

United States
Environmental Protection
Agency

Office of Environmental
Information
Washington, DC 20460

EPA/240/B-06/004
February 2006



Systematic Planning: A Case Study for Hazardous Waste Site Investigations

EPA QA/CS-1

Quality

FOREWORD

This document, *Systematic Planning Using the Data Quality Objectives Process*, shows the use of the Data Quality Objectives (DQO) Process in the form of a case study. The Environmental Protection Agency (EPA) has developed the DQO Process for project managers and planners to help them collect the appropriate type, quantity, and quality of data needed to support Agency actions. This guidance is the culmination of experiences in the design and collection environmental data in different Program Offices at the EPA.

Systematic Planning Using the Data Quality Objectives Process is one of a series of quality management documents that the EPA Quality Staff has prepared to assist users in implementing the Agency-wide Quality System. Other related documents include:

<i>EPA QA/G-4</i>	<i>Guidance for the Data Quality Objectives Process</i>
<i>EPA QA/G-5S</i>	<i>Guidance on Choosing a Sampling Design for Environmental Data Collection</i>
<i>EPA QA/G-9R</i>	<i>Data Quality Assessment: A Reviewer's Guide</i>
<i>EPA QA/G-9S</i>	<i>Data Quality Assessment: Statistical Methods for Practitioners</i>

This document provides guidance to EPA program managers and planning teams as well as to the general public as appropriate. It does not impose legally binding requirements and may not apply to a particular situation based on the circumstances. EPA retains the discretion to adopt approaches on a case-by-case basis that differ from this guidance where appropriate.

This case study is one of the U.S. Environmental Protection Agency Quality System Series documents. These documents describe the EPA policies and procedures for planning, implementing, and assessing the effectiveness of the Quality System. Mention of any copyrighted method does not constitute endorsement. These documents are updated periodically to incorporate new topics and revisions or refinements to existing procedures. Comments received on this version will be considered for inclusion in subsequent versions. Please send your comments to:

Quality Staff (2811R)
Office of Environmental Information
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue N.W.
Washington, DC 20460
Phone: (202) 564-6830
Fax: (202) 565-2441
E-mail: quality@epa.gov

Copies of the EPA's Quality System documents may be downloaded at: www.epa.gov/quality.

PREFACE

Systematic Planning Using the Data Quality Objectives Process: A Case Study of a Hazardous Waste Investigation describes the Data Quality Objectives (DQO) Process in a decision-making situation. The case study shows how application of the DQO Process leads to sound data collection techniques, sampling methods, and analysis of the results for decision making. Elementary Data Quality Assessment is used to draw conclusions from the results.

The case study is presented in two parts – a Preliminary Investigation followed by a Remedial Investigation – which correspond to the general stages of data collection and analysis in an environmental investigation. With each investigation, information is presented according to the three stages of EPA's Quality System – planning, implementation, and assessment. The case study demonstrates how the study team succeeded in establishing the nature and extent of site contamination and contaminants of potential concern during the Preliminary Investigation, which provided the data necessary to support a well focused, statistically-based, sampling campaign to complete the subsequent Remedial Investigation.

While reviewing the DQO development phase of the Preliminary Investigation, one should focus on the iterative manner used for the DQO steps (concept grounded in good sense). In the implementation phase, attention should be given to how a flexible probabilistic sampling scheme was developed. In the final phase, assessment, note how only rudimentary statistics were generated in the Preliminary Investigation but were enough to lay the groundwork for the Remedial Investigation.

In the Remedial Investigation, the DQO activity was quite abbreviated due to the efficiency of the DQO Process in the Preliminary Investigation. DQO activity in the implementation phase was equally short. The assessment phase used simple statistical techniques to analyze the data to produce a clear picture of what was happening and what subsequent activities should be.

The case study is intended for all EPA and extramural organizations that 1) have quality systems based on EPA policies and specifications, 2) may periodically assess these quality systems for compliance to the specifications, or 3) may be assessed by EPA. The use of the DQO Process conforms to the requirements of EPA Order 5360.1 A2 (EPA 2000) to use systematic planning in the collection of environmental data.

The techniques discussed in this case study are non-mandatory and the case study is intended to help project managers and staff understand how the DQO Process should be applied in practical situations. The data and techniques discussed in the case study are real; however, the location and identifying characteristics of the actual site have been modified to protect its identity.

TABLE OF CONTENTS

1. INTRODUCTION	1
1.1 Purpose and Scope	1
1.2 Case Study Background	1
1.3 Site History and Description	2
2. PRELIMINARY INVESTIGATIONS	3
2.1 Planning the Preliminary Investigations	5
2.2 Implementation of the Preliminary Investigations	14
2.3 Assessment the Preliminary Assessment	15
2.3.1 Real-Time Data Assessment and Post Stratification	15
2.3.2 Sample Size Selection for Background Comparisons	16
2.3.3 Decision to use Fixed Laboratory Data with BAP Equivalents	18
2.4 Preliminary Investigation Conclusions	18
3. REMEDIAL INVESTIGATION	21
3.1 Planning the Remedial Investigation	21
3.2 Implementation of the Remedial Investigation	21
3.3 Assessment of the Remedial Investigation	24
3.3.1 The Remedial Investigation Data	24
3.3.2 Calculation of Upper Confidence Limits	25
4. CONCLUSIONS	29
5. SUMMARY OF THE DQA STUDY	29
6. REFERENCES	30
APPENDIX A. THE PAH IMMUNASSAY METHOD	A-1
APPENDIX B. VISUAL SAMPLE PLAN	B-1
APPENDIX C. SUMMARY OF THE DOWS DEVELOPED FOR THE PRELIMINARY INVESTIGATION	C-1
APPENDIX D. FIELD IMPLEMENTATION DETAILS	D-1
APPENDIX E. SAMPLE SIZE SELECTION FOR BACKGROUND COMPARISONS	E-1
APPENDIX F. COMPARISONS OF IMMUNOSSAY AND FIXED LABORATORY RESULTS	F-1
APPENDIX G. SUMMARY OF THE PRELIMINARY INVESTIGATIONS STATISTICAL ANALYSIS OF THE COPCS	G-1

LIST OF FIGURES

Figure 1-1 M & H Ltd. Site Map.....	3
Figure 2-1. The Data Quality Objectives Process.....	4
Figure 2-2 Initial Conceptual Site Model	5
Figure 2-3 Decision Logic Diagram Part 1: Initial Steps	9
Figure 2-4 Decision Logic Diagram Part 2: Preliminary Investigation Implementation Logic ...	11
Figure 2-5 Real-Time Data Assessment Process Flow Chart.....	12
Figure 2-6 Sample Locations for Preliminary Investigation.....	13
Figure 2-7 Post Strata Showing 0-3ft. Weighted Average total PAH Concentrations	17
Figure 2-8 Locations and Depths Selected for Fixed Lab Analyses.....	17
Figure 3-1 Combined Preliminary and RI Sampling Location Map.....	23
Figure 3-2 Box-and-Whisker Plots BaP Equivalent Concentration Weighted Averages.....	26
Figure 3-3 Box-and-Whisker Plots BaP Equivalent Concentrations in the Northern (left) and Southern (right) Sample Locations	27
Figure B-1 Wilcoxon Signed Rank (One Sample) Test	B-2
Figure B-2 Hotspot Sampling of 112500 Feet.....	B-5
Figure G-1 BaP Equivalents Box Plots: Standard and Log Scales.....	G-2
Figure G-2 Arsenic Standard and Log Scales.....	G-6
Figure G-3 Benzo(a)anthracene Box Plots: Standard and Log Scales	G-7
Figure G-4 Benzo(a)pyrene Box Plots: Standard and Log Scales.....	G-7
Figure G-5 Benzo(b)fluoranthene Box Plots: Standard and Log Scales	G-8
Figure G-6 Benzo(g,h,i)perylene Box Plots: Standard and Log Scales.....	G-8
Figure G-7 Benzo(k)fluoranthene Box Plots: Standard and Log Scales	G-9
Figure G-8 Chrysene Box Plots: Standard and Log Scales	G-9
Figure G-9 Fluoranthene Box Plots: Standard and Log Scales	G-10
Figure G-10 Indeno(1,2,3cd)pyrene Box Plots: Standard and Log Scales.....	G-10
Figure G-11 Phenanthrene Box Plots: Standard and Log Scales.....	G-11
Figure G-12 Pyrrene Box Plots: Standard and Log Scales.....	G-11

LIST OF TABLES

Table 2-1	Median and Maxima for the Three Areas.....	19
Table 2-2	Number of Detects and Standard Deviation of the Three Areas.....	19
Table 2-3	Risk Assessment Meeting Notes	20
Table 3-1	Risk Assessment Meeting Notes	21
Table 3-2	Data Quality Objectives Summary Table for the Remedial Investigations	22
Table 3-3	BaP Equivalent Concentrations (ppb) for the Post Stratified Operations Area.....	24
Table 3-4	Upper 95% Confidence Intervals Northern Location.....	28
Table 3-5	Upper 95% Confidence Intervals Southern Location.....	28
Table 3-6	Upper 95% Confidence Intervals All Locations	28
Table 3-7	Upper 95% Confidence Intervals Northern Locations Weighted Data	28
Table 3-8	Upper 95% Confidence Intervals Southern Locations Weighted Data	28
Table 3-9	Upper 95% Confidence Intervals All Locations Weighted Data	28
Table B-1	Summary of Sampling Design.....	B-4
Table B-2	Number of Samples	B-6
Table B-3	Cost Information.....	B-6
Table E-1	Sample Size Required in Each Site and Background Location	E-1
Table F-1	Correlation of Fixed Laboratory total PAH with Immunoassay total PAH.....	F-1
Table F-2	Correlation of Fixed Laboratory BaP equivalents with Immunoassay total PAH	F-1

LIST OF ACRONYMS USED IN THIS CASE STUDY

BaP	Benzo(a)pyrene
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
COPCs	Contaminants of potential concern
CSM	Conceptual site model
DQOs	Data Quality Objectives
FS	Feasibility Study
GW	Ground water
HH	Human Health
MGP	Manufactured Gas Plant
PAHs	Polycyclic Aromatic Hydrocarbons
PCB	Polychlorinated Biphenyls
QAPP	Quality Assurance Project Plan
PI	Preliminary Investigation
RI	Remedial Investigation
UCL	Upper Confidence Limit
VOCs	Volatile Organic Compounds
VSP	Visual Sample Plan

1. INTRODUCTION

1.1 Purpose and Scope

The primary objective of this case study is to demonstrate the application of EPA's systematic planning guidance, associated statistical design, and assessment tools to derive sampling design and quality planning documents at each stage of an environmental investigation. This case study is presented to provide insights into the process to systematic planning applied to hazardous waste sites.

The case study focuses on a manufacturing/shipping facility situated near the downtown of Brufton, a small city located on the U.S. east coast. A recent dredging operation revealed the presence of high levels of Polycyclic Aromatic Hydrocarbons (PAHs) in offshore sediments, calling into question whether the facility was the source of the observed contamination. Research into previous uses of the property found that the site was the location of a former gas manufacturing plant. Similar to many small towns in the U.S., a manufactured gas plant (MGP) operated in the town between the late 1800s and mid-1900s. At these plants, coal or oil was heated in the absence of oxygen to drive off the volatiles that comprised the manufactured gas. The typical by-products of this process included tars, ash and other solid waste from boilers, and wastes from purifiers.

These findings of offshore sediment contamination prompted an investigation of possible on-shore contamination sources that may be continuing to release these contaminants to the Bay. Additionally, the investigation revealed that *M & H Ltd.* was previously the site of an MGP.

This case study will be presented in two parts, corresponding to the general stages of data collection and analysis in an environmental investigation. These parts are Preliminary Investigation (Chapter 2) followed by the Remedial Investigation (Chapter 3). In each chapter, information will be presented according to the three stages of EPA's Quality System – planning, implementation, and assessment.

The case study will show how the study team succeeded in quickly establishing the nature and extent of site contamination and contaminants of potential concern (COPCs) during the Preliminary Investigation (PI), which provided the data necessary to support a well focused, statistically-based, sampling campaign to complete the subsequent Remedial Investigation (RI).

1.2 Case Study Background

In a recent dredging operation of the ship channel in Buccaneer Bay, near *M & H Ltd.* (a shipping and light manufacturing company), PAHs were detected in sediment at levels exceeding ecological benchmarks. Routine bioassays and bioaccumulation studies resulted in a decision to dispose of the dredge spoils in an offshore contained area disposal facility, wherein the more contaminated sediments were capped with relatively clean sediments from other portions of the channel.

Research of the available records revealed that from the mid-1870s until approximately 1930, an oil-gas MGP operated on a property on the shore of the Bay. For a period of approximately 30 years after the MGP was shut down, the land was owned by the regional power company. During this period, records indicate that visible waste materials were removed. In the 1960s, the remaining derelict plant buildings were demolished and the city assumed ownership of the land. The property was used as a materials storage yard for the municipality until 1985, when *M & H Ltd.* purchased it from the city.

The former use of this property and the prior existence of waste materials at the site were documented during initial site inspection activities by local and state environmental agencies. Negotiations among representatives of the power company, the state environmental agency, and EPA resulted in the signing of a Consent Order under which the power company and current landowner will jointly fund investigatory activities at the site. If remediation is required, mediation will be used to determine who is financially responsible for the cleanup. The environmental investigation will be conducted in a manner consistent with CERCLA guidance, and the efforts will be overseen primarily by the State Department of the Environment (with Regional EPA and other stakeholder participation), but the site has not been included on the National Priorities List.

The Consent Order separated the site into two Investigation Units: An onshore area, and an offshore area. The spatial extent and potential impacts related to offshore contamination in the Bay are being addressed by a separate group. The purpose of the environmental investigation for onshore contamination described in the Consent Order is to determine:

1. The nature and extent of onshore contamination associated with the former MGP,
2. Whether onshore contamination is a continuing source of contaminants to the Bay environment, and
3. Whether onshore contamination presents unacceptable human or ecological risks.

1.3 Site History and Description

An MGP uses coal or oil as the feedstock to generate gas in a pyrolytic process. The raw gas typically contained impurities including tar particles, sulfur, and metals which were undesirable for transmission and for the purposes of gas illumination or heating. To remove impurities, the raw gas was subjected to cleaning using various combinations of water sprays or baths and filtering on media such as lime, wood chips, or iron. These impurities are the source of recent contamination. One characteristic waste product of some oil-gas MGPs was a powdered carbon material called lampblack. This originally nontoxic material could easily adsorb aromatic hydrocarbons if co-disposed with other waste materials.

The property owned by *M & H Ltd.* occupies an area of approximately 25 acres. Historical photographs of the area were obtained from archives of the power company and these revealed that the main structures of the MGP (generator house, boiler house, purifiers, and gas holders) were located in an area on the southern portion of *M & H Ltd.*'s property and occupied several acres of land which today is covered with graveled parking lots and offices. A rail line and spur runs along the border of the facility opposite the Bay; oil tanks were located near the

rail spur. Much of the remainder of the property is today comprised of a single large manufacturing building, shipping/dock facilities along the waterfront, and unused areas. Figure 1 shows the locations of current features of the *M & H Ltd.*'s site.

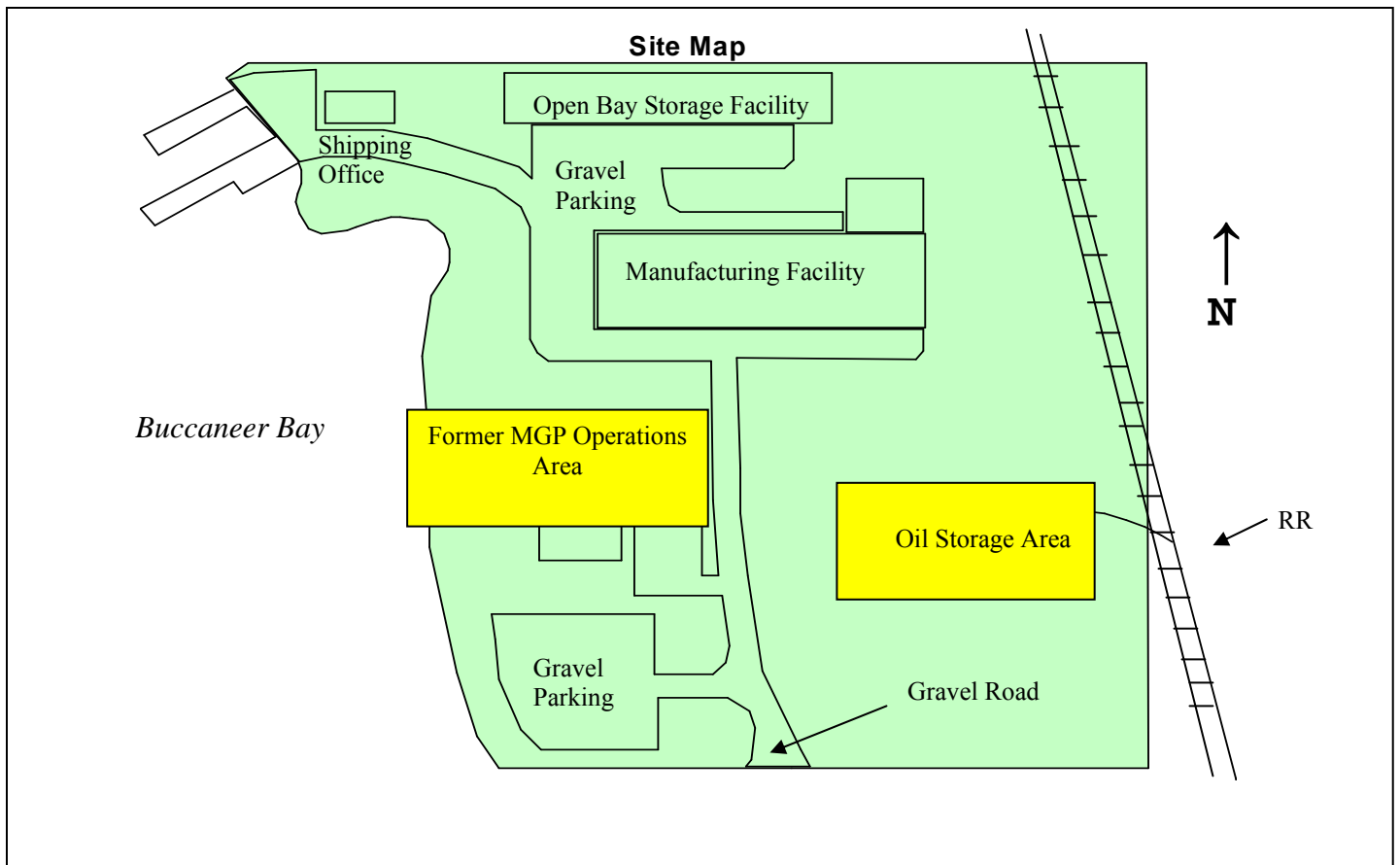


Figure 1.1. M&H Ltd. Site Map

2. PRELIMINARY INVESTIGATION

Planning for the Preliminary Investigation was conducted by representatives of the State Department of the Environment, site owners and operators, and *Topdog Inc.*, an environmental consulting firm hired as a facilitator. The investigation was led by a planning team who quickly agreed to use the EPA's recommended method for systematic planning: the seven step Data Quality Objectives (DQO) Process. The DQO Process (Figure 2-1) was used to design the Preliminary Investigation (PI), and then repeated in an abbreviated manner for the Remedial Investigation that followed. An initial conceptual site model (CSM) was developed that describes these potentially contaminated areas and the fate and transport mechanisms by which contamination may migrate or change in characteristics over time (Figure 2-2). The CSM was used to structure discussions among stakeholders regarding the overall technical approach for the investigation. The team recognized that the boundaries of the site would likely need to be refined based on empirical data at a later date.

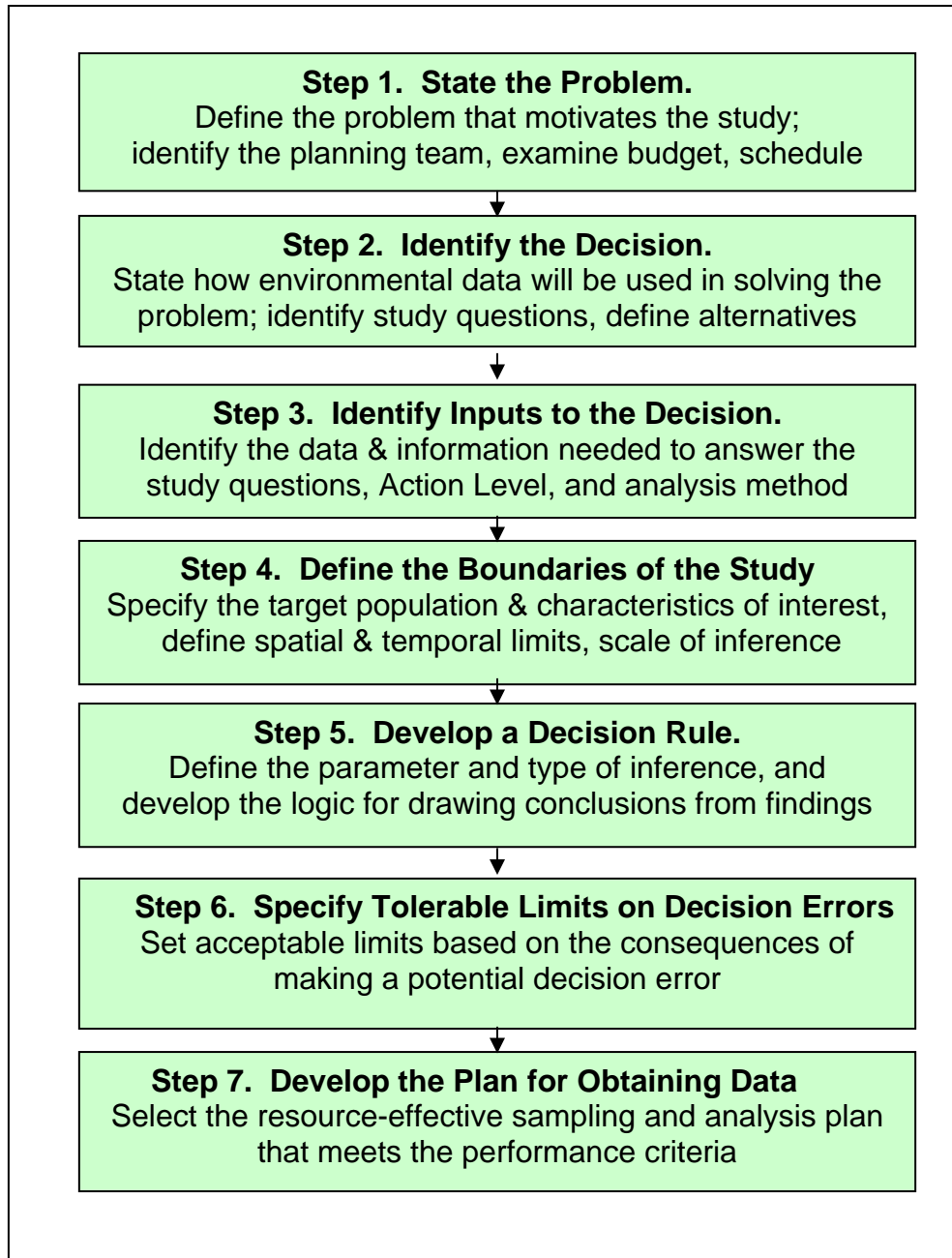
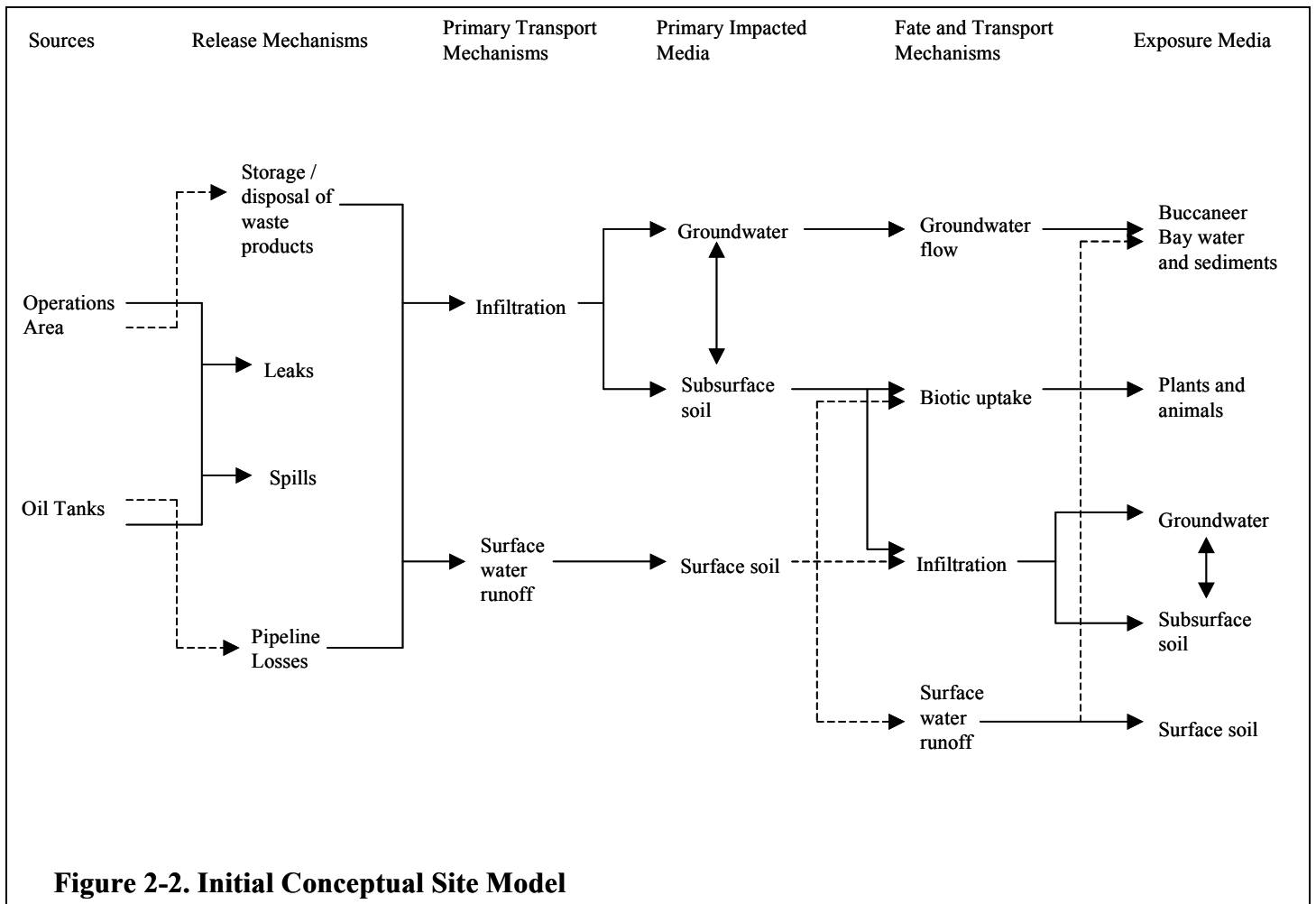


Figure 2-1. The Data Quality Objectives Process



2.1 Planning the Preliminary Investigation

Existing data for the site was limited to PAH measurements from three surface soil samples collected by the State Department of the Environment subsequent to the discovery of PAHs in Bay sediments. The team examined this in order to project expected data collection. These samples showed concentrations of one or more PAHs exceeding residential screening criteria for two of the three samples and one also exceeded industrial screening criteria.

The information from local records, the findings at other former MGP sites, and the initial soil samples indicated that residual onshore contamination was likeliest in the following subareas of the site:

- 1) The Operations Area where the generator house, boiler house, purifiers, and gas holders were located. The three initial soil samples collected by the Department of the Environment were from this general area

- 2) The area of the Oil Holding Tanks near the rail spur and along the route of associated piping leading towards the Operations Area. Although no records exist that document releases of oil from tanks or piping, leaks are known to be relatively common based on investigations of similar MGPs.

A review of findings at similar MGP facilities indicates that the analytical suites for soil and groundwater samples should include PAHs, VOCs, metals and cyanide.

Topdog Inc. facilitated three rounds of planning team meetings using the DQO Process. The key points of each meeting are grouped below according to the seven steps of the DQO Process.

Meeting 1: DQO Steps 1 – 3 Investigated

Step 1: State the Problem

- *Topdog Inc.* identified the important goals of maintaining timeliness, assuring technical defensibility, obtaining data adequate for risk-based decision-making, and controlling project costs.
- It was agreed that field-based analysis of chemical concentrations in soil could be used to establish the approximate spatial boundaries of contamination.
- The study should be completed and a report issued within six months, and should not exceed \$100K for analytical and data assessment, analysis and report generation.
- There was an extended discussion of land use assumptions. Though residential and ecological habitat uses were considered, it was agreed to focus on current and future industrial use scenarios.
- Ecological concerns are limited due to current conditions at the site (largely unvegetated disturbed soil, with extensive packed dirt and gravel roads and parking lots and buildings) and the agreed-upon future land use. Concern regarding potential migration of contaminants to the Bay was considered valid, and would be addressed as part of the planned study.
- The CSM identified the major anticipated sources, transport mechanisms, and primary exposure media. The major pathway of concern was exposure to contaminated soil with exposure to groundwater being also of potential concern.

Step 2: Identify the Decision

- The goals of the PI center on identification of the types of contaminants present at the site and their spatial boundaries of contamination. The information generated during the PI

will then be used to establish statistically rigorous data collection in the subsequent Remedial Investigation.

- These are the primary study questions the PI produced:
 - What is the nature and extent of contamination in soil and groundwater?
 - Is soil and groundwater contamination a significant source of ongoing contamination to the Bay?
 - Are concentrations of contaminants greater than background concentrations?
 - Are concentrations of contaminants in addition to PAHs greater than risk-based screening criteria?
 - What is the depth, direction, and rate of groundwater flow?
 - What is the potential for erosion of contaminated site soils?

Step 3: Identify Inputs to the Decision

- Total PAHs and Benzo(a)pyrene-(BaP) equivalent values were considered to be a good surrogate for the presence of soil contamination using immunoassay techniques for field measurements (Appendix A).
- In order to confirm that the immunoassay results would provide data of adequate quality for delineating contaminated soils, a small pilot test using SOPs from a similar investigation consisting of five surface soil samples was ordered with results to be presented by the second meeting. *Topdog Inc.*, when preparing for second meeting, made the team aware that previous steps may have to be refined.

Meeting 2: DQO Steps 2 and 3 Refined; Steps 4 – 7 Investigated.

Step 2: Identify the Decision

- Discussion of whether the proposed study questions are objectively testable led to the following revised versions:
 - What is the spatial distribution of contaminants in soil and are they greater than the screening criteria?
 - Has soil contamination reached the shallow aquifer?
 - Is soil and groundwater contamination a significant source of ongoing contamination to the Bay?
- The secondary study vs. primary study questions were then revised accordingly as follows:
 - In what areas of the site do COPC concentrations in soil exceed background concentrations or industrial risk screening levels?
 - Are soil contaminants detected in groundwater at locations of relatively high soil contamination?

- What is the potential for erosion of contaminated site soils?

Step 3: Identify Inputs to the Decision

- Analytical laboratory PAH results from the five pilot test samples were evaluated. For these pilot samples, the degree of correlation between total PAH and BaP equivalent values (a function of the individual PAHs) was reasonably high ($r^2=0.82$) thus confirming that the field immunoassay method was appropriate for establishing the boundaries of soil contamination.
- Based on information from similar former MGP sites, PAHs, VOCs, metals and cyanide were identified as potential site contaminants.
- It was decided that background data would be compared to the fixed lab confirmation data collected using a range of statistical tests sensitive to overall shifts as well as shifts in the upper tail of a distribution (such as would be related to a “hot spot”).
- In order to assure that background comparisons and estimates of mean and variance can be performed for all potential site contaminants, it was decided that a minimum number of fixed laboratory confirmation samples would be analyzed for each of the subareas.

Step 4: Define the Boundaries of the Study

- It was decided to divide (stratify) the site after the PI into regions of relatively similar contaminant concentrations. Estimates of the mean and variance of soil contaminant concentrations in these subareas would then be used to guide subarea-specific sampling in the RI with decisions made on a subarea basis.
- Well logs from properties adjoining *M&H Ltd.* indicate that the depth to groundwater of the first water-bearing zone in the area is between 15 and 25 ft. below ground surface. There is also evidence of a clay layer in the borehole logs at a depth of approximately 10 ft. below ground surface. This clay layer is ubiquitous in the alluvial deposits near the Bay and is believed to be continuous.
- The boundaries of the onshore investigation were accepted with the provision that they could be extended if soil or groundwater sampling indicated contamination outside these boundaries.

Step 5: Develop a Decision Rule

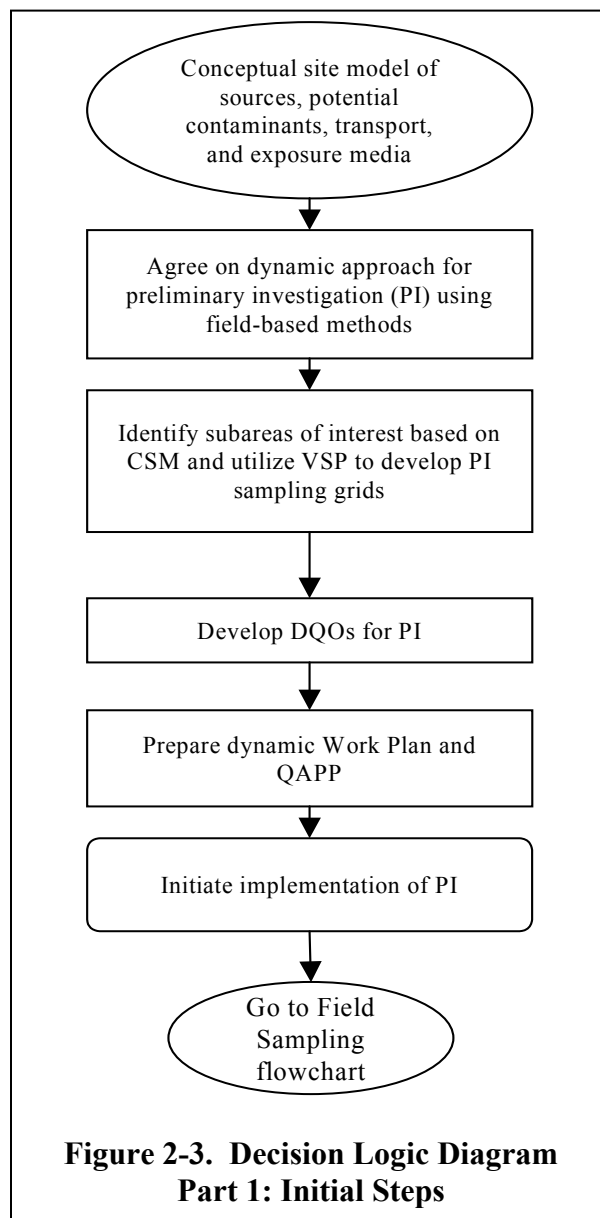
- The facilitator presented a draft decision flowchart addressing the primary study questions that had been established at the previous meeting.

Step 6: Specify Tolerable Limits on Decision Errors

- For the Oil Storage Area and Operations Areas the team will evaluate the use of different grid scales to intersect a potential hot spot. The selection of a grid size will be done by examining different options together with the probability of failing to obtain a sample from a hot spot smaller than a certain size and shape. *Visual Sample Plan (VSP)** is a software tool that may be helpful in this exercise.

Step 7: Develop the Plan for Obtaining Data

- Several draft sampling designs were generated for discussion using VSP. Denser sampling was proposed for the Operations and Oil Tanks Areas based on an increased concern of missing contaminants, the higher probability of contamination in these areas, and the potential for smaller locally elevated areas of contamination. Systematic sampling grids were selected and as part of the evaluation of grid spacing options, the size and shape of a hot spot that the grid would have a high probability of intersecting were considered. A draft design was produced with a sampling grid of approximately 150 ft. in the Operations and Oil Tanks Areas and approximately 300 ft. in the rest of the site.
- The initial sampling design resulted in 30 sampling locations within the Operations Area, 10 locations within the Oil Tanks Area, and 20 samples in the remaining site area. Samples from multiple depths were discussed. Initial cost estimates for collecting 60 samples at 3-4 depth intervals, and performing field tests on these samples was roughly \$36,000 (240 @ \$150 per sample). Fixed laboratory confirmation samples would be roughly \$1,000 per sample.



* VSP produces all the plans necessary for sound statistically defensible data, and it provides for immediate comparisons of selected sampling plans. It has the capability of allowing for different sampling and analysis costs to be compared and contrasted. For details about VSP, see Appendix B.

Meeting 3: DQO Steps 4 – 7 Refined

Step 4: Define the Boundaries of the Study

- After extended discussion of soil sampling depths intervals; the top interval was defined as 0 – 0.5 ft. (the traditional definition of “surface soil.”), and a second interval of 0.5 – 3 ft. defined (the approximate depth commonly associated with trenches for utilities and building foundations).
- Appropriate depths for lower depth intervals (3- 6 ft.; 6 -12 ft., 12 ft. to groundwater) were not easily agreed upon. The use of four intervals proposed by the State Department of the Environment was considered acceptable, with the addition of a decision rule to forego the two deepest intervals if total PAH concentrations in higher intervals were essentially consistent with background concentrations.

Step 5: Develop a Decision Rule

- Reconsideration of the draft decision flow chart lead to a decision to divide it into three separate charts, two that address the planning stages (see Figures 2-3 and 2-4) and one that addresses data assessment (Figure 2-5). Figure 2-4 focuses on decisions that will be made essentially in real time during the sampling campaign as part of a dynamic, adaptive sampling plan.

Step 6: Specify Tolerable Limits on Decision Errors

- The planning team elected to specify error tolerances based on the desire not to miss hot spots of a specified size and shape, if they are present.
- Given the site conceptual model, and experience at other MGP sites, small hotspots were not expected. The shape of a hot spot was also unknown, however an elliptical shape was agreed to be a reasonable assumption. Finding a hot spot during the preliminary investigation will result in further study to evaluate the risk associated with the elevated chemistry – however, failing to find a hot spot might result in no further action. Using VSP to investigate alternatives, the sample sizes required to have a 95% probability of detecting a hot spot of approximately 1/8 acre in the two primary areas of concern, and one acre in the rest of the site were determined. A five percent chance of missing a hot spot of this size was considered acceptable, recognizing the consequences of failing to evaluate the need for action.

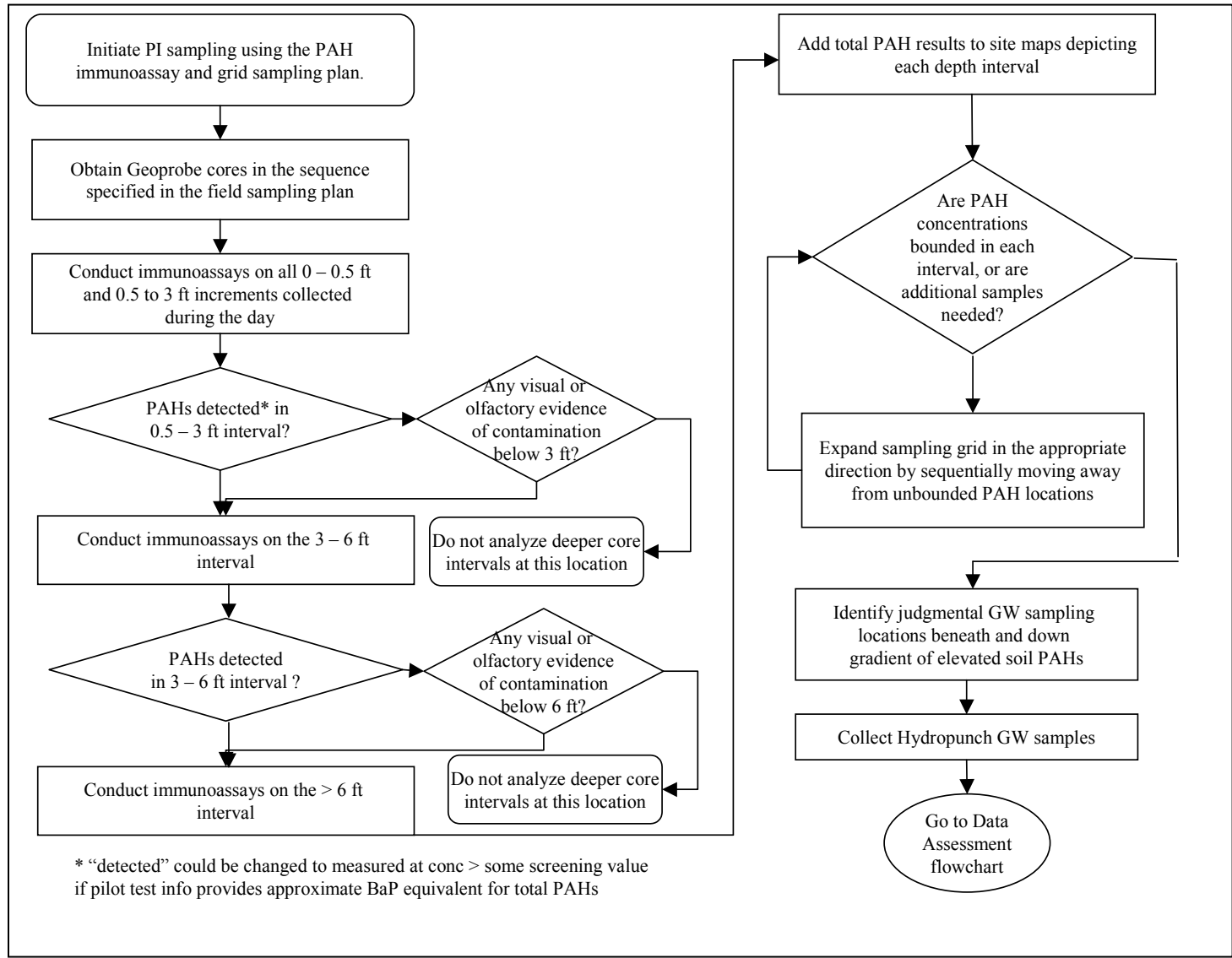


Figure 2-4. Decision Logic Diagram Part 2: Preliminary Investigation Implementation Logic

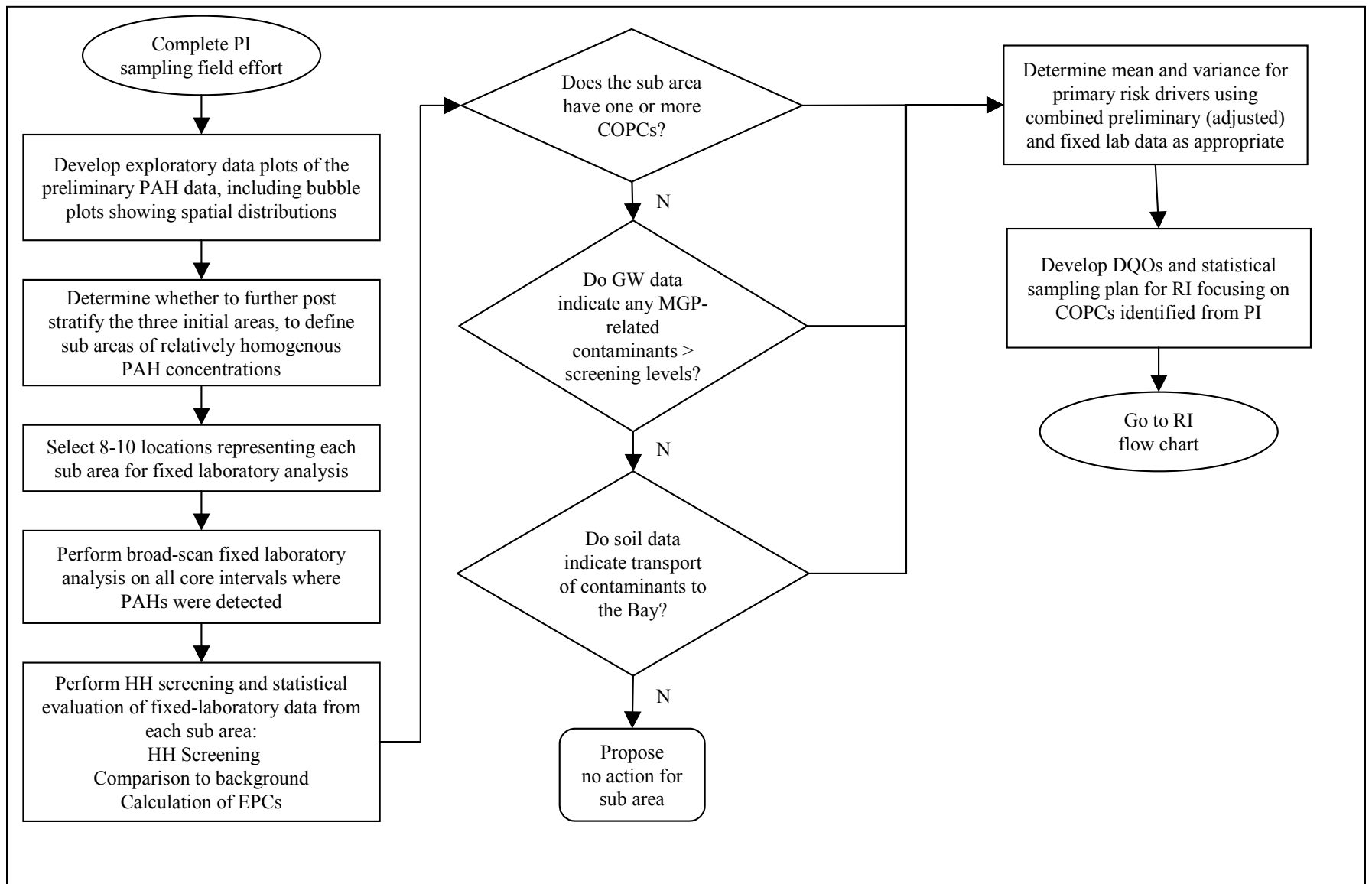


Figure 2-5. Real-Time Data Assessment Process Flow Chart

Step 7: Develop the Plan for Obtaining Data

VSP software was used to generate sampling locations for total PAHs in soil using systematic grid sampling with a random start. Within the areas where historical information suggests contamination may be likely, a triangular sampling grid of approximately 75 ft. was selected to maximize the probability of detecting any localized areas (approximately 5000 sq. ft.) of elevated PAHs. This resulted in a planned 41 sampling locations in the Operations Area and 15 samples in the Oil Tanks area. The remainder of the site was given a 200-ft. sampling grid, with a resulting 24 samples. This sampling density provided a 95% chance of hitting a hot spot approximately 35,500 sq. ft. (less than an acre). With 320 potential samples at \$150 per sample (for sample collection and immunoassay analyses), the cost of this part of the plan would be approximately \$50,000 (approximately half the original budget). The sample locations are shown in Figure 2-6 where “A” represents the Oil Storage Area, “B” represents the Operations Area, and “C” represents the general environs of the site.

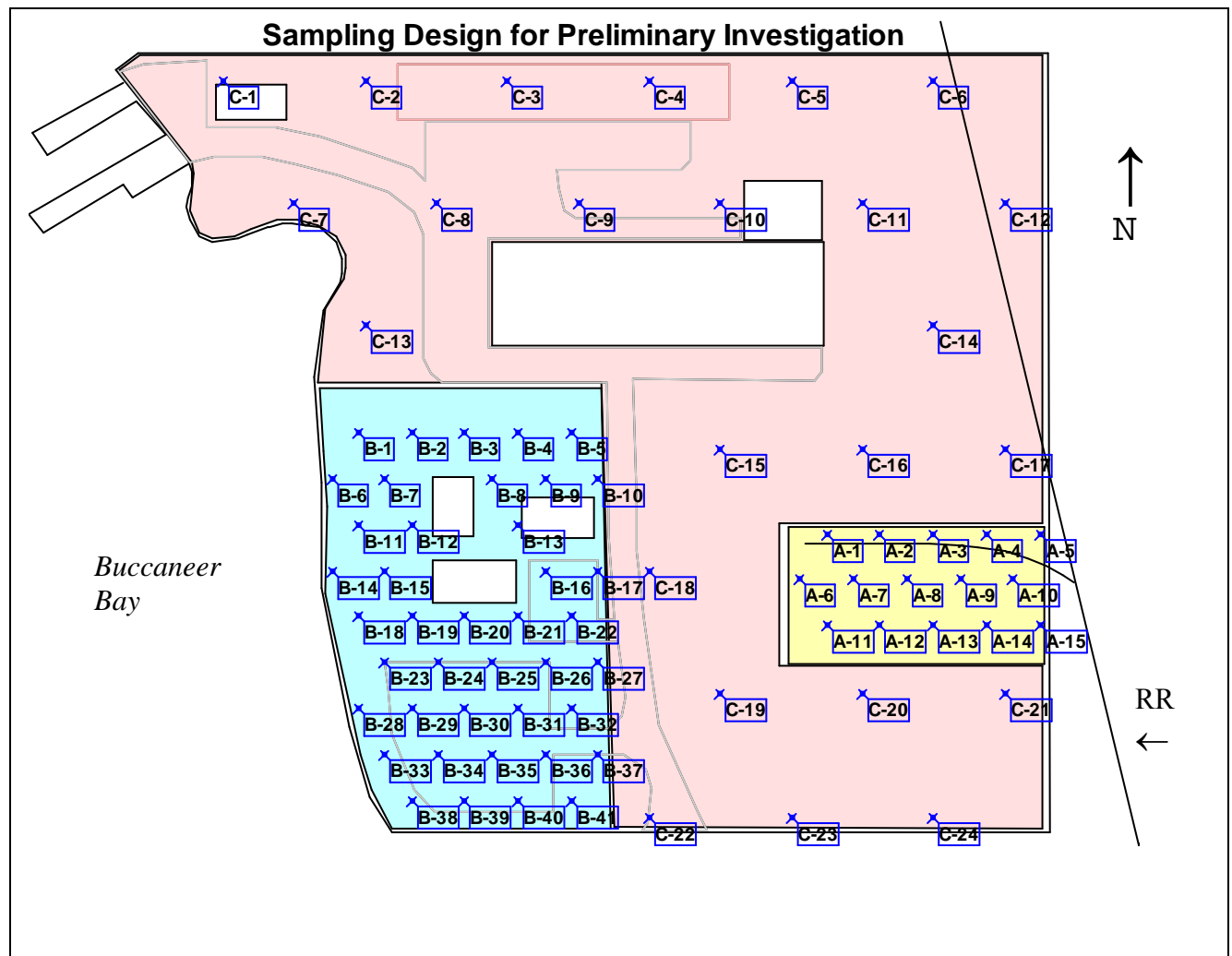


Figure 2-6. Sample Locations for Preliminary Investigation

As the PI was an adaptive sampling campaign, the actual number and location of samples were determined in the field in accordance with the following criteria:

Soil Sampling Criteria: Geoprobe® cores were obtained for all grid sampling points; sampling was conducted in Areas A, B and C sequentially. The data assessment team determined whether to extend the sampling grids beyond the northern, eastern, or southern boundaries of the site, and whether to extend the finer Area A or B grids into Area C. The team used the following considerations in making this decision:

- Consistency of data with conceptual site model;
- Evidence of decreasing total PAH concentrations at the edge of a grid;
- Relation of total PAH site data to background concentrations; and,
- Relation of total PAH site data to the risk-based screening criterion.

Soil Depth Interval Sampling Criteria: PAH immunoassay data were collected from the 0 – 0.5 ft. and 0.5 – 3 ft. intervals at all sampling locations. If total PAHs were not detected in the 0.5 – 3 ft. interval, and if there was no visual or olfactory evidence of hydrocarbon contamination in the deeper core, no additional depth intervals would be sampled. If otherwise, then PAH immunoassay data would be collected from the 3 – 6 ft. interval. These criteria were to be repeated to determine whether to sample the last depth interval of 6 ft. – groundwater.

Soil Confirmation Sampling Criteria: PI soil sampling subareas would be initially defined using the total PAH immunoassay data. Splits from a minimum of eight or a maximum of 10% of the soil samples collected in a subarea, were submitted to an analytical laboratory. Previous experience with this kind of soil contamination indicated the use of normality-based methods for calculating power would be appropriate criteria. By collecting a minimum of eight samples per subarea, and having eight or more from a reference area, comparisons to background distributions could be made, and differences of approximately 50-75% detected with reasonable power (greater than .70), and larger shifts such as 100% with greater power.

Groundwater Sampling Criteria: Shallow aquifer groundwater samples will be collected by Hydropunch™ upgradient of the site boundary, and downgradient from each soil subarea in a location of the highest total PAH measurement in the deepest soil interval. If the selected location does not yield a sufficient water sample, another location in the subarea will be selected. Three attempts per subarea will be made to obtain a groundwater sample.

The final DQOs for the PI are summarized in Appendix C.

2.2 Implementation of the Preliminary Investigation

Preliminary Investigation soil and groundwater sampling was conducted over a period of 10 days. Sampling was begun in the Operations area, moved to the Oil Storage area, and finally was completed in the remaining area of the site. Approximately 10 cores per day were collected. Batches of five cores were sent to a trailer on site for sample preparation, logging, and analysis. Extractions for all core intervals in a batch were performed simultaneously, and immunoassays

initiated as soon as possible. In this manner, the results from 10 locations and associated QA samples were generated in a given day.

The field team kept a daily log of the field screening results and resulting statistical adaptive decisions related to the study design. A brief summary follows with a more detailed discussion in Appendix C.

For the Oil Storage Area, a total of 15 cores were collected resulting in 52 immunoassays for total PAHs; seven of which were from the 3-6 ft. interval, but none of which was in the > 6 ft. interval. No highly contaminated soil in this area was found; however, total PAH concentrations on the southern side of the rail spur in the top three depth intervals were higher than north of the rail spur. In general, the surface interval was cleaner than subsurface samples.

For the larger Operations Area, cores were collected at the planned 41 sampling locations. Many of these samples were taken from beneath the gravel parking lot, south of the buildings. In these cases, the gravel layer was scraped away, and cores were taken starting at the soil interface. The original sampling grid in this area was extended by two rows on the southern boundary of the Operations Area due to the detection of high concentrations of total PAHs in the southern region of the area and a total of 201 immunoassays performed for total PAHs. Elevated PAHs continued to be found in the first new row to the south, but dropped to approximately ambient levels in the last new row. The highest concentrations were closely related to visual observations of a powdery black substance, probably lampblack, in the core.

For the remainder of the site, Site Environs, cores from the initial 24 sampling locations were collected and field-screened for total PAHs. A visual analysis of some core material from sample C-8 revealed debris mixed in with native soil, ash, slag, and lampblack. In the deeper intervals, the core contained thick, black, tarry wastes which were analyzed by immunoassay and found to contain levels of PAHs higher than the calibration scale of the kit.

After discussion it appeared that this sampling location was in alignment with a ravine-like feature extending from the parking area to a cove in Buccaneer Bay. Based on concentration, the team decided to take samples up- and down-gradient from C-8, to further evaluate what appeared to be materials dumped in a former ravine. In all, five additional cores were obtained and analyzed: two in the direction of the cove, and three upgradient. Tarry material was found in the samples surrounding C-8, especially below three ft. of cover. An additional core near the cove revealed consistent soil-fill with no evidence of tarry material, and was not analyzed.

2.3 Assessment of the Preliminary Investigation

The logic to data assessment for the PI is summarized by the following three subsections:

2.3.1 Real-Time Data Assessment and Post Stratification

On a daily basis, sample results were added to a spreadsheet and imported (as txt files) into the base map in VSP to post the results for each sample depth interval on a map. As

previously described in Section 2.2, the initial sampling plan was supplemented by extending the sampling grid two additional rows to the south of the Operations Area, and along the ravine in the Site Environs area. A split of each homogenized core interval was created, labeled, and stored in a freezer, following standard tracking and chain of custody procedures.

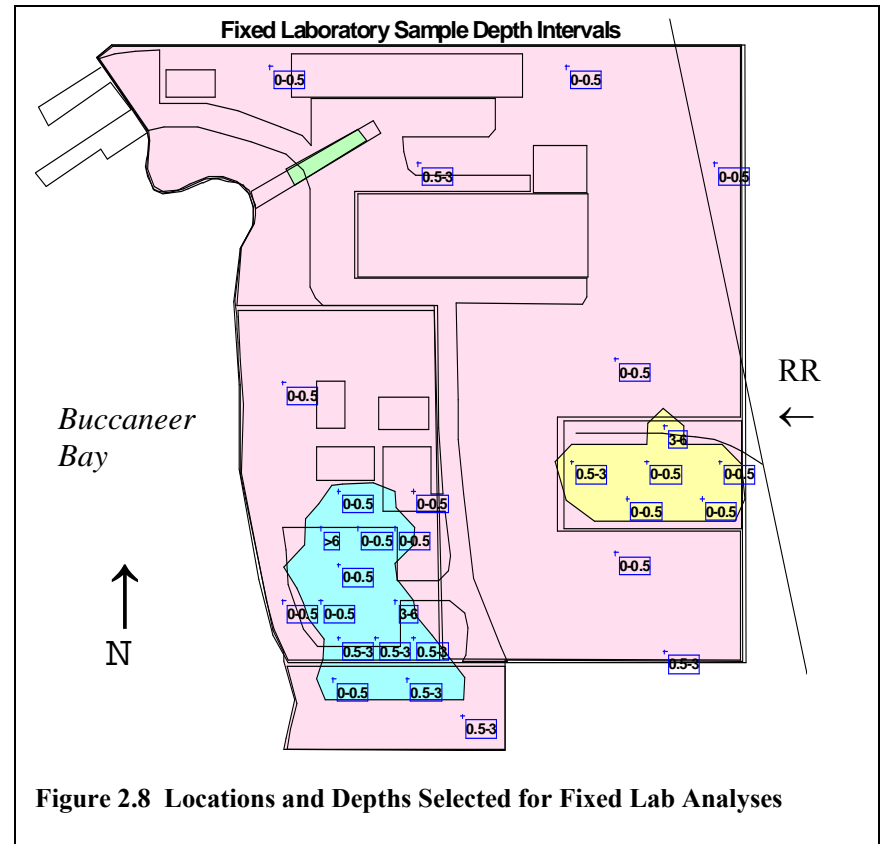
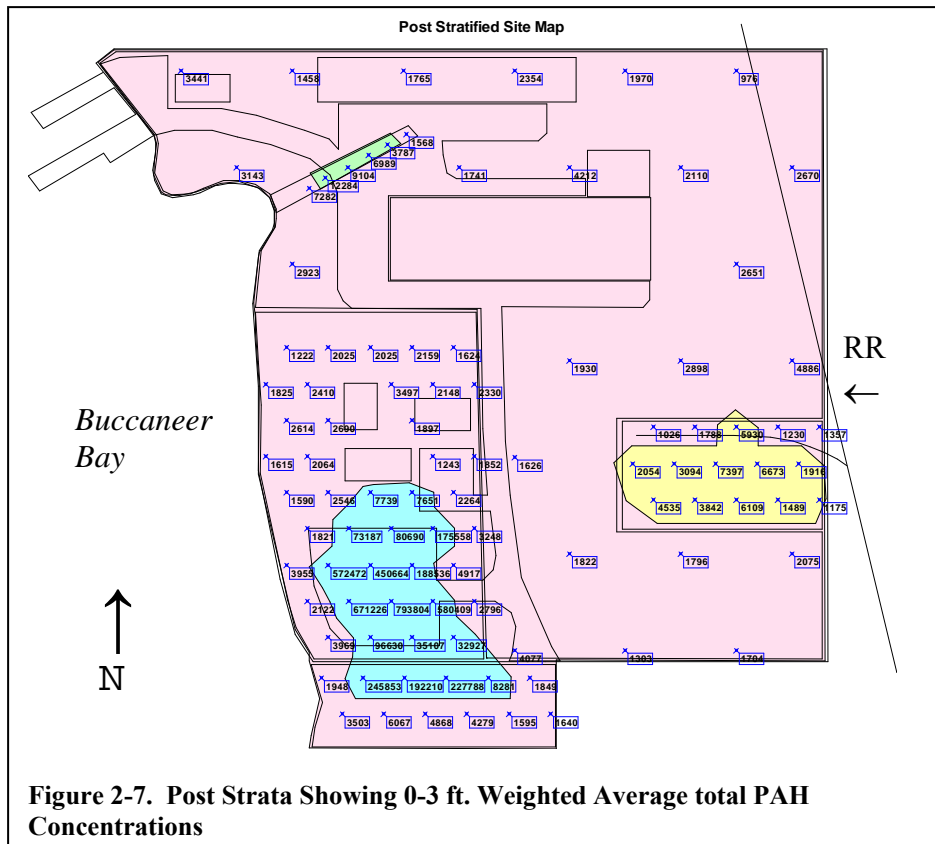
After completion of all PAH immunoassays, the concentrations at each depth interval were evaluated for post-stratifying* the site. To assist in post stratification, weighted average 0-3 ft. concentrations were created and posting plots created. A decision was made to perform an expedited removal action to excavate the heavy tarry material (that was very high in PAHs) and related ash and other debris in the ravine. This decision was made due to the high concentrations, well-defined boundaries, and potential for migration via the historical ravine to offshore sediments under flood events. The boundaries of the Oil Storage area and Operations Area boundaries were refined, with a portion of each of those areas being added to the general Site Environs, leaving those stations where concentrations were more uniformly elevated. The post stratified sub- areas identified through this process became the spatial boundaries for all subsequent data assessment, including risk screening, COPC identification, background comparisons, and calculation of the mean and variances to support the design of follow on investigations. Figure 2-7 shows the 0-3 ft. weighted average total PAH, and boundaries of the post stratified areas.

To ensure that each area was represented by fixed laboratory results, a systematic subsample of each of the post-stratified sub areas was taken, and depth intervals randomly selected from each chosen location. The final selection included seven (six systematic plus one judgment) samples from the Oil Storage Area, 12 (nine systematic plus three judgment) from the Operations Area, and 11 (all systematic) from the Site Environs (Figure 2-8). The judgmental samples were added to assist in forming fixed lab to field method regressions. Since a decision was made to perform an expedited action in the ravine, no samples from that area were selected for fixed lab analysis.

2.3.2 Sample Size Selection for Background Comparisons

The Background samples were taken from a similar area to the east of the railway line of the MGP site with approximately the same sample matrix as those samples taken from the MGP site. There were 15 existing background samples with a mean total PAH concentration of 1110.9 ppb and a standard deviation of 942.4 ppb (see Appendix D).

* Post-stratification is a re-examination of the boundaries of the original strata (Areas) to ascertain if the boundaries were appropriate and, if not, redefine them. The technique improves the precision of the estimate made.



2.3.3 Decision to Use Fixed Laboratory Data with BaP Equivalents

The Immunoassay results, Fixed Lab total PAH data, and BaP Equivalents data were compared to assess their reliability. Although there was a moderately good correlation over the entire range of concentrations it was significantly less for lower concentrations (see discussion in Appendix E). This lack of a strong correlation led to the decision to use Fixed Lab data only and not the Immunoassay data.

The results from fixed laboratory data were used to estimate potential carcinogenic risk based on the toxicity of individual chemicals. Concentrations of individual PAHs were converted into BaP equivalents in order to estimate the overall cancer risk from the combination of potentially carcinogenic PAHs. The BaP equivalent is based on the EPA 1993 toxicity equivalency factors and the concentrations of the seven individual carcinogenic PAHs. The BaP equivalent calculation is based on a BaP toxicity equivalence factor multiplied by the concentration of the PAH for each of the seven carcinogenic PAHs:

$$\text{BaP equivalents} = (0.1) \text{ benzo(a)anthracene} + (1.0)\text{BaP} + (0.1)\text{benzo(b)flouranthene} + (0.01) \text{ benzo(k)flouranthene} + (0.001)\text{chrysene} + (0.1) \text{ ideno(1,2,3cd)pyrene}.$$

The standard EPA industrial exposure scenario estimates the exposure of an outdoor worker over 25 years and converts the BaP equivalents in soil concentration into a risk level. From discussions with between risk assessors from the regulatory agencies and MGP personnel, a risk of 1×10^{-5} was determined to be acceptable which converts to 2300 ppb BaP equivalents in soil. Therefore, if the BaP equivalent for a sample is below the BaP equivalents corresponding to 10^{-5} risk (2300 ppb), there is no need to screen that sample for carcinogenic risk from individual PAHs.

2.4 Preliminary Investigation Conclusions

Concentrations of metals in soils showed a slight elevation over Background, but all (including Arsenic) were well below the applicable risk-based screening levels. Arsenic was found above the industrial screening level at each of the areas but not thought to be a site contaminant of concern. Based on the potential for risk relative to the MGP contaminants, Arsenic should be included for the RI and other metals excluded from further investigation.

For the Post-Stratified Operations Area, the BaP equivalents for a number of samples exceeded the BaP risk benchmark. These screening results indicate that, based on carcinogenic risk from PAHs under an industrial scenario, the Post-Stratified Operations Area should be the focus of any further soil investigation, while the other two areas could be proposed for no further investigation based on the results of the preliminary investigation. The maximum concentration of BaP equivalents found in one sample generates an estimate of the magnitude of the potential carcinogenic risk from PAH to be 1.1×10^{-4} ; approximately 10 times the acceptable risk level agreed to. Further investigation of carcinogenic PAHs in the RI should provide a more accurate estimate of the potential risk to human health at this site.

Tables 2-1, 2-2, and 2-3 show the statistical summaries for the three areas.

Parameter	Oil Storage Area (A)	Operations Area (B)	Site Environs (C)
BaP Equivalents	330.2	7465.3	420.0
Fixed Lab.total PAH	2538.3	57445.2	2747.8
Immuno. total PAH	3749.0	184752.6	3132.3
Benzo(a)anthracene	99.5	2472.6	133.8
Benzo(a)pyrene	189.8	4260.3	218.2
Benzo(b)fluoranthene	187.7	2762.4	167.1
Benzo(g,h,i)perylene	277.2	3823.1	270.8
Benzo(k)fluorathene	80.7	1408.0	71.7
Chrysene	140.8	2863.1	151.0
Fluoranthene	537.0	14885.1	578.9
Indeno(1,2,3cd)pyrene	291.8	3901.0	216.1
Phenanthrene	173.5	5308.9	176.7
Pyrene	407.7	11380.9	550.5
Acenaphthylene	0	425.0*	9.1*
Anthracene	71.2*	1372.4*	54.5*
Dibenzo(a,h)anthracene	81.5*	2274.4*	149.3*
Naphthalene	0	13.3*	0
Acenaphthene	0	2.4*	0
Fluorine	0	292.1*	0

Background BaP Equivalents Mean = 130.33

* assumes all nondetects = 0

Parameter	Oil Storage Area (A)		Operations Area (B)		Site Environs (C)	
	Median	Maximum	Median	Maximum	Median	Maximum
BaP Equivalents	281.2	754.6	4441.0	30167.0	349.4	1065.2
Fixed Lab.total PAH	2550.0	5430.0	20284.0	197200.0	2493.0	6906.0
Immuno. total PAH	4120.5	6697.0	60677.0	856420.0	3582.0	4070.0
Benzo(a)anthracene	97.0	200.0	1307.0	9700.0	110.0	335.0
Benzo(a)pyrene	140.0	450.0	1910.0	15000.0	190.0	559.0
Benzo(b)fluoranthene	196.0	340.0	1206.0	9300.0	151.0	335.0
Benzo(g,h,i)perylene	250.0	590.0	2111.0	11000.0	280.0	407.0
Benzo(k)fluorathene	89.0	140.0	683.0	5500.0	60.0	213.0
Chrysene	145.0	260.0	1106.0	12000.0	121.0	427.0
Fluoranthene	530.0	1200.0	3216.0	55000.0	500.0	1727.0
Indeno(1,2,3cd)pyrene	277.5	620.0	1608.0	12000.0	230.0	406.0
Phenanthrene	125.0	560.0	1100.0	21000.0	120.0	802.0
Pyrene	395.0	870.0	3417.0	45000.0	413.0	1117.0
Acenaphthylene	0	0	0*	1985.0	0*	66.0
Anthracene	11.5*	360.0	180.0*	4900.0	25.0*	259.0
Dibenzo(a,h)anthracene	14.5*	320.0	240.0*	12000.0	102.0*	400.0
Naphthalene	0	0	0*	120.0	0	0
Acenaphthene	0	0	0*	22.0	0	0
Fluorine	0	0	0*	1389.0	0	0

* assumes all nondetects = 0

Parameter	Oil Storage Area (A)		Operations Area (B)		Site Environs (C)	
	# Detects	Std. Dev.	# Detects	Std. Dev.	# Detects	Std. Dev.
BaP Equivalents	6	285.5	9	9495.9	11	280.1
Fixed Lab. total PAH	6	2073.4	9	67230.2	11	1763.0
Immuno. total PAH	6	2398.4	9	294886.6	11	1052.9
Benzo(a)anthracene	6	72.0	9	3098.3	11	90.8
Benzo(a)pyrene	6	167.4	9	4798.6	11	134.3
Benzo(b)fluoranthene	6	144.1	9	3013.4	11	94.6
Benzo(g,h,i)perylene	6	256.1	9	3760.4	11	123.8
Benzo(k)fluorathene	6	53.2	9	1708.9	11	52.5
Chrysene	6	101.1	9	3854.5	11	107.6
Fluoranthene	6	449.7	9	19450.5	11	509.4
Indeno(1,2,3cd)pyrene	6	269.8	9	4107.2	11	97.5
Phenanthrene	6	199.5	9	7732.7	11	226.3
Pyrene	6	309.6	9	14539.9	11	360.1
Acenaphthylene	0	0	3	808.1*	2	21.6*
Anthracene	4	142.5*	8	1958.5*	10	84.8*
Dibenzo(a,h)anthracene	3	128.8*	6	4089.5*	8	154.9*
Naphthalene	0	0	1	0	0	0
Acenaphthene	0	0	1	0	0	0
Fluorine	0	0	3	570.4*	0	0

Background BaP Equivalents = 15, with Std. Dev. 155.6

* assumes all nondetects = 0

The conclusions from inspection of Tables 2-1, 2-2, 2-3:

- Concentrations of BaP Equivalents and total PAH were highest in the Operations Area as shown by comparisons of means and maxima (Tables 2-1 and 2-2).
- There was almost no difference between Oil Storage Area and Site Environs averages, but a large difference between both and the Operations Area (Table 2-1).
- There was almost no difference between Oil Storage Area and Site Environs in terms of variability, but the Operations Area very much more variable (Table 2-3).
- Mean concentrations of BaP Equivalents and total PAH were significantly above Background for the Operations Area and Site Environs, but less so for the Oil Storage Area (see Appendix F for further discussion).
- Mean concentration of BaP Equivalents exceeded the industrial worker risk-based screening level of 2300ppb in only the Operations Area (Table 2-1).
- Maximum concentration of BaP Equivalents did not exceed the industrial worker risk-based screening level of 2300ppb in the Oil Storage Area and Site Environs (Table 2-2).

- Groundwater was minimally affected by MGP operations as contamination was minimal in groundwater.

The overall conclusion to be drawn for the Remedial Investigation planning phase was that only the Operations Area as defined by the post stratification needed to be the focus of attention, and that the COPCs should be the BaP equivalents and Arsenic.

With the conclusion of the Preliminary Investigation, the data are now available for the Remedial Investigation phase.

3. REMEDIAL INVESTIGATION

With the newly available preliminary data, the DQO Process was used again to develop quantitative performance and acceptance criteria for the next phase, Remedial Investigation (RI). The primary purpose of the RI was to produce data suitable for conducting a baseline human health risk assessment.

3.1 Planning the Remedial Investigation

After completion of the preliminary investigation phase, the Planning Team held a meeting to discuss issues related to development of risk estimates for exposure to COPCs in soil at the Operations Area and the principal issues and agreements summarized in Table 3-1.

Table 3-1. Risk Assessment Meeting Notes
<ul style="list-style-type: none"> • The risk assessor summarized the relevant risk parameters as: <ol style="list-style-type: none"> 1) 0.0 - 0.5 ft. interval (surface soil for an industrial outdoor worker) 2) 0.5 - 1.5 ft. interval (possible exposures during construction and trenching) 3) 1.5 – 3.0 ft. (not based on a particular exposure scenario) • The post-stratification was based on a screening comparison to the maximum concentration at each subarea followed by a screening comparison of each fixed lab sample within each subarea. • The BaP screening level of 2300 ppb was revised after discussion to a 15 year exposure with a soil ingestion rate of 25 mg/day (given the eight hour workday); this led to a site specific Remedial Action Concentration of 12,000 ppb. • A not-to-exceed BaP concentration of 15,000 ppb will be used in the RI to evaluate whether the Operations Area might contain small areas of unacceptably elevated concentrations of BaP equivalents (hot spots).

After agreeing on the basic risk assessment parameters, the planning team discussed DQOs for the additional data required to conduct the risk assessment. The hardest work had

been completed by this point; the rest of the procedure would be relatively easy. Table 3-2 presents the DQOs resulted from the planning team discussions.

Step 1: State the Problem	The preliminary investigation indicated PAHs exceeding the soil screening threshold were primarily located in the Operations Area (the focus of this DQO), and the ravine area (to be investigated later).
Step 2: Identify the Decision	Does PAH contamination in Operations Area soils pose an unacceptable risk to users of the site and need to be evaluated in a further study?
Step 3: Identify Inputs to the Decision	<ul style="list-style-type: none"> • Site-specific risk-based Remedial Action Criteria for BaP-equivalents set at 12,000 ppb • Concentration of individual PAHs and BaP equivalents from soil samples to be collected in Operations Area.
Step 4: Define the Boundaries Of the Study	The Operations Area has been redefined in the X, Y, and Z dimensions by means of post-stratification of the screening and fixed laboratory data. A single decision will be made for this area, by evaluating the PAH concentrations to several depths (0.0-0.5 ft., 0.0-1.5 ft. and 0.0-3.0 ft.).
Step 5: Develop a Decision Rule	<p>If the 95% Upper Confidence Limit (UCL) values for BaP-equivalents in the top 0.5 ft, or any weighted average down to 3 ft. exceed the applicable site-specific criteria, then conclude that the Operations Area presents an unacceptable risk, and proceed with a determination of the extent of remediation necessary to lower risk to an acceptable level, and the evaluation of remedial alternatives in a future study.</p> <p>If any single value is found to exceed 15,000 ppb, then determine the extent of the hot spot and evaluate remedial alternatives for hot spot removal.</p>
Step 6: Specify Tolerable Limits on Decision Errors	<p>Null hypothesis (baseline assumption): Average site concentrations are > 3000 ppb. A false rejection would be an indication that the average is less than 3000 ppb, when in truth it is not. A false acceptance would be an indication that the average is greater than 3000 ppb, when in truth it is not.</p> <p>The following error tolerance specifications were selected for this purpose:</p> <ul style="list-style-type: none"> • false rejection rate: 10% • false acceptance rate: 5% (in keeping with the use of 95% UCL to represent RME exposures in risk assessments) • width of gray region: 6000 ppb • estimated standard deviation based on fixed lab BaP equivalents: 8335
Step 7: Develop the Plan for Obtaining Data	21 sampling points were selected. Given the skewed distribution, the sample sizes were based on use of the nonparametric Wilcoxon Rank Sum test. A systematic grid was placed over the Operations Area, by selecting a random start that results in a grid staggered from the original grid. Figure 3-1 shows the location of the new sampling locations.

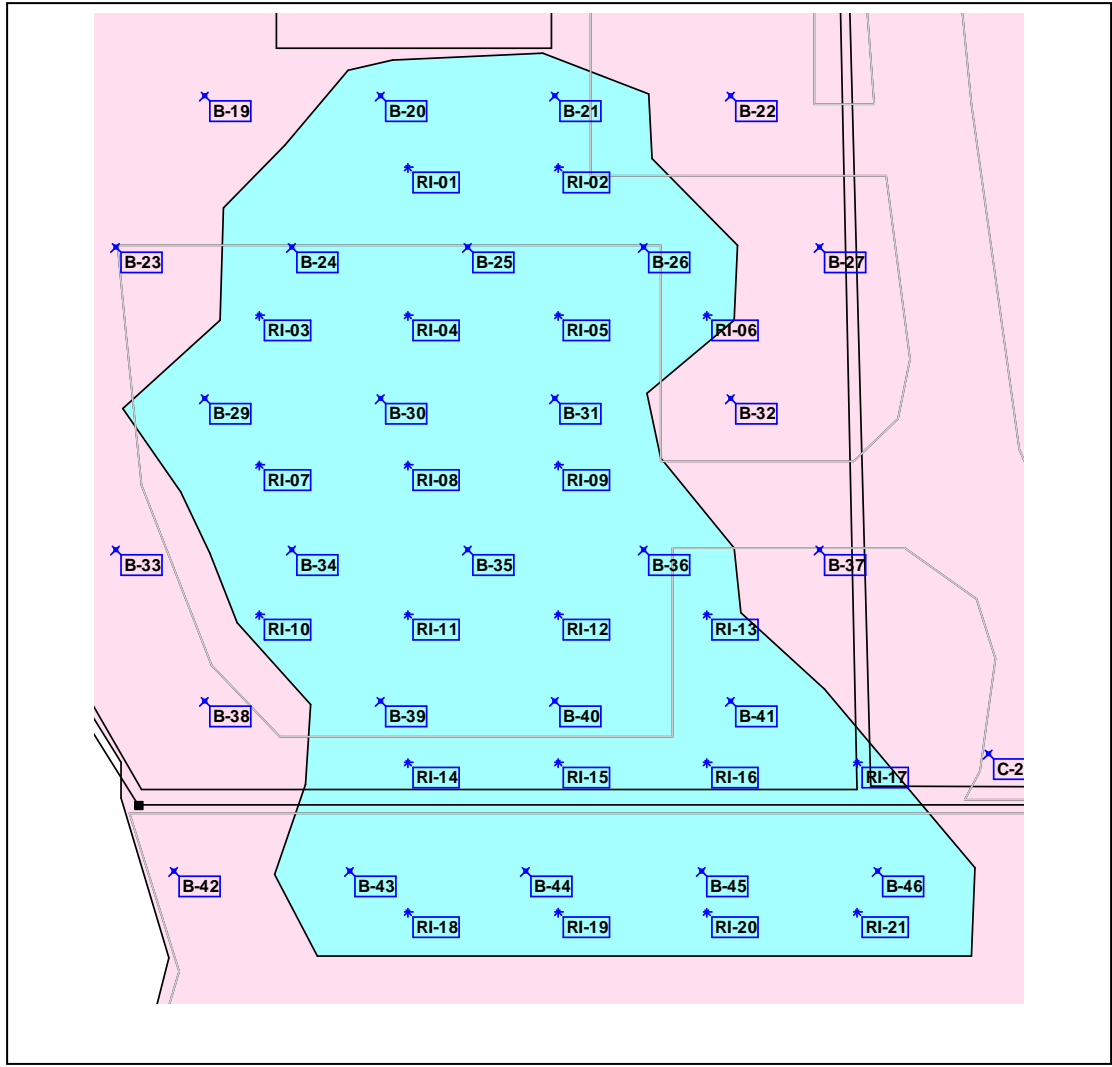


Figure 3-1. Combined Preliminary and RI Sampling Location Map

3.2 Implementation of the Remedial Investigation

Collection of the 21 core samples occurred over a three-day period. Since a large number of the samples were located under the large gravel parking lot, the following procedures were used. Hand shovels were used to move the gravels aside, down to the surface of the soil. A geoprobe sampler was then used to drive a two inch core down to the three ft. depth, and the core samples taken to an on-site trailer for preparation for shipment to the fixed laboratory. Chain-of-custody procedures as specified in the QA Project Plan were followed, and none of the samples were lost during sample preparation or shipment, and none of the holding time requirements were missed.

Fixed laboratory results were generated in one month, and electronic deliverables were sent to an independent firm for validation, which took an additional week. After completing the verification and validation process, data analysis and data quality assessment activities ensued.

3.3 Assessment of the Remedial Investigation

RI data were received from the fixed laboratory and an initial inspection of the data showed that none of the individual non-carcinogenic PAH values exceeded the screening levels used in the preliminary investigation. BaP equivalents were then calculated for the carcinogenic PAH values for each depth, at each sampling location, following the same procedure that was used for the fixed laboratory data in the Preliminary Investigation.

3.3.1 The Remedial Investigation Data

After verification and validation, the BaP equivalent values obtained from the Post Stratified Area are presented in Table 3-3.

Table 3-3. BaP Equivalent Concentrations (ppb) for the Post Stratified Operations Area (values exceeding 15,000 ppb are indicated in boldface)							
RI Data							
Label	Easting	Northing	Depth			Weighted Average	
			0.0 – 0.5 ft	0.5 – 1.5 ft	1.5 – 3.0 ft	0.0 – 1.5	0.0 – 3.0
RI-01	530.5	-776.2	1959	2785	2740	2510	2625
RI-02	594.4	-776.2	4708	3693	2908	4031	3470
RI-03	466.7	-840.1	4057	7377	3063	6270	4667
RI-04	530.5	-840.1	4392	8589	3181	7190	5186
RI-05	594.4	-840.1	6674	6434	2472	6514	4493
RI-06	658.3	-840.1	3547	5379	2256	4768	3512
RI-07	466.7	-904.0	6470	7818	3039	7369	5204
RI-08	530.5	-904.0	8420	14008	2212	12145	7179
RI-09	594.4	-904.0	7963	12428	2752	10940	6846
RI-10	466.7	-967.8	7957	9088	5631	8711	7171
RI-11	530.5	-967.8	21929	18707	15103	19781	17442
RI-12	594.4	-967.8	23198	16251	11641	18567	15104
RI-13	658.3	-967.8	7188	7597	6530	7461	6995
RI-14	530.5	-1031.7	18423	19566	14107	19185	16646
RI-15	594.4	-1031.7	26755	32720	16419	30732	23575
RI-16	658.3	-1031.7	16950	16284	14726	16506	15616
RI-17	722.1	-1031.7	7482	8492	6864	8155	7510
RI-18	530.5	-1095.6	7848	8273	5716	8131	6924
RI-19	594.4	-1095.6	15179	12976	12232	13710	12971
RI-20	658.3	-1095.6	13298	14205	10547	13903	12225
RI-21	722.1	-1095.6	6665	8332	5967	7776	6872

“Label” is the sample identification number and uses additional letters to denote the depth from which the sample was taken (0.0 – 0.5 being “A”; 0.5 – 1.5 being “B,” and 1.5 - 3.0 being “C”). Samples RI-01 to RI-09 are from the north end, samples RI-10 to RI-21 from the south end. “Easting” and “Northing” give the co-ordinates of the sample location on a standard map; these co-ordinates are then translated into actual Post Stratified Operations Area locations. “Depth” gives depth from which the sample was taken such that a three-dimensional picture of the Area may be constructed. “Weighted Average” combines the actual values proportionately, giving more weight to the deeper values than those near the surface. The weighted average for the 0.0 –

1.5 depth is given by 1/3("A" value) + 2/3("B" value); that for the 0.0 – 3.0 depth by 1/6 ("A" value) + 2/6 ("B" value) + 3/6("C" value).

Inspection of Table 3-3 values shows some interesting results. All the values from the northern end of the Post Stratified Operations Area are below 15,000 ppb, but approximately a third of the southern end values exceed 15,000 ppb. A somewhat similar pattern can be seen in the weighted average values. To visualize these patterns better a box-and-whisker plot (Figure 3-2) was developed.

From Figure 3-2 it can be seen that the overall median (horizontal line within the central box) for each of the three depths is well below the Action Level of 12,000 ppb; however, a substantial proportion of each depth's data values are above the Action Level (the upper whiskers), thus confirming the conclusion derived from inspection of the raw data in Table 3-3. Notice how the three plots are very similar in box sizes (showing the majority of the values) and the whiskers or tails which show the extreme values. This indicates the distribution of data values is consistent across the three depths with a slight diminution as the depth increases. The difference in length between upper and lower whiskers shows the data to be skewed towards the higher values but not too different from symmetry (recall, only 21 values were collected).

To further examine the difference between northern and southern locations, two additional box plots were made (Figure 3-3), dividing the data into northern and southern locations within the post stratified operations area. This plot clearly indicates that the values in the northern part of this area are below the action level while those in the southern half contain locations far in excess of the 12,000 ppb Action Level.

3.3.2 Calculation of Upper Confidence Limits

A UCL on the mean (arithmetic average) concentration is used to represent the reasonable maximum concentration in evaluating risk due to site contamination. The manner by which a confidence interval is calculated depends on several factors, the most important being the assumption of normality in data distribution. This assumption is probably not warranted given the appearance of the data in Table 3.2 but providing the normal-based UCL is carefully interpreted, can yield information useful in characterizing the area from which the samples were obtained. The UCL on the mean is given by:

$$\bar{X} + t_{n-1, 1-\alpha} \cdot \frac{s}{\sqrt{n}}$$

where \bar{X} represents the mean of the area under investigation, s the standard deviation, n the number of samples, and $t_{n-1, 1-\alpha}$ is taken from standard t-tables. Tables 3-4, 3-5, and 3-6 show the 95% confidence intervals for the Northern Locations, the Southern Locations, and for All Locations, for all three depths.

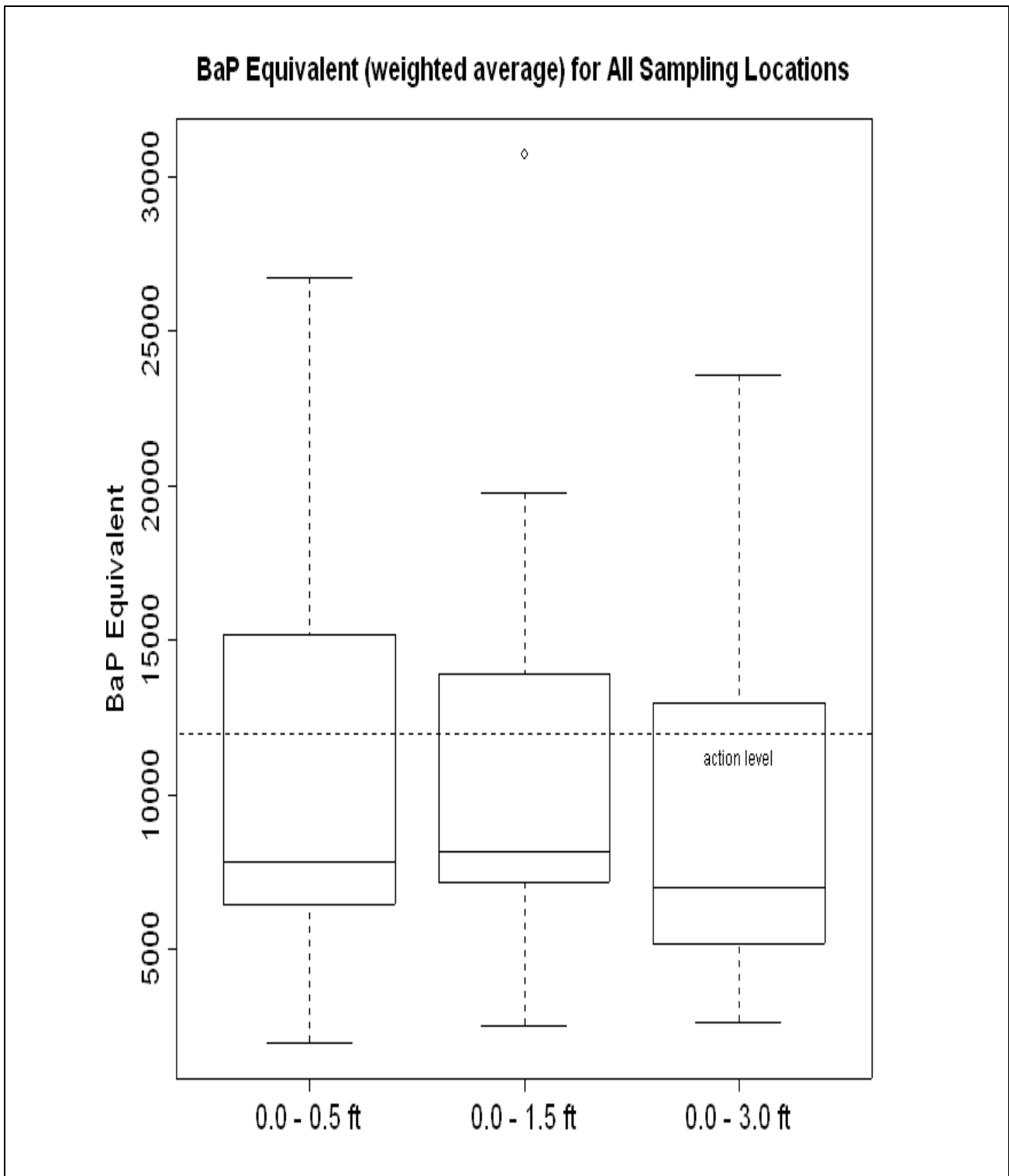


Figure 3-2. Box and Whisker Plots BaP Equivalent Concentration Weighted Averages

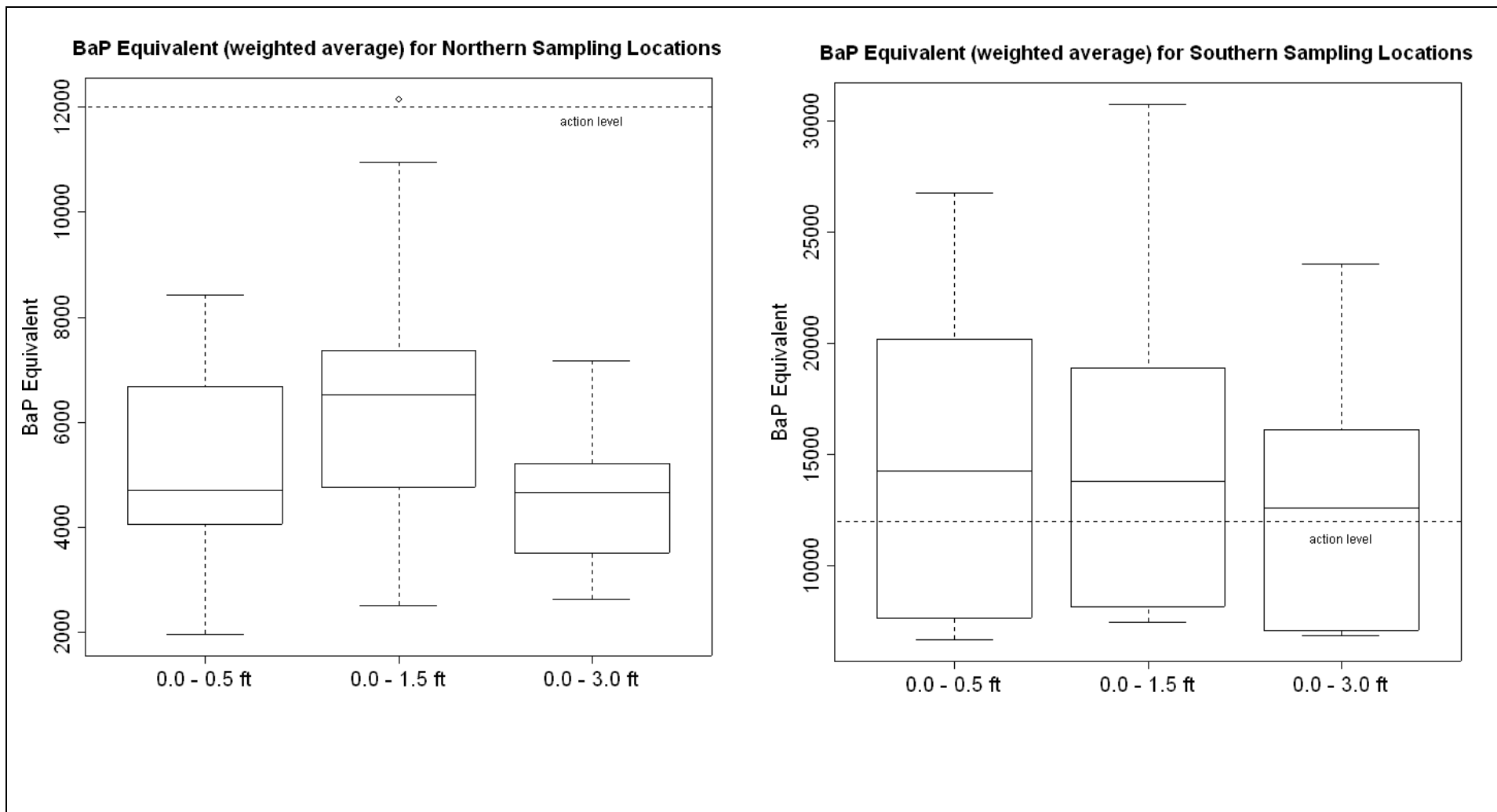


Figure 3-3. Box-and-Whisker Plots of BaP Equivalent Concentrations in the Northern (left) and Southern (right) Sample Locations

Table 3-4. Upper 95% Confidence Intervals Northern Locations					
Depth	n	Median	Mean	Std. Dev.	UCL
0.0 – 0.5 ft.	9	4708	5354	2151	6688
0.5 – 1.5 ft.	9	7377	7612	2541	9187
1.5 – 3.0 ft.	9	2752	2736	1784	3842

Table 3-5. Upper 95% Confidence Intervals Southern Locations					
Depth	n	Median	Mean	Std. Dev.	UCL
0.0 – 0.5 ft.	12	14239	14406	7106	18090
0.5 – 1.5 ft.	12	13591	15861	7632	19818
1.5 – 3.0 ft.	12	10844	10457	4129	12597

Table 3-6. Upper 95% Confidence Intervals All Locations					
Depth	n	Median	Mean	Std. Dev.	UCL
0.0 – 0.5 ft.	21	7848	10527	7120	14946
0.5 – 1.5 ft.	21	8492	11476	6780	14028
1.5 – 3.0 ft.	21	5716	7148	4975	9020

Table 3-7. Upper 95% Confidence Intervals Northern Locations Weighted Data					
Depth	n	Median	Mean	Std. Dev.	UCL
0.0 – 0.5 ft.	9	4708	5354	2151	6688
0.5 – 1.5 ft.	9	6514	6860	3097	8779
1.5 – 3.0 ft.	9	4667	4798	1518	5739

Table 3-8. Upper 95% Confidence Intervals Southern Locations Weighted Data					
Depth	n	Median	Mean	Std. Dev.	UCL
0.0 – 0.5 ft.	12	14239	14406	7106	18090
0.5 – 1.5 ft.	12	13807	14385	7031	18030
1.5 – 3.0 ft.	12	12598	12421	5455	15249

Table 3-9. Upper 95% Confidence Intervals All Locations Weighted Data*					
Depth	n	Median	Mean	Std. Dev.	UCL
0.0 – 0.5 ft.	21	7848	10527	7120	14946
0.5 – 1.5 ft.	21	8155	11160	6752	15155
1.5 – 3.0 ft.	21	6995	9154	5677	12231

* Assumes a Lognormal distribution

Table 3-4 shows the UCLs to be well under the Action Level for all three depth levels, but Table 3-5 show the UCLs to be well above the Action Level. For Table 3-6, Upper Confidence Intervals all Locations, note how the UCL for both of the shallow depth ranges exceed the Action Level (12,000 ppb), but that the deeper depth level appears to be well under the Action Level. Taking Tables 3-4, 3-5, 3-6 together leads to the general conclusion that even if the assumption of normality in data distribution was questionable, there seems to be a difference in the interpretation of level of contamination (and hence risk between the Northern and Southern Locations.

Tables 3-7 and 3-8 show the same patterns as do the raw data tables and show a consistency in interpretation. Table 3-9, however has to be treated a little differently. The reason lies in the assumption of normality made for the construction of the UCLs. When the Northern and Southern Locations are considered individually, the assumption of normality cannot be meaningfully challenged as the sample sizes are too small to show significant deviations from normality. When the data sets are merged to become All Locations, there become a sufficient number of samples to show any deviation. For the actual values (Table 3-3 “Depth,” and Table 3-6) the assumption of normality is possible although there are many significant deviations from normality. When the weighted values are considered (Table 3-3 “weighted average,” and Table 3-9) the assumption of normality is definitely not true and an alternative must be sought.

One alternative is to make the assumption that the distribution of the data is lognormal as opposed to normality. Using this assumption and Land’s Method (Gilbert, 1987) the UCL can be calculated and Table 3-9, Upper Confidence Intervals All Locations Weighted Data, shows the UCL to be well past the Action Level of 12,000 ppb.

4. CONCLUSIONS

The RI data indicate that concentrations of PAHs exceed the site-specific remedial action concentrations in the operations area, and therefore potentially pose an unacceptable risk due to exposure of industrial workers to residual contamination. The next phase after a Remedial Investigation is a Feasibility Study (FS) designed to evaluate remedial alternatives in the top three ft. of soil in the Operations Area. Based on the RI, it is evident that the FS should focus on the southern part of Operations Area, where elevated PAH concentrations are highest.

5. SUMMARY

Two important parts of the Consent Order (the Preliminary Investigation followed by a Remedial Investigation) were carried out in an efficient and timely manner because of the DQO Process. It enabled a flexible, graded approach to the problem, resulting in a minimal necessity to rework and repeat previous conclusions. The structure of the DQO Process culminated in two fully documented data collection studies that allowed for the development of remedial action of a potentially hazardous site.

6. REFERENCES

Gilbert, R.O., 1987. *Statistical Methods for Environmental Pollution Monitoring*, Wiley, N.Y.

US EPA 2000. *Guidance for the Data Quality Objectives Process (EPA QA/G-4)*.

APPENDIX A

THE PAH IMMUNOASSAY METHOD

Environmental immunoassay techniques use the ratio of a light-absorbing chemical-enzyme conjugate and the free chemical from the environmental sample to estimate the concentration of a chemical in the environmental sample. The light absorption of various samples is compared to the light absorption from a set of standards; the absorption decreases as the concentration from the environmental sample increases. This technique is available for a wide array of chemicals, as well as for classes of chemicals such as PCBs and PAHs. General PAH kits are sensitive to 2,3 and 4 ring compounds. These kits use phenanthrene as the target compound and are preferred for sites with fuel oil #2, diesel, and kerosene. Carcinogenic PAH kits are for 3,4,5, and six ring PAHs and are preferred for JP-4, and mixtures of fuel oil coal tar, and creosote.

PAH kits that are designed to detect a class of compounds cross-react more with compounds that are not the PAHs of interest but have similar structures. These kits give results that tend to be “biased high” compared to fixed lab results. In general, the detection limit for total PAHs in a soil sample is 600-700 ppb. Detection limit for target compound (phenanthrene or pyrene) in a soil sample is 300-400 ppb.

All immunoassay techniques run samples in batches. Reported times to run immunoassay: (samples per batch/batch process time in minutes) 5/30, 10/70 and 20/120, not including preptime. Anecdotal evidence shows that 35-200 samples per person per day can be run depending on extent of preparation (soil samples would probably be on the low end of this range).

At this site, a preliminary set of samples was taken to determine the expected range of PAHs at site. This preliminary test was needed to determine the required dilutions and the site – specific correlation between the immunoassay and fixed laboratory analysis. With multiple standard calibration curves for each sub area of site (because of potential differences in constituent PAHs or soil chemistry), the correlation coefficient for calibration should be 0.990-0.995. Continuing calibration samples should be run every 4-6 samples and blank samples should be run with each batch analyzed by immunoassay.

QC requirements are specific to the PAH kit, and obtained from the manufacturer. Field duplicates should be run at a rate of one per 10 samples, or one per batch analyzed (some kits run 20 samples per batch), whichever is the higher rate. Field duplicates should fall within 50% of original for soils. Performance Evaluation samples (i.e., split samples) are needed throughout the sampling to determine the comparability acceptance criteria. The comparability acceptance criteria for soils is 50% relative percent difference for soil samples. Recommended frequency for split samples is conventionally 10%, but may need to be adjusted for the project based on meeting comparability acceptance criteria.

APPENDIX B

VISUAL SAMPLE PLAN

Overview

Visual Sample Plan (VSP) is a software tool for selecting the right number and location of environmental samples such that the results of statistical analyses of the resulting data have the desired confidence for decision making. Sponsors of this public domain software include the EPA, Department of Energy, Department of Defense, and Department of Homeland Security; it was developed by Battelle Pacific Northwest National Laboratory. It provides simple, defensible tools for defining an optimal sampling scheme for any two-dimensional contamination problem including surface soil, building surfaces, water bodies, or similar applications.

Reports generated by VSP can be exported directly into a QA Project Plan or Sampling and Analysis Plan. VSP uses the seven Data Quality Objectives steps and is especially useful in resolving technical and statistical issues arising from steps 6 (Specify tolerable limits on decision errors), and 7 (Develop a plan for obtaining data). In particular, VSP can be used to generate different scenarios involving different decision error rates and statistical assumptions. VSP is easy to use and contains many graphics, including help and tutorial guides.

VSP utilizes state-of-the-art statistical and mathematical algorithms applicable to environmental statistics and presents the results in plain English. It provides the projected number of samples needed to meet DQO specifications, total sampling costs, and actual locations of the samples on an actual map of the site. VSP is designed for the non-statistician and is upgraded at various intervals to include more functions and methodologies. It is available at no cost from the website <http://dgo.pnl.gov/vsp>.

VSP and Comparing a Mean Against a Fixed Threshold

This is the basic option of VSP and follows the development of the DQO Process as described in Guidance for the *Data Quality Objectives Process (EPA QA/G-4)* (US EPA 2000). VSP takes the information developed in Step 6 (Specify Tolerable Limits on Decision Errors) and calculates the number of samples needed assuming a normal distribution. The formula used to calculate the number of samples is:

$$n = \frac{s^2 (Z_{1-\alpha} - Z_{1-\beta})^2}{\Delta^2} + \frac{Z_{1-\alpha}^2}{2}$$

where

- n is the number of samples,
- s is the estimated standard deviation of the measured values including analytical error,
- Δ is the width of the gray region,
- α is the acceptable probability of incorrectly concluding the site mean is less than the threshold,
- β is the acceptable probability of incorrectly concluding the site mean exceeds the threshold,

$Z_{1-\alpha}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha}$ is $1-\alpha$,
 $Z_{1-\beta}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\beta}$ is $1-\beta$.

VSP then increases the number of samples to be taken by a further 16% to account for the probable skewness of the actual distribution and to allow for nonparametric tests such as the Wilcoxon Signed Rank Test. The resulting Performance Goal Diagram (figure B-1) shows the performance curve based on the input information. The vertical line is shown at the threshold (action limit) on the horizontal axis. The gray region is the shaded area; the upper horizontal dashed line is positioned at $1-\alpha$ on the vertical axis; the lower horizontal dashed line is positioned at β on the vertical axis. A vertical line is positioned at one standard deviation below the threshold.

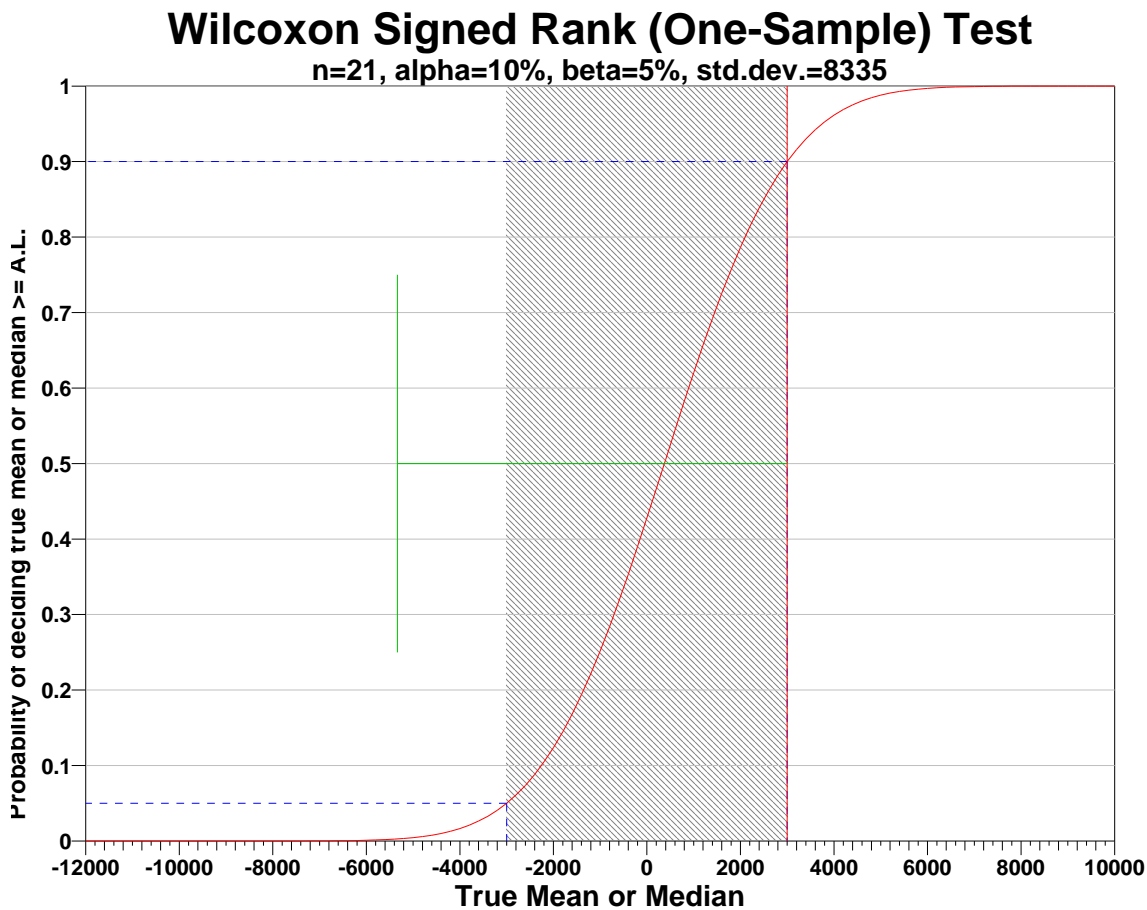


Figure B-1: Performance Goal Diagram using the Wilcoxon Signed Rank Test

VSP also generates different scenarios for the specified variables and so can enable the project team to conduct a sensitivity analysis of the problem and compare costs of different scenarios.

VSP and Hot Spot Detection

For the location of a hot spot, four variables need to be considered: size of the potential hot spot, the desired chance of hitting a hot spot (usually considered in terms of the risk of missing a hot spot), the cost of sampling, and finally, the size of the grid spacing proposed. VSP can be used to create statistically sound sampling designs that reflect combinations of each of these variables by specifying three of the variables:

- Using a predetermined grid spacing, VSP can calculate the chance of finding a hot spot of a specified size.
- For a specified probability with pre-determined grid spacing, VSP can calculate the smallest size hot spot that can be detected.
- For a specified probability and specified hot spot size, VSP can calculate the minimum number samples to be taken in order to hit the hot spot.
- For a predetermined cost (which dictates the grid size), VSP calculate the probability of finding a hot spot of a given size.

VSP can be used to investigate the effects of changing the grid type (square, rectangular, or triangular), grid spacing (distance between sampling nodes), and shape of hot spot (circular or elliptical of some type).

VSP does, of course demand some reasonable assumptions: the hot spot is elliptical or circular and not a strange shape with contours, the definition of what constitutes a hot spot is known, and that the distance between grid points is much greater than the dimensions of the area from which the physical sample will be taken.

It is useful to note that for hot spot detection, there is only one decision error to consider: the false acceptance (false negative) error (saying a hot spot doesn't exist when in reality it does). The other error (false rejection or false positive) does not apply as when we obtain a "hot" sample we will automatically assume it really is a hot spot (i.e. we are saying with certainty that it really is a hot spot).

It is common practice to make VSP generate a set of different scenarios for different specified variables and so enable the project team to conduct a sensitivity analysis of the problem, and finally select one arrangement that best satisfies the project needs.

For example, an abbreviated report from sampling the Operations Area would have the following appearance:

Systematic sampling locations for detecting an area of elevated values (hot spot)

This report summarizes the sampling design used, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan

components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design developed.

Primary Objective of Design	Detect the presence of a hot spot that has a specified size and shape
Type of Sampling Design	Hot spot (elliptical ratio 0.8)
Sample Placement (Location) in the Field	Systematic (Hot Spot) with a random start location
Formula for calculating number of sampling locations	Singer and Wickman algorithm Probability of detection $(1-\beta) = 0.9566$
Calculated total number of samples	15
Number of samples on map ^a	14
Number of selected sample areas ^b	1
Specified sampling area ^c	112500 ft ²
Grid pattern	Triangular
Size of grid / Area of grid cell ^d	75 feet / 2435.625 ft ²
Total cost of sampling ^e	\$9000.00

^a This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.

^b The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.

^c The sampling area is the total surface area of the selected colored sample areas on the map of the site.

^d Size of grid / Area of grid cell gives the linear and square dimensions of the grid used to systematically place samples.

^e Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.

The following graph shows the relationship between number of samples and the probability of finding the hot spot. The dashed blue line shows the actual number of samples for this design (which may differ from the optimum number of samples because of edge effects).

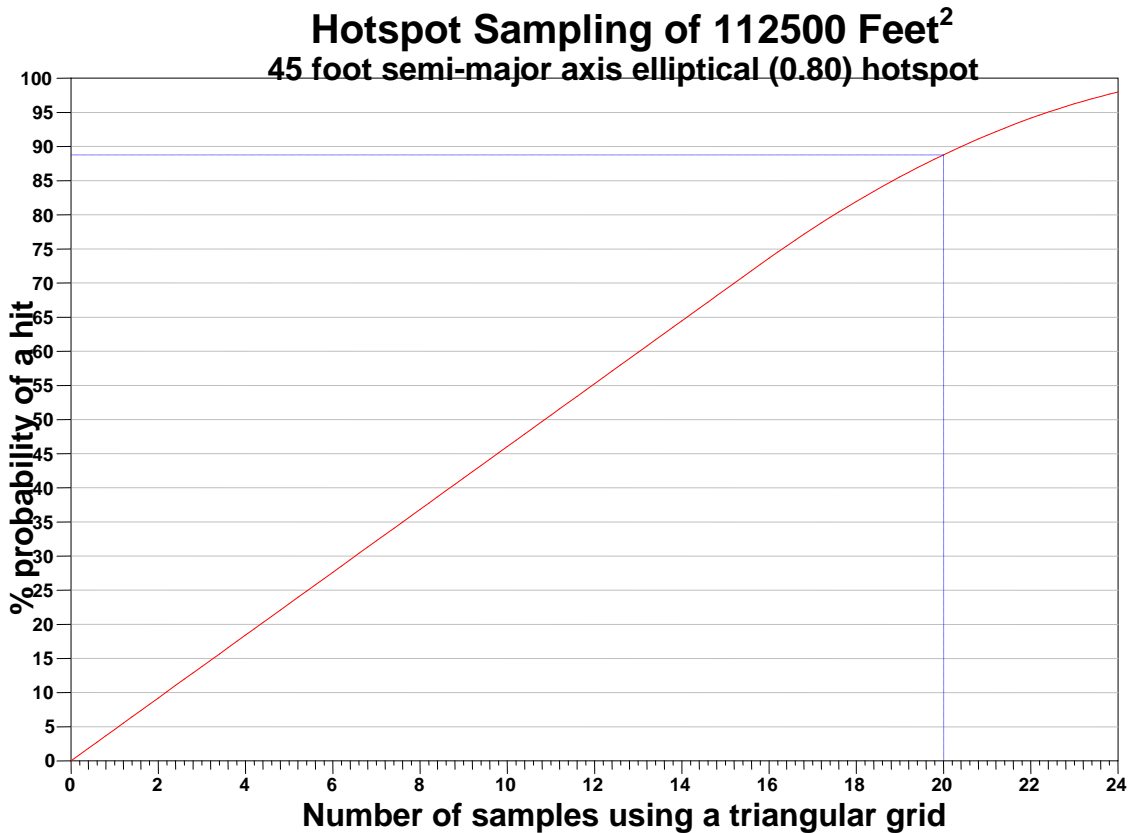


Figure B-2: Hotspot Sampling of 112500 Feet

Statistical Assumptions

The assumptions associated with the sample spacing algorithm are that:

1. the target hot spot (its projection onto the coordinate plane) is circular or elliptical,
2. samples are taken on a square, rectangular, or triangular grid,
3. a very small proportion of the area being studied will be sampled (the sample is much smaller than the hot spot of interest),
4. the level of contamination that classifies a hot spot is well defined, and
5. there are no misclassification errors (a hot spot is not mistakenly overlooked or an area is not mistakenly identified as a hot spot).

These assumptions cannot be validated through data collection. The size and shape of a hot spot of interest are well defined prior to determining the number of samples and the measured value that defines a hot spot is well above the detection limit for the analytical methods that will be used. Grid sampling will be carried out to the level achievable; topographic, vegetative, and other features that prevent sampling at the specified coordinates will be noted and their influence recognized.

Sensitivity Analysis

The sensitivity of the calculation of number of samples was explored by varying Area and Side and examining the resulting changes in the number of samples. The following table shows the results of this analysis.

	Side=37.5	Side=75	Side=112.5
Area= 56250	31	8	5
Area=112500	60	15	8
Area=168750	91	23	12

Area = Total Sampling Area

Side = Length of Grid Side

Cost of Sampling

The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is \$9000.00. The following table summarizes the inputs and resulting cost estimates.

Cost Details	Per Analysis	Per Sample	15 Samples
Field collection costs			
Analytical costs	\$150.00	\$600.00	\$9000.00
Sum of Field & Analytical costs		\$600.00	\$ 9000.00
Fixed planning and validation costs			
Total cost			\$ 9000.00

APPENDIX C

SUMMARY OF THE DQOS DEVELOPED FOR THE PRELIMINARY INVESTIGATION

Step 1: State the Problem	<ul style="list-style-type: none"> • The Preliminary Investigation (PI) will focus on establishing the types of contaminants present at the site and the approximate spatial distribution of contaminant concentrations. The Remedial Investigation will use the information generated in the PI to estimate the quantity of additional data needed to make remedial decisions for the site with a specified level of confidence.
Step 2: Identify the Decision	<p>Preliminary Investigation Questions</p> <p><i>Primary Questions</i></p> <ul style="list-style-type: none"> ➤ What is the spatial distribution of contaminants in soil and are they greater than the screening criteria? ➤ Has soil contamination reached the shallow aquifer? ➤ Is soil and groundwater contamination a significant source of ongoing contamination to the Bay? <p><i>Secondary Questions</i></p> <ul style="list-style-type: none"> ➤ In what areas of the site do COPC concentrations in soil exceed background concentrations or industrial risk screening levels? ➤ Are soil contaminants detected in groundwater at locations of relatively high soil contamination? ➤ What is the potential for erosion of contaminated site soils?
Step 3: Identify Inputs to the Decision	<ul style="list-style-type: none"> ▪ Field immunoassay method will be used for establishing the boundaries of soil contamination. ▪ Based on information from similar former MGP sites, PAHs, VOCs, metals and cyanide are contaminants of potential concern (COPCs). ▪ Background data will be compared to the fixed lab confirmation data. ▪ A minimum number of fixed laboratory confirmation samples would be analyzed for each of the post-stratified subareas.
Step 4: Define the Boundaries of the Study	<ul style="list-style-type: none"> ▪ The top soil sampling depth interval will be 0 – 0.5 ft. (the traditional definition of “surface soil.”), and the second interval will be 0.5 – 3 ft.. ▪ Lower depth intervals (3- 6 ft.; 6 -12 ft., 12 ft. to groundwater) were defined with the addition of a decision rule to forego the two deepest intervals if total PAH concentrations in higher intervals were essentially consistent with background concentrations. ▪ Three initial subareas of interest were defined: 1) Operations Area, 2) Oil Storage Area, and 3) Surrounding Environs (the remaining portion of the site). ▪ Based on patterns of total PAH concentrations in each of the three initial subareas and each of the four soil depth intervals, the boundaries of the subareas will be refined for the RI.

<p>Step 5: Develop a Decision Rule</p>	<ul style="list-style-type: none"> • <i>Overall Decision Rule:</i> If concentrations of MGP-related contaminants in surface soil, subsurface soil, or groundwater in a subarea exceed PI risk-based screening criteria, or if the pattern of COPC concentrations in surface soil indicate transport of COPCs to the Bay, determine the additional data needs to complete a Remedial Investigation for this subarea. • If concentrations of PAHs in a subarea or depth interval are found to be elevated relative to background PAH concentrations, then MGP-related contaminants will be presumed to be present and the area will be further evaluated to determine the need for remedial action. • If a constituent is present in subarea soils at concentrations significantly higher than background and if the maximum detected concentration is greater than its screening criterion, then the constituent will be considered a COPC, and will be evaluated further in the Remedial Investigation. • If the maximum detected concentration of a chemical in groundwater beneath a subarea exceeds the groundwater screening criterion, then the chemical will be identified as a COPC. • If surface soil data show decreasing concentrations of COPCs along an axis perpendicular to the shoreline, and if surface soil concentrations at sampling locations nearest the Bay are below screening criteria, then conclude that transport of contaminants in site soils to the Bay is not a significant process.
<p>Step 6: Specify Tolerable Limits on Decision Errors</p>	<ul style="list-style-type: none"> • In both the Oil Storage and Operations Areas, a 75 ft. triangular grid will be used as it provides a 95% chance of hitting an elliptical hot spot with the length of the semi-major axis of 45 ft. In the Site Environs, a 200 ft. grid will be used as it provides a 95% probability of hitting an elliptical hot spot with the length of the semi-major axis of 119 ft. A systematic triangular grid will be utilized using a “find a hot spot” sampling goal.
<p>Step 7: Develop the Plan for Obtaining Data</p>	<ul style="list-style-type: none"> • Oil Storage Area (A) has 15 sample locations, Operations Area (B) has 41 locations and Site Environs (C) 24 locations, for a total of 80 locations. If samples are analyzed for all four depths at each location, 320 analyses will be conducted. Since the sampling design is adaptive, a finer grid may be applied over a larger area and it is assumed an additional 10% (32 analyses) may be required. The total projected costs for the PAH immunoassay survey sampling and analysis is \$52.8K. Additional costs will be incurred for fixed laboratory analysis, estimated at \$1K/sample. Based on experience with other PAH-contaminated sites, it is assumed that 30 fixed lab samples will be collected. In addition, a standard number of f QC samples will be collected. • Groundwater sampling design will be biased toward the locations in each subarea with the highest observed concentration of PAHs in the subsurface. Assuming the three initial areas are retained, and two upgradient samples will be collected, the team budgeted for five groundwater samples. These samples will be sent to the laboratory to perform a PAH, SVOC and VOC analyses at a projected cost of \$700/sample, plus the cost of obtaining the hydropunch samples, estimated at \$300/sample. <p>The total projected costs for the PI is therefore \$87K, to include soil and groundwater sampling representative of the site. The remaining \$13K is to be held as a contingency fund to complete the field effort.</p>

APPENDIX D

FIELD IMPLEMENTATION DETAILS

Oil Storage Area – A total of 15 cores in the Oil Storage Area were collected resulting in 52 immunoassays for total PAHs; seven of which were from the 3-6 ft. interval, but none of which was in the > 6 ft. interval. No highly contaminated soil in this area was found, however, tPAH concentrations on the southern side of the rail spur in the top three depth intervals were higher than north of the rail spur. In general the surface interval was cleaner than subsurface samples. Due to slightly elevated concentrations at depth in station A-13, the team considered expanding sampling at the 75 ft. spacing to the south. However, due to the fact that on both sides of A-13 concentrations were much lower, and the moderate concentrations observed, the team decided not to expand this grid. If sampling in the site environs (Area C) indicates elevated PAHs at C-20 or C-21, additional samples in this area will be considered.

Coring logs and notes indicate that the clay layer noted in sampling of the Operations Area was present in all samples. Slight discoloration, possibly due to oil contamination was observed in the two samples with the highest PAH concentrations from just below the ground surface to the clay layer, but not beneath the clay layer.

A groundwater sample was collected by Hydropunch™ at a location just to the west of A-13: the location of the highest subsurface concentration in the 3-6 ft. interval (Figure 2-6) and no free product was observed in the sample.

Operations Area – Cores were collected at the planned 41 sampling locations. The original sampling grid in this area (Figure 2-7) was extended by two rows on the southern boundary of the Operations Area (12 samples) due to the detection of high concentrations of total PAHs in the southern region of the area. Following the adaptive analysis rules, a total of 201 immunoassays were performed for total PAHs. Elevated PAHs continued to be found in the first new row to the south, but dropped to approximately ambient levels in the last new row. The highest concentrations are correlated with visual observations of a powdery black substance, probably lampblack, in the core.

Coring logs indicated a one to three ft. thick clay layer was present in all samples at a depth of approximately 10 ft. below ground surface. Observations of the cores where high total PAH concentrations were measured indicated that visible staining and petroleum odors were occasionally noted immediately above the clay layer or at shallower depths but never below the clay layer. This finding appears to confirm the team hydrogeologist's hypothesis that the clay layer may have served as a barrier to vertical migration of hydrocarbons.

A groundwater sample was collected west of B-45, the location of the highest subsurface PAH concentration (Figure 2-13). No free product was observed in any sample.

Site Environs – Cores from the initial 24 sampling locations were collected and field-screened for total PAHs. During the first day, the field team met core refusal numerous times when attempting to obtain a core from C-8. Gravel from the parking area was scraped away from

a three ft. area, and eventually a reasonably intact core was collected. A visual analysis of failed cores revealed debris mixed in with native soil, ash, slag, and lampblack. In the deeper intervals, the core contained thick, black, tarry wastes. The tarry waste was analyzed by immunoassay and found to contain levels of PAHs higher than the calibration scale of the kit. After discussing these findings with the Power Company representative, and looking more closely at the surrounding areas, it appeared that this sampling location was in alignment with a ravine-like feature extending from the parking area to the cove. The team decided to take samples up- and down- gradient from C-8, to further evaluate what appeared to be materials dumped in a former ravine. In all, five additional cores were obtained and analyzed: two in the direction of the cove, and three upgradient. Tarry material was found in the samples surrounding C-8, especially below three ft. of cover. An additional core near the cove revealed consistent soil-fill with no evidence of tarry material, and was not analyzed.

APPENDIX E

SAMPLE SIZE SELECTION FOR BACKGROUND COMPARISONS

This appendix describes the determination of the necessary number of fixed lab samples in the post-stratified areas to make comparisons to background levels. Under the null hypothesis of equal site and reference means, it is assumed that the populations are identical and so the reference mean and standard deviation will be used to determine sample size requirements. Table D-1 shows the necessary sample sizes for a two-sample, one-sided *t*-test for various significance levels, power levels, and difference-to-be-detected as a percentage of the mean. The sample sizes listed are for each population.

Table E-1. Sample Size Required in Each Site and Background Location											
		Significance level = 0.05					Significance level = 0.10				
		Difference-to-be-detected					Difference-to-be-detected				
		50%	75%	100%	125%	150%	50%	75%	100%	125%	150%
Power	0.80	37	17	10	7	5	27	12	7	5	4
	0.85	43	20	12	8	6	32	15	9	6	4
	0.90	50	23	14	9	7	39	18	10	7	5
	0.95	63	29	17	11	8	50	23	13	9	6

It was decided to use a significance level of 0.05, a power level of 0.90, and a difference-to-be-detected of 150%. These values reflect the planning team's assumption that if a contaminant is related to the MGP, it is likely to be present at concentrations well in excess of the Background site – that is at least 200% higher. By ensuring that a shift of 150% can be detected with good power, it was thought that larger shifts would be detected with even better power. The choice of a 10% chance of missing a shift of 150% (the power level of 0.90) was based on experience with similar situations. In addition, since multiple statistical tests are to be used, and not simply a two sample *t*-test, the probability of seeing other shifts, such as a shift in the upper tail, would be greatly enhanced.

From Table D-1, with power of 0.90 and a difference to be detected at 150%, it can be seen that at least seven samples would be required from both the sampled area and background location. This minimal requirement was increased from seven (Oil Storage Area), to 11 (Site Environs), to 12 (Operations Area) after discussion of the potential of each area to contain contamination and the probable use of a non-parametric test (Wilcoxon Rank Sum) in the final analysis (the use of a non-parametric test increases the sample size by about 15%). The background requirement of seven samples was easily met by the existence of 15 samples being available.

APPENDIX F

COMPARISON OF IMMUNOASSAY AND FIXED LABORATORY RESULTS

Total PAH and BaP equivalents from Fixed Laboratory analyses were compared to the immunoassay results to assess the reliability of the less expansive immunoassay readings. It was thought that high immunoassay results would have a disproportionate influence on correlation and so several correlation studies were made by eliminating some of the higher values to ascertain their effect. Table E-1 shows the correlation of Fixed Laboratory total PAH with Immunoassay total PAH; Table E-2 shows the correlation of Fixed Laboratory BaP equivalents with Immunoassay total PAH.

Data Used	Correlation
All data	0.9085
2 values > 200,000 ppb omitted	0.8135
6 values > 50,000 ppb omitted	0.8581
11 values > 10,000 ppb omitted	0.8486

Data Used	Correlation
All data	0.6127
2 values > 200,000 ppb omitted	0.6216
6 values > 50,000 ppb omitted	0.6523
11 values > 10,000 ppb omitted	0.7827

Table E-1 shows a relatively high degree of correlation between the immunoassay and total PAH values (with $r > 0.8$) in all cases, whereas the correlation to BaP equivalents (Table E-2) is not as good, and improves only once the higher screening values are removed. This indicates that there is more variability in the concentration of those PAHs contributing to the BaP equivalents and indicates that immunoassay is a better predictor of total PAH than it is of the carcinogenic subset of PAHs. Based on this observation, it was decided that Fixed Laboratory results will be necessary to complete the baseline risk evaluation in the Remedial Investigation.

APPENDIX G

SUMMARY OF THE PRELIMINARY INVESTIGATION STATISTICAL ANALYSIS OF THE COPCS

Three analyses were conducted: BaP equivalents, Arsenic, and Individual PAHs.

The BaP Equivalents Analysis

Graphical comparisons and formal statistical tests were used to compare the three investigation sites (Oil Storage Area, Operations Area, and Site Environs) and Background for the principal COPC: BaP equivalents. These tests are described more fully in *Data Quality Assessment: Statistical Tools for Practitioners* (EPA QA/G-9S).

Graphical Comparisons

In addition to the straightforward interpretation of Tables 2A, 2B, and 2C, further investigation of the distribution of contamination may be made through examination of Box-and-Whisker plots. These are a simple way of condensing the data to a visual image and are composed of a central box divided by a horizontal line representing the median, and two lines extending out from the box called whiskers. The length of the central box indicates the spread of the bulk of the data (the central 50%) while the length of the whiskers show how stretched the tails of the distribution are. The sample mean is displayed using a “+” sign and any unusually small or large data points are displayed by a “*” on the plot. If the distribution is symmetrical, the box is divided in two equal halves by the median, the whiskers will be the same length and the number of extreme data points will be distributed equally on either end of the plot for symmetric data. Values that are unusually large or small can be easily identified, and a side-by-side comparison of box-plots for different sets of data can be made to ascertain similarities in distribution.

Example of a Box-and-Whiskers Plot

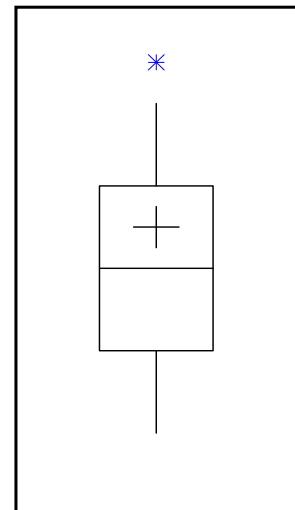


Figure F-1 shows the BaP Equivalents for the three areas and Background on a standard scale followed by the same comparison using the logarithmic scale. Note how only data from the Operations Area exceed the screening level with the median level well above the screening level of 2300 ppb; the Oil Storage Area, Site Environs, and Background being all well below 2300 ppb. For the Operations Area, the largest value (30167.0 ppb) has been noted (asterisk) as a potential outlier as it is well larger than the majority of the data. The highly “squashed” appearance of the Oil Storage Area, Site Environs, and Background data being due to the scale being used to illustrate the data. Conversion of the data to a logarithmic scale allows the data to be seen a little better with less influence from the variability of the data. Again, it is clear that most of the contamination from the Operations Area exceeds the screening level. The long “tail”

of the Background is indicative of there being mostly low levels of contamination present outside the MGP operations area.

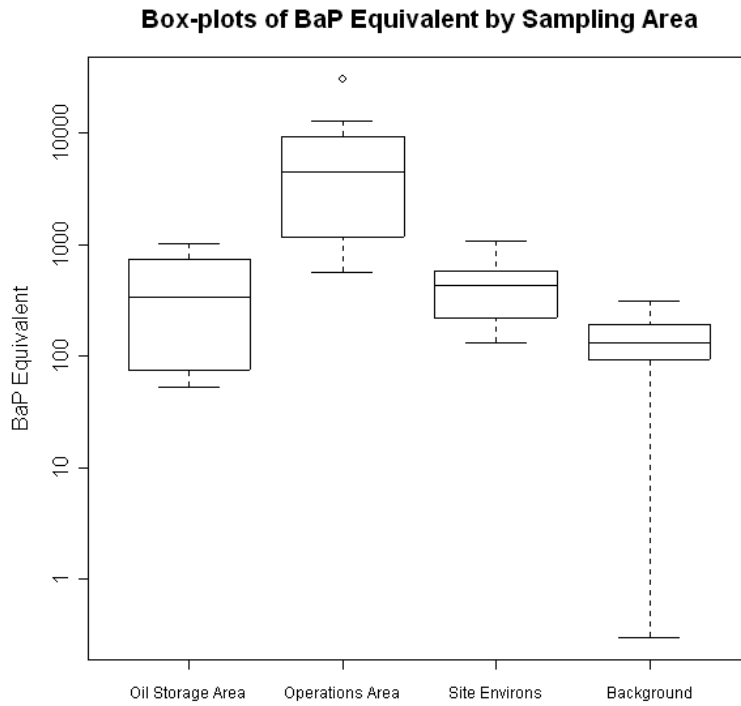
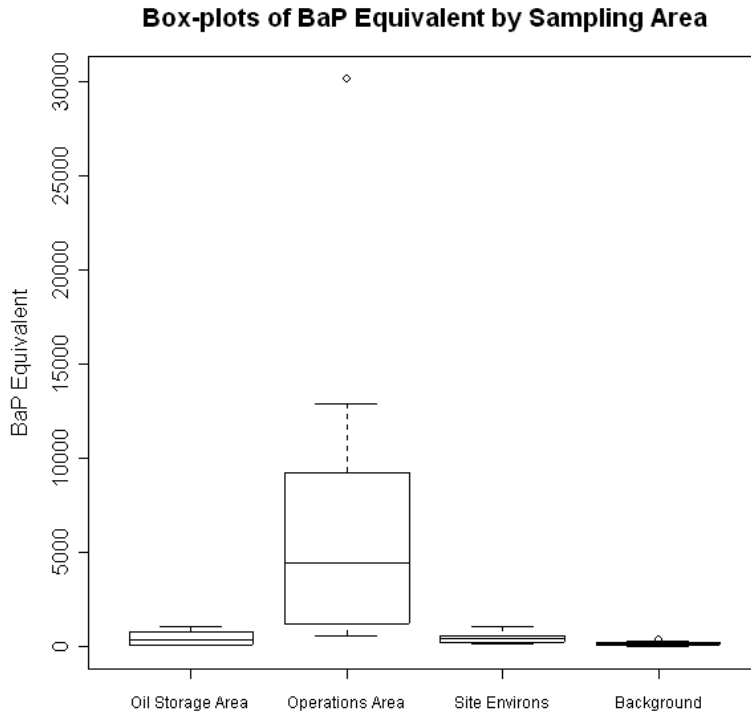


Figure G-1. BaP Equivalents Box Plots: Standard and Log Scales

Statistical Tests on the BaP Equivalents Data

Four statistical tests were made on the data from the three areas with respect to the Background data: Two-Sample t-test, Wilcoxon Rank Sum test, Quantile test, and the Slippage test. Each test considers a different aspect of the data.

The Two-Sample t-test

This test compares the mean of each area to the mean of the Background area. Due to the large differences in standard deviations between the three areas and Background, Satterthwaite's Two-Sample t-test is to be preferred over the more commonly encountered standard t-test (see section 3.3.1.1.2 of *Data Quality Assessment: Statistical Tools for Practitioners EPA QA/G-9S*). The assumption of approximate Normality in distribution of each data set used in the comparisons is unlikely to be true (for the Operations area, note the large discrepancy between mean [7465.3ppb] and median [4441.0ppb]; the mean and median should be quite close if the data were normally distributed) and so the test may not be particularly informative.

Area	Mean	p-value
Oil Storage	330.2	0.0745
Operations	7465.3	0.0246
Site Environs	420.0	0.0033

Background Mean = 130.3

Operating on the base-line assumption (null hypothesis) that there is no difference between the individual area means and Background, and using the p-value as a guide, it is clear that there is a significant difference between the Operations Area and Background, also between Site Environs and Background, but not between the Oil Storage Area and Background. It would be expected that the difference between the Operations Area and Background should be far more pronounced (compare the two p-values, Operations Area is not as strong as Site Environs) but the effect is being greatly obscured by the lack of Normality, especially with the large data values. The difference between Site Environs and Background suffers less because the lack of Normality is in the low values. The test shows there is a difference but is not really too useful.

The Wilcoxon Rank Sum Test

This test may be regarded as the non-parametric analogue of the Two Sample t-test in that the assumption of Normality is not required. The distributions of the two areas in the comparison should be approximately the same and the test essentially looks to see if one distribution differs from the other by a fixed (yet unknown) amount (see section 3.3.2.1.1 of *Data Quality Assessment: Statistical Tools for Practitioners EPA QA/G-9S*).

Area	p-value
Oil Storage	0.1336
Operations	<0.0000
Site Environs	0.0005

Operating on the base-line assumption (null hypothesis) that there is no difference between the individual areas and Background, and using the p-value as a guide, it is clear that there is a significant difference between the Operations Area and Background, also between Site Environs and Background, but not between the Oil Storage Area and Background. The extremely small p-value for Operations Area indicates clearly that the difference clearly could not be due to chance and this reinforces the conclusion drawn from inspecting the Box-and-Whiskers plot. A similar conclusion can be reached for the difference between Site Environs and Background (although not quite as strong). The failure to find a difference between the Oil Storage Area and Background is a little surprising but all the tests may be affected by the lack of similarity in distributional shape (the differences can be seen clearly in Figure F-1).

The Quantile Test

The Quantile test is useful in detecting instances where only parts of the data are different as opposed to a complete shift in the data. It is often run together with the Wilcoxon test (see section 3.3.2.1.2 of *Data Quality Assessment: Statistical Tools for Practitioners* (EPA QA/G-9S)).

Area	p-value
Oil Storage	0.1146
Operations	0.0006
Site Environs	0.0020

Using the base-line assumption (null hypothesis) that there is no difference between the individual areas and Background, and using the p-value as a guide, it is clear that there is a significant difference between the Operations Area and Background, also between Site Environs and Background, but not between the Oil Storage Area and Background. The extremely small p-value for Operations Area indicates clearly that the difference clearly could not be due to chance and this reinforces the conclusion drawn from the Wilcoxon test. A similar conclusion can be reached for the difference between Site Environs and Background (although not quite as strong). As in the Wilcoxon test, there seems to be no difference between the oil Storage Area and Background.

The Slippage Test

This test concentrates on the larger values of each data set and essentially compares the largest values of an area against the maximum value of the Background data set (see section 3.3.2.1.3 of *Data Quality Assessment: Statistical Tools for Practitioners* EPA QA/G-9S).

Area	p-value
Oil Storage	0.0150
Operations	< 0.0000
Site Environs	0.0020

With the null hypothesis that the larger values of an area are not significantly larger than the maximum of the Background data set, inspection of the p-values indicates that there is indeed a significant difference between the MGP areas and Background.

Arsenic

Graphical comparisons only are presented (Figure F-2) because all statistical tests on the data (Two-Sample t-test, Wilcoxon Signed Rank, Quantile, and Slippage) failed to be significant at even the 0.3 level. Inspection of the box-and-whisker plots shows only the Background area to be even close to the screening level with the distribution of Arsenic approximately the same across all four areas.

The Individual PAHs

Graphical comparisons only (Figures F-3 to F-12) were used to compare the three investigation sites (Oil Storage Area, Operations Area, and Site Environs) and Background for the individual PAHs as it had been decided to focus resources on the BaP Equivalentents.

Inspection of the individual standard scale box-and-whisker plots indicates the large predominance of PAHs in the Operations Area; large values in the Operating Area often being orders of magnitude larger than the other areas. Inspection of the individual logarithmic scale box-and-whisker plots show a marked similarity in the distribution of PAHs in the Operating Area with the exception of fluoranthene and indeno(1,2,3cd)pyrene (which are very similar in distribution themselves). The other areas (Oil Storage, Site Environs, and Background) show marked similarity across all PAHs.

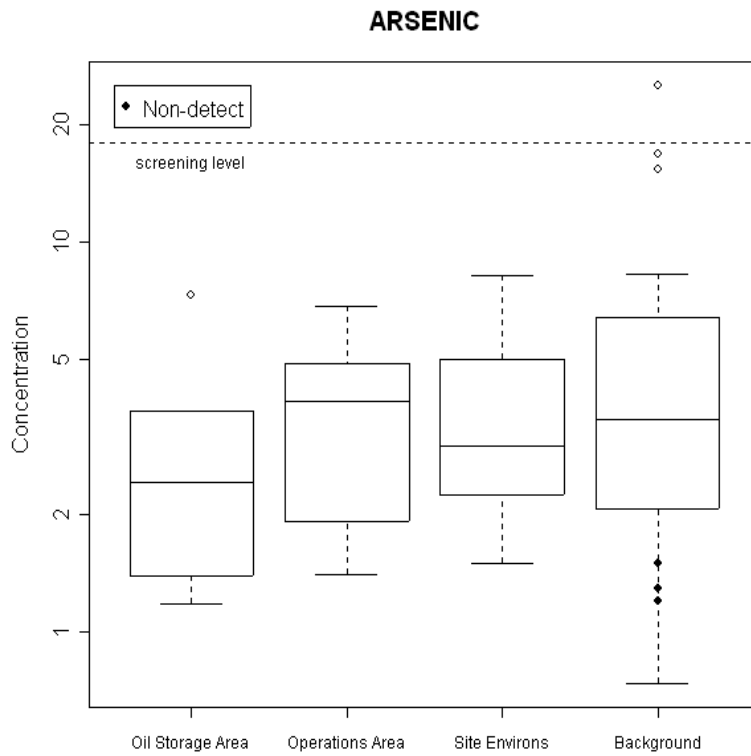
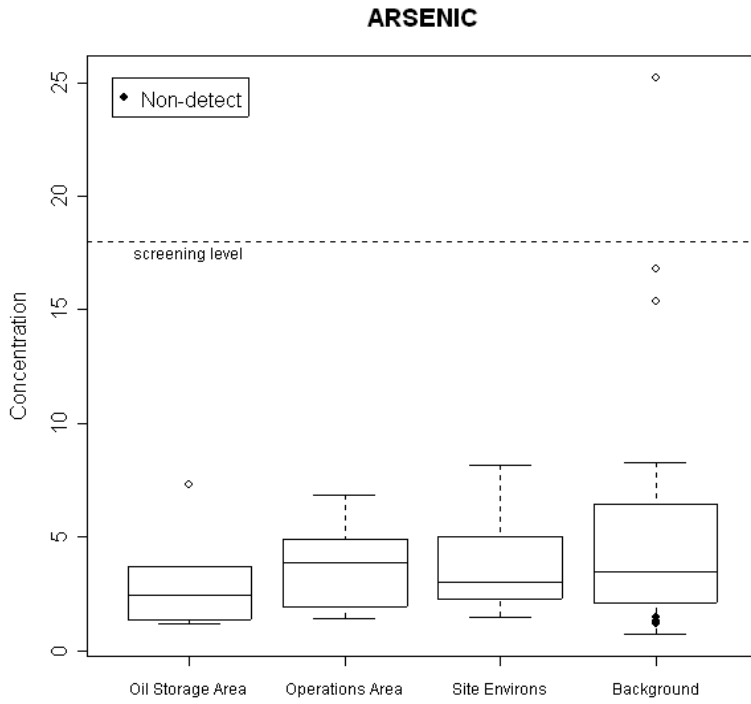


Figure G-2. Arsenic Standard and Log Scales

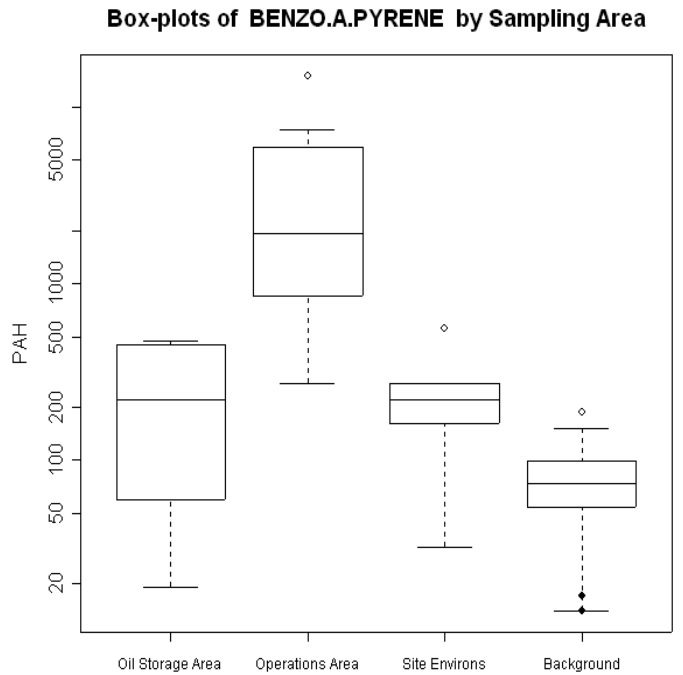
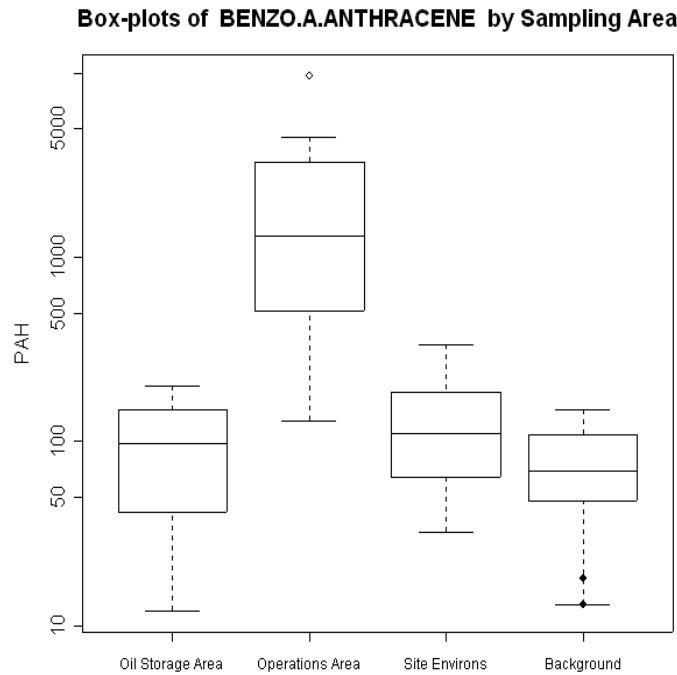
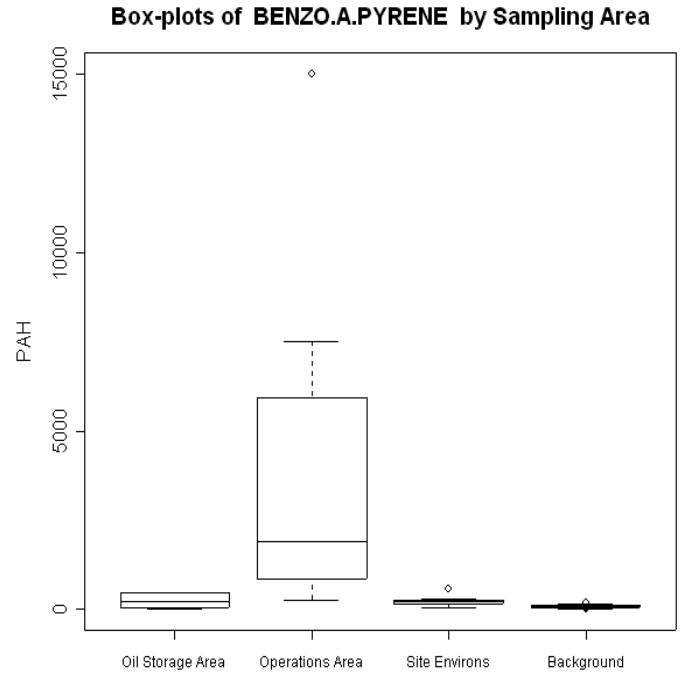
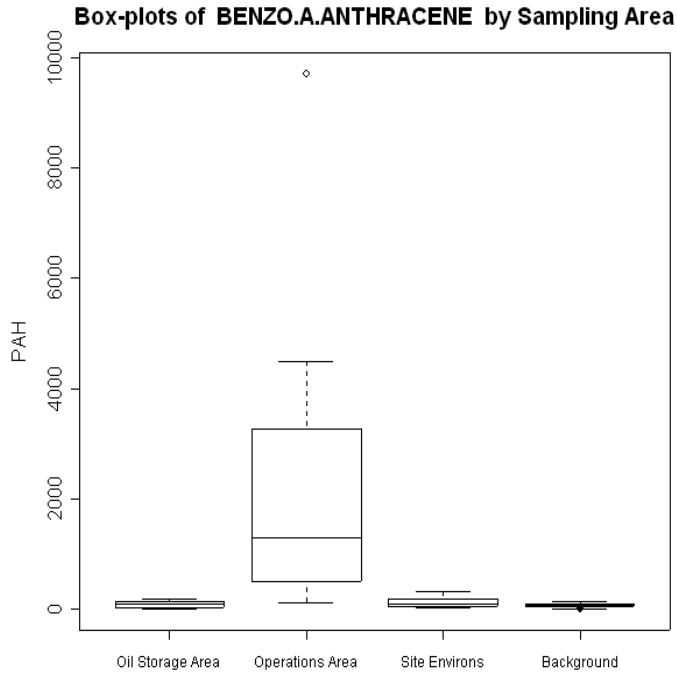
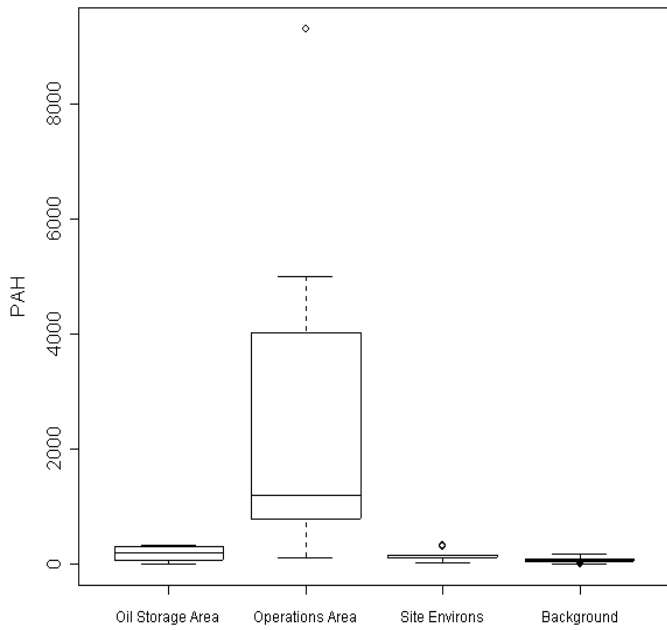


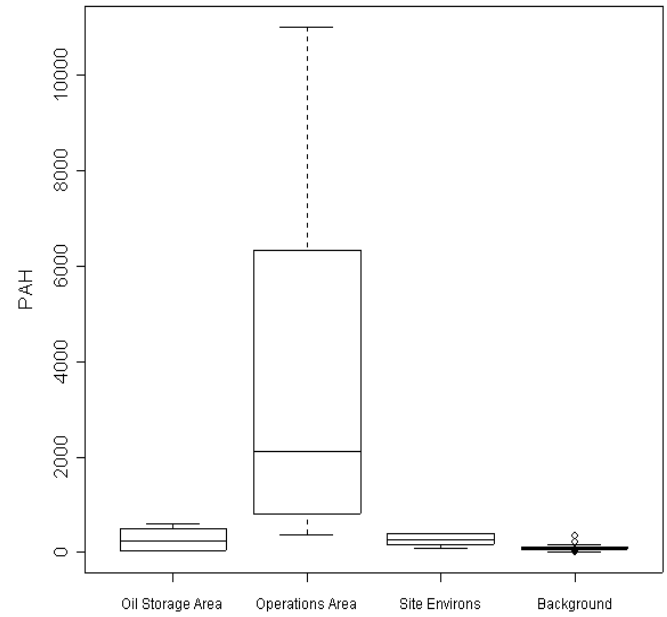
Figure G-3. Benzo(a)anthracene Box Plots: Standard and Log Scales

Figure G-4. Benzo(a)pyrene Plots Standard and Log Scales

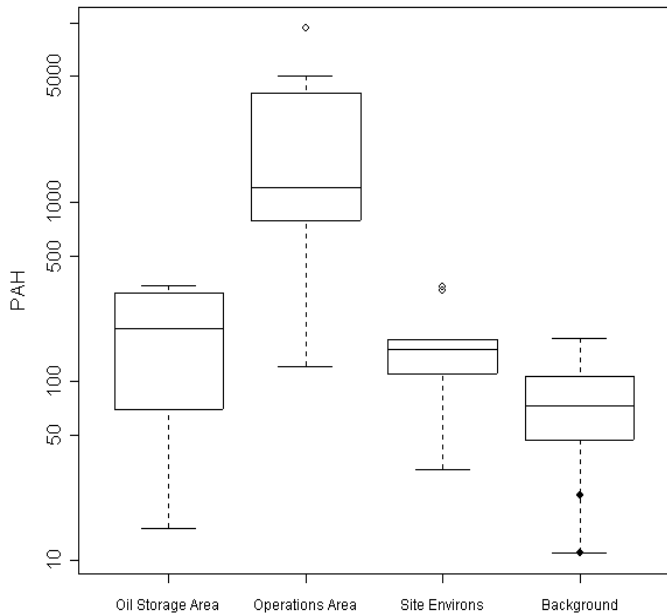
Box-plots of BENZO.B.FLUORANTHENE by Sampling Area



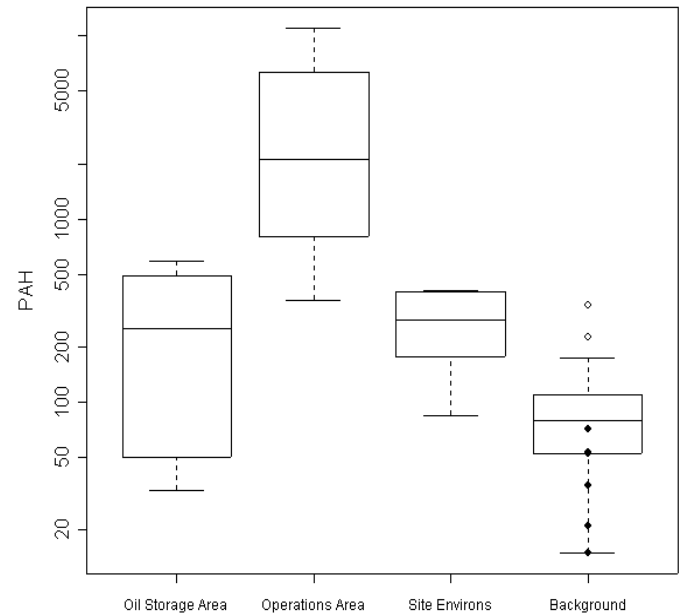
Box-plots of BENZO.G.H.I.PERYLENE by Sampling Area



Box-plots of BENZO.B.FLUORANTHENE by Sampling Area



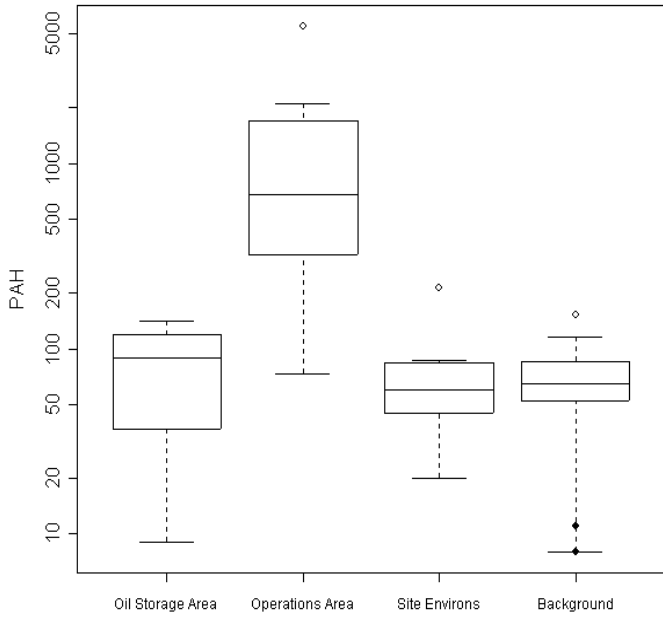
Box-plots of BENZO.G.H.I.PERYLENE by Sampling Area



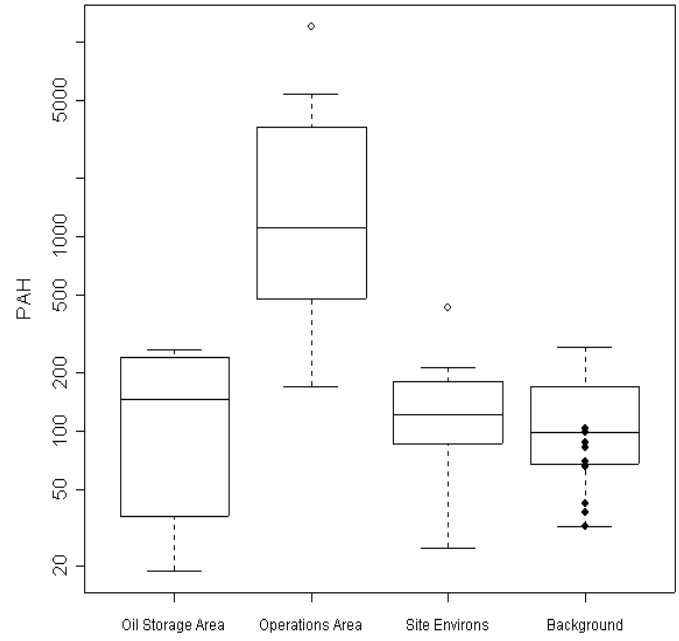
**Figure G-5. Benzo(b)flouranthene Plots:
Standard and Log Scales**

**Figure G-6. Benzo(g,h,i)perylene Plots
Standard and Log Scales**

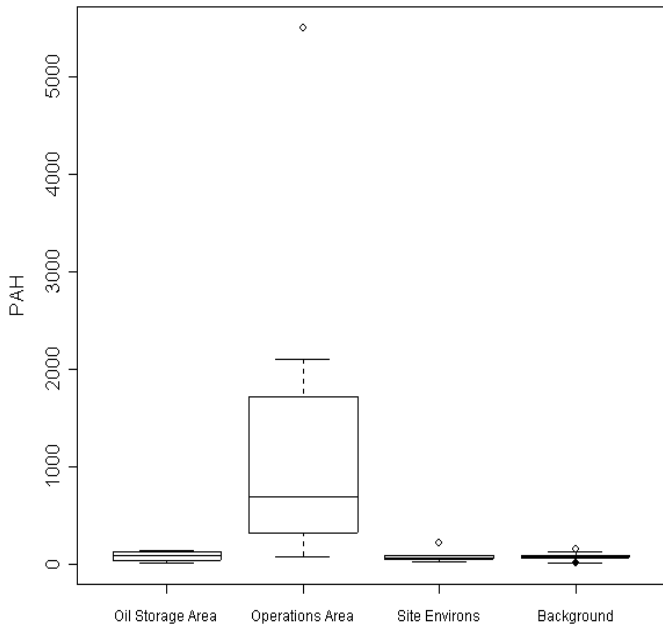
Box-plots of BENZO.K.FLUORANTHENE by Sampling Area



Box-plots of CHRYSENE by Sampling Area



Box-plots of BENZO.K.FLUORANTHENE by Sampling Area



Box-plots of CHRYSENE by Sampling Area

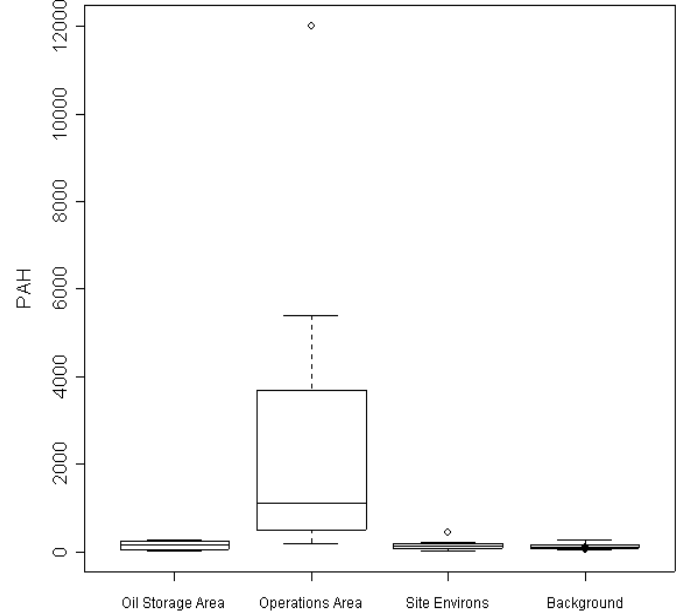
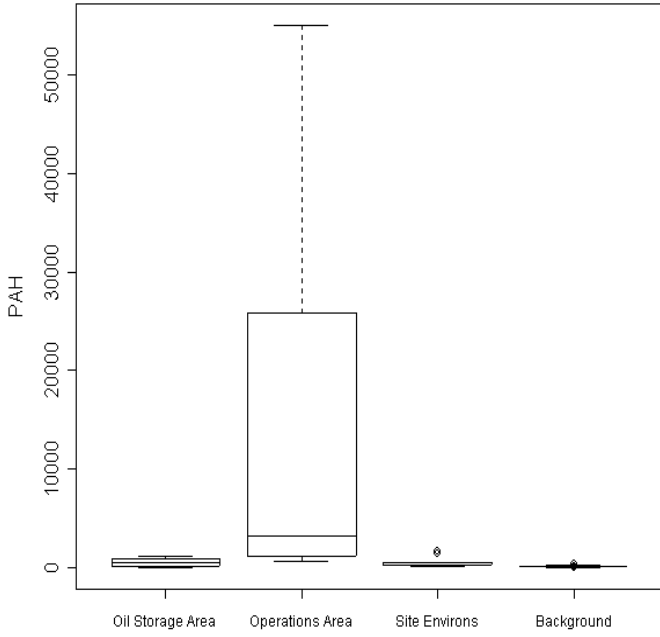


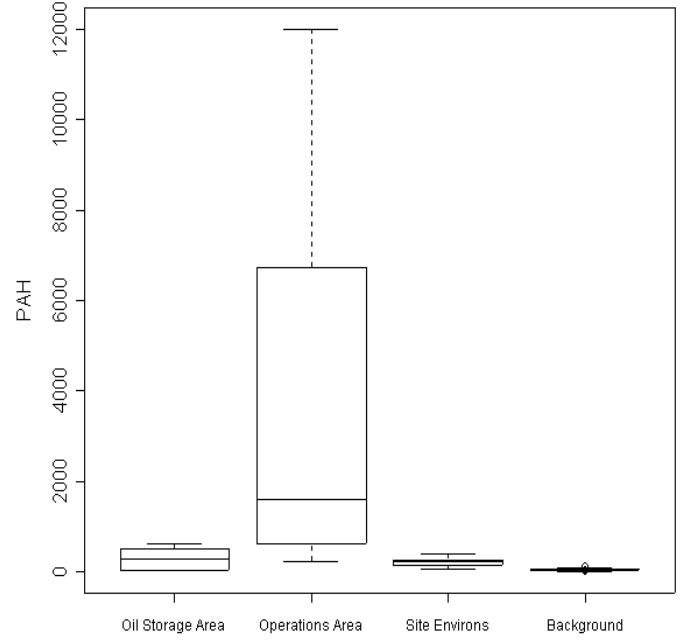
Figure G-7. Benzo(k)fluoranthene Plots: Standard and Log Scales

Figure G-8. Chrysene Box Plots: Standard and Log Scales

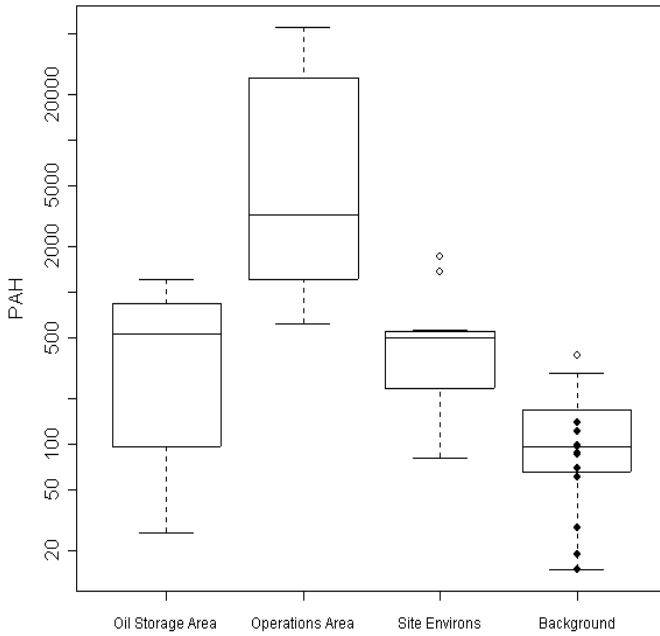
Box-plots of FLUORANTHENE by Sampling Area



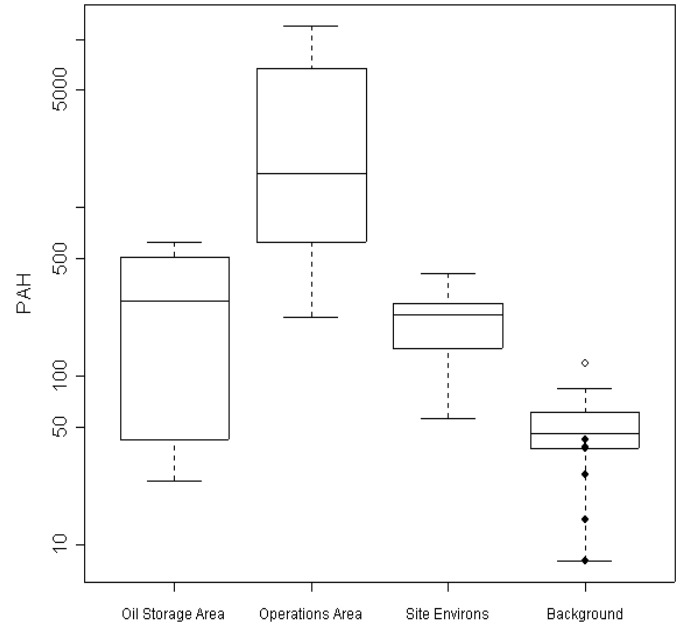
Box-plots of INDENO.1.2.3.CD.PYRENE by Sampling Area



Box-plots of FLUORANTHENE by Sampling Area



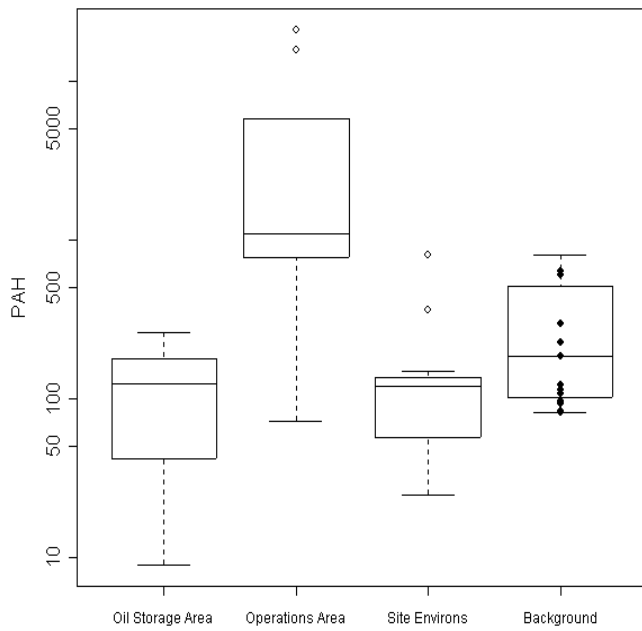
Box-plots of INDENO.1.2.3.CD.PYRENE by Sampling Area



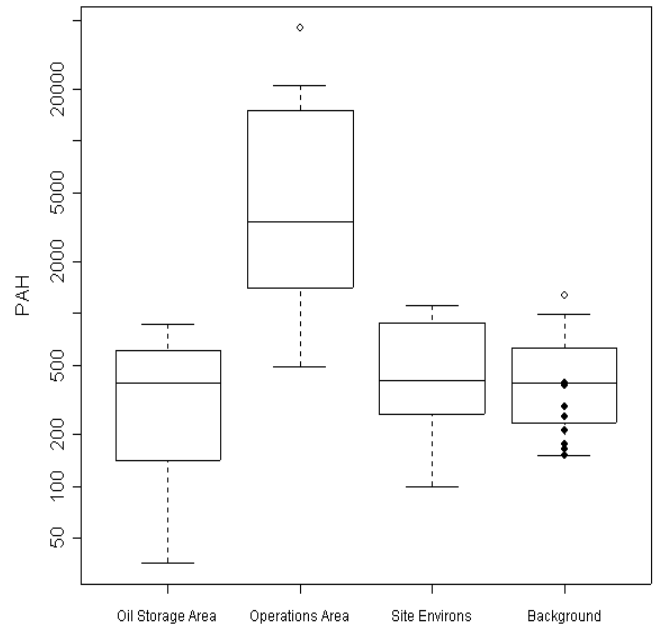
**Figure G-9. Fluoranthene Box Plots:
Standard and Log Scales**

**Figure G-10. Indeno(1,2,3cd)pyrene Plots
Standard and Log Scales**

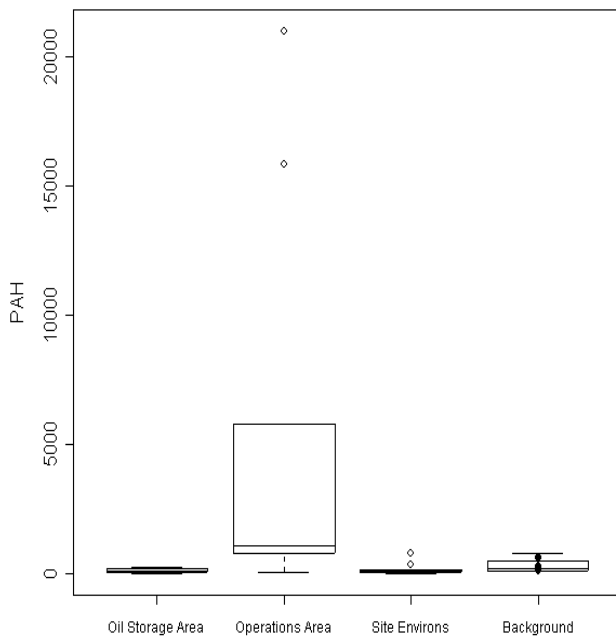
Box-plots of PHENANTHRENE by Sampling Area



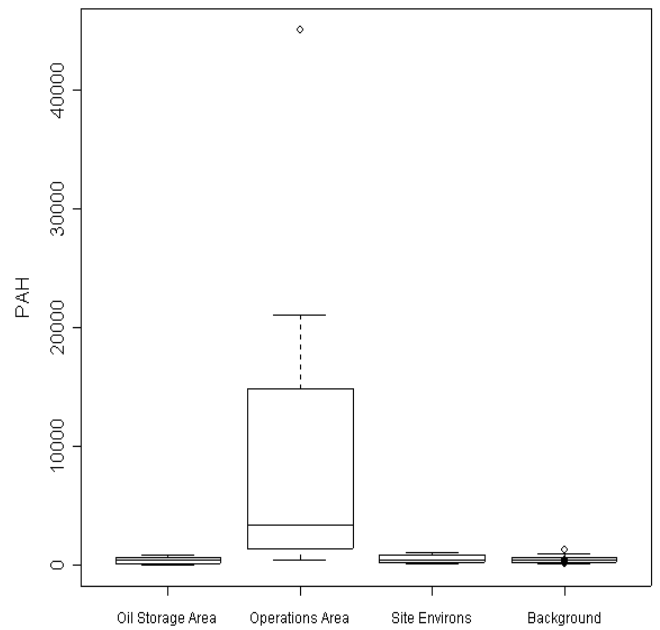
Box-plots of PYRENE by Sampling Area



Box-plots of PHENANTHRENE by Sampling Area



Box-plots of PYRENE by Sampling Area



**Figure F-11. Phenanthrene Plots:
Standard and Log Scales**

**Figure G-12. Pyrene Box Plots:
Standard and Log Scales**