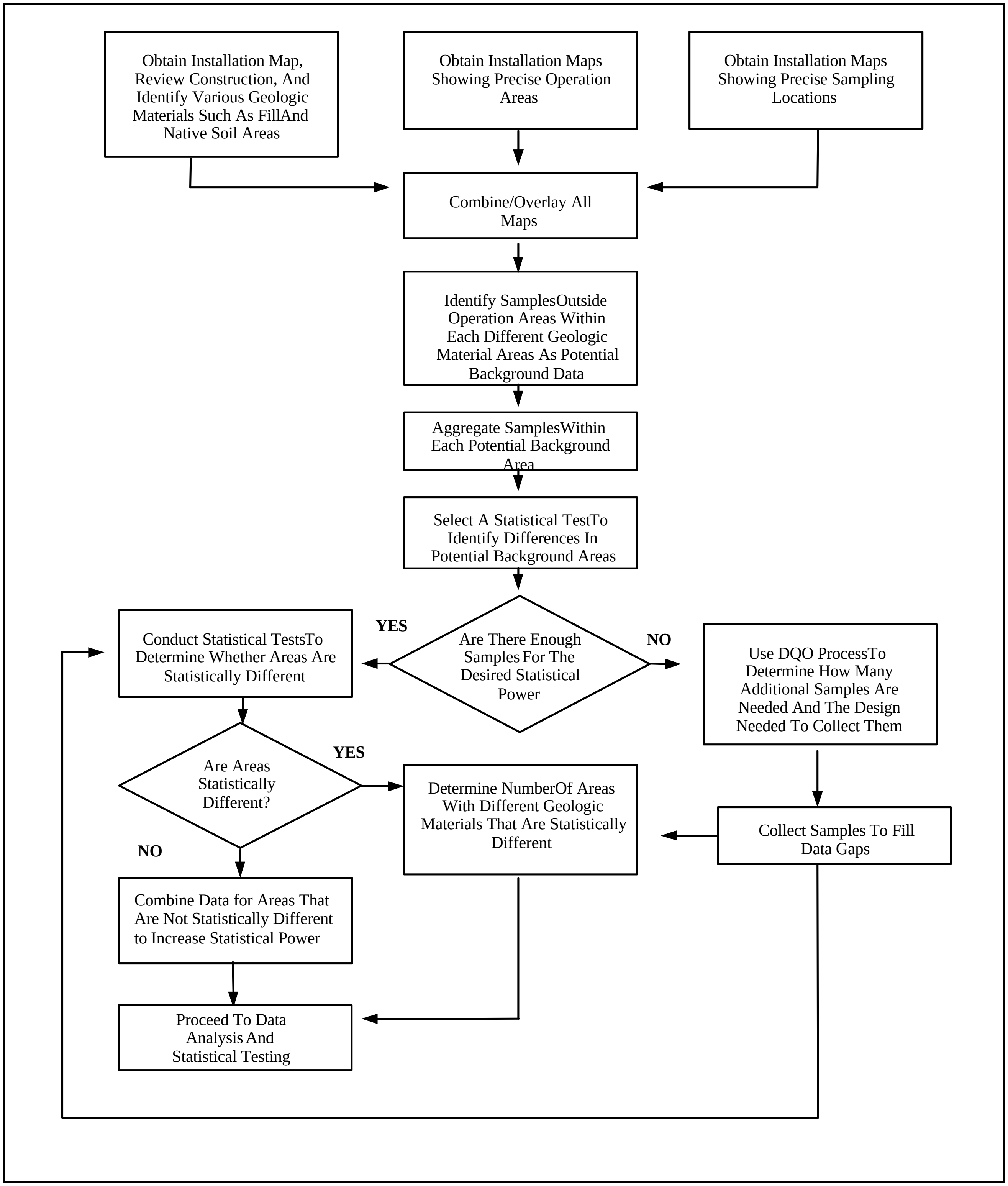
Many installations adjacent to coastlines were built upon, or constructed with fill comprised partly of dredged sediments. Fill was used to build up the ground elevation above sea level, create landfill boundaries, and increase the area of the installations. Non-native fill composition will likely be significantly different from native soils with regard to naturally occurring background chemicals. Additionally, sediments may have been contaminated with anthropogenic organic chemicals from historical non-point sources before they were dredged for construction fill. Therefore, project teams need to establish the levels of both native and non-native populations. The general steps involved in this process are presented in Figure 9.

For installations constructed primarily with dredged fill material, differentiating the anthropogenic fraction will present a challenge. However, the effort is necessary because the concentrations of many anthropogenic inorganic and organic chemicals may pose risks that will trigger remedial action unless they are correctly characterized as background.

For Navy installations where different sources of dredged fill were used for different areas of the base, historical records and maps should be used to identify different areas of anthropogenic background conditions. This is consistent with EPA guidance (EPA 1989b), which states:

> "If the variability is great, it may be necessary to determine the mean background concentration (and variance) for each of these areas separately, rather than estimating one overall mean background concentration (and variance). In such cases, the site could be partitioned into several distinct sectors to compare the measured site-related concentrations against the background concentrations on a sector-by-sector basis."

FIGURE 9
EXTRACTING BACKGROUND DATA SETS
FROM EXISTING DATABASES



Areas of non-native fill can typically be identified through base records and aerial photographs taken to document the chronological development of the installation. Data from different regions should be evaluated both horizontally and vertically. For instance, when a release occurs at the soils surficial level, the relative concentration typically decreases with increasing depth. Increasing concentrations with depth may indicate it is of anthropogenic origin associated with dredged fill material. The highest concentrations may be detected in deeper soil profiles because the sediment floor upon which the dredged material is placed would typically represent the most concentrated stratum. Horizontal differences in anthropogenic chemicals from dredged sediments can also result from different dredged material from different sources used during construction. Evaluating the chronological sequence of how the installation was constructed (where different areas may have been constructed many decades apart) may provide important clues as to the type of anthropogenic comparisons that should be made.

When the background composition of different areas of the installation are similar (not significantly different), data sets should be combined to increase the power of data analysis and statistical testing. The two data sets should be inspected in order to verify that approximately the same Quality Assurance measures were used in order to generate the data. If different measurement methods were used to analyze the data, reference to an expert may be necessary to assure that the data are comparable. The actual method by which the physical specimens were collected prior to analysis should be examined to assure that the data collection methodologies are comparable. For example, if one data set consisted of single readings from soil cores, while the other consisted of a single value from a series of composite sample cores, merging the data sets may not be advisable. The same parametric and non-parametric statistical tests that are used to determine whether two populations are different in background comparisons should be used to evaluate the appropriateness of combining two data sets. When there is no statistical difference between data sets and sufficient representative samples were used in the statistical tests, it can be assumed that the two data sets represent the same geochemical conditions and can be combined.

Project teams will also gain valuable insight if they can determine the depth of the original construction. This information can serve two important functions. First, it may provide the baseline concentrations of anthropogenic chemicals in the bay before construction began. Secondly, it may allow soils associated with subsequent fill activities to be characterized. Fate and transport processes should be considered during this phase. For example, because polycyclic aromatic hydrocarbons (PAHs) are not water soluble, leaching to lower depths following a release at soil surface will be negligible. Therefore, PAHs detected at lower depths would be more likely associated with the original fill material rather than indicating a release. The working hypothesis in this instance should be that the PAHs represent anthropogenic background associated with the fill material and project teams should collect information to evaluate this hypothesis.

Calculating the ratios of different organic and inorganic chemicals can also provide useful information regarding anthropogenic chemicals. For example, the ratio of chemicals concurrently released into the bay could be relatively constant and may differ from those that are typical of naturally occurring geologic conditions in regional soils.

After identifying non-native fill areas that have different baseline chemical concentrations, data may be extracted from the database using the same techniques as for naturally occurring background levels. That is, locations within the fill soil areas that have not been affected by site activities can be used to represent anthropogenic conditions. For example, if trees or other permanent vegetation were planted soon after the region of the base was built, samples from beneath or near the vegetation unlikely to have been disturbed by subsequent site activity could represent anthropogenic background conditions for the fill.

The same selection criteria for determining the appropriateness of samples for naturally occurring background conditions can be applied to anthropogenic samples. That is, any sample containing organic chemicals should be further evaluated to determine if the chemicals are potentially site-related, are due to ambient conditions from upgradient sources, or will affect inorganic chemical concentrations. Organic chemicals detected in fill soil that cannot be attributed to site activities should be evaluated to determine anthropogenic origin. Site history and operation records will provide vital information in this effort. The background conditions for anthropogenic chemicals should be documented and presented with the results of all background analyses. Furthermore, since sediment contamination is directly correlated to total organic carbon (TOC) and grain size, background samples must represent the full range of TOC and grain size that the site samples contain. The project team must be careful to compare background and site samples with similar TOC and grain size qualities or risk making false positive conclusions about the chemical concentrations in the sediments.

2.4.2 Identifying Data Gaps

Prior to or in conjunction with the DQA process, project teams should identify any data gaps. Data gaps should be evaluated within the existing framework of the site conceptual model (SCM). The SCM is developed with information on historical uses of the site; site operations, storage, and releases; upgradient sources of chemicals; chemical fate and transport; weather patterns; rate and direction of groundwater movement; and other site-specific information. The SCM should not be based solely on the number of samples collected, but should be developed to evaluate whether the data accurately represent the site and that all environmental media are adequately characterized.

An incomplete or incorrect SCM can result in erroneous conclusions. For example, if a chemical is unexpectedly detected in a sample it may be from an off-site source, and it may be mistakenly concluded that a past chemical release occurred at the site. Similarly, sites that were historically used for mining or agricultural purposes where residual heavy metals or pesticides may have previously contaminated soils or groundwater can be mistakenly attributed to Navy operations.

In contrast, when chemicals associated with known past site operations are not detected, it may indicate that historical records based on interviews, site operations, or chemical inventories are inaccurate. It may also indicate that sampling is incomplete or the sampling design is flawed. For example, if heavy metals are not detected in surficial soils at a site used as a plating facility, it may indicate that clean fill soil was added to the site after site operations ceased.

Data sets are sometimes considered incomplete simply because there are too few available samples for a particular site. Although it is usually preferable to have a large number of samples during site characterization, it is not always correct that a site will be better characterized, and statistical tests more powerful, if more samples are collected. Accurate site characterization depends on many factors, including the confidence level selected for statistical testing, data variability, sampling design, and spatial arrangement of sampling locations at the site. For example, a few samples collected from a relatively homogeneous site may result in more accurate site characterization than collecting many more samples from a heterogeneous site with high chemical variability.

Project teams should consider the desired level of statistical confidence. EPA guidance (EPA 1996b) for determining the number of soil and groundwater samples required to achieve a given level of statistical accuracy.

In summary, identifying data gaps is largely based on the project team's knowledge of the site, professional judgment, and statistical requirements. Cost associated with additional sampling and analysis, and data validation and evaluation should be considered when potential data gaps are identified. If, for example, there is uncertainty in the existing database but only a few COPCs with relatively low toxicity and low estimated risks have been identified, additional sampling may not be necessary because remediation may not be warranted. However, if there is uncertainty in the data set and several toxic COPCs have been tentatively identified which may pose unacceptable risk, additional sampling may be warranted. Particularly if remediation costs associated with those COPCs are likely to be high. Sites should never be targeted for remediation based on uncertainty.

3.0 ANALYZING DATA AND STATISTICAL TESTING

This chapter presents the data analysis and statistical testing methods that should be used by all project teams to establish background conditions. The purpose and activities involved in this process are presented in Figure 10.

Before beginning data analysis, project teams should carefully review the quality and representativeness of both site and background data sets which is an iterative process. During this review, data will likely be discarded due to inadequate data quality, or because the data are outliers and do not accurately represent the site conditions. After identification as a potential outlier, attention should be given to the magnitude of the observation with respect to the underlying model assumed for the entire distribution to characterize representativeness. For example, an observation identified as an outlier may be one if the underlying assumption of data distribution is Normal, but not be one if the underlying assumption is that of Lognormality. In addition, the sampling design should be reviewed to determine what areas are represented by the data sets and whether the objectives of the sampling design were met.

Brief descriptions of individual analytical methods and statistical tests are provided in the following sections. Additional details and examples for each method are presented in the companion handbook. Project teams are also strongly encouraged to refer to EPA's DQO/DQA guidance for further detail and examples of how to conduct statistical tests. To facilitate statistical testing (and to ensure consistency) project teams should acquire EPA's software programs DataQUEST (EPA 1996a) and Scout (EPA 1993)

Figure 11 presents the decision-making flowchart that integrates data analysis and statistical testing. The flow chart is designed to assist project teams in selecting the correct statistical test for the background analysis. This process should be followed at the earliest part of the investigation to identifying COCs which will focus and streamline all subsequent investigative phases.

3.1 ANALYZING DATA SETS

To reiterate, before statistical testing is initiated it is important that data sets be analyzed to ensure that background and potential release sites have been adequately characterized. It is also important that the characteristics of the data set be established to select the most appropriate and robust statistical test to determine whether a release has occurred.

The following subsections provide a brief overview of analytical methods for identifying data gaps, combining or pooling data sets, developing descriptive summary statistics, evaluating outliers, handling censored (non-detect data), goodness-of-fit tests, identifying spatial and temporal trends, evaluating stoichiometric (geochemical) correlations, and identifying hot spots.

FIGURE 10
STEP 3-ANALYZE SITE AND BACKGROUND DATA AND CONDUCT STATISTICAL TESTS

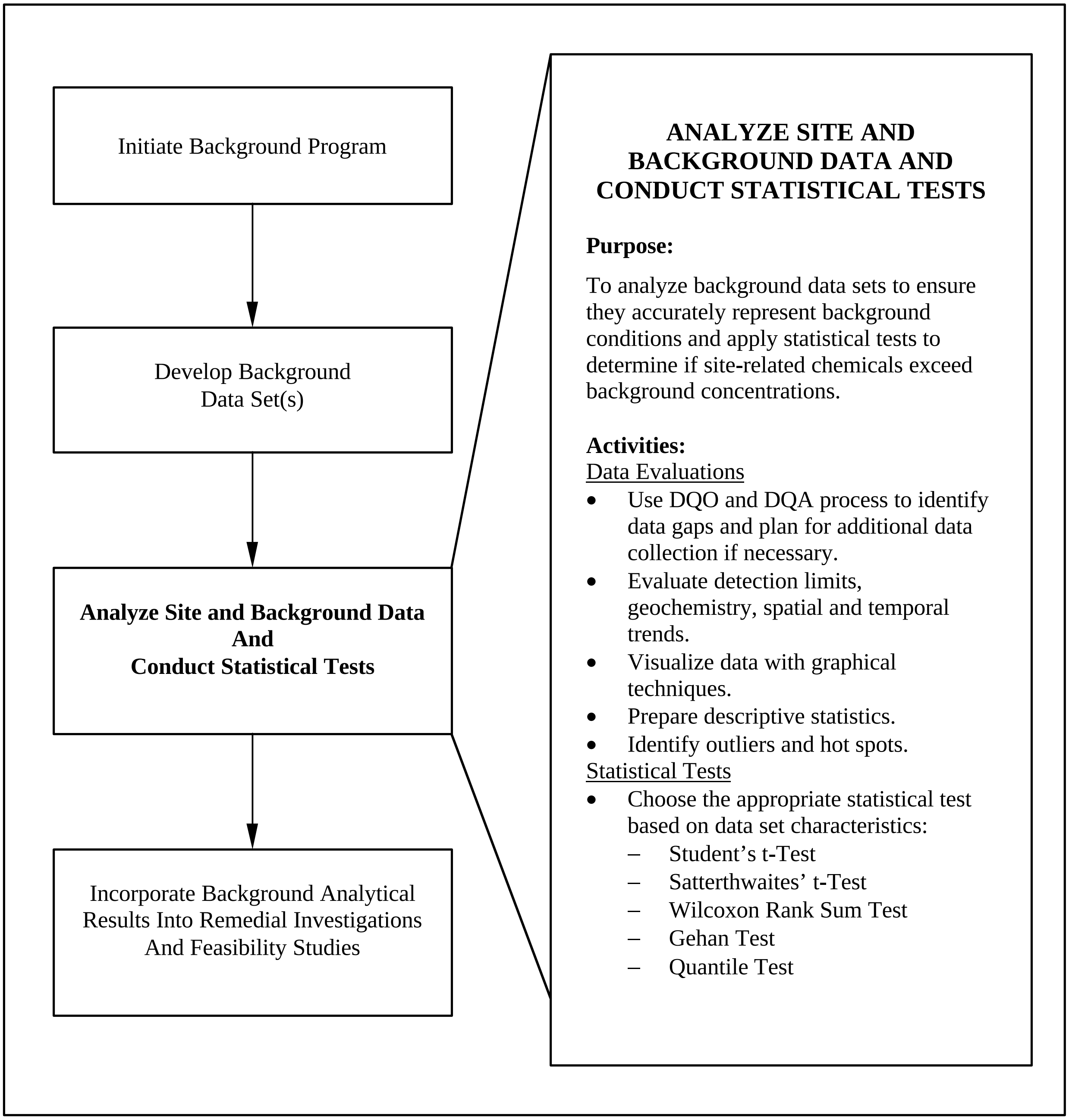
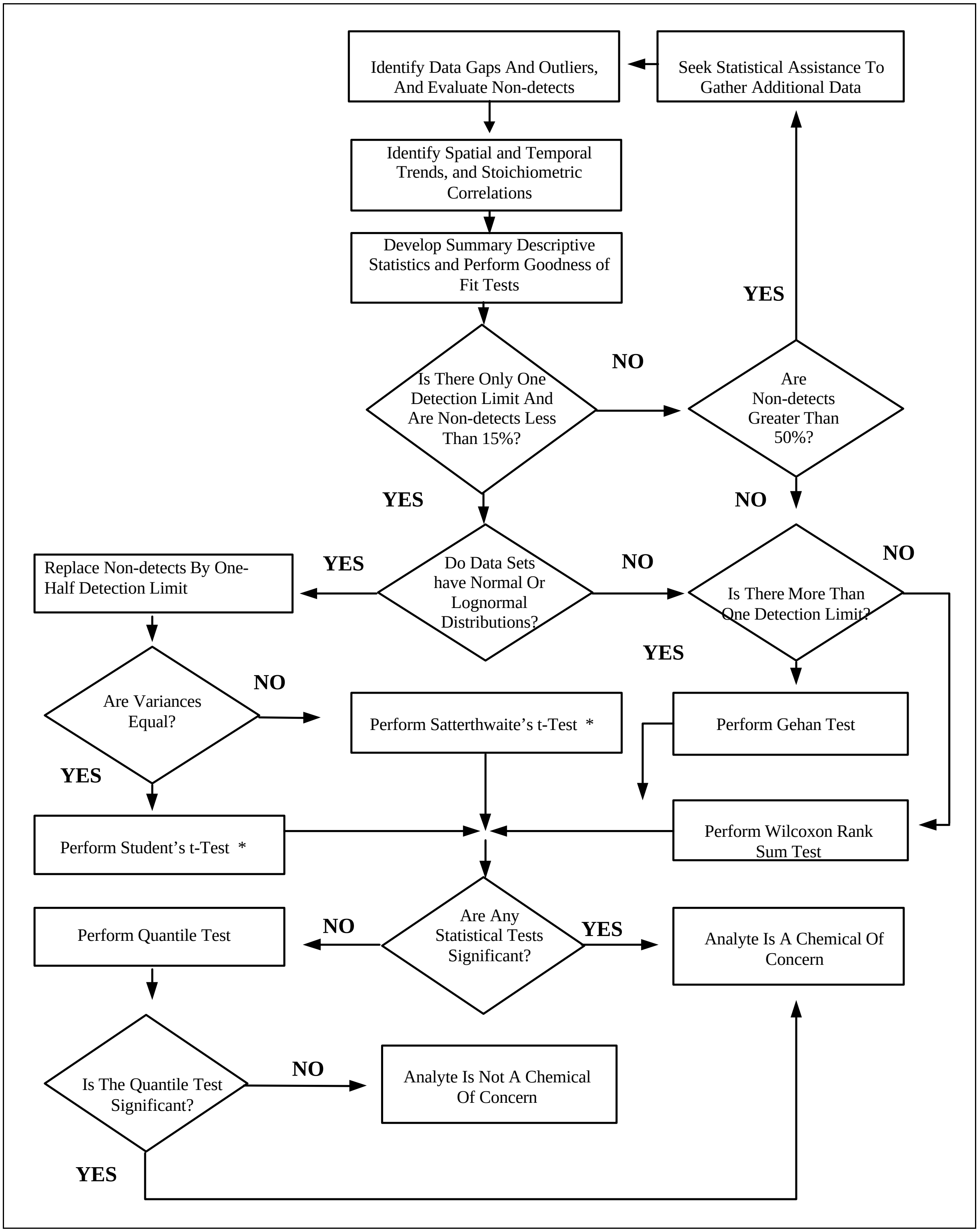


FIGURE 11
ANALYZE SITE AND BACKGROUND DATA AND CONDUCT STATISTICAL TESTS



* If the data are lognormally distributed then the t-test and satterthwaite's t-test are conducted on the natural logarithms of the data

3.1.1 Combining Data Sets

Samples have been historically collected in support of PA, SI, RI/FS, EBS, USTI, petroleum remediation and air and groundwater monitoring. All data sets contain potentially valuable information about the site under study. When appropriate, data sets should be combined to increase the power of statistical testing. However, it may not always be mathematically or statistically correct to combine data sets. Before combining data sets and conducting statistical testing, it should be confirmed that the data sets are similar. Project teams should review all data sets to be combined to ensure:

- Data sets have similar chemical composition representing the geographic or hydrologic units;
- Data sets are of similar type with regard to sampling and analysis protocols;
- Data sets have been analyzed with similar analytic methods and similar detection limits;
- Data quality (validated and nonvalidated data) is relatively constant among data sets;
- Descriptive statistics, (means, medians, etc.) and graphical plots such as the cumulative frequency distribution plots (also see Section 3.1.4) are similar for each data set.

When these criteria are met then it is acceptable to combine the data sets.

3.1.2 Evaluating Outliers

Data analysis and statistical testing will be confounded by the presence of outliers in both site and background data sets. Outliers are extreme high or low measurements that are sometimes referred to as "spurious" data because they are highly divergent from the main population of data. Outliers may arise from matrix interferences or errors in transcription, sampling technique, datacoding, analytical methods, or instrument calibration. Alternatively, what may appear to be outliers may simply represent inherent variability in the regional background geochemistry. This will be particularly true for background areas in which the geochemistry is heterogeneous. Apparent outliers may also represent hot spots in site data sets. When outliers are not identified and removed from data sets, they can disproportionately affect the statistical descriptors of the data set. That is, the mean can be biased toward the direction of the outlier(s) and artificially increase data variability and standard deviation. Ultimately, outliers can lead to flawed statistical testing and erroneous conclusions about background conditions. Therefore, it is important to identify outliers in both background and site data sets and eliminate them before conducting further statistical analysis.

Outliers that are obvious mistakes should be corrected whenever possible. Outlier tests can provide objective evidence that a suspect data point does not conform to the assumed underlying distribution of the data set. However, these tests should only be used to identify potential outliers that require further evaluation (EPA 1996b). After identification as a potential outlier, attention should be given to the magnitude of the observation with respect to the underlying model assumed for the entire distribution to characterize representativeness. For example, an observation identified as an outlier may be one if the underlying assumption of data distribution is Normal, but not be one if the underlying assumption is that of Lognormality.

According to EPA (EPA 1996b), five steps should be followed when evaluating data sets for outliers:

1. Identify extreme values that may be potential outliers
2. Apply statistical test
3. Scientifically review statistical outliers and decide on their disposition
4. Conduct data analyses with and without statistical outliers
5. Document the entire process

Outliers can be identified by visually inspecting graphical representations of the data set. Box and whisker plots, scattergrams, and ranked data plots are useful in identifying data that are much higher or lower than the main data set population. When potential outliers are identified through the use of one or more graphical techniques, a statistical test should be conducted to confirm that the data are outliers. Care must be taken when dealing with multivariate (multi contaminant) data sets as the identification of potential outliers may be difficult. An observation that is truly an outlier may be obscured by the masking effect of the multi contaminants and pass all individual single variate statistical tests. The EPA 1993 SCOUT program is useful in the identification of multivariate outliers.

Several statistical tests have been developed for evaluating the probability that extreme high or low data values are outliers. These tests have broad application for many different types of data sets. EPA (EPA 1996b) recommends applying one of four statistical tests: (1) Dixon's Extreme Value test (Dixon, 1953), (2) Discordance test, (3) Rosner's test (Rosner, 1983), or (4) Walsh's test (Walsh, 1958).

When the underlying probability density function (PDF; sometimes referred to simply as the "distribution") is normal, EPA (1996b) suggests applying Dixon's Extreme Value test when the sample size is equal to or less than 25, and Rosner's test when the sample size is equal to or greater than 25. When evaluating a single potential outlier, the Discordance test may be substituted for either the Extreme Value or Rosner test. When data do not conform to a normal or lognormal PDF, Walsh's test (which is a non-parametric test) should be used. If the data, other than the suspected outliers, are believed to have a lognormal PDF, then the tests are performed on the natural logarithms of the data.

It is important to emphasize that no datum should be discarded as an outlier based solely on the results of one of these statistical tests. There is always the possibility that the suspect outlier is an accurate measurement. Before the outlier is deleted from the data set, one of the above four statistical analyses should be performed on both the original data set containing potential outliers and the truncated data set minus the outlier. It is important to independently evaluate each potential outlier because the presence of two or more outliers may indicate that the data conform to a PDF not yet considered. For example, several unusually large measurements may indicate that the data set is more appropriately modeled by a lognormal distribution which is skewed to the right rather than a normal distribution which has a symmetrical appearance. The outliers may also represent a hot spot indicating a release at the site.

When suspect data are identified and subsequently confirmed as outliers, the scientific rationale for this identification should be fully documented regardless of whether or not they are deleted from the data set. Documentation should include the approach that was used to identify the outlier. This information should include the statistical outlier test and decision criteria, the results of the outlier test, and the results of the statistical analysis with and without the outlier.

Detailed instructions for performing Dixon's Extreme Value test, the Discordance test, Rosner's test, and Walsh's test are presented in the handbook and in EPA guidance (EPA 1996b).

3.1.3 Establishing Probability Density Functions With Goodness-Of-Fit Tests

PDFs are used to graphically model the data distribution. Common PDFs used to model environmental data include normal, lognormal, and Weibull distributions. Determining the PDF that best fits a particular data set is important for selecting the statistical test best suited for the data set to provide optimal statistical performance. The statistical tests presented in this procedure document were originally developed to be used with particular types of data sets. One of the most important characteristics of a data set is the underlying distribution of the data. For example, Student's t-test may be quite useful for testing data that are distributed normally or lognormally. The Student's t-test may not provide the needed statistical power to determine differences between site and background populations if the underlying distributions are not normal. Hence, to determine whether site and background data sets are significantly different it may be useful to first conduct a goodness-of-fit test to determine the best statistical test.

Two of the most important distributions for tests involving environmental data are the normal distribution and the lognormal distribution (EPA 1996b). Non-parametric tests should be used for data sets that do not follow either of these two PDFs.

The Shapiro-Wilk W-test and the D'Agostino test are goodness-of-fit statistical tests recommended for testing if the underlying PDF for a data set has a normal (or lognormal) distribution. Additional goodness-of-fit tests can be found in EPA (1994b, 1996b) and Gilbert (1987). In addition to these statistical tests, graphical techniques such as histograms and cumulative frequency plots can be used to visualize the data set distribution. Although statistical tests and graphical techniques separately yield information about the PDF, greater confidence in the modeled data set distribution is achieved when they are used in combination.

3.1.3.1 Shapiro-Wilk W-test

The Shapiro-Wilk W-test Gilbert (1987, Section 12.3.1) is a powerful and highly recommended test that can be used to evaluate whether a data set PDF is normally or lognormally distributed. The W-test can be used if the total number of samples is less than or equal to 50. When the test is used to evaluate if the data set is likely to have been drawn from a lognormal distribution, the test is applied to the natural logariithms of the data.

3.1.3.2 D'Agostino Test

D'Agostino's test (D'Agostino, 1971) is a goodness-of-fit test that can be used for data sets larger than 50, which is the upper limit for the W-test. The D'Agostino test can be used with data sets numbering between 50 and 1000 samples. An explanation of this test along with an example is presented in the handbook.

3.1.4 Graphing Data Sets

3.1.4.1 Histograms

Histograms provide project teams with a graphical technique to visualize background and site data sets. The histogram is essentially a column plot of the frequency of occurrence of data that fall within intervals. The frequency of occurrence is plotted against concentration. This graph allows for a rapid evaluation of symmetry and variability of data sets. An example of a histogram is presented in the companion handbook.

3.1.4.2 Box Plot

A box plot (sometimes referred to as a box and whisker plot) is a simplified representation of a data set that provides a summary view of the entire data set. A box plot shows the location (mean and median), degree of symmetry, range of variation, and potential outliers of the data set. The data set is shown as a rectangular box that represents the main population of the data set (middle 50 percent). The upper and lower quartiles (75th and 25th percentiles) define the top and bottom of the rectangle, respectively. The median is represented by the line bisecting the rectangle to indicate the middle point of the data set. The data set mean is represented by a '+' sign. The lines (or whiskers) extending from the box show how far the tails of the distribution extend. Data that fall beyond the main population, which may represent outliers, are presented as '*'. Box plots allow project teams to rapidly assess single or multiple data sets at a glance to compare their means, medians, variability, and skewness.

3.1.5 Evaluating Censored Data

Many environmental data sets contain non-detected data. These data sets are said to be censored. While the precise concentration of a non-detect is unknown, it lies between zero and the detection limit. In other words, it may not be present at the site or it may be present slightly below the detection limit. The detection limit (DL) represents the lowest amount of a chemical that can be detected above the noise of an analytical instrument or method and is based on the signal-to-noise ratio. There are two types of DLs, namely, the instrument detection limit (IDL) and the method detection limit (MDL). The IDL represents the lowest concentration that can be 'seen' or detected by an analytical instrument while the MDL takes into account reagents, sample matrix, and preparation steps. The IDL and MDL are specific for each analytical run and can very from run to run and day to day. Consequently, it is necessary to develop a quantitation limit (QL) which represents the chemical concentration which can reliably and reproducibly be detected. In other words, the DL represents the lowest amount a chemical can be 'seen' whereas QL represents the levels at which measurements can be 'trusted.' Generally, a factor of three to five is applied to the DL to derive a QL (EPA 1989b).

Two types of QLs may be reported in data summaries. Contract required quantitation limits (CRQLs) and sample quantitation limits (SQLs). CRQLs are the detection limits EPA requires laboratories that are certified in the Contract Laboratory Program (CLP) to routinely achieve and are standardized. CRQLs are not necessarily the lowest detectable levels possible. In contrast, SQLs are specific for the particular analytical run and are the most relevant DLs to include as part of background and site data sets. It should be noted that even when the SQL is reported in data summaries it is highly likely that the laboratory could have detected lower concentrations for a particular chemical in a sample. Laboratories report the upper limit concentration that they are confident the chemical in the sample does not exceed. Consequently, the actual chemical concentration is likely far below the reported value. This is an important consideration for project teams when developing ‘proxy values’ for use in data analysis and statistical testing. Because the reported SQL is likely to overestimate the actual concentration present in the sample, the proxy value, which is typically based on the SQL, should be viewed as a higher estimate of the concentration in the sample. This information should be communicated to stakeholders and factored in all analyses where the potential exists for proxy values (for non-detects) to influence the analytical result. The data set report will typically include both detected and non-detect values for a particular chemical. Non-detected values, will conventionally be reported as the sample quantitation SQL for a particular laboratory method. The SQL is specified for a particular chemical and day of analysis. SQLs can differ between analyses for the sample chemical and provides more information than CRQLs, which are not sample-specific. SQLs should always be used in background analyses because they provide more information.

Data sets containing non-detect values are called censored data sets. When the detection limits (DLs) are low relative to detected concentrations, data sets are often referred to as left-censored data. This is because observations to the left of the detected values have no assigned values. However, the particular SQL assigned to a non-detect may frequently exceed detected measurements. This may result from matrix interferences in the sample. High SQLs can greatly confound data characterization and interpretation because they are typically used to generate proxy values. In some instances the proxy values can greatly exceed the detected concentrations. When this occurs, project teams should consider reanalyzing the sample (as allowed by analytical protocols) or resampling. Unusually high proxy values can bias the data set and result in misidentifying background analytes as COCs.

The most common techniques used to derive proxy values for censored data sets involve deletion and substitution techniques. EPA (EPA 1996b) has developed general guidelines for these procedures based on the number of non-detected data in the data set (Table 2). The analytical approaches include: (1) replace non-detects with one-half the SQL (not the CRQL), (2) Cohen's Adjustment, Trimmed Mean, Winsorized Mean and standard deviation, (3) and the test for proportions.

Although choosing the most applicable approach is primarily based on the percentage of non-detects, professional judgment should also be applied. For example, in addition to percentage of non-detects, the number of data in the data set should be a factor in the decision. A data set where 1 sample out of 4 is not detected should be treated differently from a data set where 25 out of 100 samples are not detected.

When only a small percentage of samples in a data set are non-detects, EPA recommends either a deletion or substitution approach. The deletion techniques estimate the mean and standard deviation using only the observations that are above the SQL (that is, only detected values). There may be justification in concluding that the chemical concentration is zero when data are collected purposively based on historical or site operation information and are biased toward hot spots. For example, if a relatively large number of samples collected from a small, well-defined site, are collected and analyzed with low limits of detection, it may be correct to use zero as a proxy value. However, deletion techniques should be used

TABLE 2

GUIDELINES FOR ANALYZING DATA WITH NON-DETECTS

Percent Non-detects	**Statistical Analysis Method**
< 15	Replace Non-detects with DL/2, DL, or a Very Small Number
15 to 50	Trimmed Mean, Cohen's Adjustment, Winsorized Mean and Standard Deviation
>50 to 90	Use Test for Proportions

For data sets where the non-detects are greater than 90 percent the number of samples with detected values becomes important. Additional samples may be required after a review of data quality analytical procedures.

(DL = detection limit)

only when site information is exhaustively reviewed to determine that it is highly unlikely that the chemical is present in the sample. In contrast to the deletion approach, the full SQL or one-half the SQL should be used as a proxy value for each of the non-detected samples when there is reason to suspect the chemical is present in the sample. EPA guidance (EPA 1989b) states:

> "If there is a reason to believe that the chemical is present in a sample at a concentration below the SQL, use one-half of the SQL as a proxy concentration. The SQL value itself can be used if there is reason to believe the concentration is closer to it than one-half the SQL. Unless site-specific information indicates that a chemical is not likely to be present in a sample, do not substitute the value zero in place of the SQL (i.e., do not assume that a chemical that is not detected at the SQL would not be detected in the sample if the analysis was extremely sensitive)."

There are several problems associated with the deletion and substitution techniques that introduce bias when the mean or standard deviation of censored data sets are estimated. For example, when considering two data sets with the same mean and SQL but different standard deviations, substitution of zero or one-half the SQL for the non-detected samples shifts the estimated mean of one sample (larger standard deviation) downward more than that of the other (small standard deviation). Using the full SQL or deleting any non-detected samples has the opposite effect: estimated means are larger than the true mean. Additionally, when considering two hypothetical data sets with different mean concentrations but the same standard deviation and SQL, deletion or substitution techniques generate a standard deviation estimate much lower for the sample with the small mean than for the sample with the large mean.

When the data set is distributed normally or lognormally, and contains between 15 and 50 percent of samples that are at non-detectable levels, EPA (EPA 1996b) recommends applying Cohen's adjustment method to the data set. This method allows the mean and standard deviation to be adjusted based on the data below the detection limit. Cohen's adjustment method is based on the statistical technique of maximum likelihood estimation of the mean and variance. This method provides a more robust estimate of the mean and variance and does not rely on simple substitution techniques. Cohen's adjustment method should not be used when more than 50 percent of the data set samples are not detected or when the PDF, minus the non-detected samples, is not normal or lognormal. Detailed information on how to apply Cohen's adjustment method is presented in the handbook, along with an example.

Data sets can also be trimmed to derive an unbiased estimate of the mean concentration. In this approach, both tails of the PDF are trimmed. The proportion that is trimmed depends on the number of samples in the data set. For symmetrical distributions, 25 percent of the data set can be trimmed and still provide a good estimate of the mean concentration (EPA 1996b). The number of non-detect samples can also be used as a benchmark to determine the appropriate amount of trimming. This approach can only be used to adjust the mean concentration. It should not be used to adjust the standard deviation. The trimming approach is detailed in EPA (1996b) and the handbook.

Data sets having between 15 and 50 percent non-detect data can be Winsorized. Winsorizing replaces data in the tails of the data set then the mean and standard deviation are computed on the new data set. Winsorizing data sets is described in the handbook, along with an example.

For data sets having between 50 and 90 percent censored data, EPA (EPA 1996b) recommends that the test of proportions be applied. In this approach, the parameter of interest, such as the mean, is switched to a percentile to allow statistical testing of the data set. For example, EPA recommends that when 67 percent of data are censored, the parameter of interest may be switched to the 75th percentile and the test of proportions applied. The test of proportions is detailed in the handbook to this document.

3.1.6 Identifying Spatial and Temporal Trends in Data Sets

Spatial and temporal trends in a data set indicate increases or decreases in chemical concentrations over an area, or time, respectively. Although spatial variation can be due to inherent variability of environmental data, it can also indicate the presence of hot spots or source areas. Temporal trends can result from direction of groundwater flow, or season. It is important to identify spatial trends and biases so that both site and background data can be collected, pooled, and interpreted correctly.

The three general sampling strategies described in EPA guidance (EPA 1989b)—random, purposive, and systematic—are not equally effective in detecting spatial trends. Random sample collection has the potential of leaving large areas unsampled, decreasing the likelihood of detecting a spatial trend. Authoritative (purposive) samples are collected with the intention of identifying areas of contamination and may fail to provide a clear picture of conditions across an entire site. Systematic (grid) sampling is preferable to other types of sampling for locating spatial trends. Also, geostatistical methods for estimating spatial trends provide the most reliable results when based on data collected on a grid. Geostatistical procedures can also be helpful to facilitate further sample design if subsequent sampling becomes necessary. A stratified-grid sampling scheme is a good, cost-effective option (Entz and Chang, 1991).

Field conditions, and consequently chemical concentrations, in the environment are not constant over time. Sampling plans should be developed to account for seasonal changes. Collecting samples at only one point in time may result in biased conclusions. For example, although the mass of the chemical in soil or groundwater may remain constant, the concentration can be significantly altered during periods of heavy rainfall or snowmelt. If samples are collected only during one time of the year, chemical concentrations may not be representative of true site conditions.

To investigate temporal variability, a full annual sampling cycle should be conducted. If it is not possible to collect quarterly samples, at least two sampling events should be considered during periods of opposite seasonal extremes. Examples of opposite seasonal extremes presented in EPA guidance (EPA 1989b) include the following:

- High water and low water
- High recharge and low recharge
- Windy and calm conditions
- High suspended solids and clear water

Gilbert (1987) provides a detailed discussion of the importance of recognizing spatial and temporal patterns and provides methods for estimating and testing for trends.

3.1.7 Geochemical Correlations Between Naturally Occurring Elements

A geochemical analysis to determine whether correlations exist among analytes should be conducted as part of characterizing both site and background data sets. The concentration ratios of naturally occurring elements are often constant. Determining the ratios between analytes may yield important information about possible releases. For example, an independent analysis for one analyte may show it to be present at an elevated level but the ratio of that analyte with another naturally occurring analyte remains constant. In this case, the constant ratio indicates a release may not have occurred and that although the mass of the chemical is increased it is representative of the underlying natural geochemistry. A classic example is the correlative relationship between nickel and chromium which form a solid matrix in many minerals. An elevated concentration of nickel at a site (compared to ‘background’) may not indicate a release but rather there is more of the nickel containing mineral at the site. This can be confirmed by calculating and comparing site and background ratios for other pairs of analytes (such as iron and manganese, iron and cobalt, aluminum and lead), rather than the absolute concentrations of individual analytes. If the element ratios remain fairly constant between background and site data sets, the elevated concentrations may simply indicate an increased concentration of a naturally occurring mineral or adsorption of the element to a particle.

Metal concentrations slightly above the background values do not necessarily indicate that the soil has been affected by the site operations. Such metals above the background values can be treated as merely "possible contaminants" and be subject to further evaluation in terms of their concentration, geochemical distribution, and association with potential sources.

Since the distribution of many heavy metals in soil is controlled by the concentration of specific adsorbents, scatterplots and simple linear regression methods can be used to show the association of metals and their adsorbents. For example, a scatterplot of chromium and iron concentrations using data from the entire facility usually can quickly identify the outliers and the trend between the two metals. Also, when used in conjunction with statistical normal probability plots, the geochemical scatterplots often can confirm the inflection points on the normal probability curves. For sites with multiple ambient background conditions, these techniques can be used to define the underlying geochemical conditions so that site data can be appropriately compared to the correct background population. Some metals, such as lead and copper often display unusual concentration distributions across sites due to different sorption or binding mechanisms (i.e. inorganic versus organic sorption). These metals are best represented by more than one population. The interpretation of normal probability plots should take the geochemical characteristics of the elements and geologic conditions into consideration.

The distribution of metals in soils is controlled by several mechanisms, including precipitation, dissolution, coprecipitation, and sorption. Adsorption, the accumulation of matter at the solid-water interface, is the basis of most surface chemical processes. The common adsorbents for heavy metals in soils include clay minerals, organic matter, and metal oxides and hydroxides. The retention of cationic metals in soil has been correlated with such soil properties as pH, redox potential, surface area, cation exchange capacity, organic matter content, clay content, iron and manganese oxide content, and carbonate content. Clay minerals are known to be effective in controlling heavy metals in soils because of the negative surface charges on the surface of clay minerals. Soil organic matter is also known to be effective in adsorbing heavy metals. Soil organic matter forms coating on inorganic mineral surfaces. Humus, a series of high molecular weight polymers in soil organic matter, is responsible for the sorption of cationic species. For example, lead and copper tend to accumulate in surface soils due to adsorption to organic matter. However, in a soil sample, metals showing strong affinity with soil organic matter are also adsorbed by inorganic adsorbents to a lesser degree.

To fully explore the distribution of metals in soil, appropriate geochemical analyses and comparisons should be performed as part of all background analyses and may be particularly important for identifying hot spots. Chemical speciation and geochemistry should be performed to confirm that the same chemical forms (valency) are being compared and that apparent hot spots are not the result of unusually dense clay formations to which inorganic elements tightly bind.

3.1.8 Descriptive Summary Statistics

EPA guidance (EPA 1989b) suggests that data be presented in descriptive summary statistics tables, stating that "a table that includes all chemicals detected in a medium can be provided for each medium sampled at a hazardous waste site or for each medium within an operable unit at a site." Summary statistics can provide valuable insight into the data set. The following summary statistics should be developed to summarize both site-specific and background data sets:

Total Number of Samples. The total number of samples collected from each environmental medium for each data set should be reported.

Standard Deviation. The arithmetic and geometric standard deviation of each data set should be computed to provide information on data variability.

Proportion of Detected Samples (frequency of detection). The proportion of samples in which the analyte was detected should be presented. The frequency of detection is defined by EPA (EPA 1989b) as the "number of times a chemical was detected over the total number of samples considered." Frequency of detection should be determined and reported since "chemicals that are infrequently detected may be artifacts in the data due to sampling, analytical, or other problems, and therefore may not be related to site operations or disposal practices." Artifacts (which represent spurious data) should be eliminated.

Quantitation Limits. The range of the reported quantitation limits (QLs) should be provided for each chemical in the various samples. Data summary tables should present SQLs and CRQLs.

Probability Density Functions (PDFs). The PDF should be reported for each data set. The PDF for an environmental data set will frequently follow a normal, lognormal, or Weibull distribution. In some cases, the data will not be of sufficient quantity or quality to determine the PDF with adequate confidence. In those cases non-parametric statistical tests should be used.

The Normal PDF: A bell-shaped curve symmetric about the mean
The Lognormal PDF: A curve that starts at zero, rises very rapidly to a maximum, then tapers off to a long tail on the right hand side
The Weibull PDF: In many cases, similar to a Lognormal but the right hand tail is much smaller (thinner)
The Gamma PDF: In most cases, similar to the Lognormal but the right hand tail is much greater (fatter)

Minimum, Maximum, Mean, and Median Chemical Concentrations. The range (maximum and minimum) of detected concentrations should be included in the descriptive summary statistics tables. The lower and upper concentrations will be reported as the minimum and maximum detected values, respectively. Additionally, the arithmetic mean, geometric mean and median chemical concentration should also be obtained and reported for each data set. Finally, the 25^{th} and 75^{th} percentiles of each data set should be reported.

Coefficient of Variation. The coefficient of variation (CV), defined as the standard deviation divided by the mean concentration, should be presented to indicate the relative variability in the data set.

3.1.9 Analyzing Data Set Variability

It is necessary to evaluate the variance of site and background data sets that follow a normal PDF (distribution) in order to select the most appropriate parametric statistical test. As will be discussed in Section 3.2.4, selecting between the Student's t-test and Satterthwaites's test depends on whether the variabilities for two data sets are equal or unequal. This determination is made with the F-test (Iman and Conover, 1983, pg 275; EPA 1996b, pg 4.5-2).

In conducting the F-test, the variances for the individual data sets are determined and the variance ratios calculated. The variance ratio is compared to a critical value which is based on the sample size. If the calculated variance ratio is less than the critical value, it is assumed that there is no difference between data set variances. In other words, the variances are assumed to be equal and the Student's t-test should be used to test whether there is a statistical difference between the site and background means.

3.2 APPLYING STATISTICAL TESTS TO EVALUATE BACKGROUND CONDITIONS

This section describes the statistical methodology that should be used to compare site chemical concentrations to background levels to: 1) identify COPCs, 2) identify chemical releases (COCs), and 3) establish background levels should remediation become necessary.

As indicated in Figure 11, the statistical paradigm for identifying Navy chemical releases is a comprehensive and iterative process. It involves identifying discrete areas of elevated concentrations often referred to as "hot spots" as well as widely-distributed high chemical concentrations.. As shown, Figure 3-6 of the handbook, a hot spot is identified as a distinct subpopulation that appears as a "right tail bump" in the data set. However, high concentrations of a chemical over a wide area may appear as right tail bump or as a shift of the entire data set population to the right relative to the background population. Comprehensively evaluating both situations assisted by geographical location data of the samples with high concentration values will ensure that Navy releases have not been overlooked. Evaluating a site only for potential hot spots does not provide information about potential population differences that may exist between site and background conditions. Likewise, statistical testing of population means or medians may not identify potential hot spots alone. Project teams should include both types of analyses in the background investigation.

3.2.1 Definitions and Purpose of Statistical Hypothesis Testing

Statistical hypothesis testing, as defined by EPA guidance (EPA 1989b), "is a rule used for deciding whether or not a statement (i.e., the null hypothesis) should be rejected in favor of a specified alternative statement (i.e., the alternative hypothesis)." In relation to background analyses, the null hypothesis states that there is no difference between site and background chemical concentrations. Conversely, the alternative hypothesis states that a difference between site and background concentrations does exist which indicates a release may have occurred. Generally, the alternative hypothesis for Navy applications is "site concentrations are higher than background." This implies a one-tailed test of hypothesis (that site concentrations are higher), rather than a two-tailed test (that site concentrations are either higher or lower than background).

In hypothesis testing, two types of decision errors may occur by incorrectly accepting or rejecting the null hypothesis. A Type I decision error occurs when the null hypothesis is rejected when, in fact, it is true. With a Type I error, Navy project teams will incorrectly conclude that a chemical release has occurred. A Type I error can lead to unnecessary remediation. In contrast, a Type II decision error occurs when the null hypothesis is accepted when, in fact, it is false. With a Type II error, Navy project teams will incorrectly conclude that a release has not occurred. A Type II error can lead to a decision that no remediation is necessary when remediation may be required.

The probability that a Type I error will occur is typically denoted as alpha (α), while the probability of a Type II error is designated by beta (β). The power of a statistical test to detect that a release has occurred is equal to 1-β, and is defined as the probability that the test procedure correctly rejects the null hypothesis and accepts the alternative hypothesis. The α is referred to as the level of significance of a statistical test.

Reducing the probability of Type I or Type II errors simultaneously is difficult because the confidence and power of statistical tests are related (EPA 1996b). For a particular data set and a fixed number of samples, decreasing the Type I error rate of the statistical test increases the Type II error rate. According to EPA (EPA 1992): "Because there is an inherent trade-off between the probability of committing a Type I or Type II error, a simultaneous reduction in both of these can only occur by increasing the number of samples."

Acceptable Type I and II error rates should be set during the DQO and DQA process by the project team. The consequences of making Type I and II decisionerrors should be important criteria for setting α and β. The choosing of a tolerable error rate is not easy and requires substantive input from the DQO development team. Evidence of potential threat to human or ecological health, wasteful expenditure of scarce resources, and socio-political consequences must be translated into a numerical value or range of values. To assist analysts in the interpretation of how different tolerable error rates affect the overall number of samples needed, the use of the interactive software DEFT (EPA 1994c) is recommended. EPA guidance (EPA 1989b) states that for risk assessment purposes, the minimum recommended confidence should be 80 percent ($\alpha = 0.20$) and power should be 90 percent ($\beta = 0.10$). This is because EPA generally believes that if an error is to be made, errors that lead to unnecessary cleanup (Type I) are preferable to leaving chemicals at a site at levels greater than background (Type II). However, project teams should exercise scientific judgment in developing decision criteria and make adjustments if it is deemed necessary through stakeholder consensus. When project teams are extracting background data from an existing data set and there are insufficient samples to meet predetermined α and β levels, project teams should consider cost associated with additional sampling in relation to potential health hazards associated with the chemicals under investigation. For example, if there is low toxicity (corresponding to low risk) associated with the target chemicals, a higher Type I decision error rate may be acceptable.

EPA guidance (EPA 1989b) states that statistical tests are the appropriate analytical tools for background analyses and that "the medium sampled influences the kind of statistical comparisons that can be made with background samples." The important first step in this process is selecting the appropriate statistical test based on data set characteristics to optimize statistical power and reduce decision errors to the lowest levels possible, consistent with the values of α and β specified during the DQO process. The statistical tool box approach described in the next section will allow project teams to achieve this goal.

3.2.2 Statistical "Tool Box" Approach

The statistical tool box approach provides great flexibility for statistical testing in background investigations. With the statistical methods selected for this tool box, project teams will be able to statistically analyze most any type of environmental data set that is composed of representative data.

Selecting the appropriate statistical method requires matching the strengths and weaknesses of the statistical method with the data set under investigation. In other words, data should not be "force fit" into an inappropriate test or inappropriately manipulated to fit the requirements of the statistical method. In order to conduct statistically robust background comparisons, matching the correct statistical method with a data set is a critical, but often overlooked, first step.

The statistical methods in the tool box include two parametric tests: Student's "two-sample" t-test and Satterthwaites's t-test; and two non-parametric tests: Wilcoxon rank sum test and Gehan's test. In addition, the non-parametric Quantile test is included to identify sites with pockets of elevated concentrations. These statistical methods in conjunction with graphical analyses provide a wide range of application and, with very few exceptions, will likely be the only methods project teams will need to conduct statistical background comparisons. The following sections present an overview of the tests and selection criteria. A more detailed description of these tests is presented along with example calculations in the companion handbook. Additional information and examples are presented in EPA guidance for data quality assessment (EPA 1996b).

3.2.3 Selecting The Appropriate Statistical Test

All conclusions regarding background conditions should be based on statistical comparisons between data sets from operation (or site) and background areas. The appropriate statistical test is selected after data are collected, analyzed and the data sets fully characterized. Statistical testing is used to determine whether the operation data set is "shifted" to the right relative to the background data set.

The appropriate statistical test is selected based how much information is available about the site and background PDFs, frequency of detection, and sample size of the data set.

As shown in Figure 11, the Student's t-test is a parametric statistical test that should be
used to detect differences in the background and site means when both background and site data sets follow a normal PDF, have a frequency of detection of 100 percent, and have equal variances. Many environmental data sets are lognormally distributed which requires natural log transformation of the data before computing statistical tests. Satterthwaites's t-test, which is also a parametric test, should be used to detect differences in means when both data sets follow a normal distribution.

For data sets that follow a normal PDF but the frequency of detection is significantly less than 100 percent or the data set has multiple detection limits (for non-detect samples) non-parametric tests should be used because they are better able to handle the non-detects and are expected to provide greater statistical power.

Non-parametric statistical tests should also be used for data sets that do not follow a normal PDF. When there is a single detection limit (for non-detect samples), the Wilcoxon rank sum (WRS) test should be used. For non-normal data sets with multiple detection limits, the Gehan test should be used. For data sets that follow a lognormal distribution, either the non-parametric tests or the t-test computed on the natural logarithms of the data may be used.

3.2.4 Parametric Statistical Tests

Parametric tests such as the Student's two-sample t-test and Satterthwaites's t-test should be used for testing whether the means of normally distributed data are statistically different. To use these tests, both the site and background data should have been collected at locations identified using simple random sampling or at the nodes of a systematic grid pattern placed at random over the site. Also, unless the number of samples is large, at least 30 or 40, for both site and background, then both data sets must also follow a normal distribution. If the number of samples is large for both data sets, then the t-test may be used because the estimated mean will then be approximately normally distributed. The power of the test increases with increasing number of samples. Selecting between these two tests depends on an accurate estimate of variance in both site and background data sets. The two tests are generally not appropriate for extremely small data sets (less than five samples) and should not generally be used for data sets with more than a few non-detect values.

If the background and site data are believed to be lognormally distributed, then the t test and Satterthwaite's t test can still be used if they are computed on the natural logarithms of the data. However, it should be noted that when this procedure is used, the null hypothesis that is being tested is that the medians (not the means) of the background and site data are equal. This distinction is important because the equality of medians of lognormal distributions does not imply equality of the means of those distributions. If testing for the equality of means is important, and if the background and sit distributions are strongly believed to be lognormal, then the Z-score method (a modified t test) described in Zhou, Gao and Hui (1997) may be used.

3.2.4.1 Student's "Two-Sample" t-test

Student's two-sample t-test is a well studied and verified statistical test used to compare the population means of two sets of continuous data with normal PDFs having equal variances and very few if any non-detects.

To conduct the test, a t-statistic is calculated for the data sets and compared to a critical t-value taken from a t-distribution table (included in the companion handbook). The critical t-value selected from the table is based on the number of samples and the selected level of significance (Type I error rate). If the calculated t-value exceeds the corresponding critical tabulated t value, the hypothesis that there is no difference between site and background means must be rejected. This result indicates that a release may have occurred and that the analyte should be identified as a site-specific COC. Detailed instruction for performing this test are presented in the companion handbook.

3.2.4.2 Satterthwaites's t-test

Satterthwaites's t-test is a modified Student's two-sample t-test that should be used when data sets follow a normal distribution but have unequal variances. There should be very few or no non-detected values in the data sets.

Although the calculations are slightly different from those in the Student's t-test, a “t” statistic is calculated and compared to a tabulated critical t-value taken from a t-distribution. The t-value selected from the table is based on the number of samples and level of significance (Type I error rate). If the calculated t-value exceeds the critical t-value from the table the null hypothesis of equal means is rejected and it is concluded that the site mean is significantly greater than the background mean indicating a release may have occurred. Directions for performing this test are presented in the companion handbook.

3.2.5 Non-Parametric Statistical Tests

When data sets do not follow a normal or lognormal distribution non-parametric statistical tests should be used. Non-parametric tests are not dependent on a particular type of PDF and test for differences between medians of site and background data sets. Parametric tests are only slightly more powerful than non-parametric statistical tests for data sets that follow a normal distribution. Therefore, non-parametric tests can also be used to test for differences between medians and means of normally distributed data. When the actual distribution of the data sets is uncertain, project teams can use the WRS and the Gehan non-parametric tests discussed in the following sections.

3.2.5.1 Wilcoxon Rank Sum (WRS) Test

The WRS test is a non-parametric version of the two-sample Student's t-test. According to EPA (EPA 1996b), when the WRS test is applied with the Quantile test, the combined tests are most powerful for detecting true differences between two population distributions. However, the WRS test should not be used if more than 40% of the background or sit data set are non-detects (Marssim 1997).

The WRS test is conducted by first ranking the combined site and background data from smallest to largest. Ranks are then assigned to each datum, starting with the rank of one for the lowest value and continuing until all data have been assigned a corresponding rank. The ranks of the site data are then summed and compared to a critical value corresponding to a specified level of significance (e.g. α = 0.05). If the sum of the site ranks is less than the critical value, then the null hypothesis (that the site and background data are similar) is not rejected. If the sum of the site ranks exceeds the critical value, then the median site concentration is statistically greater than the median background concentrations (Gilbert 1987).

This approach can be used even when some data points are tied (equal in rank). In that case, the tied values are each assigned the mean value of the tied ranks. For example, if three data points are equal and correspond to the ranks of 3, 4, and 5, each of the data points is ranked as 4 (Gilbert 1987). The next largest data point has the rank of 6. However, if the number of tied ranks becomes large, then the WRS test loses statistical power and may lead to the erroneous conclusion there is no difference between data sets when in fact there really is.

3.2.5.2 Gehan Test

The Gehan test is a generalized version of the WRS test. When there are no non-detect values in the data set, the Gehan test is identical to the WRS test. However, when there are large numbers of non-detect values with multiple detection limits (SQLs) the Gehan test may be more powerful. As it is a non-parametric test, it can be used for data sets that follow any distribution.

Similar to the WRS test, site and background data sets are combined and ranked from smallest to highest. The ranks are converted to scores and a Z statistic is determined. This statistic is then compared to a tabulated critical Z value. If the calculated Z value exceeds the critical Z value, is should be assumed that the chemical is present above background levels. This testing procedure is appropriate if the number of data in both the background and site data sets are greater than 10. Additional information is presented along with examples in the companion handbook.

3.2.6 Identifying Hot Spots

Hot spots have been generally defined by EPA (1989b) as areas of very high chemical concentrations. The tests described above cannot be relied upon to detect small hot spots because they are designed to look for differences in average (mean or median) values. The Quantile test described in the next section is more effective at detecting areas of elevatd concentrations, as long as those areas are not real small and a sufficient number of locations are sampled. Small hot spots are not likely to be "hit" unless the number of locations sampled is large and a triangular or square grid-sampling plan is used. Gilbert (1987) describes how to determine the grid spacing that is needed to detect with specified probability a hot spot of specified circular or elliptical size and shape. However, the methods in Gilbert (1987) assume no information is available about where to look for hot spots. Hence, grid spacing determined using the method in Gilbert (1987) is smaller than would be needed if reliable information about the likely location of hot spots is available. The method in Gilbert (1987) is now available as a Windows software package ELIPGRID-PC that can be downloaded from the Internet. The address for this site is: http://terrassa.pnl.gov:2080/DQO/home.html. The operation of the software is described in Davidson (1995).

Although graphical techniques are useful in identifying hot spots, project teams should be cautious about automatically identifying a cluster of data with high concentrations as representing hot spots. This is because background data sets will normally contain a wide range of concentrations with some of them expected to be relatively high. A normal range of background concentrations may span two to three orders of magnitude depending on the regional geochemical heterogeneity. Background conditions can only be represented by the entire range of concentrations in the data set. No single concentration can adequately represent the entire data set population. EPA guidance (EPA 1994b) discusses this concept in relation to statistical testing to determine whether site concentrations are different from reference or background concentrations, stating:

> "It should be understood that the use of the hypotheses in Equation 2.1 will, in general, allow some site measurements to be larger than some reference-area measurements without rejecting the null hypotheses that the reference-based cleanup standard has been achieved. The real

question addressed by the statistical tests in this document (Chapters 6 and 7) is whether the site measurements are sufficiently larger to be considered significantly (statistically) different from reference-area measurements."

In summary, project teams will likely detect several high values in site data sets. This finding should not automatically trigger alarm or remediation activities. It should, however, prompt project teams to more fully investigate the data set using the statistical test described in the following section. Sometimes it may be wise to resample in the vicinity of the high values. Details on this type of sampling technique may be found in Thompson and Seber (1996).

3.2.6.1 Quantile Test

As indicated in Figure 11 the Quantile test should be implemented as a last step in the background analysis when parametric and non-parametric tests do not detect a difference between site and background data sets. The Quantile test has been included to help ensure that all Navy releases are identified. Parametric and non-parametric tests are superior to the Quantile test in detecting differences in population means and medians. However, the Quantile test is a more powerful method to identify when the site contains areas of elevated concentrations if a sufficient number of samples are taken. Software package "EnvironmentalStats for S-Plus" is available to compute the Quantile test (Millard, 1992).

The Quantile test is a non-parametric statistical test that can be used on any data set regardless of the underlying data set PDF.

To conduct the Quantile test, data from both the site and background are ranked from largest to smallest. The number of site data that are among the largest "r" measurements of the combined data set are counted. The value for "r" is taken from tabled reference values at the specified level of significance. If the number of these measurements exceeds or equals a predetermined number (k, selected from reference tables at the specified level of significance), it can be concluded that a hot spot may exist at the site. Project teams should combine this information with spatial information to ensure all data representing the potential hot spot represent the sample area. What may appear to be a discrete hot spot (using graphical techniques such as cumulative frequency plots and the Quantile test), may simply represent independent and unrelated data that represent disparate locations.

The Quantile test is not as powerful as the WRS test when the site concentration distribution is shifted in its entirety to the right of the background distribution. However, the Quantile test is more powerful when a small group of data at high concentrations are present in the site data set. The performance of the Quantile test depends on the highest concentrations in site and background data sets. Therefore, it is important to verify that those values are not outliers. Furthermore, when there are multiple non-detect values, some of which are greater than the largest measurement, the data should be ranked using an alternate procedure. One procedure assumes that no information is available about what the actual concentration might have been, had the quantitation limits been lower. This procedure assumes, for example, that if the non-detect value is < 10, there is an equal likelihood that the concentration is any number less than 10. Another approach to ranking multiple non-detects is to set all non-detects equal to their detection limits. This approach is not recommended, however, unless it can be justified based on site-specific information. For example, if sufficient information exists to conclude that the analyte is present in samples at concentrations just below the detection limit, the detection limit can be used as a proxy value. See Section 4.7 in EPA (1996b) for additional information.

3.2.7 Testing Methods Not Recommended

Some methods for comparing background and site data sets may seem logical, but in fact can lead almost inevitably to incorrectly declaring that the site concentrations are greater than background concentrations. One such method is to compare the maximum site measurement to a background threshold value. The threshold value might be the estimated 95th percentile of background or perhaps an upper 95% confidence limit on the 95th percentile (i.e., the 95/95 tolerance limit). If the maximum site measurement equals or exceeds the threshold, then the idea is to conclude that the site distribution is shifted to the right of the background distribution. However, as discussed and illustrated in the companion handbook to this procedures document, the probability of incorrectly coming to that conclusion becomes very large (approaches certainty) as the number of site measurements become large. Hence, if you sample the site enough times, you will eventually conclude the site distribution is shifted to the right of the background distribution, incorrectly identifying chemicals of concern.

Another method that is not recommended is to compare the maximum site measurement with the maximum background measurement. Indeed, if the site and background distributions are identical (not different) and the same number of samples are collected at random in both areas, the probability is 0.50 (50%) chance) that the maximum site measurement will exceed the maximum background measurement. In other words the Type I error rate is 0.50 (50%). This large an error rate is not usually acceptable. Further discussion of this flawed testing approach is found in O'Brien and Gilbert (1997). The Slippage test (Rosenbaum 1954) should be used to compare site measurements to the maximum background measurement.

4.0 SUMMARY

This Navy procedural guidance document is intended to provide the decision-making framework and analytical tools to project teams enabling them to conduct scientifically defensible background analyses. It will promote consistency and reduce the uncertainty that exists in many background investigations. The Navy expects that this background analytical approach, summarized in Figure 12, will be used by project teams at all installations in the State of California to:

- Identify site-specific COCs
- Identify chemical releases
- Estimate and communicate the results of human health and environmental risk assessments
- Evaluate the feasibility of meeting compliance standards during remedial alternative selection

This Navy background analytical approach will increase confidence in all these activities.

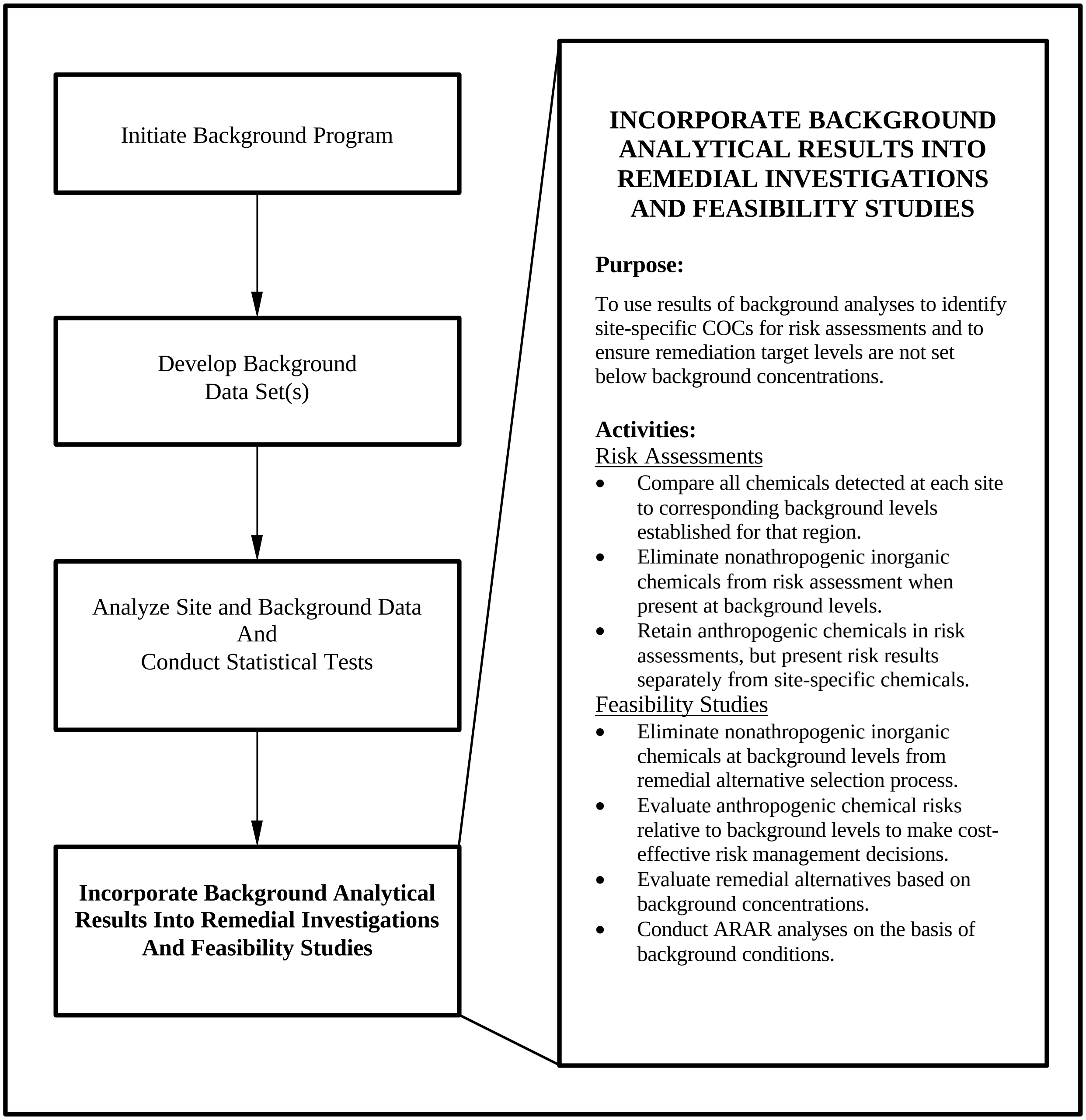
There are many Navy installations currently undergoing diverse types of environmental investigations. While it would be ideal for all project teams at all installations to incorporate this background analytical approach, complete implementation may be impractical because investigations are nearly complete. However, when evaluating the impact of these procedures on the existing program, project teams should carefully weigh the short term inconvenience against the long term gains this approach may provide. Certain elements of the background analysis are important even at the last stage where remediation has commenced. For example, when the cost of remediation is high and cleanup is expected to be protracted and extensive, the statistical approach in this document can be used to confirm that remediation is necessary and that appropriate cleanup goals have been developed.

This procedural guidance document clarifies the types of background conditions that project teams should evaluate. The definitions of naturally occurring and anthropogenic background conditions are consistent with regulatory guidance. Anthropogenic background chemicals typically result from off-site sources, and dredged fill material. Considerable effort may be necessary to accurately define each of these background conditions, but results will be used in every phase of investigations.

Developing a comprehensive site conceptual model is important when determining whether a data set is complete and accurately represents the site. This information is used to determine whether data gaps exist and if further sampling is necessary.

Data characterization is an important first step in the statistical analysis. Data characterization includes preparing descriptive summary tables and graphing data sets. This information is used to characterize the site and select the appropriate statistical test. The statistical tests in the tool box have wide-ranging applicability and can be used to evaluate most types of data sets.

FIGURE 12
STEP 4-INCORPORATE BACKGROUND ANALYTICAL RESULTS INTO REMEDIAL INVESTIGATIONS AND FEASIBILITY STUDIES



5.0 REFERENCES

D'Agostino, R.B., 1971. An omnibus test of normality for moderate and large size samples. *Biometrika 58:341-348*.

Davidson, J. R, December 1995. ELIPGRID-PC: UPGRADED VERSION, ORNL/TM-13103, Oak Ridge National Laboratory, Oak Ridge, TN http://terrassa.pnl.gov:2080/DQO/home

Dixon, W.J., 1953. Processing data for outliers. *Biometrics. 9:74-89.*

DTSC 1992. Supplemental Guidance for Human Health Multimedia Risk Assessments of Hazardous Waste Sites and Permitted Facilities. Department of Toxic Substances Control, Office of the Science Advisor. July 1992.

DTSC 1994. Preliminary Endangerment Assessment Guidance Manual. State of California Environmental Protection Agency, Department of Toxic Substances Control, January 1994.

DTSC 1996. Interim Final Policy. Determining Ambient Concentrations of Metals For Risk Assessments at Hazardous Waste Sites and Permitted Facilities. State of California Environmental Protection Agency. Department of Toxic Substances Control. June 1996.

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. May 1991.*

EPA 1988a. Guidance for Conducting Remedial Investigations and Feasibility Studies Under CERCLA. Interim Final, OSWER Directive 9355.3-01. October 1988.

EPA 1985. National Oil and Hazardous Substances Pollution Contingency Plan. Final Rule. 50 Federal Register 47912. November 1985.

EPA 1988b. Preliminary Assessment Guidance for FY 1988. OSWER Directive 9345.0-01. January 1988.

EPA 1989a. RCRA Facility Investigation (RFI) Guidance-Volume 1. Interim Final. EPA 530/SW-89-031. May 1989.

EPA 1989b. Risk Assessment Guidance for Superfund, Volume 1, Human Health Evaluation Manual (Part A) [HHEM], Interim Final. EPA/540/1-89/002. Office of Emergency and Remedial Response. December 1989.

EPA 1990. Guidance for Data Useability in Risk Assessment. Interim Final. EPA/540/G-90/008. October 1990.

EPA 1991. GEO-EAS 1.2.1. User's Guide. EPA 600/8-91/008. Environmental Monitoring Systems Laboratory, Office of Research and Development.

EPA 1992. Options for Addressing High Background Levels of Hazardous Substances at CERCLA Sites. Draft Final Issue Paper. Office of Emergency and Remedial Response. January 1992.

EPA 1993. SCOUT: A Data Analysis Program. http://www.epa.gov/crdlvweb/databases/scout Environmental Monitoring Systems Laboratory, Office of Research and Development.

EPA 1994a. Guidance for the Data Quality Objectives Process. EPA QA/G-4. Office of Research and Development. EPA/600/R-96/055. September 1994.

EPA 1994b. Statistical Methods for Evaluating the Attainment of Cleanup Standards, Volume 3: Reference-Based Standards for Soils and Solid Media. Office of Policy, Planning, and Evaluation. EPA 230-R-94-004. June 1994.

EPA 1994c, Data Quality Objectives Decision Error Feasibility Trials (DQO/DEFT), EPA 600/R-96/056. EPA G-4D Office of Research and Development. Washington, DC.

EPA 1995, Determination of Background Concentrations of Inorganics in Soils and Sediments at Hazardous Waste Sites, Technology Innovation Office, Office of Soil Waste and Emergency Response, EPA/540/S-96/500, December 1995.

EPA 1996a. DataQUEST (Data Quality Evaluation Statistical Tools, EPA QA/G-9D). http://es.epa.gov/ncerqa/qa/qatools.html Office of Research and Development. Washington, DC.

EPA 1996b. Guidance for Data Quality Assessment. EPA/600/R-96/084. EPA QA/G-9. Office of Research and Development. Washington, DC. July 1996.

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

Goglak, 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, U. S. Nuclear Regulatory Commission, Washington DC.

Iman, R.L. and Conover, W.J., 1983. A Modern Approach to Statistics. Wiley

Marssim 1997, Multi-Agency Radiation Survey and Site Investigation Manual *(MARSSIM)*, NUREG-1575, EPA 402-R-97-016, U.S. Nuclear Regulatory Commission, Washington DC

Millard, S.P., July 1997. EnvironmentalStats for S-Plus User's Manual, Version 1.0, Probability, Statistics & Information, 7723 44th Ave, N.E., Seattle WA 98115-5117

O' Brien, R. F. and Gilbert, R. O. 1997. Comparing Sites to Background, Part 1, *Environmental Testing & Analysis*, pages 12-13, September/October 1997.

Rosenbaum, S. 1543. "Tables for a Nonparametric Test of Location," *Annuals 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 Seber, G.A., 1996 *Adaptive Sampling*. Wiley, New York
Walsh, J.E., 1958. Large sample nonparametric rejection of outlying observations, Annals of the Institute of Statistical Mathematics 10:223-232.

Zhou, Xiao-Hua, Sujuan Gao, and Siu L. Hui, September 1997. Methods for Comparing the Means of Two Independent Log-Normal Samples. Biometrics 53:1129-1135.