ESTABLISHING BACKGROUND ARSENIC IN SOIL OF THE URBANIZED SAN FRANCISCO BAY REGION

A thesis submitted to the faculty of
San Francisco State University
In partial fulfillment of
The Requirements for
The Degree

Master of Science
In
Geosciences

by

Dylan Jacques Duvergé

San Francisco, California

December 2011

CERTIFICATION OF APPROVAL

I certify that I have read *Establishing Background Arsenic in Soil of the Urbanized San Francisco Bay Region* by Dylan Jacques Duvergé, and that in my opinion this work meets the criteria for approving a thesis submitted in partial fulfillment of the requirements for the degree: Master of Science in Geosciences at San Francisco State University.

Dr. Jason Gurdak
Assistant Professor of Geology

Dr. Jerry Davis
Professor of Geography and Environmental Studies

Mary Rose Cassa

Mary Rose Cassa, P.G.
Professional Geologist, San Francisco Bay Regional Water Quality Control Board

ESTABLISHING BACKGROUND ARSENIC IN SOIL OF THE URBANIZED SAN FRANCISCO BAY REGION

Dylan Jacques Duvergé
San Francisco, California
2011

Soil analysis data within the nine-county San Francisco Bay Area was compiled from the State Water Quality Control Board's Geotracker online database to determine the background levels and variability of arsenic concentrations across four Quaternary geologic units. Arsenic analyses of 1,454 soil samples across 77 sites were screened from Geotracker for inclusion in a JMP 7.0 database. Mean arsenic concentrations within Holocene alluvium (5.10 mg/kg) were determined to be statistically greater than those within Pleistocene alluvium (3.65 mg/kg) and "other" Quaternary units (3.30 mg/kg); and no significant relationship was found between arsenic concentrations and sampling depth. The proposed upper estimate for background arsenic (99^{th} percentile) within undifferentiated flatland soils of the study area—11.00 mg/kg—is markedly lower than commonly cited sources in the literature. These findings present the first regional estimates of background arsenic concentrations in the San Francisco Bay Area.

I certify that the Abstract is a correct representation of the content of this thesis.

Chair, Thesis Committee

12-5-2011

Date

ACKNOWLEDGEMENT

First and foremost, I would like to thank my advisor, Dr. Jason Gurdak, for his receptiveness, thoughtful guidance and support. I consider it an honor to have worked with such a respected professional, and I admire his ingenuity and commitment to this research. I would also like to thank my thesis committee, Dr. Jerry Davis and Mary Rose Cassa for their unique perspectives, and invaluable insight. I am immensely thankful for my friends and family, who have provided moral support throughout my studies; as well as the workload flexibility and tuition assistance offered by my employer, Environmental Science Associates.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF APPENDICIES

1.0 INTRODUCTION AND BACKGROUND

1.1 Introduction

The San Francisco Bay Area (Bay Area) has an enormous natural and cultural diversity where concern for environmental quality permeates regional and local land use planning decisions on a daily basis. Among the broad spectrum of environmental problems is the potential for contaminated soil and groundwater to adversely affect human health. Arsenic—recognized as a potential contaminant by local, state and federal environmental agencies—is a trace metal that is present in low levels in all environmental media (soil and rock, water, and air). Over our lifetimes, we ingest trace levels of arsenic that are naturally occurring in food, drinking water, and (to a lesser extent) air without suffering adverse health effects. However, long-term exposure to elevated levels of inorganic arsenic is known to decrease production of red and white blood cells, cause damage to blood vessels, and lead to characteristic effects such as changes in skin pigmentation, appearance of warts or bruises, skin irritation, and increased risk of skin and other cancers (ASTDR, 2009). In the Bay Area, arsenic is present both as a natural component of soil and rock, as well as a byproduct of human activities such as historical pesticide applications and the presence of copper chromated arsenate- (CCA) treated wood. In this context, environmental investigators face the difficult challenge of determining whether arsenic detections at a site reflect the local soil type or anthropogenic inputs, particularly when detected in the upper range of arsenic concentrations thought to occur naturally.

By screening from an extensive database of soil sampling results that have been submitted to regulatory agencies in the course of environmental cleanup activities, this thesis aims to estimate background arsenic concentrations in soil around the Bay Area, and investigate whether spatial variability in arsenic concentrations can be at least partially explained by variations in soil type. Estimates of the mean and range for trace metals have been developed nationally (7.2, <0.1 – 97 mg/kg) (Shacklette and Boerngen, 1984), for California (3.5, <0.2 -11 mg/kg) (UCR, 1996), and for localized areas in the San Francisco Bay region (5.5, <DL – 42 mg/kg) (LBNL, 2002; Scott, 1991); but there has thus far been no effort to characterize background arsenic on a regional scale or investigate the effect of soil type on arsenic concentrations in the Bay Area. An improved understanding of background concentrations of trace metals and their variability across soil types could help regulators make informed decisions on whether trace metal detections on a property reflect site-related contamination.

1.2 Definitions

This thesis uses several terms and concepts that may have various meanings in other works depending on their topic, scope, and purpose. The meaning of commonly used terms in this thesis is clarified below:

Arsenic: Toxicological profile sheets distributed by the U.S. Department of Health and Human Services (ASTDR, 2009) define arsenic as a naturally occurring element that is widely distributed in the Earth's crust. Arsenic is a chemical element (As) classified as a metalloid, having both properties of a metal and a nonmetal; however, it is frequently referred to as a metal. Elemental arsenic (sometimes referred to as metallic arsenic) is a steel grey solid material. However, arsenic is usually found in the environment combined with other elements such as oxygen, chlorine, and sulfur. Arsenic combined with these elements is called inorganic arsenic. Arsenic combined with carbon and hydrogen is called organic arsenic.

Background: Both U.S. Environmental Protection Agency (EPA) and U.S. Navy describes "background" as substances or locations that are not influenced by existing site-related sources of contamination and is often specified as either representing the naturally occurring background or the anthropogenic background (NAVFAC, 2002; EPA, 2002):

- The *natural background* refers to substances present as a result of geochemical processes that have not been influenced by human activity. Naturally occurring organic and inorganic background substances in soil are solely attributable to the natural geological characteristics of the area.
- The *anthropogenic background* (sometimes referred as the "ambient" levels of a substance) refers to substances present at concentrations that potentially exceed the natural background as a result of human activities, but that cannot be attributed to a specific land-use activity or contaminated area.

Soil: The term "soil" as used in this thesis is broadly defined as loose, unconsolidated clay, silt, sand, and gravel found from the ground surface down to the depth of bedrock. This meaning is consistent with the usage in the field of engineering and environmental geology, and is broader than the definition used by soil scientists and agronomists.

Soil Type: Soils are classified in this thesis based on mapping of quaternary geology, which distinguishes soils by age (e.g., Holocene or Pleistocene) and depositional process (e.g., fluvial, marine, estuarine, or lacustrine). References to soil type contained herein are not synonymous with U.S. Department of Agriculture soil series or surveys, which are more specific and focused on the upper 200 cm of soil for agricultural and other resource management purposes.

Source Rock/Parent Material: The source rock or parent material of a soil refers to the bedrock upon which the soil formed (for residual soils), or from which the soil material was originally derived (for transported soils).

Censored Data: Censored data refers to analytical values that are determined by the laboratory, but that are lower than limits deemed reliable enough to report as numerical values. These observations are reported as seminumerical values that contain qualifiers indicating that the analyte is below the limits of reliability for accurate quantification. Typically, these values are expressed as ''nondetects'' or ''less thans'' such as <0.5.

Environmental Screening Levels (ESLs): ESLs are a compilation of screening levels specific for use at sites overseen by San Francisco Bay Regional Water Quality Control Board for a number of different environmental concerns. ESLs for chemicals in soil are developed for protection against direct exposure (ingestion, dermal absorption, inhalation of vapors and dust in outdoor air), protection of groundwater quality (leaching of chemicals from soil), protection of terrestrial (nonhuman) biota; and protection against nuisance concerns (odors, etc.).

1.3 Arsenic Variability in Soil and Rock

Arsenic (atomic number 33 and relative atomic mass 74.92) belongs to a group of elements often referred to as "trace" elements because its concentration does not normally exceed 1,000 mg/kg (0.1%) while a small group of ten "major" elements make up over 99% of the earth's crust (Alloway, 1990). Trace elements are initially introduced into igneous rocks by substituting for the more common cations that form the crystal matrix of minerals. Typically, substitution occurs when arsenic has similar elemental properties and atomic radii of the more common heavy element. Numerous arsenic containing minerals have been identified, the most common of which are arsenopyrite ($FeAsS$), realgar (AsS), orpiment (As_2S_3), and enargite (Cu_3AsS_4). Arsenic in sedimentary rocks is related to the source and absorptive properties of the sedimentary material that was lithified, the properties of secondary minerals and clays, and the arsenic content of the water that deposited the sedimentary material (Alloway, 1999).

Alloway (1990) reported global mean concentrations of arsenic in different types of rock (Table 1). Mean arsenic content is generally consistent among the major rock types—about 1–1.5 mg/kg—except for some argillaceous sedimentary rocks (shales, mudstones, slates) and phosphorites, which have mean arsenic concentrations from 10–15 mg/kg and have locally been reported to have natural concentrations as high as 900 mg/kg. Separating by rock type, the typical range of concentrations is <1–15 mg/kg for various igneous rocks, <1–20 mg/kg for sandstones and limestones, and <1–200 mg/kg for phosphate rocks. The arsenic content of metamorphic rocks usually reflects the arsenic content of the original, unmetamorphosed rock type.

The typical range of concentrations for arsenic in soils is 1–40 mg/kg with most soils being on the lower end. Kabata-Pendias (1985) reported the mean and range of arsenic background concentrations for several different types of soil, including alluvial soils (8.2, 1.2 to 22 mg/kg), clay and clay loamy soils (7.7, 1.7 to 27 mg/kg), light loamy soils (7.3, 0.4 to 31 mg/kg), and granitic soils (3.6, 0.7 to 15 mg/kg). The type of parent rock is only one of the factors that control metal concentrations in soils. Weathering, biological chemical reactions, and other natural geochemical processes can significantly enrich or deplete the concentrations of certain metals. Due to the high capacity of clay and organic matter to adsorb metallic ions, arsenic concentrations tend to be highest in

Table 1 – Worldwide mean and range of arsenic for major rock types (mg/kg)

	Earth's Crust	Igneous Rocks			Sedimentary Rocks			Alluvial Soils
		Ultra Mafic	Mafic	Granitic	Limestone	Sandstone	Shale	
Mean	1.5	1.5	1	1	1.5	1	9	8.2
Range	--	1-15			1-20		1-900	1.2-22

Sources: Alloway, 1999; Kabata-Pendias, 1985 soils that contain high percentages of clay and organic material (e.g., clay and clay loamy soils, organic light [or rich] soils) (NAVFAC, 2002; Alloway, 1990). Therefore, it is expected that finer-grained depositional environments within the Bay Area would likely have higher natural concentrations of arsenic relative to sandy or gravelly soils.

1.4 Geologic Sources of Elevated Arsenic

As discussed above, soil and rock rich in clay and organic material have been reported to contain elevated concentrations of arsenic relative to other rock types and sandy soils. In addition, unusually high arsenic concentrations have been attributed to highly mineralized geologic environments and zones of hydrothermal alteration. For this reason, arsenic concentrations are commonly used as a pathfinding tool in mineral resource prospecting because high concentrations can indicate the presence of mineralized areas containing valuable commodities such as silver and gold (Alloway, 1990). Further, a national study by Welch et al. (2000) associated thermal waters (e.g. Yellowstone and the Mono Basin), presence of sulfide minerals (e.g. pyrite and marcasite), and areas of high evapotranspiration to high levels of arsenic in groundwater. Welch et al. (2000) did not identify the Bay Area as a region with high arsenic concentrations associated with these processes; however, due to the coarse scale of their study, the possibility that the Bay Area contains localized "hot spots" of naturally-occurring arsenic cannot be ruled out.

Hydrothermally altered mineral zones and coal deposits are relatively rare in the Bay Area, but such environments are locally present in the hills of the region. For example, abundant sulfide bearing rocks are present in the Mt. Diablo district in Contra Costa County, where mercury sulfides and copper were mined in scattered locations on and off throughout the latter half of the 18^{th} century until about the mid-1950s (USGS, 1940; USGS, 2005). Sulfide minerals, including pyrite, marcasite, cinnabar and metacinnibar were formed via hydrothermal deposits within Franciscan Complex rocks bounded on the east by the Great Valley Sequence (USGS, 1940). In addition, locations on either side of the central and southern Santa Clara Valley (including the historic New Almaden Quicksilver District) have also been reported to contain mercury and other precious metals, although associated sulfide minerals are less abundant than in the Mt. Diablo District (USGS, 2005: Bailey and Everhart, 1964). Aside from the mineral districts around Mt. Diablo and New Almaden, the USGS mineral resources data system indicates widely scattered locations within the hills and mountains of the region that contain occurrences of mercury, copper, gold, silver and other mineral commodities that are possible indicators of hydrothermally altered zones and sulfide minerals (USGS, 2005). The only extensive deposits of coal in Northern California are located in the Black Diamond Mines area, north of the Mt. Diablo district. The coal originates from lignite coal beds in the Domengine Formation, and was extensively mined from the 1860's to the beginning of this century (Mount Diablo Interpretive Association, 2009).

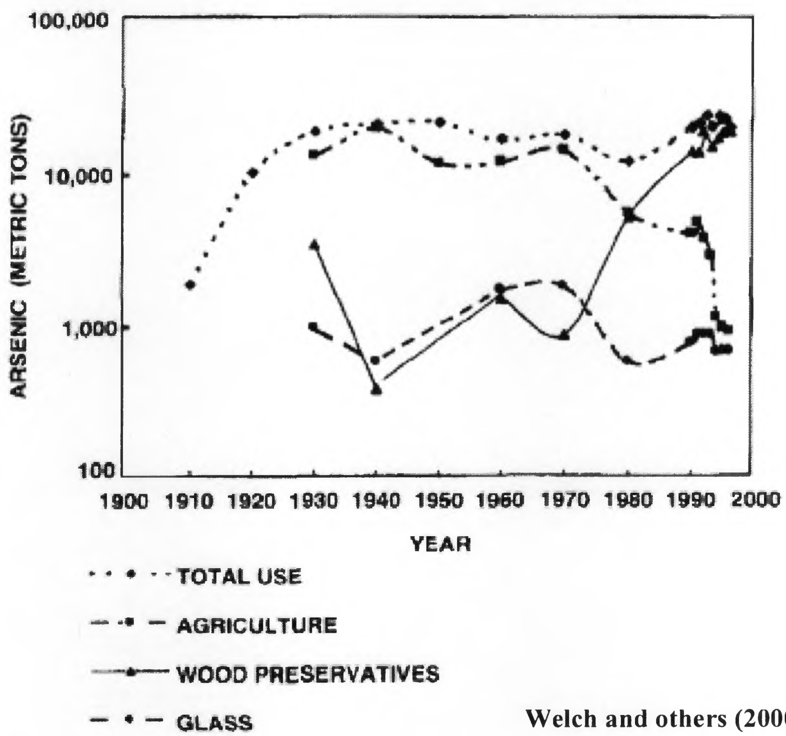
There are no studies specifically aimed at confirming or quantifying the presence of arsenic “hot spots” within the aforementioned locales; but similar geologic environments have been reported in the literature to contain naturally high concentrations of arsenic (USGS, 1940; Alloway, 1990; NAVFAC, 2002; Welch et al., 2000). Such mineralized areas and coal-bearing deposits are confined to a few localized areas in the hills and mountains of the region. Moreover, mercury deposits and associated sulfide minerals in these areas are concentrated in narrow fracture zones within the host rock (Bailey and Everhart, 1964; USGS, 1940). The predominant bedrock and the Quaternary-age deposits of the region are likely to have arsenic concentrations that reflect the more typical concentration ranges discussed in Section 1.3.

1.5 Anthropogenic Sources of Arsenic

Arsenic has a long history of use as a poison dating back to ancient times, but there are several modern uses of arsenic that has made primary production of arsenic compounds, primarily arsenic trioxide, commercially viable. Arsenic trioxide or elemental arsenic is no longer produced in the United States, but it continues to be imported in large quantities, primarily for use as a wood preservative using copper chromated arsenate (CCA). CCA-treated wood, also referred to a "pressure" treated wood, currently accounts for over 50% of domestic consumption of arsenic trioxide; though prior to 2004, it accounted for over 90% of consumption (USGS, 2010). CCA is a water-based product that protects several commercially available species of western lumber from decay and insect attack and is widely used in treating utility poles, building lumber, and wood foundations. The use of CCAs in the wood industry has experienced more recent declines, owing to voluntary elimination of CCA in residential wood products in 2004 (USGS, 2010). However, CCA continues to be used in commercial and industrial applications, and is present in residential structures built prior to 2004 (e.g., in wood needing all-weather proofing). None of the major manufacturers of CCA-treated wood are located in the Bay Area.

Arsenic has also been used in the agricultural industry for pest and weed control. As shown in Figure 1, the use of arsenic in the agricultural industry has experienced a significant decline since the early 1900s. Prior to the introduction of organic pesticides (such as DDT) in the 1940s, inorganic arsenic was the primary pesticide used by orchard growers and farmers. Inorganic arsenic compounds continued to be used as an herbicide, fungicide, growth regulator, desiccant, and/or as a weed control agent along railroad right-of-ways, in potato fields, on grape vines, on lawns, cotton crops, in industrial areas, as well as in baits and to debark trees (ASTDR, 2009). As a result of voluntary industry

Figure 1 – Arsenic consumption in the United States by industrial sector (1900-2000)



Welch and others (2000)

phase-outs and regulatory decisions by the EPA, the use of inorganic arsenic in agriculture has virtually disappeared since the 1980's and 1990's (ASTDR, 2009). The only remaining allowable uses are as ant baits and wood preservatives. Arsenic is also used in the manufacture of glass products, as an alloying element in ammunition and solders, and in semiconductors that are broadly used in computer, biomedical, communications, solar cells, space research and electronics applications (USGS, 2010). Certain industrial processes and mining activities release arsenic as a byproduct, such as stack emissions from copper smelting, coal combustion, and waste incineration; and from mine tailings (Alloway, 1990).

Of particular relevance to the Bay Area is that many of the flatlands surrounding the San Francisco Bay (in particular the Santa Clara Valley) have historically supported irrigated agriculture, such as orchards and other crops that are likely to have utilized arsenic-based agricultural chemicals (Anderson, 1998). As urbanization has encroached on formerly agricultural land, the underlying soils may continue to have elevated arsenic levels representative of its past agricultural use. However, the extent to which former land owners actually applied arsenic-based pesticides, and whether or how much of the arsenic has since leeched out of soils is usually unknown. Generally, anthropogenic sources of arsenic which cannot be attributed to a specific waste discharge, disposal activity, or emission source can be considered “non-point” sources. As defined in Section 1.2, the natural background combined with the anthropogenic background (i.e., non-point anthropogenic sources) makes up the regional or “ambient” levels of background arsenic.

Arsenic may also have been released to the environment from current or former smelters, coal-fired power plants, and municipal incinerators; but very little is known about arsenic atmospheric deposition rates to Bay Area soils. Coal combustion is commonly cited as a source of atmospheric emissions of arsenic, although review of the USGS mineral resources data system shows no current or former coal mines or natural geological occurrence of coal in the Bay Area aside from the Black Diamond Mine area discussed in Section 1.3 (USGS, 2005). Alloway (1999) reports that the annual rate of increase in arsenic concentrations in soil due to atmospheric deposition is minor—about 0.05% for the northern hemisphere. Further, energy production facilities in the Bay Area use natural gas, solar, wind, geothermal, and landfill gas as energy sources rather than coal (EIA, 2009). Prior to the availability of natural gas, manufactured gas plants, primarily concentrated in San Francisco and Oakland, used coal and oil to produce gas for lighting, heating and cooking., these gas plants have all been closed and operated for a short time in the early 1900s (PG&E, 2011). Today, air pollution control technologies used in the Bay Area for stationary sources are advanced and tightly regulated by the EPA and the California Air Resources Board (BAAQMD, 2011). Given the lack of coal combustion or copper smelting facilities in the Bay Area, atmospheric deposition is not likely to be a significant contributor to arsenic concentrations in soil. However, current and former stack emissions cannot be entirely ignored as a possible contributor to the anthropogenic background level of arsenic in the Bay Area.

Disposal of arsenic-containing products, including CCA-treated wood and electronic-wastes (for arsenic-containing products such as semiconductors), can cause locally concentrated levels of arsenic in regulated landfills; or if improperly disposed of, in undocumented areas on private or public property. Numerous former military bases located around the margins of the San Francisco Bay have been closed and identified as hazardous waste or superfund sites (EPA, 2011). Since arsenic was used for munitions and other military applications (USGS, 2010), areas on these bases that formerly stored munitions may also have elevated arsenic concentrations. Numerous state and federal laws—such as the Resource Conservation and Recovery Act and Title 22, Division 4.5 of the California Code of Regulations—regulate the generation, treatment, and disposal of solid and potentially hazardous wastes such that most arsenic-containing products are likely to end up in a landfill. However, such regulations originated in the 1970s and thus improper disposal of arsenic-containing wastes could have occurred prior to that time, and may still occur as a result of negligent or unlawful activities. Generally, anthropogenic sources of arsenic such as these, which can be traced back to an identifiable source, can be considered "point" sources.

1.6 Arsenic Background Studies

Many environmental scholars, managers, and regulators have recognized the need to characterize the source and distribution of trace metals in the soil environment. Previous work has focused on a) the association between groups of trace metals and their potential to predict other geochemical properties, b) the effect of rock type and land use on metal concentrations at the scale of cities, and c) the advantages and disadvantages of various methods for characterizing geochemical background environments (Facchinelli et al, 2001; Yesilonis et al., 2008; Zhang and Selenius, 1998; Li et al., in press). Traditionally, classical statistics and multivariate analysis have been used to characterize trace metal populations in soils; however, researchers have increasingly recognized the value of using Geographic Information Systems (GIS) and geostatistics to map the spatial pattern and variability of trace metals in soil and visualizing relations with geology and land use (Zhang and Selenius, 1998). While researchers often carry out their own sampling and laboratory analyses for local studies, regional studies have increasingly utilized publically available geochemical databases for the study of trace metals (Rawlins et al., 2003; Zhao et al., 2007; and Lado et al., 2008). Existing studies of background concentrations of arsenic that are relevant to the Bay Area are listed in Table 2 and described below.

Table 2 – Scope and findings for existing background studies of arsenic

Author(s)	Geographic Scope	Number of Samples / Depth	Average, Range (mg/kg)
Shacklette and Boerngen (1984)	National, along major roads, average of 1 sample / 6000 km^2	1,318 / 20 cm (7.9 inches)	7.2, <0.1 – 97
UCR (1996)	Statewide, agricultural soils, primarily Central Valley	50 / 50 cm (1.6 feet)	3.5, <0.2 -11
Lawrence Berkeley National Laboratory (2002)	Local, Berkeley Hills, large cleanup site	1,397 / up to 60 meters (197 feet)	5.5, <DL - 42
Scott (1991)	Local, northern Santa Clara Valley, urbanized light industrial and research land uses	108 / up to 10 meters (33 feet)	2.86, <DL – 20

On a national level, the most comprehensive study of naturally-occurring trace metals in the environment has been performed by Shacklette and Boerngen (1984). The study collected 1,318 soil samples from around the country at depths of 20 cm below the ground surface (bgs) from locations about 80 km apart that, insofar as possible, had surficial materials that were very little altered from their natural condition and that supported native plants. The mean concentration of arsenic for the western conterminous United States was 7.2 mg/kg, with a maximum concentration of 97 mg/km. Three Bay Area samples that were collected near Stanford University, in the City of San Francisco, and near Mill Valley were all within the upper 40% of the frequency distribution plot (Shacklette and Boerngen, 1984). Gustavvson et al., (2001) later re-interpolated the results and produced a colored surface map of arsenic distribution in the United States, which indicated broad regional variability in arsenic concentrations. For example, high arsenic concentrations in northern Idaho and the Appalachian Basin were at least in part coincident with base- and precious-metal mining, coal-bearing deposits and coal-fired power plants (Gustavvson et al., 2001). An area in north-central Nebraska with low concentrations of arsenic corresponds to the Nebraska Sand Hills, the largest dune field in the Western Hemisphere. The authors observe that many of the geochemical abundance patterns reflect regional geological characteristics. However, they acknowledge that the low spatial density of the dataset means that some of the observed patterns may be due to random chance rather than geologic source controls.

On the state level, an important source of information on background trace metals is from the Kearny Foundation Special Report on Background Concentrations of Trace and Major Elements in California Soils (UCR, 1996). The study selected 50 samples from 22 benchmark soils from a collection of soil profiles held at the University of California, Berkeley (the soil profiles were collected in 1967). The profiles were taken from sites distant from known point sources of contamination throughout the state at 50 cm bgs, primarily within agricultural fields. Arsenic concentrations across the 22 "benchmark" soils had an average of 3.5 mg/kg, a standard deviation of 2.5 mg/kg, and values ranging from 0.6-11 mg/kg. The report authors used the W test for normality, finding arsenic to be neither normally nor lognormally distributed.

In the Bay Area, private consulting firms, local governments, and academic researchers have used a variety of methods to characterize background concentration of trace metals. A study by researchers at the Lawrence Berkeley National Laboratory (LBNL) (2002) characterized the natural background metal concentrations on the LBNL property in the Berkeley Hills by compiling the results of previous environmental investigations, eliminating outliers, evaluating the probability distributions of metal samples, and deriving summary statistics. The LBNL (2002) study determined that the 1,257 soil samples at various depths less than 60 meters (179 feet) bgs had a mean arsenic concentration of 5.5 mg/kg and standard deviation of 5.4 mg/kg. The authors also observed that naturally occurring concentrations of arsenic in samples taken from the Great Valley Sequence were elevated relative to those within Tertiary-age sedimentary rocks. Similar to the Kearny Foundation Special Report, the authors found that arsenic concentrations did not appear to be either normally or lognormally distributed, even after separating the dataset by rock type. The LBNL study determined the upper limit of background concentrations for arsenic to be 42 mg/kg for the Great Valley Sequence, and 24 mg/kg for other geologic units.

Anderson's (1998) literature review of natural concentrations of selenium, nickel, and arsenic in soil and groundwater of the South Bay identified that that certain geologic environments are naturally enriched in nickel (from serpentinite) and selenium (from marine shales and sulfides), but found no evidence or areas with naturally enriched with arsenic. A master's thesis by Scott (1991) characterized background soil metals in an area within a two mile (3.2 kilometer) radius in Mountain View and Sunnyvale in northern Santa Clara County. An analysis of 108 samples up to 10 meters (33 feet) bgs revealed mean arsenic concentration to be 2.86 mg/kg with a standard deviation of 2.61 mg/kg. Neither of the studies proposed upper limits for background concentrations of arsenic.

The existing studies of background arsenic vary greatly in scale, geographic scope and data source. While all the studies described above took measures to avoid obvious sources of anthropogenic arsenic contamination, the studies at the local and state scale (UCR, 1996; LBNL, 2002; Scott, 1991) were located in agricultural fields or heavily urbanized settings previously disturbed by humans, whereas the study by Shacklette and Boerngen (1984) targeted natural areas supporting native plants (although many were located close to roads, and not necessarily outside of urban areas). While all of the studies reported high variability and ranges, even in relatively localized study areas; interestingly, arsenic concentrations detected in the national study were generally higher than those reported in the local and state studies which have high degree of anthropogenic disturbance relative to the undisturbed areas sampled in the Shacklette and Boerngen study. Due to differences in study design, such as variability in number and depth of samples, land-use setting, and geographic scope, the available background studies report inconsistent arsenic concentrations and thus may provide misleading benchmarks of background arsenic for use by environmental managers and regulators in the Bay Area or any other specific location.

1.7 Problem Statement and Purpose

Arsenic found in soil—either naturally occurring or from anthropogenic releases—forms insoluble complexes with iron, aluminum, and magnesium oxides found in soil surfaces, and in this form, arsenic is relatively immobile. However, under certain reducing conditions, arsenic can be released from the solid phase, resulting in soluble mobile forms of arsenic, which may potentially leach into groundwater or result in runoff of arsenic into surface waters (Alloway, 1990). Thus, in addition, to arsenic in soil representing a direct exposure hazard (i.e., via inhalation of dust during construction activities; children eating soil in open-space areas; or consumption of food crops grown on contaminated soils), elevated arsenic levels in soil could also lead to elevated levels of arsenic in surface and groundwater used as sources of drinking water—leading to the potential for more dispersed and widespread exposure to the public. It is in this context that regulators seek to control and minimize potentially hazardous levels of arsenic in soil.

To protect the human health and/or the environment, state environmental regulators have established environmental screening levels (ESLs) for arsenic, above which trace metal concentrations are considered potentially hazardous (San Francisco Bay RWQCB, 2008). Generally, locations with arsenic concentrations below ESLs are assumed to not pose a significant, long-term (chronic) threat to human health and the environment. Locations with arsenic concentrations above ESLs usually require some form of action which may range from additional sampling and analysis to contaminant removal. The ELSs used in the San Francisco Bay Region are risk based screening values, which are derived from equations combining exposure assumptions with toxicity data, and are not related to background levels of a substance in the environment. The risk-based screening level for arsenic in soil varies based on exposures assumptions. For example, the ESL for arsenic in shallow residential soil is 0.39 mg/kg (San Francisco Bay RWQCB, 2008). Due to regional background concentrations of arsenic, ESLs are nearly always exceeded, even in locations where no anthropogenic arsenic contamination has occurred (UCR, 1996; San Francisco Bay RWQCB, 2008). Therefore, as a practical matter, regulators have generally accepted the background levels of arsenic found in the environment as an appropriate screening criteria, because property owners are not considered liable for arsenic concentrations that are representative of background (ITRC, 2005).

When environmental contamination is reported or suspected on a site, or in the course of non-residential real estate transactions, environmental site assessments (ESAs) are typically performed to identify potential sources of contamination and guide further cleanup efforts. When laboratory analyses of soil are performed as part of an ESA, detections of arsenic above ESLs may require no further action (with respect to arsenic) if there is a reasonable basis to conclude that arsenic concentrations are representative of background (San Francisco Bay RWQCB, 2008). In the absence of site-specific background control samples, it is my experience that ESAs in the Bay Area frequently use data from Shacklette and Boerngen (1984), LBNL (2002) or other studies of uncertain applicability to conclude trace metal concentrations found on a property are non-anthropogenic in origin. Most Bay Area properties that require ESAs are located on the urbanized bay plain whereas the LBNL site is located in the east bay hills where the rock type, geomorphology and soil forming processes differ. As such, use of LBNL background concentrations, or nationwide estimates, may not be appropriate given that geology may be a significant control on background arsenic concentrations.

Despite the abundance of soil analytical data publically available through environmental agencies, in particular the State Water Resources Control Board (SWRCB), there has been little attempt to discern whether there are geochemical patterns of arsenic that correlate with soil types or source rocks. Sites undergoing environmental investigations and cleanups often submit soils for laboratory analysis of a standard suite of trace metals (referred to as CAM 17 metals) that are incidental to the primary contaminants of concern on the site, such as motor fuels or organic solvents. As such, the data represents a potentially valuable source of background information that has thus far been underutilized. By incorporating regional information on soil type to these analytical results, this thesis will assess whether there are statistically significant differences in arsenic concentrations across the different soil environments in the Bay Area.

The purpose of this thesis is to provide regulators and environmental investigators with a locally relevant study of background arsenic in the Bay Area. The findings herein can be used as a tool to make informed decisions about whether arsenic detections on a property are indicative of background in cases where site-specific background reference samples are infeasible or cost-prohibitive. Rather than the current reliance on background arsenic data of low-resolution and questionable applicability, this thesis provides a regional and geologic context to the question of "what is background arsenic in soil?"

2.0 STUDY AREA

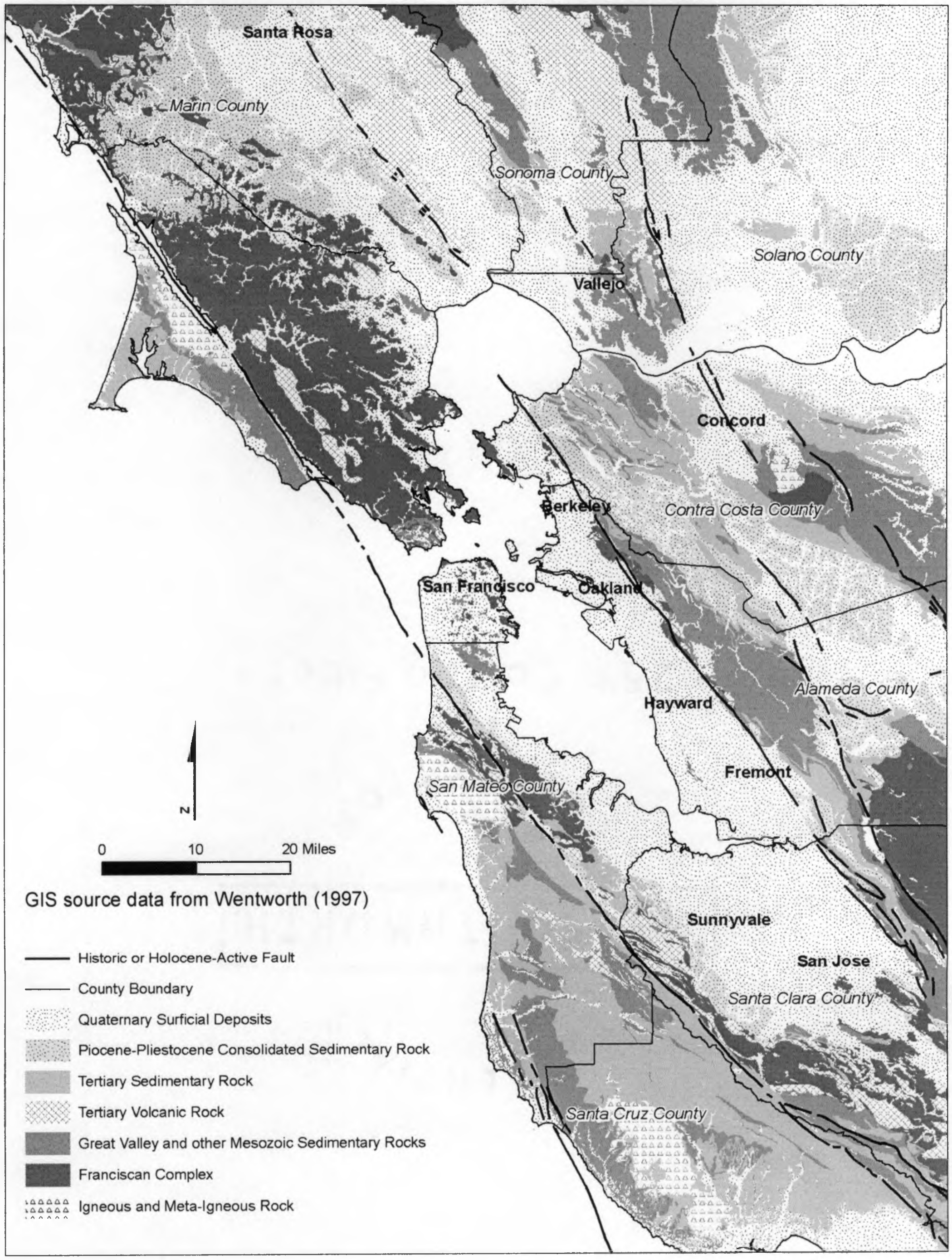
2.1 Study Area

The geographic scope of this study encompasses the nine-county Bay Area (Figure 2), including Alameda, Contra Costa, Marin, Napa, San Mateo, San Francisco, Santa Clara, Solano, and Sonoma Counties. Because environmental investigations are most often done in urban land use settings where the potential for contamination is greatest, the study area is restricted to the urbanized flatlands of the bay region underlain by Quaternary-age geologic units. By restricting the geographic scope of analysis in this way, the resulting background estimates are more likely to be directly applicable to future ESAs.

2.2 Regional Geologic Setting

The Bay Area is part of the Coast Range geomorphic province characterized by northwest-trending mountain ranges and valleys that are subparallel to the general structural trend of the San Andreas Fault System (CGS, 2008). The Coast Range is predominantly composed of thick Mesozoic and Cenozoic sedimentary strata. In the northern Bay Area, the Coast Ranges are dominated by the irregular, knobby, landslide topography of the Franciscan Complex, which is overlain in several regions by volcanic cones and flows of the Quien Sabe, Sonoma and Clear Lake volcanic fields. In the eastern Bay Area, the Coast Ranges are characterized by strike-ridges and valleys in Tertiary and Upper Mesozoic sedimentary strata. The southern Bay Area is characterized by a mix of Franciscan Complex rocks on the east side of the Santa Clara Valley, and both Tertiary sedimentary rocks as well as granitic rocks of the Salinian Block west of the valley in the Santa Cruz Mountains.

Figure 2 – Geologic map of the San Francisco Bay Area

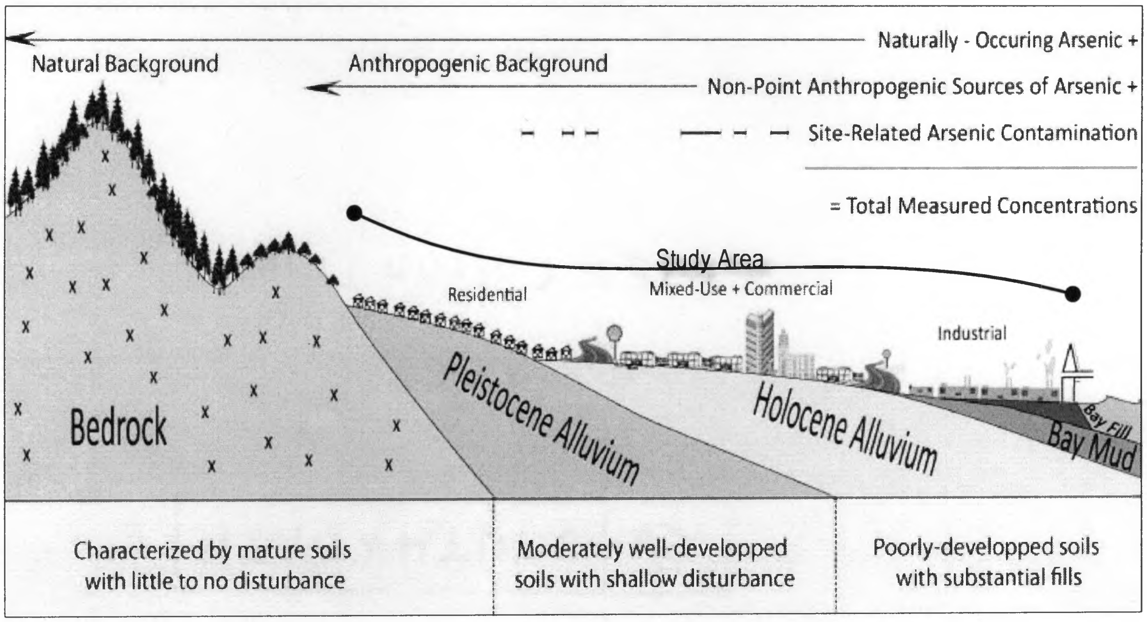


The core of the urbanized Bay Area is located on Quaternary-age surficial deposits that have formed the flatlands around the margins of the bay. These flatland deposits include Pleistocene and Holocene alluvial fans emanating out from the hills and mountains, as well as floodplain, basin and bay mud deposits located closer to the bay margins (Helley et al., 1979).

2.3 Land Use Setting

With 7.1 million residents, the Bay Area is the fifth most populous metropolitan area in the United States (ABAG and MTC, 2011). In 2000, approximately 16 percent (or about 700,000 acres) of the region's total acreage was developed for urban use (ABAG and MTC, 2011). The majority of the land areas developed for urban use consists of flatlands that surround the San Francisco Bay and which create several large inland valleys in the east bay. Generally the most intensely developed areas, including ports, airports, former military bases, and major industrial areas, are located close to the bay margins, whereas the urban fringes and foothills of major mountain ranges generally support low-density residential development. Mixed use, high-density residential areas, and commercial districts are concentrated in urban centers and along major highway corridors. Figure 3 presents a conceptual cross section of the east bay, showing the general relationship between geology, land use, and the components of total measured arsenic concentrations. Because this study is regional in scope, a rough understanding of the interplay between geology, land use and their possible effects on measured arsenic concentrations can help frame the discussion of results.

Figure 3 – Representative cross section of land use and geology and the relation to total measured arsenic concentrations



Geologic cross section based on Helley et al., 1979, not to scale.

3.0 METHODS

Results of soil chemical analyses from previous environmental investigations in the Bay Area were used to derive summary statistics and investigate the variability of arsenic concentrations across different Quaternary soil types. The source data, site selection criteria, database compilation, analysis and treatment of data, and statistical tests are described below.

3.1 Source Data

The data used in this study was retrieved from the State Water Resources Control Board (SWRCB) "Geotracker" database. Geotracker is a data system for managing sites that impact groundwater, especially those that require groundwater cleanup. In September 2004, the SWRCB formally adopted regulations that require Electronic Submittal of Information (ESI) for all groundwater cleanup programs, although parties responsible for cleanup of underground storage tanks had already been required to submit groundwater analytical data, surveyed locations of monitor wells, and other data to the Geotracker database for several years (since about 2001). As of January 1, 2005, ESI has been required by all groundwater cleanup programs including underground storage tanks, non-tank site cleanups, military sites, and land disposal sites. ESIs include site location information, soil and groundwater analytical data, monitoring well and boring log information, and electronic (pdf) copies of site investigation reports prepared by responsible parties and/or their consultants. GeoTracker's ESI module is the largest receiving system nationally for analytical and field data for cleanup sites (SWRCB, 2010). Geotracker has about 4,500 sites from the Bay Area.

Other sources of publically accessible soil analytical data exist—namely from the Department of Toxics Substances Control (DTSC), which regulates sites that handle, treat, store, or dispose of hazardous wastes, and some limited data from the U.S. Geological Survey (USGS) geochemical database. Geotracker has advantages over other data sources because of the large volume of data and the common reporting protocol required for Geotracker, which includes specific guidelines for preparing datasets and a defined set of valid values for each database field. This protocol helps ensure that the various laboratories that analyze soil samples report data in a consistent manner. Direct electronic reporting also avoids the need for manual re-entry of hard-copy laboratory data, minimizing data entry errors and inconsistencies. Data in Geotracker is generally less than five years old, which means that the laboratory methods used are consistent and reflect the current industry standard. This is important because analysis procedures and method detection limits have frequently changed over the past decades, which can present problems in obtaining reliable or comparable statistics.

Finally, the majority of the sites regulated by the SWRCB are those that have underground storage tanks that have leaked or are potentially leaking their contents, or that for other reasons have groundwater contaminated with motor fuels or organic solvents. For most sites, there will be little or no correlation between the metal and organic compound distributions (NAVFAC, 2010). Chemical releases that contain both types of contaminants are relatively uncommon and, more importantly, organic compounds and metals have very different fate and transport properties. It also is important to note that the presence of organic co-contaminants has no effect on metal concentration background ranges (NAVFAC, 2002). In most cases, the soils analyzed for arsenic are done so as a precautionary measure to demonstrate the absence of arsenic contamination, and are generally incidental to the primary contaminants of concern. In this context, such analyses have value as a potential source of background data.

3.2 Site Selection Criteria and Database Compilation

The user interface on the Geotracker website allows for site or location queries, allowing users to search by address, site name, or other identifying information. In order to perform a custom query of the database, raw ESI data was downloaded as a tab-delimited file for each of the nine counties in the Bay Area (called an electronic data file, or EDF). Geotracker is always being updated as additional sampling and analytical data is generated at regulated sites. As such, the arsenic-related data presented in this thesis should be considered as representative of the database as of March 2010, which is the date the EDFs were downloaded. The EDF contains raw laboratory analytical data for the numerous cleanup sites in the Bay Area that are associated with the specific locations using a Global ID field. Concurrently with the download of ESI data, an excel file of regulated site information was downloaded, which contains site names, addresses, coordinates, cleanup/regulatory status, potential contaminants of concern, and other site information fields that are also associated to a Global ID field. Additional information on the database structure is available on the Geotracker website (geotracker.waterboards.ca.gov).

Using JMP 7.0, a statistical software package, the ESI data and regulated site information were linked by Global ID and queried to return all sites that have analytical data for arsenic from soil samples. The criteria for including sites in the database were as follows:

- Arsenic is not identified as a contaminant of potential concern
- ESI data includes arsenic analyses on soil samples (analyte=AS and Matrix=Soil).
- At least 5 samples per site (N>=5)
- At least 25% of the data is above the reporting limit.

These criteria were applied in order to eliminate sites that were identified in Geotracker as being potentially contaminated with arsenic, or that contained insufficient data to reliably estimate a central tendency or derive other simple statistics. Sites that otherwise would have satisfied the criteria were excluded from the database due to one or more of the following reasons: 1) laboratory notes indicated excessive interference or other problems with the analysis, 2) arsenic contamination was suspected based on detected data, and 3) duplicate ESI entries (i.e. the same laboratory report was submitted to Geotracker more than once). The level of effort and approach taken to avoid sites with metals contamination used in this study is consistent with other works that have utilized existing data (LBNL, 2002; Scott, 1994; Yesilonis et al., 2008; Lado et al., 2008).

Following ESI data download and site selection, the JMP 7.0 database was expanded to include fields that were not a part of the original ESI, including sample depth and geologic unit. Using ArcMap 9.2, the site locations were overlain onto a regional geologic map of the Bay Area to assign geologic units to each site. The geologic map is a digital database containing a GIS shapefile for the general distribution of geologic materials in the San Francisco Bay Region released by the USGS (Wentworth 1997). Geologic materials are categorized in the database by general age and lithology. The cleanup sites used in this study were predominantly underlain by Pleistocene alluvium, Holocene alluvium, and Holocene bay mud deposits. Other geologic units included undifferentiated Quaternary units such as terrace deposits, colluvium, and dune sands. The fields included in the database, their definitions and source are provided in Table 3. All the samples in the database were analyzed by either inductively coupled plasma (ICP) atomic emission spectroscopy or ICP mass spectrometry.

Table 3 - Explanation of database fields by name, type and source

Field Name	Type	Source / Description
Global ID	text / nominal	Geotracker
Site name	text / nominal	Geotracker cleanup site database
City	text / nominal	Geotracker cleanup site database
County	text / nominal	Geotracker cleanup site database
Site N	numeric / ordinal	Number of samples per site; derived in JMP
Sample ID	text / nominal	Geotracker EDF download
Field point class	text / nominal	Identifies sample collection method. Geotracker EDF download
Depth	numeric / continuous	Site investigation reports
Depth class	text / nominal	Shallow or subsurface, based on Navy guidance. See section 3.3
Value (mg/kg)	numeric / continuous	Geotracker EDF download
RL	numeric / continuous	Reporting Limit, Geotracker website, site by site search
MDL	numeric / continuous	Method Detection Limit, Geotracker website, site by site search
D_Arsenic	numeric / ordinal	Censored data identifier
Arsenic	numeric / continuous	Arsenic value field with censored data estimates. See Section 3.4
Substitution method	text / nominal	See Section 3.4
Comments	text / nominal	Optional field for comments
Age-Lith	text / nominal	Age / lithology ID from USGS
Geologic Unit	text / nominal	Geologic unit name from USGS
Geology Class	text / nominal	Geologic units grouped into four categories for this analysis

In order to gather information on sample depths, pdf or scanned hard copies of the site investigation report(s) for each site were reviewed and pertinent information was then transferred to the database. Site investigation reports were not available for approximately 17 sites in the Geotracker database, in which case information on sample depth was either left blank, or assumed based on the sample ID (i.e., if the sample ID was "B-2@2'" the depth was recorded as 2 feet in the database). Sample depths and field collection method were reviewed to classify samples as being either surface or subsurface samples. Boring equipment used to collect soil samples may not be capable of collecting samples over discrete intervals less than 2 feet long. In addition, the boring action may mix soil from near the surface with deeper soils. Therefore, as recommended in a Navy guidance document for environmental background analysis (NAVFAC, 2002), each of the following were considered as surface soil samples:

- soil samples collected with hand tools ("grab samples") between the surface and 0.5 foot bgs
- soil samples collected from borings between the surface and 2 feet bgs
- soil samples explicitly identified as surface samples

All other samples were considered subsurface soil samples. Composites or samples without depth information were not assigned depths or depth classes.

As discussed in Chapter 2, the study area is within urban areas whose soils have likely been disturbed and reworked within several feet of the ground surface due to grading, soil moving, construction activity and utility work. It is possible that anthropogenic inputs of arsenic, if present, have been mixed down to the historical depth of disturbance. To account for this possibility, soil depths were also classified as shallow (≤ 6 feet bgs) or deep (> 6 feet bgs). Six feet (1.8 meters), while somewhat arbitrary, was considered a reasonable depth based on common depths of excavation needed for utilities, roads, building foundations and site leveling within flatland soils.

3.3 Assessment of Site Data, Treatment of Censored Data, and Identification of Outliers

Due to the broad geographical area, geological diversity, and land-use setting of the study area, before conducting an assessment of background arsenic concentrations within the Bay Area as a whole, outliers and censored data for each individual site were evaluated. For sites that contain censored data (i.e., nondetects, or values that are less than the laboratory reporting limit), normal quantile plots of site data were generated in JMP to characterize the distribution of arsenic concentrations. Where neither a normal nor lognormal model fit the data, non-parametric statistical methods were used to conduct further analyses. For several sites, especially those with a low number of samples, the graphical methods were insufficient to determine the type of population distribution. In such cases, goodness-of-fit tests available in ProUCL were used to best estimate the distribution type of the data. ProUCL is a statistical application released by the U.S. EPA that is designed specifically for environmental datasets with nondetects. Most of the statistical methods described and recommended in EPA's guidance on assessing background concentrations at contaminated sites (EPA, 2002) are incorporated into ProUCL. Either the Shapiro-Wilk test or the Lilliefors test, depending on sample size was used to determine the distribution type. Information on the distribution type was used to estimate the values of censored data, as described below.

The predominant method in the environmental field to incorporate nondetects data into statistical analysis is to replace censored data with artificial values, such as the reporting limit or half of the reporting limit (i.e., simple substitution). However, Helsel and Hirsch (2002) found that summary statistics obtained using the simple substitution method do not perform well even when the percentage of nondetect observations is low, such as 5%-10%. Therefore, rather than handling non-detect values in the conventional way, the regression on order statistics (ROS) method recommended by Helsel and Hirsch (2002) was used to estimate the values of censored data. An ROS estimation function in ProUCL was used to generate estimated values for the censored data based on the most likely distribution type at each site. For censored data at sites where no discernable distribution was apparent, simple substitution using half the reporting limit was used to substitute for nondetects.

Outliers—defined as sample values that are unusually large (or small), and that are obvious deviations from the background distribution—may result from analytical errors, transcriptions errors, or the presence of contaminated samples in the background dataset. To identify outliers, box plots were generated for each site in JMP. Any values beyond the upper (or lower) quantile +/- 1.5x the inter-quartile range (IQR) were identified as outliers and eliminated from the database. The purpose of this evaluation was to minimize the effect of outliers on the background statistical analysis.

3.4 Summary Statistics and Statistical Tests

Using JMP, summary statistics were derived for each site in the database, including the number of observations (N), percent of observations that were nondetects, mean, median, standard deviation, and IQR. A frequency distribution and a normal probability plot were also generated to graphically display the site medians. The spatial autocorrelation tool in ArcMap 9.2 was used to assess the degree to which site medians were spatially clustered. The sites were then grouped based on the mapped geology, and summary statistics were derived for each soil type. Using analysis of variance (ANOVA) and the Tukey Kramer HSD test (Tukey test) available in JMP, each geological grouping was used to test for significant differences among their means. The sites were also grouped by depth class, and Wilcoxon test was used to determine if the group medians are significantly different. A 95% confidence level was used for all statistical tests to determine statistically significant differences among group means. ProUCL provides a number of statistical options for calculating background threshold values (BTVs). Because of the large, well distributed nature of the dataset, and consistent with federal guidance (EPA, 2002; NAVFAC, 2002), the 99th percentile was selected as the appropriate measure of the upper range of background concentrations within the study area.

3.5 Limitations and Assumptions

Because this research is based on existing data and does not involve field sampling or field verification of geologic mapping, the statistical analyses and associated findings presented herein must be viewed in the context of several assumptions and associated limitations:

- It is assumed that the geology of a Geotracker site is representative of the lithology mapped by Wentworth (1997), and that Geotracker has recorded accurate locations for each of the sites. Locations close to the bay or in dense urban settings are likely to be underlain by a variable thickness of artificial fill soils used to prepare sites for development. Due to the relatively flat topography of Quaternary geologic units, it is standard practice to balance cuts and fills onsite during construction-related grading. Therefore, for sites mapped as being within Pleistocene or Holocene alluvium, it is assumed that fill soils are representative of the same geologic unit. For sites within bay mud, which are highly compressible and (from a geotechnical standpoint) not suitable for most urban development projects, fills from offsite sources have commonly been placed over the surface of the bay mud (Helley at al., 1979). For sites in Geotracker mapped as being on bay mud, boring logs of the site were reviewed, if available, to verify the accuracy of the mapped geology. Samples identified as being within artificial fills were removed from the database because their origin and lithology is unknown.

- It is assumed that the Geotracker database fields identifying potential contaminants of concern are accurate and represent the full range of contaminants thought to be the result of site-related activities. If arsenic or metals were not identified as a contaminant of concern in the database, it is assumed that arsenic detections are generally representative of the anthropogenic background. This assumption was verified to the extent possible through review of the site investigation reports available in Geotracker.
- A generic method (see Section 3.2) was used to identify outliers for each site in the database that may not effectively identify outliers that are part of a second distribution. EPA guidance (NAVFAC, 2002) generally recommends identifying outliers through observation of log-transformed data on a probability plot. Data points that are not near the line or do not fit a continuous distribution are generally considered as outliers or belonging to a second, contaminated population. However, due to the high number of sites in the database, any values exceeding the upper quartile + 1.5*IQR were considered outliers for the purpose of efficiency.

This study does not attempt to characterize the geochemical behavior of arsenic in soil or explain the influence of small-scale geochemical processes on total arsenic concentrations. Rather, sufficient data is being collected to reasonably characterize arsenic concentrations representative of background at a regional scale, and to determine whether differences in flatland geology represents a statistically significant variable. The results of this study are most relevant to areas underlain by quaternary geologic units within the Bay Area, and should not be used outside of the relevant geographical area.

4.0 RESULTS

4.1 Summary of the Database

Based on the site selection criteria (Section 3.2), 77 sites were selected from the Geotracker database for inclusion in this study. The 77 sites represent 2 percent of the total number of Geotracker sites within the 9-county Bay Area. The number of arsenic samples at each site ranges from 5 to a maximum of 139, totaling 1,454 samples across the 77 sites. All 1,454 records, including the fields described in Table 3 are included in a Microsoft Excel or JMP 7.0 file, which may be obtained by request (dylanduv@gmail.com). Approximately 65 percent of the data selected from Geotracker comes from soil borings, which were made for the purpose of collecting soil samples or as part of the installation of groundwater monitoring or remediation wells. The remaining 35 percent of the data consists of 1) soil samples collected from the walls or pits of excavated areas that formerly contained underground storage tanks or soils impacted by petroleum hydrocarbons or organic solvents, 2) surface samples collected by hand or hand-auger as part of an environmental investigation, or 3) soil stockpile samples for the purpose determining an appropriate off-site disposal method. Reporting limits in the database were generally below 1 mg/kg, although approximately 16 samples (less than 0.5 percent of the database) had high reporting limits over 5 mg/kg.

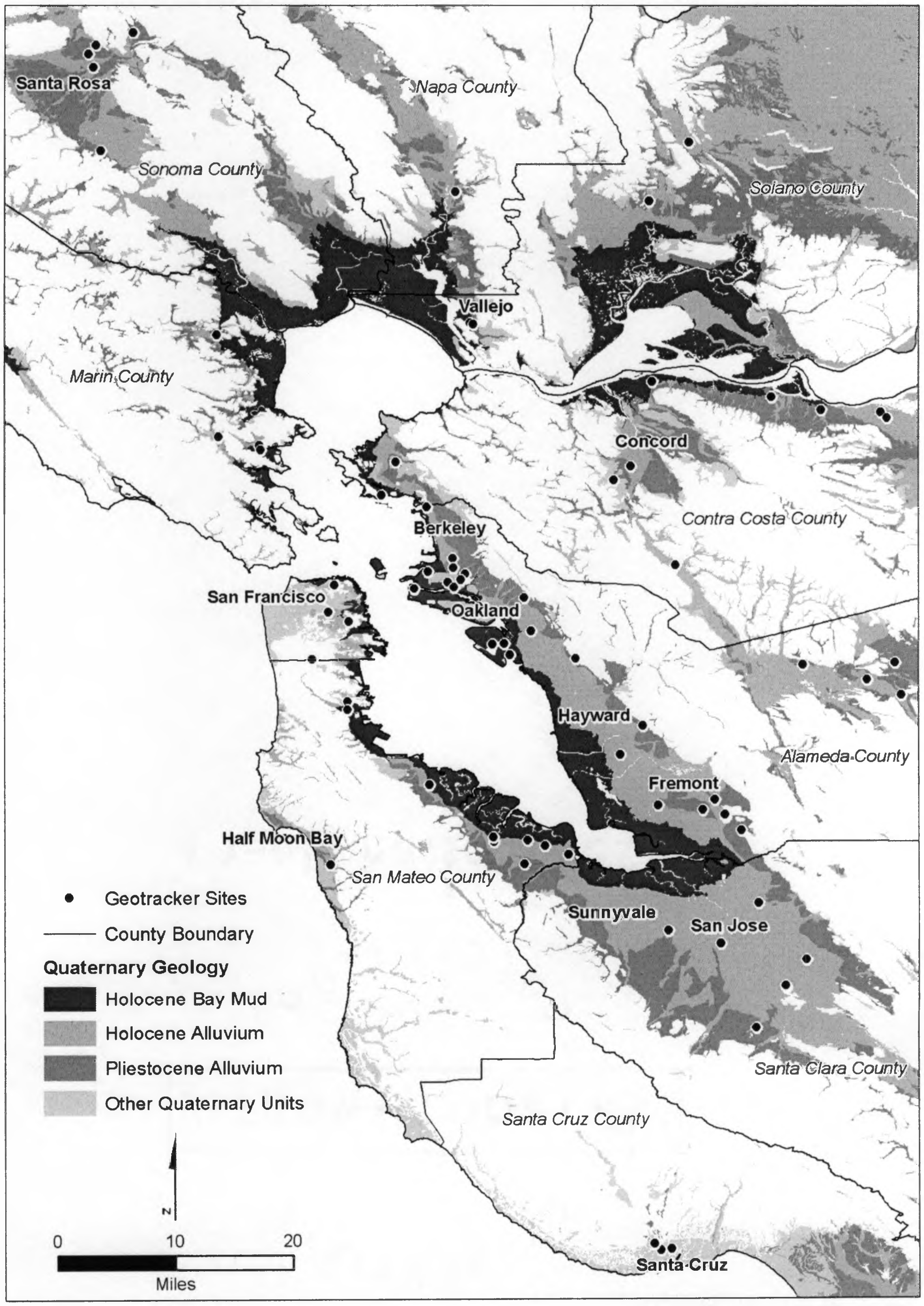
A summary of the 77 sites by county and soil type is provided in Table 4 and illustrated in Figure 4. All of the sites are located within urbanized portions of the Bay Area and consist primarily of industrial, military, transportation, and service commercial facilities, including numerous gas stations. A handful of sites consist of housing developments proposed on formerly industrial sites or residential properties. The sites are located across relatively flat Quaternary surficial deposits (Figure 4).

Table 4 – Summary of the 77 sites from the Geotracker database, sorted by county and soil type

	No. of Sites	No. of Samples
By County		
Alameda	30	745
Contra Costa	11	145
Marin	4	48
Napa	1	9
San Francisco	4	30
San Mateo	10	140
Santa Clara	5	154
Santa Cruz	3	81
Solano	4	48
Sonoma	5	54
By Soil Type		
Holocene Bay Mud	14	192
Holocene Alluvium	27	694
Pleistocene Alluvium	24	369
Other Quaternary Unit	11	190
Residual Soil, Franciscan Complex	1	9
TOTAL	**77**	**1,454**

Sites located within Pleistocene alluvium tend to be located on large alluvial fans extending out from the base of hills, whereas sites underlain by Holocene bay mud are located along the margins of the bay (Figure 4). Sites underlain by Holocene alluvium are generally found on large flat plains between the Pleistocene alluvial fans and bay muds (see Figures 3 & 4).

Figure 4 – Location of the 77 selected sites selected from the Geotracker database

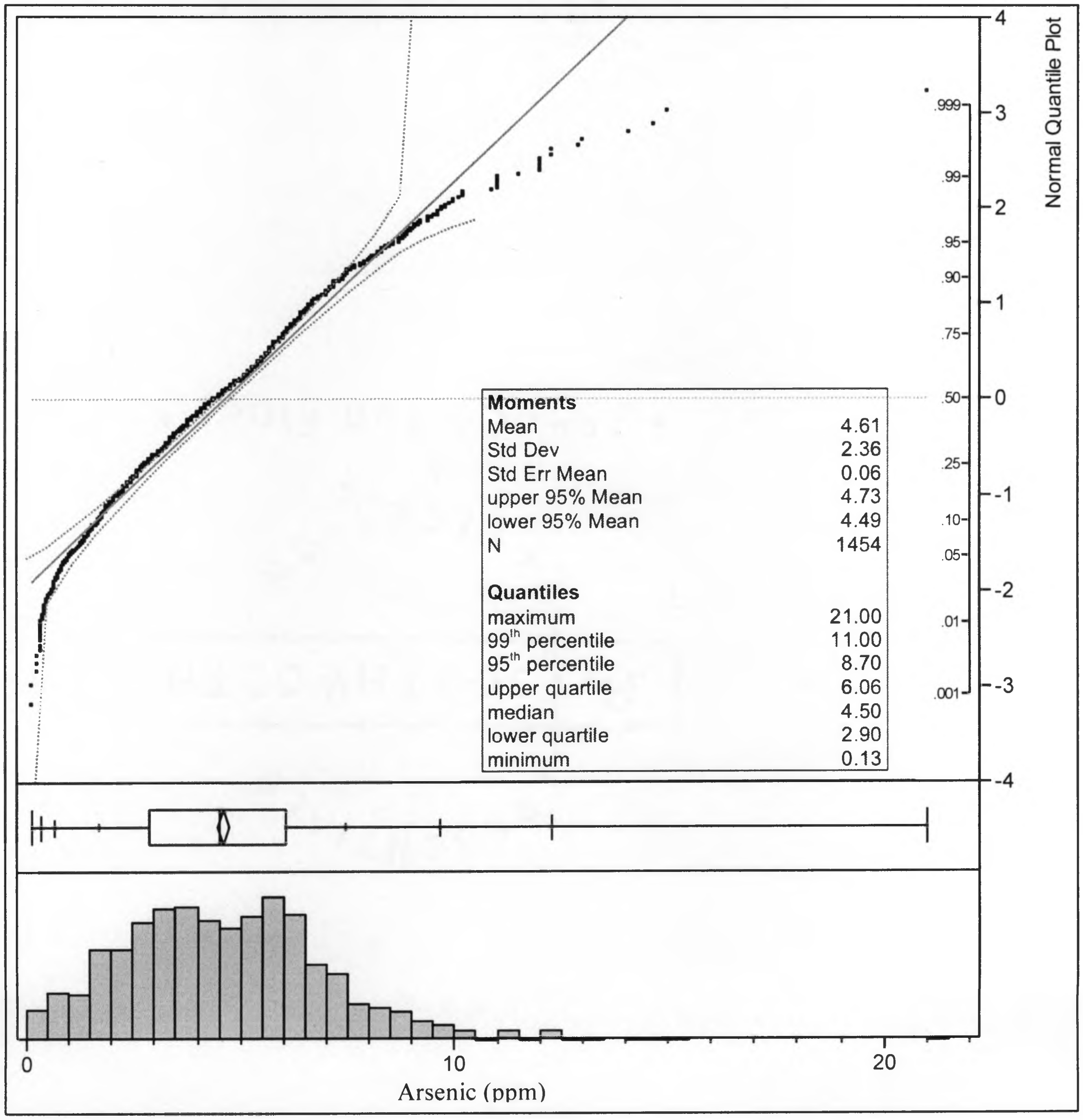


Adapted from Wentworth (1997)

The geographic distribution of the 77 sites is not uniform; rather, they are more concentrated within the more heavily urbanized portions of the Bay Area, particularly in Alameda County and other parts of the east bay (Figure 4). The relatively high density of sites in certain areas may indicate the general intensity of industrial and commercial development and the efficiency with which local enforcement agencies impose electronic reporting to Geotracker. Over 50 percent of the data in the database comes from Alameda County. Further, a relatively small number of sites make up a large fraction of the database—50 percent of the data comes from about 16 of the 77 sites. For the above reasons, the background dataset is biased both in terms of the number of samples per site and due to geographic clustering. Given the Bay Area has a developed land area of about 2,800 square kilometers (ABAG and MTC, 2011); the average density of sites is approximately one site per 36 square kilometers.

The database contains all arsenic data that is considered representative of background. Using the methods described in Section 3.3, the ROS method was used to replace 77 nondetects with estimated values, and 60 outliers were identified and eliminated. Figure 5 presents a histogram, quantile box plot, normal quantile plot, and summary statistics for the arsenic concentrations within the database. The data includes all samples from the 77 sites, thereby skewing the distribution pattern and overall summary statistics in favor of sites with a high number of samples, and combining multiple background populations into one distribution. As such, the visual analysis of the histogram and probability plot are unlikely to point to regional-scale influences on arsenic concentrations such as geography or geologic unit. In addition, the right tail of the probability plot shows several data points that might be interpreted as outliers; however, these did not meet the criteria for excluding outliers discussed in Section 3.3.

Figure 5 – Distribution of arsenic concentrations, including a histogram, a quantile box plot, and a normal quantile plot



Summary statistics of arsenic concentration in the database include a mean of 4.61 ppm, median of 4.50 ppm mg/kg, standard deviation of 2.36 mg/kg; and an IQR of 3.16 mg/kg (Figure 5). The concave shape of the normal quantile plot indicates that arsenic concentrations are not normally distributed. Based on the Lilliefors tests in ProUCL, the arsenic concentration data do not follow a discernable distribution and thus non-parametric methods are used when comparing groups (i.e., sample depth) within the database. The 95th percentile is commonly used to best represent the anthropogenic background for arsenic (EPA, 2002; NAVFAC, 2002). Based on this data, the upper estimate of arsenic concentrations considered as background is **11 mg/kg**.

4.2 Summary of Background Arsenic Concentrations by Site

Appendix A lists the 77 sites selected from Geotracker, their location, the soil type and basic summary statistics, including quantiles. Figure 6 and Figure 7 include a normal quantile plot, a quantile box plot, and a histogram of median values from the 77 sites first in original values (Figure 6), and as log-transformed data (Figure 7). The Lilliefors test was used on both distributions to test the null hypothesis that the data come from a normally (or log-normally) distributed population. The test, which used untransformed data to test both the normal and lognormal model, failed to reject the null hypothesis in either case, indicating the data can be characterized as being normally or log-normally distributed. The correlation coefficients (R) for both tests were nearly identical—R values were 0.981 and 0.983 for the normal and the log-normal data, respectively; however, the lognormal distribution has a better visual fit to the data. Site medians range from 0.61 mg/kg to 11 mg/kg, and the data display a positively skewed distribution. The mean of the dataset is 4.23 mg/kg, the median is 3.9 mg/kg, and additional summary statistics are shown in the box in Figure 6.

Figure 6 – Normal quantile plot, box plot, and histogram of site medians

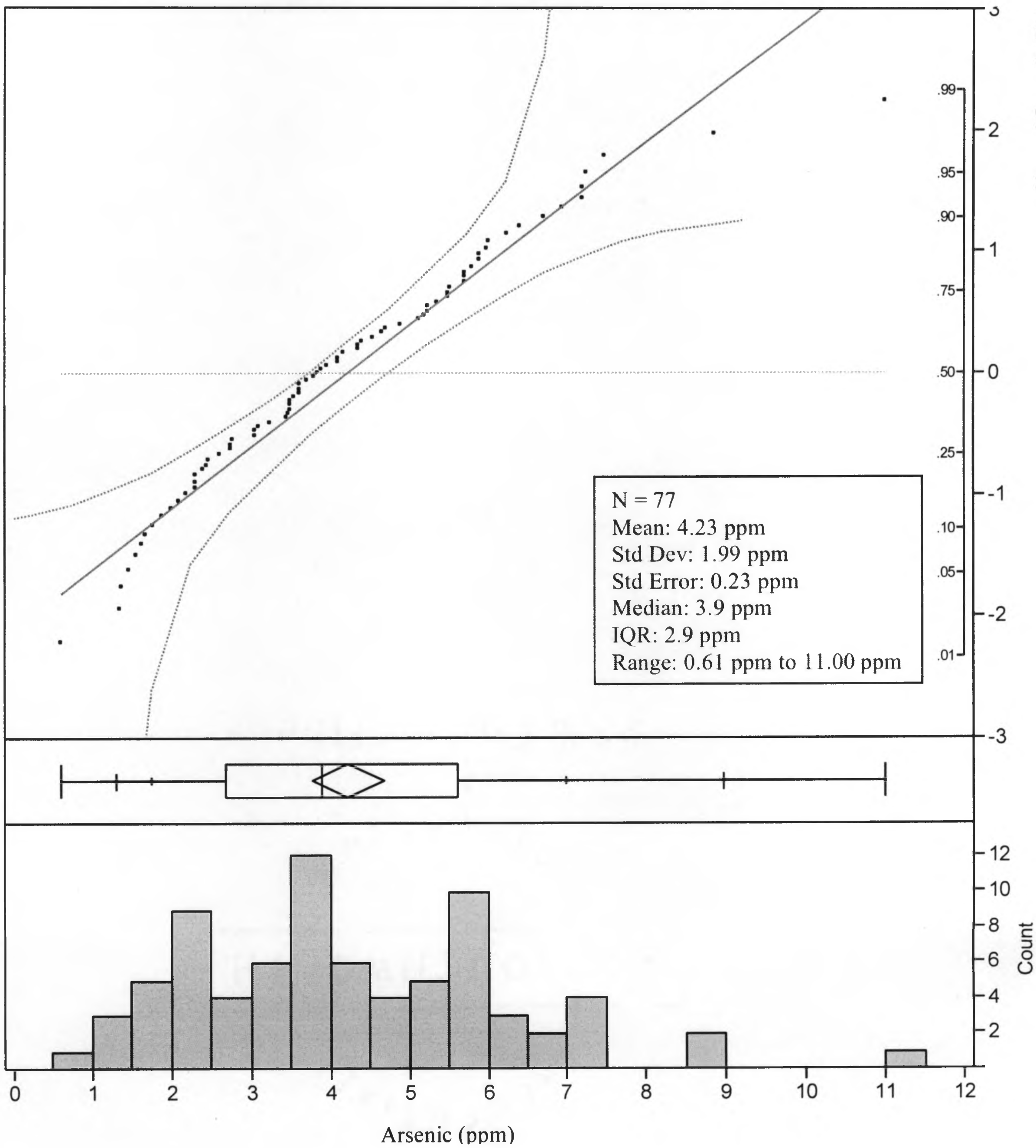
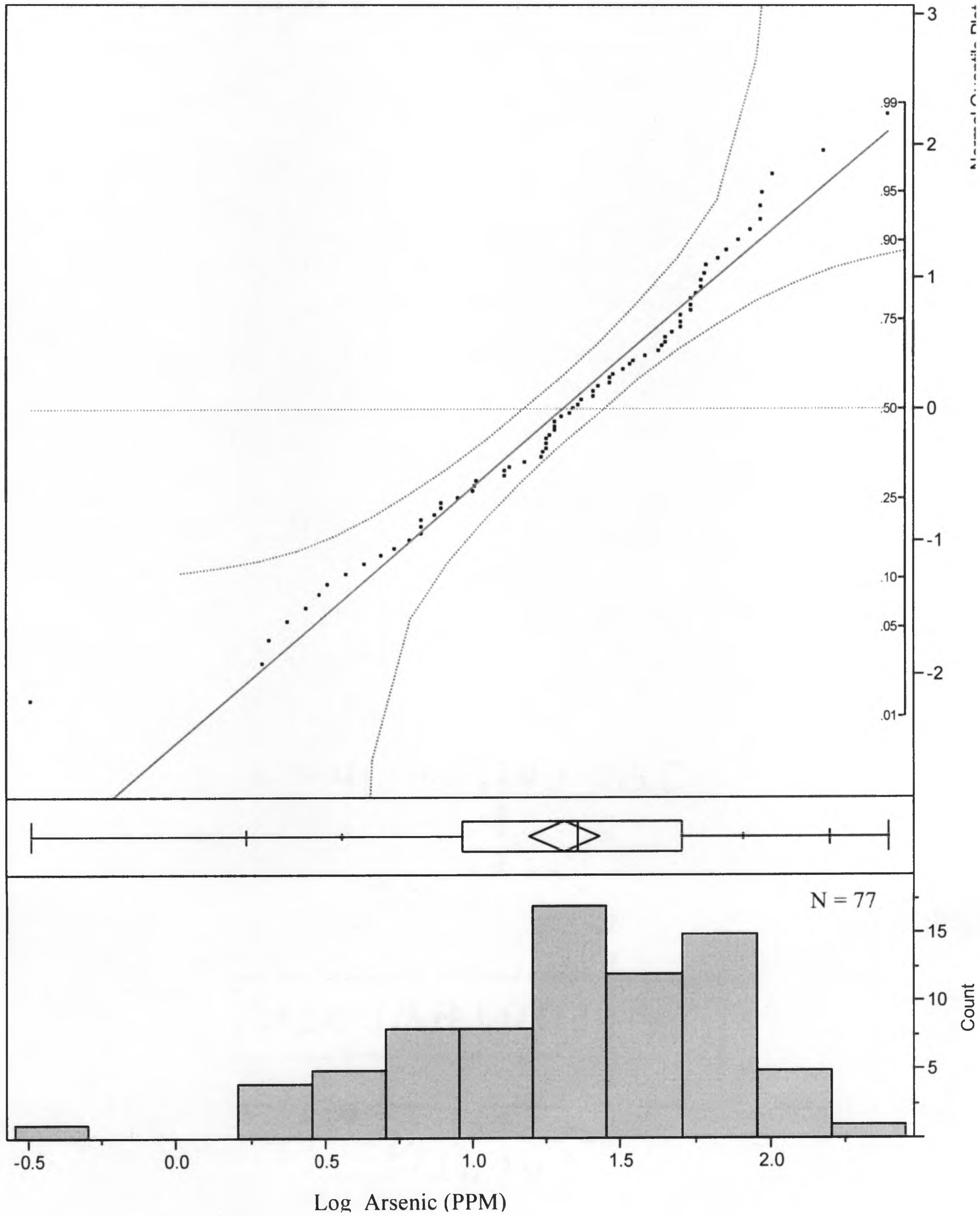


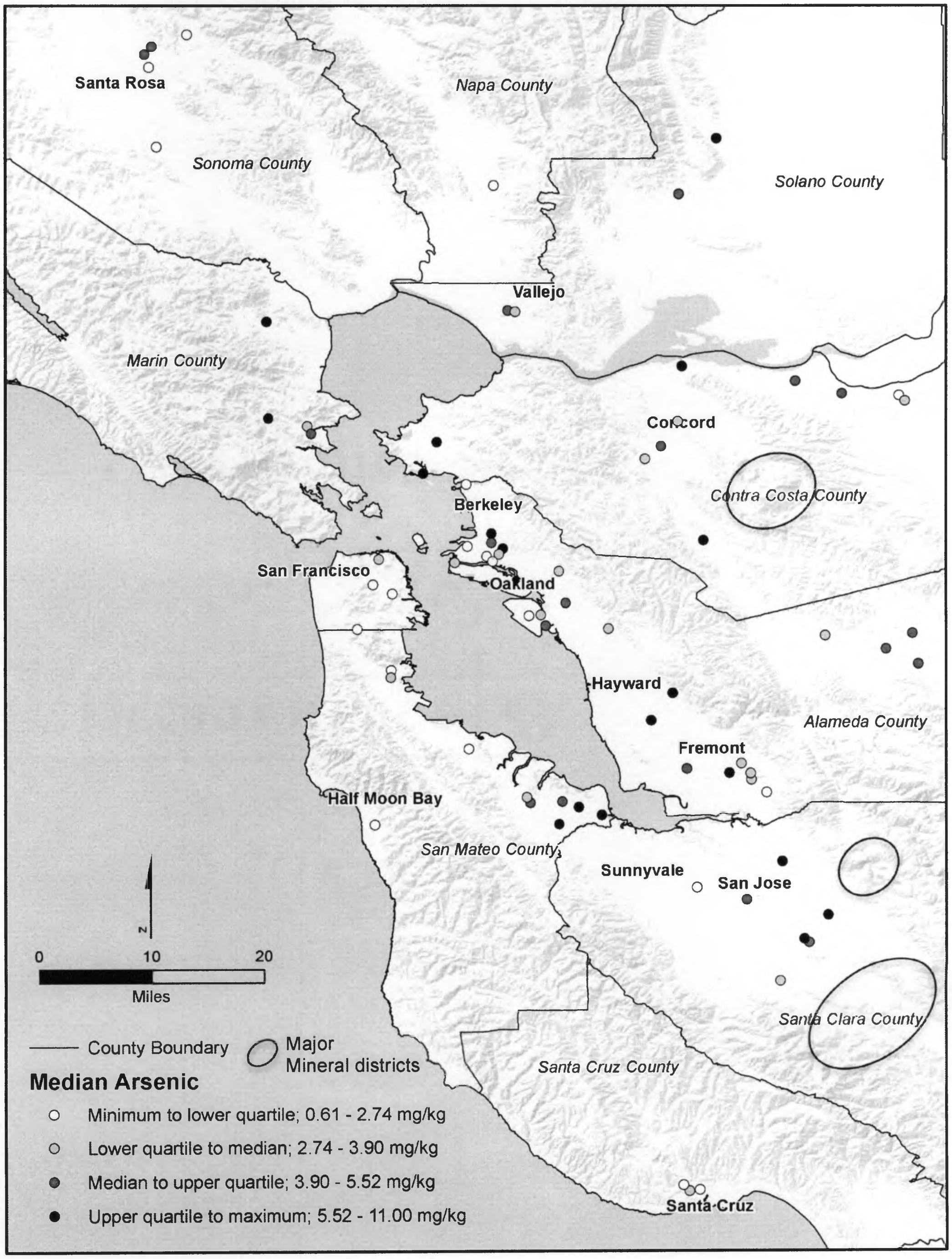
Figure 7 – Normal quantile plot, box plot, and histogram of the natural log of site medians



Three sites contributed to a large portion of the spread in the data. On the low end, site no. 22 has a mean of 0.8 mg/kg and 75 percent of the data, including the median is below the reporting limit (Appendix A). The detected data for site no. 22 ranged from 1.0 mg/kg to 2.0 mg/kg. Review of the site investigation report for the property did not indicate a reason (such as laboratory analysis problems or site-related sources of arsenic contamination) that the site should be excluded from the dataset. On the high end, site nos. 18 and 25 have median arsenic concentrations of 11.0 mg/kg and 8.9 mg/kg, respectively. The samples collected from both sites were deep (>1.8 meters bgs) and the site investigation reports contained no evidence to indicate site related arsenic contamination has occurred. As such, the sites were not eliminated from the dataset as outliers.

Figure 8 illustrates the spatial pattern of median arsenic concentrations by site across the study area. The 77 sites are colored by value, with white and black dots representing the lowest fourth and highest fourth of median values, respectively. Median values for sites in the northern San Francisco Peninsula and along the Pacific coastline appear to be consistently on the low end of the range, whereas sites on the high end of the range do not appear to dominate a single geographic region. Certain areas, such as central Marin County, the City of Hayward, the east side of San Jose, and the west end of Contra Costa County, have sites with high median arsenic concentrations. However, a clear geographic pattern cannot be discerned due to the low spatial density of the data in those places. Areas where the spatial density of sampling sites is high, such as the Berkeley, Oakland, Fremont, and north of Sunnyvale areas, have median arsenic concentrations that vary from the low to the high end of the range within relatively short distances. Figure 8 also shows general regions where rocks have been historically mined for Mercury.

Figure 8 – Map of median arsenic by site



To evaluate whether the spatial pattern expressed is clustered, dispersed, or random, the spatial autocorrelation tool in ArcMap 9.2 was used to calculate the Moran's I Index value and a Z score. A Moran's I value near +1.0 indicates clustering while a value near –1.0 indicates dispersion. The Z score value indicates whether or not the null hypothsis that there is no spatial clustering can be rejected. The Moran's I Index for the site medians is 0.18 and the Z score is 1.2 standard deviations. These scores confirm the visual observation that while somewhat clustered, the observed pattern of median arsenic concentrations may be due to random chance.

4.3 Arsenic Concentrations by Sample Depth

Because releases of arsenic are most likely to occur above ground (NAVFAC, 2002), sample depths (bgs) were classified as surface or subsurface as described in Section 3.2 to evaluate whether arsenic contamination within surface soils should be suspected. Soil samples were excluded from this analysis if the sample depth was not reported or if a composite depth was reported, which represents a range of depth rather than discrete depth. To compare depth classes, a univariate plot of arsenic concentrations for surface samples vs. subsurface samples was examined (Figure 9). Arsenic concentrations in the database plotted on a normal quantile plot indicate the distribution pattern is non-parametric in nature (Figure 5). As such, the non-parametric Wilcoxon test was used to determine if the group medians are significantly different. The test resulted in a significance probability (probability > |z|) of 0.40. Because the observed significance probability is not less than 0.05, there is no significant difference between surface and subsurface soil concentrations at the 95% confidence level. A univariate plot of arsenic concentrations for shallow (≤ 1.8 meters) vs. deep (> 1.8 meters) samples was created (Figure 10), and the non-parametric Wilcoxon test resulted in a significance probability

Figure 9 – Univariate plot of arsenic concentrations vs. sampling depth (surface and subsurface)

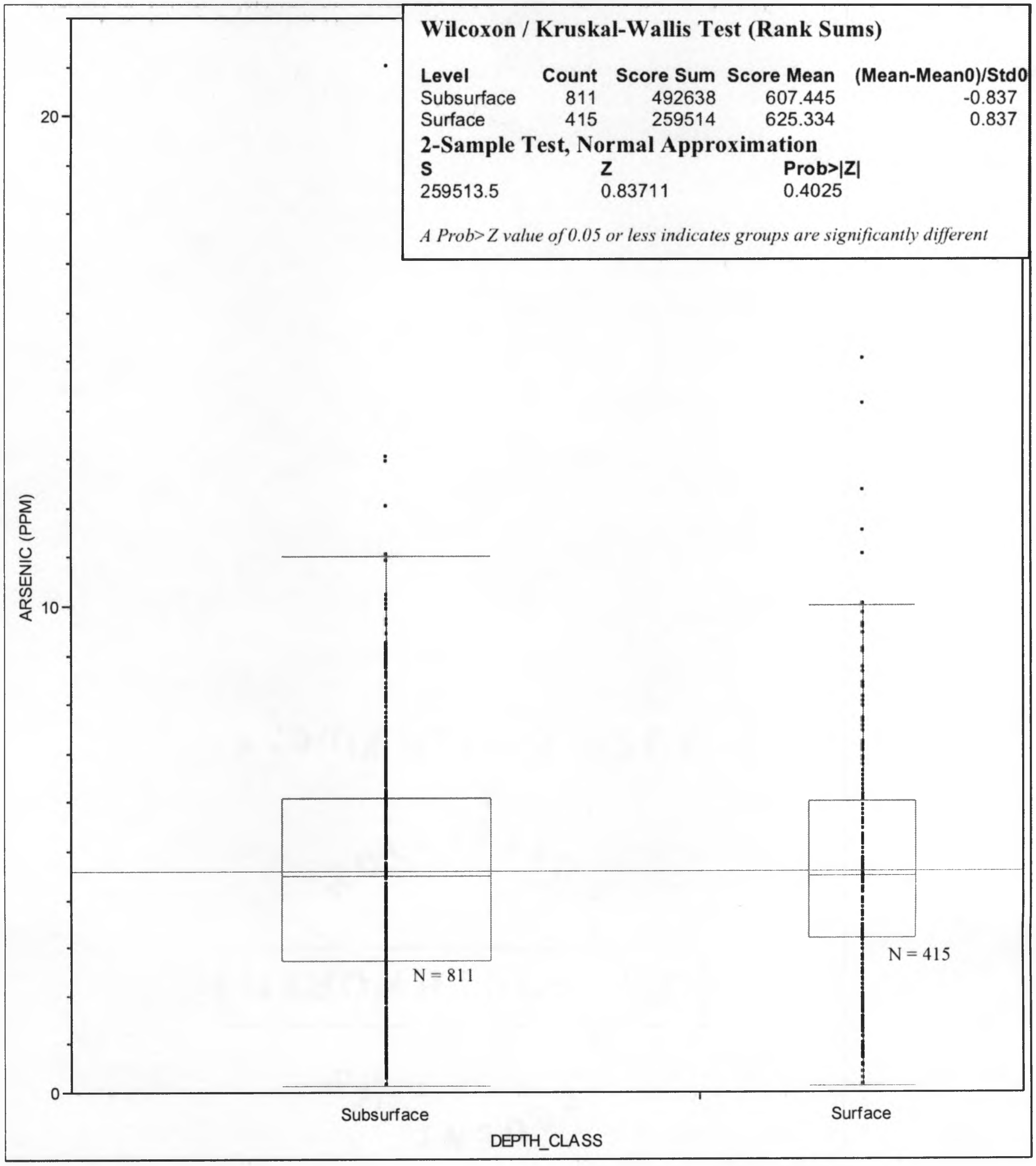
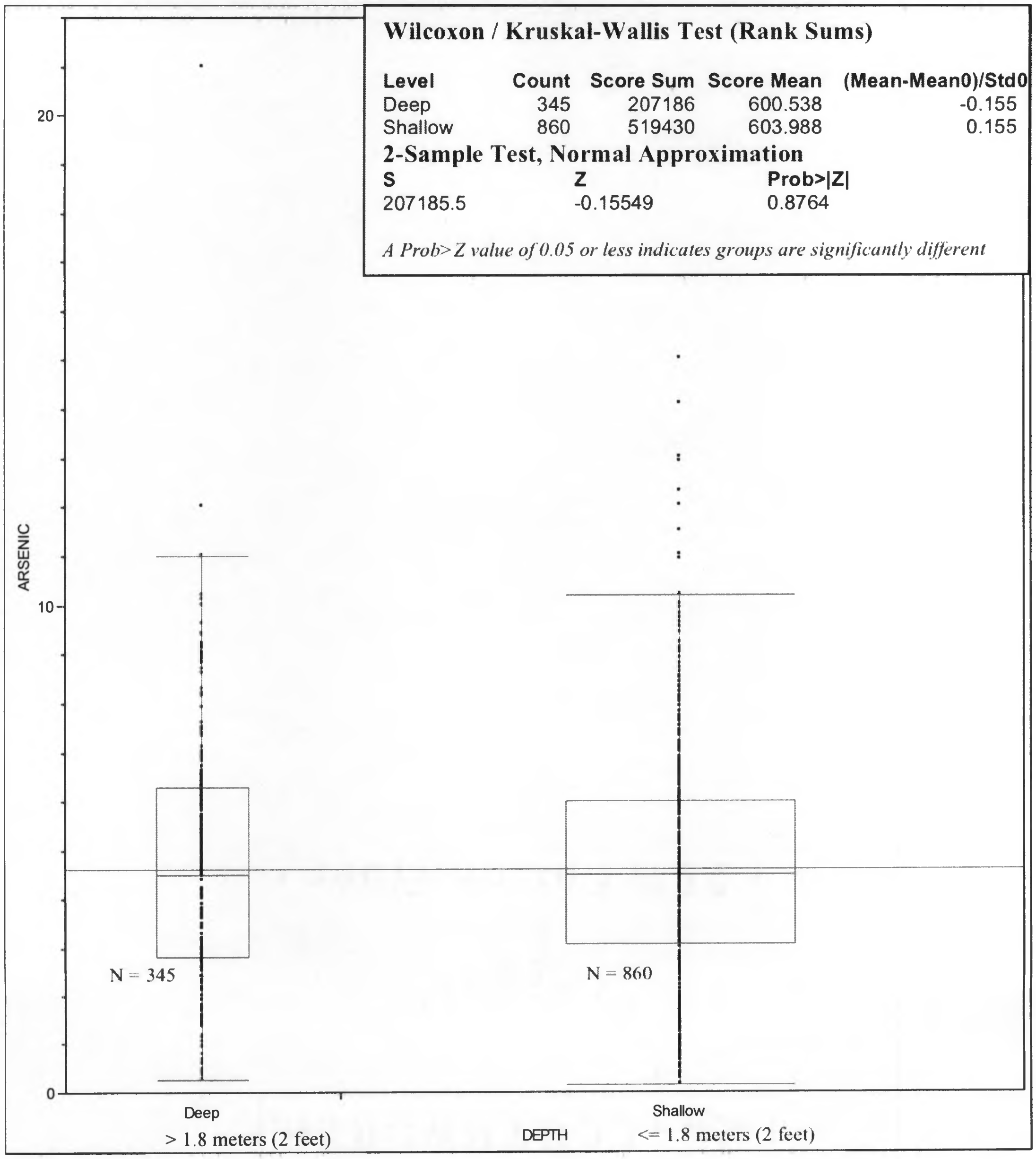


Figure 10 - Univariate plot of arsenic concentrations vs. sampling depth (shallow and deep)



(probability > |z|) of 0.88, likewise indicating there is no significant difference in arsenic concentrations between shallow and deep soil samples (95% confidence level).

4.4 Statistics by Soil Type and Significance Tests

The 77 sites are underlain by several geologic units, including Holocene bay mud, Holocene alluvium, Pleistocene alluvium, other Quaternary units, and Franciscan Complex bedrock (Table 4). Because only one site is located within the Franciscan Complex, it was excluded from this analysis. Eleven sites are underlain by several different Quaternary units that are not alluvial in origin. These sites were either underlain by dune sands, coastal/marine terrace deposits, or colluvium, and were grouped together as one category. Table 5 lists summary statistics for arsenic concentrations by soil unit. An ANOVA was performed to test whether grouping by soil type can explain some of the variation in background arsenic concentrations. An ANOVA was considered appropriate because the site medians follow a normal distribution, and because variances are equal. The F Ratio obtained from the ANOVA (3.85) indicates that the model fits the data at a 95% confidence level (probability > F is 0.013), and that group means are statistically different from the overall response mean.

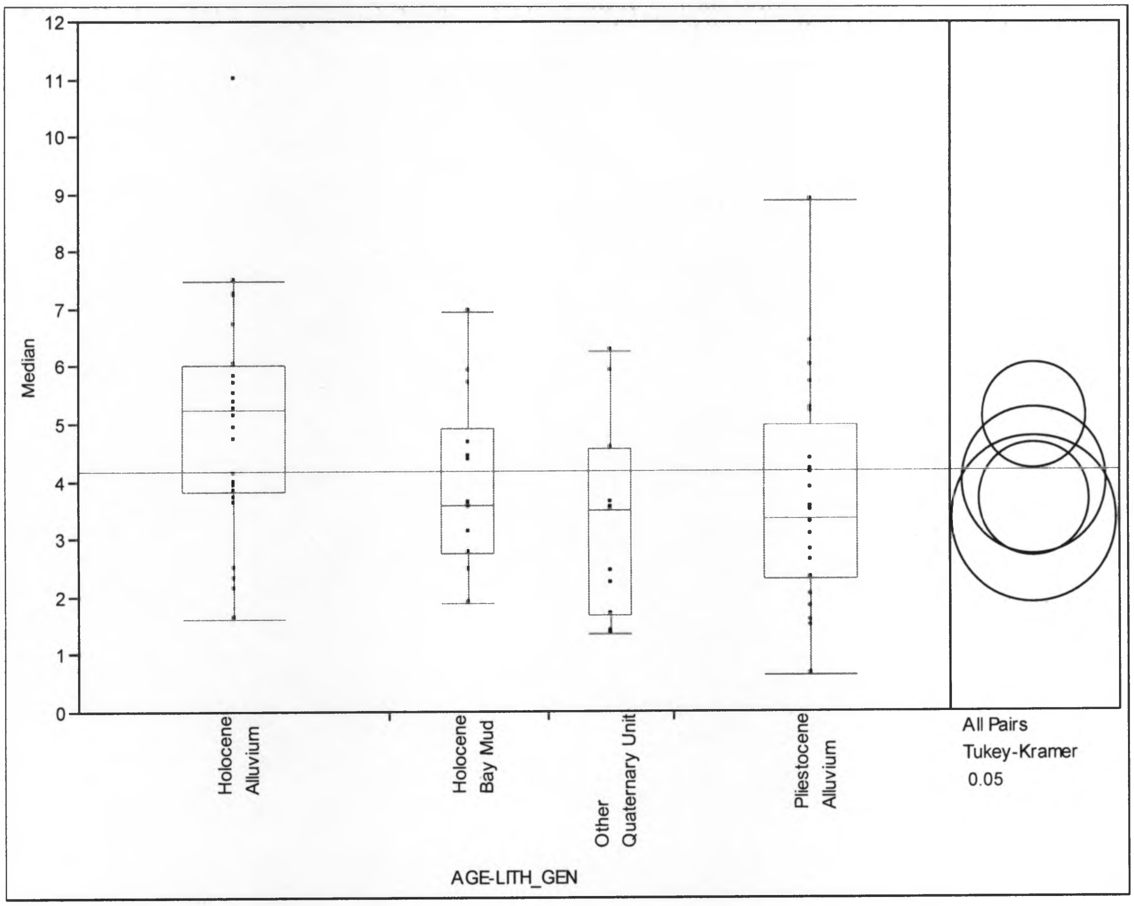
Table 5 – Statistics by soil unit and means comparison using Tukey test

Soil Type	Number	Mean	Min (mg/kg)	Median (mg/kg)	Max (mg/kg)	Tukey test*
Holocene alluvium	27	5.10	1.62	5.25	11	A
Holocene bay mud	14	3.97	1.89	3.58	6.94	A B
Pleistocene alluvium	24	3.65	0.61	3.35	8.86	B
other Quaternary unit	11	3.30	1.34	3.47	6.25	B

* *Levels not connected by same letter are significantly different*

To make multiple comparisons between soil types, a Tukey test was performed, as shown in Figure 11 and Table 5. The comparison circles plot on the right side of Figure 10 is a visual representation of group mean comparisons. Circles for means that are significantly different either do not intersect or intersect slightly so that the outside angle of intersection is less than 90 degrees. If the circles intersect by an angle of more than 90 degrees or if they are nested, the means are not significantly different. Group means for Holocene alluvium, Holocene bay mud, Pleistocene alluvium, and other Quaternary units were 5.10 mg/kg, 3.97 mg/kg, 3.65 mg/kg, and 3.30 mg/kg respectively. According the Tukey test, Holocene alluvium has a group mean that is significantly higher than both Pleistocene alluvium and other Quaternary units, but there is no significant difference between Holocene bay mud and any other unit. In addition, there is also no statistically significant difference between group means of Pleistocene alluvium and other Quaternary units. The R^2 value of the ANOVA model is 0.14, indicating the groupings explains 14 percent of the overall variability of the sample group.

Figure 11 – Distribution of arsenic concentrations by soil type and Tukey-Kramer HSD comparison of means



5.0 DISCUSSION AND CONCLUSION

5.1 Summary of Results

The results of this study indicate that background arsenic concentrations within the urbanized San Francisco Bay Region are lower than many of the estimates found in the literature, and are only weakly correlated with the underlying Quaternary geologic unit. Based on the data screened from Geotracker, the mean and upper estimate (the 99th percentile) for the regional background level of arsenic is 4.61 mg/kg and 11.00 mg/kg, respectively. In increasing order, the mean concentration of site medians grouped by soil type are 3.30 mg/kg ("other" Quaternary units), 3.65 mg/kg (Pleistocene alluvium), 3.97 mg/kg (Holocene bay mud), and 5.10 mg/kg (Holocene alluvium). Arsenic concentrations within Holocene alluvium were found to be statistically greater than Pleistocene alluvium and "other" Quaternary units; but no statistically significant difference was found between Holocene Bay Mud, Pleistocene alluvium, and "other" Quaternary units. The ANOVA and Tukey test revealed that the differences between group means are not pronounced, accounting for only 14 percent of the variation in median values across the 77 sites included in this study.

While it was anticipated that non-point anthropogenic sources of arsenic might result in higher concentrations of arsenic within surface samples, there was no statistically significant difference found between surface and subsurface soil samples. Further accounting for the soil mixing and reworking that takes place in urban settings, there was likewise no statistically significant difference found between shallow ($\leq$ 1.8 meters) and deep ($>$ 1.8 meters) soil samples. These results provide further evidence that the anthropogenic influence on the sample sites as it relates to arsenic is minimal.

The relative differences in mean arsenic concentrations across the four soil types, despite being subtle, correlate well with the general expectation that finer grained soils would result in elevated arsenic concentrations relative to coarse grained soils (NAVFAC, 2002; Alloway, 1990). Helley et al. (1979) explains that Pleistocene alluvium, which extends out from the base of hills in the Bay Area, is generally a coarser-grained unit than Holocene alluvium on the bay plains. Further, the "other" Quaternary units—predominantly composed of dune sands, colluvium, and shallow marine terrace deposits—had the lowest mean arsenic concentration. These "other" units are generally clean sandy units, and/or coarse-grained as a result of their depositional environment. This is generally consistent with finding made by Gustavvson et al. (2001), who associated the Nebraska Sand Hills, the largest dune field in the Western Hemisphere with low concentrations of arsenic.

5.2 Other Potential Sources of Variability in the Regional Background

As explained in Chapter 2, there are a number of other factors besides Quaternary soil type that likely contribute to regional variability in background arsenic concentrations, including the geologic source material for the Quaternary soils, the anthropogenic background, and/or ongoing geochemical processes (e.g., weathering, leaching, or enrichment). Localized areas in the hills and mountains of the region may produce unusually high concentrations of arsenic due to favorable geologic environments such as ore deposits (i.e. former mercury mines) and presence of organic-rich shales or coal. It is reasonable to expect that depositional settings sourced from these regions may result in naturally elevated concentrations of arsenic within Quaternary-age sediments.

However, there are several limitations, both in this study's dataset and in the existing geologic environment, that limit the ability to test this idea. Firstly, there are inherent difficulties in associating alluvial soils to specific bedrock sources, especially when the watershed is large and geologically complex. The influence of arsenic-rich geologic environments would become decreasingly detectable as a greater portion of the watershed is underlain by other bedrock units (i.e., distance from source). Secondly, while the general locations of former mercury and coal mines are known, shale and/or mudstones often occur in repeating sequences along with other sedimentary lithologies (e.g. sandstone). These sequences are frequently mapped together in the same formation, making it difficult to reliably estimate the portion of the watershed underlain by a specific lithology. Lastly, rather than being concentrated in one geographic location, shale-rich lithologies are fairly widespread throughout the Bay Area, making it unlikely a clear geographic pattern would be detected.

These limitations, along with the low geographic resolution of the data make correlations of high arsenic concentrations within Quaternary soils to specific source rocks speculative at best. If there were a strong source-rock influence on arsenic concentrations in Quaternary soils, it would be expected that sites with high arsenic values would be clustered and coincident with similar source regions. As discussed in Section 4.2 and shown in Figure 8, while the data in this study appears slightly clustered, it is also possible that it is the result of random chance. The lack of evidence for strong clustering or a striking geographic pattern may have more to do with the geographically sparse nature of the dataset than the absence of a source rock influence. The two general observations of 1) low median values along the northern end of the San Francisco Peninsula and the San Mateo and Santa Cruz County coastlines, and 2) high values in the eastern and southern Bay Area beg for a geologic explanation. Further study aimed at greater understanding of the relationship between arsenic concentrations found in Quaternary soils, and the geologic characteristics of their source regions would be valuable in further explaining natural variability in arsenic, and could possibly lead to the development of predictive tools.

An additional consideration which might influence the regional variability in background arsenic concentrations is broad land-use patterns and associated non-point sources of anthropogenic arsenic. It is important to recognize that regional land-use patterns often coincide with major changes in the underlying soil type. As illustrated in Figure 3, Holocene alluvium and bay muds underlie some of the most intensely developed urban and industrial areas, whereas Pleistocene alluvium more often underlies low-density residential areas. In addition, Holocene alluvium commonly supports prime agricultural soils and is likely to have supported agricultural uses prior to urban development, particularly in the eastern and southern Bay Area. Despite findings of no significant difference between surface and subsurface samples, it is difficult to fully dismiss the possibility that higher arsenic concentrations within Holocene alluvium are associated with concurrent variations in the anthropogenic background (e.g., the general type, intensity, and history of land development).

Due to its considerably greater age, it is also possible that Pleistocene alluvium in the study area was derived from different source rocks, or that geochemical processes that remove arsenic from alluvial soils have had a longer time to take place. It should also be recognized that the datasets for two of the geologic groups are small (n=11 for bay mud, n=14 for "other" Quaternary units), so the differences may also reflect a lack of a representative dataset. While a statistical correlation was identified between soil type and arsenic concentration, the actual processes governing those relations remain elusive.

5.3 Comparison of Findings with Other Background Studies

Despite the difficulties in clearly explaining sources of variation in the background dataset, statistics derived from the database provide defensible global estimates for background concentrations of arsenic within the flatland deposits of the Bay Area. The screening criteria avoided obvious sources of contamination and the

Geotracker database by nature contains primarily sites where fuels and organic solvents are the primary contaminant of potential concern. As discussed in Section 3.1, there is little to no correlation between metal and organic compound distributions, and the presence of organic co-contaminants has no effect on metal concentration background ranges.

The location and type of sample sites in this study is especially appropriate given they are representative of the geological and land-use settings where future environmental investigations are likely to be performed. A map of Geotracker site locations in the Bay Area instantly reveals that the vast majority of sites undergoing investigation and/or cleanup are located on urbanized flatland underlain by Quaternary-age geologic units. The commercial, industrial, institutional, and transportation-related land uses that are most often the subject of environmental investigations will continue to be predominantly located in such settings. Thus, the regional background estimates derived in this study may actually be more appropriate than background estimates derived from a pristine natural area, particularly if derived from bedrock units that naturally have anomalously high levels of arsenic.

The mean of 4.61 mg/kg and the proposed upper estimate of 11 mg/kg for the regional background concentration of arsenic found in this study are noticeably lower than upper limits from several other background studies of various geographic scope and scale (see Sections 1.3 and 1.5). The most obvious difference is with the background threshold value of 42 mg/kg for the Great Valley Sequence and 24 mg/kg for the "other" bedrock discussed in the LBNL (2002) study. The approach to screening sites/samples and the number of samples for this study was comparable to the LBNL study, though the approach to identifying outliers differed. The LBNL study used a uniform criterion of 50 mg/kg to eliminate outliers, whereas this study performed a site-by-site evaluation of outliers based on individual site distributions (the lowest value identified as an outlier, for example, was 7.1 mg/kg). While this difference in approach may have resulted in some of the disparity between background estimates, it is not sufficient to account for the substantially higher background threshold value found in the Berkeley hills.

In the LBNL case, the difference can be reasonably explained by differences in the geological setting. The LBNL area is underlain by tertiary-age sedimentary rocks (the Moraga and Orinda Formations) which are highly variable in their lithology, but commonly contain repetitious layers of shale and sandstone. The Great Valley Sequence in the area is mapped as the Claremont Shale of the Monterey Group, which is a fine-grained organic-rich shale and mudstone formation. As discussed in Section 1.3 (see Table 1), there is general consensus in the literature that shales and fine-grained soils tend to have naturally higher levels of arsenic than other types of rocks. The fact that the LBNL found a significant difference between different bedrock types on-site, and their finding of relatively high background threshold values support this notion. The national study by Shacklette and Boerngen (1984) also reported noticeably higher background arsenic concentrations (7.2, <RL to 93 mg/kg), although this isn't unexpected based on the coarse scale of the study.

Several studies in flatland geologic environments found similar or lower background levels than reported in this study. Scott's (1994) study area was located in an urban portion of the Santa Clara Valley underlain by Quaternary alluvium and bay muds, and she found a lower mean background concentration of arsenic, although a similar range of values (2.28, <DL to 20 mg/kg). The study area for the UCR study focused on alluvial/agricultural soils—primarily in the central valley—and found a comparatively lower mean and range of arsenic concentrations (3.5, <RL – 11 mg/kg) (UCR, 1996).

The findings of this thesis suggest that the most commonly cited background estimates of arsenic—namely from LBNL (2002) and Shacklette and Boerngen (1984)—are too high and do not represent the flatland soils of the Bay Area. Based on the findings herein, ESA's performed in the urbanized Bay Area should not automatically conclude that arsenic detections are representative of background so long as they are within the ranges found in the prior literature. Future metals analyses in areas underlain by flatland soils of the Bay Area should carefully examine arsenic detections in exceedance of 11 mg/kg as possibly exceeding background levels. In such cases, additional tests (such as the Wilcoxon ranked-sum test) should be performed to answer with a set confidence level whether the data exceeding background truly represents a different population. If no other defensible geological or geochemical reason for the high concentrations is provided, then site related contamination should be suspected. The background threshold value of 11 mg/kg is nearly double the 99th percentile value of 6 mg/kg found in a recent arsenic background study of the urbanized flatlands of the Los Angeles area for LA Unified School District sites (CalEPA, 2005). The Los Angeles basin is surrounded by large granitic mountain ranges that are geologically distinct than those in the Bay Area, producing much sandier flatlands than the watersheds of the Bay Area. Thus, the lower value is reasonable from a geologic standpoint, given that arsenic is thought to be elevated in mudstones and shales.

The method used herein to obtain, compile, and analyze background data on arsenic can be repeated for a suite of other constituents of concern whose natural background level often exceeds risk-based screening thresholds. Examples include other naturally-occurring metals such as cadmium, selenium, or nickel, to name a few. The increasing accessibility of environmental data in multiple, easily queried formats presents opportunities to develop better background information. Publically available environmental databases such as Geotracker provide an efficient and cost-effective means of establishing defensible regional background estimates.

6.0 REFERENCES

Alloway, B.J. (1990). Heavy Metals in Soils. *Glasgow: Blackie and Son*, 339 pp.

Anderson, D. (1998), Natural Levels of Nickel, Selenium, and Arsenic in the Southern San Francisco Bay Area, M.S. Thesis, *San Jose State University, Department of Geology*, 15 pp.

Association of Bay Area Governments (ABAG) and Metropolitan Transportation Committee (MTC) (2011), *Bay Area Focused Growth Website*, available at http://www.bayareavision.org/bayarea/, accessed November 1, 2011.

ATSDR (2009), Toxicological Profile Information Sheets, U.S. Department of Health and Human Services, *Agency for Toxic Substances and Disease Registry*, [online] Available from: http://www.atsdr.cdc.gov/toxpro2.html, Accessed May 13, 2009.

Bailey, E.H., and Everhart, D.L., *Geology and Quicksilver Deposits of the New Almaden District, Santa Clara County, California*, U.S Geological Survey Professional Paper 360, 1964.

Bay Area Air Quality Management District (BAAQMD), 2011, *Rules and Regulations Website*, available at http://www.baaqmd.gov/?sc_itemid=D39A3015-453E-4A0D-9C76-6F7F4DA5AED5, accessed November 1, 2011.

California Environmental Protection Agency (CalEPA) (2005), *Final Report Background Metals at Los Angeles Unified School Sites – Arsenic*. California Department of Toxic Substances Control. California Environmental Protection Agency. June 6, 2005

California Geological Survey (CGS), 2008, *California's Geomorphic Provinces*, CGS Note 36, 2008.

Facchinelli, A., E. Sacchi, and L. Mallen (2001), Multivariate statistical and GIS-based approach to identify heavy metal sources in soils, *Environmental Pollution*, 114(3), 313-324, doi:10.1016/S0269-7491(00)00243-8.

Gustavsson, N., Bølviken, B., Smith D.B., and Severson R.C. (2001), Geochemical Landscapes of the Conterminous United States—New Map Presentations for 22 Elements, *U.S. Geological Survey Professional Paper 1648*, November, 2001.

Helley et al., (1979), *Flatland deposits of the San Francisco Bay region, California, their geology and engineering properties, and their importance to comprehensive planning*, U.S. Geological Survey, U.S. Govt. Print. Off., viii, 88 p. : ill., maps (3 fold. col. in pocket).

Helsel D.R. and Hirsch R.M. (2002), *Statistical Methods in Water Resources, Chapter 13: Methods for Data Below the Reporting Limit*, U.S. Geological Survey, Techniques of Water-Resources Investigations Book 4, Chapter A3, 2002, pp 357-376

The Interstate Technology & Regulatory Council (ITRC) (2005), *Examination of Risk-Based Screening Values and Approaches of Selected States*, White Paper, November 2005.

Kabata-Pendias, Alina, Henryk Pendias (1985), *Trace Elements in Soils and Plants*, CRC Press Inc. 313 pgs.

Lado, L. R., T. Hengl, and H. I. Reuter (2008), Heavy metals in European soils: A geostatistical analysis of the FOREGS Geochemical database, *Geoderma*, 148(2), 189-199, doi:10.1016/j.geoderma.2008.09.020.

Lawrence Berkeley National Laboratory (LBNL) and Parsons Engineering Science, Inc (2002), Analysis of Background Distributions of Metals in the Soil at Lawrence Berkeley National Laboratory, 27 pp.

Li, J., M. He, W. Han, and Y. Gu (in press), Analysis and assessment on heavy metal sources in the coastal soils developed from alluvial deposits using multivariate statistical methods, *Journal of Hazardous Materials*, In Press, Corrected Proof, doi:10.1016/j.jhazmat.2008.08.112. [online] Available from: http://www.sciencedirect.com/science/article/B6TGF-4TFW9FC-1/2/f3081f2c54b75fe7d53bb82e0e5f46e6, accessed February 20, 2009.

Mount Diablo Interpretive Association (2009), Guide to the Geology of Mount Diablo State Park, website http://www.mdia.org/Geology/geology%20guide.htm, accessed November 11, 2011. Last Updated April, 2009.

Naval Facilities Engineering Command (NAVFAC) (2002), Guidance for Environmental Background Analysis; Volume I: Soil, Guidance Document, prepared by *Battelle Memorial Institute, Earth Tech Inc., and NewFields Inc.*, UG-2049-ENV. 169 pp.

Office of Emergency and Remedial Response (2002), *Guidance for Comparing Background and Chemical Concentration in Soil for CERCLA Sites*, Guidance Document, US Environmental Protection Agency (EPA). 89 pp.

Pacific Gas & Electric Company (PG&E), 2011, *PG&E's Manufactured Gas Plant Program Website*, available online at http://www.pge.com/about/environment/taking-responsibility/mgp/, accessed October 28, 2011.

Rawlins, R. W. (2003), The influence of parent material on topsoil geochemistry in eastern England, *Earth Surface Processes and Landforms*, 28(13), 1389-1409.

SF Bay Region RWQCB (2008), Screening For Environmental Concerns at Sites with Contaminated Soil and Groundwater, *Technical Report, California Environmental Protection Agency, Regional Water Quality Control Board, San Francisco Bay Area Region*, September 2007. 319 pp.

Shacklette, H.T., and Boerngen, J.G. (1984), Element concentrations in soils and other surficial materials of the conterminous United States: *U.S. Geological Survey Professional Paper 1270*, 105 p.

State Water Resources Control Board (SWRCB) (2010), Office of Public Affairs Fact Sheet: Geotracker, April 2010. 2 pp.

UCR (1996), Background Concentrations of Trace and Major Elements in California Soils, *University of California (Riverside), Division of Agriculture and Natural Resources*, March 1996. 33 pp.

USGS, 2005. Mineral resources data system. US Geological Survey, Reston, VA, http://www.tin.er.usgs.gov/mrds/.

United States Geological Survey (USGS), 2009, Minerals Yearbook, Arsenic, Advanced Release, October 2010. 6 pp.

United States Geological Survey (USGS), 1940, *Quicksilver Deposits of the Mount Diablo District, Contra Costa County, California*, Strategic Minerals Investigations, Bulletin 944-B pp. 31-54.

U.S. Energy Information Administration (EIA), 2009, *Inventory of Electric Utility Power Plants in the United States*, database available at http://www.eia.gov/cneaf/electricity/ipp/ipp_sum.html, accessed November 5, 2011.

U.S. Environmental Protection Agency (EPA), 2011, *Cleanups In My Community Map*, available at http://iaspub.epa.gov/apex/cimc/f?p=255:63:3625198173186524, accessed November 1, 2011.

U.S. Environmental Protection Agency (EPA), 2002, *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites*. EPA/540/R-01/003-OSWER 9285.7-41, September, 2002.

Welch, A. H., Westjohn, D. B., Helsel, D. R., and Wanty, R. B. (2000), Arsenic in Groundwaters of the Unites States – Occurrence and Geochemistry, United States Geological Survey, *Groundwater, 38(4), pp. 589-604.*

Wentworth, C.M., *General Distribution of Geological Materials in the San Francisco Bay Region, California: A Digital Map Database*, based on the work of E.E. Brabb (1989), S.E. Ellen and C.M. Wentworth (1995), and E.J. Helley and K.R. Lajoie (1979), United States Geological Survey Open File Report 97-744, 1997.

Yesilonis, I., R. Pouyat, and N. Neerchal (2008), Spatial distribution of metals in soils in Baltimore, Maryland: Role of native parent material, proximity to major roads, housing age and screening guidelines, *Environmental Pollution*, *156*(3), 723-731, doi:10.1016/j.envpol.2008.06.010.

Zhang, C., and O. Selinus (1998), Statistics and GIS in environmental geochemistry -- some problems and solutions, *Journal of Geochemical Exploration*, 64(1-3), 339-354, doi:10.1016/S0375-6742(98)00048-X.

Zhao, F., S. McGrath, and G. Merrington (2007), Estimates of ambient background concentrations of trace metals in soils for risk assessment, *Environmental Pollution*, 148(1), 221-229, doi:10.1016/j.envpol.2006.10.041.

APPENDIX A

SUMMARY OF SOIL TYPE AND ARSENIC CONCENTRATIONS BY SITE

Site No.	Site Name	City	County	Soil Type	No. of Samples	Percent Above Reporting Limit	Mean (mg/kg)	St. Dev (mg/kg)	Min	Quartile, 25%	Median	Quartile, 75%	Max
1	2236 B NORTH TEXAS STREET	FAIRFIELD	Solano	Other Quaternary Unit	10	100%	4.3	0.8	3.0	3.3	4.6	5.0	5.3
2	Alameda Naval & Marine Corps Reserve Ctr. - Naval and Marine Corps Reserve Center, Alameda	Alameda	Alameda	Other Quaternary Unit	17	100%	6.9	2.7	2.4	5.5	5.9	8.6	12.0
3	Alameda Naval Air Station - Alameda NAS Bldg 594, Tank 594-1, 2	Alameda	Alameda	Holocene Bay Mud	46	100%	4.1	3.0	1.3	1.9	2.8	5.8	14.7
4	ARCADIA PARK	Oakland	Alameda	Holocene Alluvium	139	100%	5.4	1.6	0.6	4.2	5.5	6.4	9.4
5	Bay Division Pipeline	Fremont	Alameda	Pleistocene Alluvium	23	96%	2.3	0.6	<RL	1.7	2.3	2.7	3.3
6	BECK PROPERTY	PLEASANT HILL	Contra Costa	Other Quaternary Unit	7	100%	3.7	0.5	3.1	3.2	3.5	4.2	4.5
7	Bell Gas	Pittsburgh	Contra Costa	Pleistocene Alluvium	30	100%	4.4	1.8	1.2	3.2	4.1	5.8	8.3
8	BELTRAMO PROPERTY	MENLO PARK	San Mateo	Pleistocene Alluvium	9	100%	5.7	0.3	5.1	5.4	5.7	6.0	6.1
9	BP #11184 (FORMER)	San Francisco	San Francisco	Holocene Alluvium	11	100%	3.9	0.6	3.2	3.3	3.8	4.2	5.4
10	BP RICHMOND TERMINAL (formerly ARCO)	RICHMOND	Contra Costa	Holocene Bay Mud	14	100%	7.0	3.3	1.8	5.0	5.9	9.0	13.0
11	CALIFORNIA LINEN SUPPLY CO	Oakland	Alameda	Holocene Alluvium	81	100%	6.8	1.7	3.5	5.6	6.7	7.9	12.0
12	Call Mac Transportation	Livermore	Alameda	Pleistocene Alluvium	43	100%	5.4	1.7	0.8	4.1	5.2	6.6	9.4

Site No.	Site Name	City	County	Soil Type	No. of Samples	Percent Above Reporting Limit	Mean (mg/kg)	St. Dev (mg/kg)	Min	Quartile, 25%	Median	Quartile, 75%	Max
13	CALTRANS MAINTENANCE STATION	SOUTH SAN FRANCISCO	San Mateo	Pleistocene Alluvium	6	83%	2.4	0.9	<RL	1.9	2.6	3.0	3.4
14	CHEVRON	CONCORD	Contra Costa	Pleistocene Alluvium	9	78%	4.5	3.4	<RL	1.1	3.4	8.1	9.0
15	CHEVRON #9-0020	Oakland	Alameda	Other Quaternary Unit	9	100%	3.4	0.6	2.6	2.8	3.5	3.9	4.0
16	CHEVRON 9-1374	REDWOOD CITY	San Mateo	Holocene Bay Mud	5	100%	3.5	0.4	2.9	3.2	3.6	3.8	3.8
17	Chevron No 2510	Fremont	Alameda	Pleistocene Alluvium	11	100%	6.4	0.6	5.7	5.7	6.4	6.6	7.3
18	CHEVRON No. 1570	UNION CITY	Alameda	Holocene Alluvium	5	100%	11.6	5.8	5.5	7.1	11.0	16.5	21.0
19	Chrisp Company	Fremont	Alameda	Holocene Alluvium	11	100%	3.9	0.4	3.1	3.5	3.9	4.2	4.7
20	CHUNG PROPERTY / LANE METAL FINISHERS	OAKLAND	Alameda	Pleistocene Alluvium	16	100%	5.5	1.4	2.8	4.3	5.3	6.3	8.1
21	DANVILLE SQUARE SHOPPING CENTER	DANVILLE	Contra Costa	Holocene Alluvium	15	100%	5.8	0.6	4.8	5.2	6.0	6.4	6.7
22	DOWNEY PROPERTY	SANTA ROSA	Sonoma	Pleistocene Alluvium	16	25%	0.8	0.5	<RL	<RL	<RL	1.1	2.0
23	EXXON 7-4135 SM	SAN MATEO	San Mateo	Holocene Bay Mud	9	67%	2.4	1.4	<RL	<RL	1.9	3.6	4.8
24	FORMER CHEVRON SITE #301949 (9-7093)	RICHMOND	Contra Costa	Pleistocene Alluvium	6	100%	5.5	1.4	3.2	4.2	6.0	6.5	6.7
25	FORMER CHEVRON STATION # 21-3230	HAYWARD	Alameda	Pleistocene Alluvium	14	100%	8.3	2.0	5.1	7.0	8.9	10.2	11.0
26	Former Chevron-Mills Square Park	Livermore	Alameda	Holocene Alluvium	29	100%	5.3	1.2	2.2	4.5	5.4	6.2	7.6
27	FORMER MONTGOMERY WARDS SITE	PLEASANT HILL	Contra Costa	Holocene Alluvium	8	100%	5.6	0.3	5.3	5.3	5.5	6.0	6.1

Site No.	Site Name	City	County	Soil Type	No. of Samples	Percent Above Reporting Limit	Mean (mg/kg)	St. Dev (mg/kg)	Min	Quartile, 25%	Median	Quartile, 75%	Max
28	FORMER SHELL SERVICE STATION	SAN FRANCISCO	San Francisco	Holocene Alluvium	7	100%	2.3	0.6	1.5	1.9	2.3	3.0	3.2
29	Francis Plating	Oakland	Alameda	Other Quaternary Unit	19	100%	2.6	0.8	1.8	2.1	2.4	3.0	4.9
30	FREISMAN RANCH	LIVERMORE	Alameda	Pleistocene Alluvium	14	100%	4.2	2.6	2.5	2.6	3.1	4.9	10.0
31	GE IMATRON / CARAL MANUFACTURING	ALBANY	Alameda	Pleistocene Alluvium	38	100%	2.6	0.8	1.4	2.0	2.3	3.1	4.6
32	GOODYEAR TIRE AND RUBBER COMPANY	VALLEJO	Solano	Holocene Bay Mud	10	100%	5.1	1.2	3.9	4.1	4.7	6.0	7.4
33	HAVEN AVENUE INDUSTRIAL CONDOMINIUMS	MENLO PARK	San Mateo	Holocene Alluvium	17	82%	3.8	1.2	>RL	3.1	4.0	4.8	5.6
34	Jack London Square Area	Oakland	Alameda	Other Quaternary Unit	12	79%	2.1	2.0	0.1	0.7	1.9	4.2	6.3
35	KUNG PROPERTY	EAST PALO ALTO	San Mateo	Holocene Alluvium	7	86%	5.8	4.1	<RL	1.5	7.2	9.1	11.0
36	M. Toich and Sons	San Francisco	San Francisco	Holocene Alluvium	5	100%	1.9	0.4	1.4	1.5	2.1	2.1	2.2
37	Magnetics, Inc.	Sunnyvale	Alameda	Holocene Alluvium	41	46%	2.4	2.2	<RL	<RL	<RL	3.2	8.9
38	MAIN STREET & ARNOLD WAY	HALF MOON BAY	San Mateo	Pleistocene Alluvium	16	56%	1.9	1.3	<RL	<RL	1.6	2.5	4.3
39	MAZZEI AUTOMOBILE DEALERSHIP (FORMER)	ANTIOCH	Contra Costa	Holocene Alluvium	19	100%	5.2	1.5	3.2	4.1	4.9	6.3	8.8
40	Meikle Property	Santa Cruz	Santa Cruz	Other Quaternary Unit	16	50%	1.8	0.8	<RL	<RL	1.7	2.4	3.6
41	Milpitas Senior Housing Project	Milpitas	Alameda	Holocene Alluvium	61	97%	5.5	1.6	<RL	4.9	5.7	6.2	9.6

Site No.	Site Name	City	County	Soil Type	No. of Samples	Percent Above Reporting Limit	Mean (mg/kg)	St. Dev (mg/kg)	Min	Quartile, 25%	Median	Quartile, 75%	Max
42	Oakland Army Base - USTs 11A/12A/13A	OAKLAND	Alameda	Holocene Bay Mud	6	100%	2.4	0.3	1.8	2.2	2.5	2.6	2.7
43	Oakland International Airport	Oakland	Alameda	Holocene Bay Mud	8	100%	3.8	1.6	1.3	2.5	3.6	5.3	6.1
44	Oakland International Airport, S. Field Tank Farm	Oakland	Alameda	Holocene Bay Mud	31	100%	4.3	1.1	2.3	3.4	4.4	4.9	6.4
45	PACIFIC COAST TRANSPORTATION SERVICES	NEWARK	Alameda	Holocene Alluvium	8	100%	5.3	1.1	3.4	4.3	5.5	6.2	6.7
46	Parking Corporation of America	South San Francisco	San Mateo	Holocene Bay Mud	25	96%	3.5	2.0	<RL	1.9	3.1	4.7	9.7
47	PG&E ANTIOCH NATURAL GAS TERMINAL	OAKLEY	Contra Costa	Holocene Alluvium	9	100%	2.4	0.3	1.8	2.1	2.5	2.6	2.8
48	PGE Stone Substation	San Jose	Santa Clara	Holocene Alluvium	40	98%	5.5	1.9	<RL	4.2	5.1	6.7	9.1
49	Quality Tune-Up No. 6	San Jose	Santa Clara	Holocene Alluvium	10	100%	7.2	2.5	2.9	5.5	7.3	9.1	11.0
50	RAB MOTORS/CALTRANS	SAN RAFAEL	Marin	Holocene Bay Mud	6	100%	4.5	0.7	3.8	3.8	4.4	5.2	5.5
51	RAIN FOR RENT	OAKLEY	Contra Costa	Holocene Alluvium	21	100%	3.7	1.5	1.8	2.6	3.6	4.0	7.5
52	RUST PROPERTY	REDWOOD CITY	San Mateo	Holocene Alluvium	32	100%	5.0	2.0	1.2	3.4	5.3	6.3	9.6
53	Salz Leather Inc.	Santa Cruz	Santa Cruz	Other Quaternary Unit	8	88%	2.0	0.6	<RL	1.5	2.2	2.5	2.9
54	Santa Clara Former Maintenance	Santa Clara	Santa Clara	Holocene Alluvium	15	100%	4.1	1.1	1.8	3.4	4.1	4.9	5.9
55	Seeger Property	VACAVILLE	Solano	Holocene Alluvium	8	100%	7.6	0.9	6.2	7.0	7.5	8.2	9.1
56	SHELL	SANTA ROSA	Sonoma	Pleistocene Alluvium	6	100%	2.3	0.7	1.7	1.8	2.0	2.9	3.4

Site No.	Site Name	City	County	Soil Type	No. of Samples	Percent Above Reporting Limit	Mean (mg/kg)	St. Dev (mg/kg)	Min	Quartile, 25%	Median	Quartile, 75%	Max
57	SHELL #13-6019	San Leandro	Alameda	Pleistocene Alluvium	16	100%	5.8	4.7	1.4	1.9	2.8	11.0	14.1
58	SHELL / 7-ELEVEN #20009	Oakland	Alameda	Holocene Bay Mud	5	100%	6.8	0.7	5.8	6.1	6.9	7.4	7.7
59	Shell Equilon San Jose	San Jose	Santa Clara	Holocene Alluvium	14	100%	7.2	0.7	6.2	6.5	7.2	7.7	8.3
60	SHELL NAPA	NAPA	Napa	Pleistocene Alluvium	9	78%	2.2	1.5	<RL	<RL	1.8	4.0	4.4
61	SHELL NOVATO	NOVATO	Marin	Franciscan Complex	9	100%	7.7	3.0	3.4	4.3	8.6	9.8	12.0
62	SHELL SANTA ROSA	SANTA ROSA	Sonoma	Pleistocene Alluvium	13	92%	4.1	1.5	<RL	3.1	4.2	5.6	6.0
63	Shell Service Station	Cotati	Sonoma	Pleistocene Alluvium	9	100%	1.4	0.5	0.8	0.9	1.5	1.8	2.2
64	SHELL STATION	SANTA CRUZ	Santa Cruz	Other Quaternary Unit	57	89%	3.9	2.4	<RL	2.7	3.6	5.2	10.1
65	Shell Station #4003	San Francisco	San Francisco	Other Quaternary Unit	7	100%	1.4	0.3	0.9	1.3	1.3	1.7	1.7
66	Site A	Oakland	Alameda	Holocene Bay Mud	13	100%	2.9	1.2	0.4	2.1	2.7	3.8	5.8
67	SOUTHERN PACIFIC TRANSPORATION CO - FRANCES ST	SANTA ROSA	Sonoma	Pleistocene Alluvium	10	100%	4.5	1.0	3.0	3.7	4.4	5.3	6.1
68	Standard Oil Bulk Terminal	Fremont	Alameda	Pleistocene Alluvium	16	100%	3.2	0.6	1.9	2.8	3.3	3.4	4.5
69	TERMINAL AVE HOUSING DEVELOP.	MENLO PARK	San Mateo	Holocene Alluvium	14	100%	5.8	1.0	4.0	5.1	5.8	6.4	7.9
70	UNOCAL	SAN RAFAEL	Marin	Other Quaternary Unit	7	100%	4.4	2.8	4.6	5.7	6.3	6.8	8.3

Site No.	Site Name	City	County	Soil Type	No. of Samples	Percent Above Reporting Limit	Mean (mg/kg)	St. Dev (mg/kg)	Min	Quartile, 25%	Median	Quartile, 75%	Max
71	Unocal	San Anselmo	Marin	Holocene Bay Mud	26	100%	6.3	0.9	1.5	1.9	3.5	7.4	7.6
72	UNOCAL #4921	SAN JOSE	Santa Clara	Pleistocene Alluvium	14	100%	3.2	1.0	1.6	2.6	3.1	3.9	4.8
73	UNOCAL #5781	Oakland	Alameda	Pleistocene Alluvium	5	100%	4.2	1.3	3.2	3.3	3.5	5.5	6.2
74	UNOCAL 7499	Fremont	Alameda	Holocene Alluvium	9	100%	3.6	0.5	2.4	3.4	3.7	4.0	4.0
75	US Army MOTCO (formerly Concord NWS Tidal Sites) - CONCORD NWS - E-111	CONCORD	Contra Costa	Holocene Bay Mud	7	100%	6.2	1.2	5.1	5.4	5.7	6.8	8.7
76	Vallejo Unified School District - Adminstration	VALLEJO	Solano	Pleistocene Alluvium	20	95%	5.1	3.4	<RL	2.9	3.9	7.3	15.0
77	Wente Winery	Livermore	Alameda	Holocene Alluvium	58	100%	4.6	1.1	2.5	3.6	4.7	5.6	7.4