Spring 2015

SAMPLING OF SOILS WITH METALLIC RESIDUES COLLECTED FROM MILITARY SMALL-ARMS RANGES

Jay L. Clausen
University of New Hampshire, Durham

Recommended Citation

Clausen, Jay L., "SAMPLING OF SOILS WITH METALLIC RESIDUES COLLECTED FROM MILITARY SMALL-ARMS RANGES" (2015). *Doctoral Dissertations*. 2176.
https://scholars.unh.edu/dissertation/2176

SAMPLING OF SOILS WITH METALLIC RESIDUES COLLECTED FROM MILITARY SMALL-ARMS RANGES

BY

JAY L. CLAUSEN

Master of Science, University of Maine, 1990

Bachelor of Science, University of Nebraska-Omaha, 1987

DISSERTATION

Submitted to the University of New Hampshire
in Partial Fulfillment of
the Requirements for the Degree of

Doctor of Philosophy
in
Earth and Environmental Sciences

May 2015

This dissertation has been examined and approved in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Earth and Environmental Sciences.

Dissertation Director, Dr. Kevin Gardner
Professor of Civil and Environmental Engineering

Dr. Phil Ramsey
Adjunct Professor of Mathematics and Statistics

Dr. Thomas Douglas
U.S. Army Corps of Engineers, ERDC-CRREL

Dr. Nancy Kinner
Professor of Civil and Environmental Engineering

Dr. James Malley
Professor of Civil and Environmental Engineering

On 19 December 2014

Original approval signatures are on file with the University of New Hampshire Graduate School.

ACKNOWLEDGEMENTS

I would like to thank Dr. Terrance Sobecki with Cold Regions Research and Engineering Laboratory (CRREL) for his encouragement and support to pursue a Doctoral degree as well as his willingness to entertain numerous discussions on the various facets of my research. I would also like to thank Dr. Kevin Gardner for his interest in serving as my dissertation director and advisor and his willingness to pursue a research topic aligned with my ongoing research activities at CRREL. Additionally, I would like to thank my dissertation committee of Dr. Thomas Douglas, Dr. Phil Ramsey, Dr. Nancy Kinner, and Dr. James Malley for their interest and input during the development of the dissertation. Finally, I would like to thank my wife, Valerie, and my parents Don and Mary for their encouragement and occasional needling to complete the dissertation.

Funding to support my doctoral program came from a variety of sources. Mr. Clausen was selected by the Department of Army's Long Term Training program in Fiscal Year (FY) 09 and took a one-year sabbatical from his current job with the US Army Corps of Engineers (USACE), Engineer Research and Development Center (ERDC), CRREL in Hanover, NH to attend classes at the University of New Hampshire (UNH). ERDC provided the financial support while attending classes at UNH. Ongoing enrollment financial support was provided by CRREL during FY10 to FY14. Although the research was not directly funded by the Department of Defense, Environmental Science and Technology

Certification Program (ESTCP) this dissertation-utilized information gathered in support of the ESTCP ER-0918 project, *Demonstration of the Attributes of Multi-Increment Sampling and Proper Sample Processing Protocols for the Characterization of Metals on DoD Facilities*. Additional research, data analysis, and preparation of this dissertation was self-funded through the commitment of my own time with no funds provided by UNH, ESTCP, USACE, ERDC, or CRREL.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

Page

ABSTRACT

SAMPLING OF SOILS WITH METALLIC RESIDUES COLLECTED FROM MILITARY SMALL-ARMS RANGES

BY

Jay L. Clausen
University of New Hampshire, May 2015

Small-arms, artillery, mortar, and pyrotechnic military training introduces solid metallic residues into the environment. Previous energetic residue research showed conventional judgmental soil sampling and associated sample preparation methodologies are inadequate to address the heterogeneous distribution and still yield representative and reproducible results for the area of interest. Metallic residue deposition at military small-arms ranges occurs in a heterogeneous distributional pattern, similar to solid particulate energetic residue deposition.

One of the primary objectives of an environmental investigation is to estimate the mean contaminant concentration to evaluate human and ecological risk in an exposure unit (i.e., an area where the receptors are exposed to contamination). One assumption of the risk assessment is that the receptors spend an equal amount of time in every portion of the exposure unit. Evaluation of the appropriateness of conventional soil sampling for military small-arms ranges with solid metallic residues indicated the conventional approach of judgmental sampling was inadequate for providing representative and reproducible mean estimates for the exposure unit. Consequently, a series of experiments evaluated possible changes to the field sampling and laboratory sample preparation procedures to improve the representativeness and reproducibility of the sample results. The outcome of this study found that a collective series of changes involving soil sampling and sample preparation procedures provided superior estimates of the population (exposure unit) relative to conventional grab sampling. Overall, the dissertation results indicate a sampling approach referred to as the Incremental Sampling Methodology (ISM) provided an improved estimate of the population (exposure unit) relative to conventional grab sampling and has the potential to reduce sampling costs by 5 to 50 percent. Incorporation of the dissertation findings into suggested modifications for United States Environmental Protection Agency SW-846 Method 3050B for digestion of soils or sediments undergoing metals analysis is in progress; including the addition of an appendix outlining the procedures for collection of systematic random surface soil samples in the field using the incremental approach.

CHAPTER 1

INTRODUCTION

1.1 Background

The objective in soil sampling is to acquire information about the properties of a population of "soils" with the focus typically on estimating the mean. In environmental characterization investigations, a primary objective of soil sampling is to obtain a representative estimate of the mean for the constituents of interest to assess human or ecological risks. Soils have been defined as both natural bodies at the earth's surface, named variously pedons, representative elemental volumes, or elemental soil areas (Soil Survey Staff 1975, Journel and Huijbregts 1978) and as a continuum whose properties are a function of location (Holmgren 1988). Some soil properties do co-occur as a function of landscape factors (i.e. Jenny's five factors of soil formation (Jenny 1941) and can be associated with discrete, mappable areas of soil, which forms the basis of soil surveys. Other properties vary across soil forming factor-dictated boundaries and reflect a true continuum. The former are stable-static soil properties: such as particle size distribution (texture) that is relatively persistent over time and independent of anthropogenic manipulation. The latter soil properties are use-dependent and have temporal properties. These properties are often the result of anthropogenic activity or ecosystem disturbances and may vary systematically in time and space, such as with land use differences, and be spatially auto correlated, or appear randomly distributed (Moran 1950) as a function of a particular land use or over time, or both.

As in all sampling, the goal is to obtain a sample representative of the population by making observations or measurements on experimental units (i.e. sampling units) selected from the population under some defined sampling scheme. The ultimate goal in most soil sampling is to make a decision, typically a decision about risk, future land use, or need for remediation.

Over the last two decades, it has become apparent in cases where a soil amendment (e.g. fertilizers, pesticides) or contaminant (e.g. explosives, heavy metals) is released into the soil environment as a solid particulate that conventional grab sampling (also referred to as discrete, discriminatory, or non-probabilistic [judgmental, authoritative, biased], sampling) yields neither a representative nor reproducible result of the conditions in the field (ITRC 2012, Clausen and Korte 2009, Clausen *et al.* 2007, Hewitt *et al.* 2005a, Walsh *et al.* 2004, Jenkins *et al.* 2005, 2004a,b, 1999, 1997a,b, 1996, Pitard 1993). The issue is an important one because the United States (US) Department of Defense (DoD) has responsibility for 1,400 sites amounting to 10 million acres (DSB 2003). The high cost of collecting sufficient data for a comprehensive environmental assessment coupled with data uncertainties hinders sound management of these sites.

The US Army's Military Munitions Response Program (MMRP) established under the Defense Environmental Restoration Program in 2001 is responsible for the environmental, health, and, safety associated with unexploded ordnance (UXO), discarded military munitions, and munitions constituents on non-operational ranges located on active installations, on Defense Base Closure and Realignment sites, and formerly used defense sites. Under the MMRP, the DoD is required to: 1) inventory non-operational ranges that contain or are suspected to contain munitions-related material released before September 2002; 2) identify, characterize, track, and report data on MMRP sites and clean-up activities; and 3) develop a process to prioritize site cleanup and estimate costs. In addition, established directives mandate all active DoD facilities implement procedures to assess environmental impacts from munitions on training and testing ranges (DoD Directive 4715.11 and DoD Instruction 4715.14). Consequently, the DoD spends approximately $2 billion dollars annually on their environmental program (GAO 2003) with a substantial portion spent on site characterization and investigation activities as well as remediation. As discussed in Chapter 6, implementation of the dissertation findings could reduce DoDs sampling costs by 5 to 50%.

Historically, the focus on improving the quality of environmental data has centered on analytical error, however this only represents a minor portion of the total sample error (Ramsey 2006, Jenkins *et al.* 1999, Cline 1944). The U.S. Environmental Protection Agency (USEPA) has recognized that sub-sampling can lead to variable and biased analytical results (Gerlach and Nocerino (2003). Rasemann (2000) and Markert 1990) indicate sample preparation accounts for 100% to 300% of the total sample error. In contrast, field sampling comprises the greatest source of error (Gy 1999, 1992, Pitard 1993, Peterson and Calvin

1982, Cline 1944) with relative errors as large as 1,000% (Markert 1990). Recognizing the biased nature of environmental data and its resulting skewness (Reimann and Filzmoser 2000, Singh *et al.* 1997, Moore, 1995, Ahrens 1954) the question is whether our conventional sampling techniques yield representative samples or if there is a problem with our current field sampling and laboratory sample processing methodologies?

> *The reasons why sampling is overlooked are because it's a costly operation and an error-generating process that leaves no apparent tracks of its shortcomings.*[1] *Indeed, it is easier to track down analytical errors than sampling errors. Consequently, bad sampling practices plague the entire industry and are probably responsible for losses of many millions of dollars.*[2]

Recognition that sampling accounts for most of the total sample error implies that any improvements in data quality should focus on field sampling (Ramsey 2009, Rasemann 2000, Jenkins *et al.* 1999, Gy 1993, 1992, 1982, Pitard 1993, Ingamells and Pitard 1986). However, *in the past, sampling has been the weakest feature of resource survey and field research*[3]. Since most of the total sampling error resides in field sampling, what has the history of sampling shown and could application of practices from other industries, such as mining and agricultural, be redirected for the process of collecting soil samples for environmental assessments?

---

[1] Ramsey 2009

[2] Ingamells and Pitard 1986

[3] Crepin and Johnson 1993

1.2 History of Modern Sampling

The practice of sampling in general owes its roots to many fields such as science, commerce, agriculture, and mining. Sampling developed due to the need to draw conclusions about a population of interest without assessing the entire population (i.e., conducting a census). The earliest examples of field sampling (soil or rock) within the agricultural and mining industries occurred in the early 1900s (Herzig 1914, Wright 1910, Rickard 1906, Clarkson 1894). Most of the initial sampling efforts involved judgmental or grab sampling. This approach slowly changed when Bowley (1906) introduced the concept of systematic random sampling. Systematic random sampling consists of collecting an initial sample from a random starting location, typically within a grid, and then systematically collecting a sample from the random location within each subsequent grid. Tippet's (1927) introduction of tables of random numbers facilitated random sampling, although this method was slow to gain recognition (Stephan 1949). As discussed in Section 1.2.2, the mining industry by the early 1900s was using a systematic random approach referred to as transect sampling to infer the amount of a mineral in an ore-body. Similarly, there were three agricultural papers at this time, which were instrumental in introducing the concept of systematic random sampling for assessing the heterogeneity of an agricultural field's fertility (Mercer and Hall 1911, Student 1909-10, Wood and Stratton 1910).

From the 1900s to the 1940s, little changed in the approaches used for actual soil/rock sampling. Most of the effort in the agricultural industry during this period focused on sampling to develop crop yield estimates (Stephan 1949). The other topic of interest was assessing the variability in soil constituent concentrations (Cline 1944, Lyon 1932, Waynick and Sharp 1919, Robinson and Lloyd 1915).

It was recognized in the 1940s that many academic fields had developed methods for sampling, but most were not scientifically or statistically based. The statistical texts and papers of the time ignored sample design due to lack of training of statisticians in the process of sampling (Stephan 1949). Incremental improvements in sampling approaches and statistical methods for sampling occurred over the next 20 years as cross-discipline communication increased. The needs of statisticians and social scientists drove the changes in sampling, which slowly spilled over into other industries and other academic disciplines (Stephan 1949). These included the adoption of composite soil sampling techniques by the agricultural industry in the 1940s to determine the yield or concentration of a nutrient over a given area of acreage. From the 1940s until the 1970s, sampling research for the agricultural and mining industries focused on sampling designs for the collection of soil/rock samples (Cameron *et al.* 1971, Bicking 1967, Duncan 1962, Blaut 1959, Hansen *et al.* 1953). The development of the variogram for geostatistics within the soil science field occurred in the late 1970s (Journal and Huijbregts 1978) with development occurring simultaneously within the mining industry.

Beginning in the early 1960s and continuing into the early 1990s, the mining industry began to recognize the role of heterogeneity and its influence on obtaining representative results (Table 1-1). Geostatistics, developed by the mining industry in the early 1960s, addressed the heterogeneity issue. Its use by the environmental industry has increased, but it is still not widely employed by geologists or engineers. Similarly, Gy's sample theory, although published in English in 1982, 1989, and 1999 and summarized by Pitard (1993) was largely unknown until recently, and even then its acceptance and application within the environmental industry has been limited (ITRC 2012).

In contrast, the USEPA from the late 1980s through the present has focused on the collection of judgmental samples using various sampling designs (USEPA 1995). By the late 2000s, several states, namely Hawaii and Alaska had rewritten their environmental soil sampling guidance incorporating the concept of Incremental Sampling Methodology (ISM) for all constituents. In 2012, the Interstate Technology Regulatory Council published guidance (ITRC 2012) on the theory and application of the Incremental Sampling Methodology (ISM), based on Gy's sample theory. More recently, Florida issued guidance on the use of ISM (Florida 2013).

The principle difference between current environmental soil sampling and mining/agricultural soil sampling is the idea of an area as the sample unit, in the later, as opposed to an individual point in space or time, with the former. The definition of a sampling unit varies but generally, it is an area of interest, e.g. farm, field, (Reed and Rigney 1947, Cline 1944, King and Simpson 1944, Yates and Zacopany 1935) ore-body, or contamination zone.

Table 1-1. Chronological summary of publications discussing the role of heterogeneity in sampling and introduction of the incremental sampling approach.

Time Period	**Activity**	**References**
1960s−1990s	Recognition of the role of heterogeneity in distribution of metals in mining samples and development of methods to obtain representative samples	Duncan 1962, Ingamells 1974, Journal and Huijbregts. 1978 Johanson 1978, Elder *et al.* 1980 Gy 1992, 1999, Wallace et al. 1985 Pitard 1993 Leutwyler 1993, Studt 1995
Early 1990s−2004	Demonstration of presence of energetic residues on ranges	Racine *et al.* 1992 Jenkins *et al.* 1997a, b, 1998, 2001 Walsh and Collins 1993, Walsh *et al.* 1997 Thiboutot *et al.* 1998, 2000a, b, 2003 Ampleman *et al.* 2003a, b Clausen *et al.* 2004 Pennington *et al.* 2004 Taylor *et al.* 2004
1990s	Demonstration of heterogeneity issues associated with environmental samples	Pitard 1993, Jenkins *et al.* 1996
Mid 1990s−Early 2000s	Recognition of heterogeneity issues for energetic constituents on military ranges	Racine *et al.* 1992 Jenkins *et al.* 1997a, b, 1999, 2000 Taylor *et al.* 2004 Walsh and Collins 1993, M.E. Walsh *et al.* 1997
2001−2009	Development of sampling and sample processing methods for soils containing energetic constituents	Jenkins *et al.* 2001, 2004a,b, 2005a, 2006 Thiboutot *et al.* 2002 Walsh *et al.* 2002, 2003, 2005, 2006 Hewitt and Walsh 2003 Hewitt *et al.* 2005, 2007, 2009
2004−2009	Demonstration and comparison of ISM with traditional grab sampling approach for soils with energetic constituents	Jenkins *et al.* 2004 Walsh *et al.* 2004 Hewitt *et al.* 2005, 2009 Nieman 2007
2007−2010	Demonstration of heterogeneous distribution of metals in soils from military ranges	Clausen *et al.* 2007, 2010 Clausen and Korte 2009a, b
2008−present	Adoption of ISM for soils	Hawaii 2008, Alaska 2009, ITRC 2012, Florida 2013
2009−present	Evaluation of ISM for soils with metallic residues	Clausen *et al.* 2012a,b

Also different is the concern in the agricultural and mining industry with the average concentration over an area and not an individual point result. As a consequence, both industries developed sampling protocols based on the concept of collection of a composite sample as compared to the environmental industry's focus on judgmental grab samples.

A composite sample is one in which multiple increments of a material are collected over a given area or time and combined together to yield a single analytical sample representative of the area under consideration. However, the definition of what constitutes a composite sample often varies. Dorfman (1943) is considered to have been the first to introduce the concept of composite sampling for use in blood testing to determine the presence of the syphilis bacterium in army recruits. Blood samples collected from a variety of individuals were combined into groups so that many individual samples could be assessed with a single analysis. Since this point in time, both the mining and agricultural communities have extensively published results on a variety of composite sampling topics up until the present (Barbizzi *et al.* 2004, Boswell *et al.* 1996, Garrett and Sinding-Larsen 1984, Cameron *et al.* 1971, Duncan 1962, Onate 1953, Reed and Rigney 1947).

1.2.1 Agricultural Soil Sampling

The first mention of soil sampling for agricultural purposes occurred in the paper of Smith and Prentice (1929) where the focus was on sub-sampling error. Most of the papers at the time focused on sampling to determine crop yields and did not focus on soil sampling to determine mineralogy or nutrient needs.

Stephan (1949) suggests the main driver behind agricultural sampling was the need for data on crop and livestock numbers so statisticians could develop estimates on trends for government reporting.

Soil testing, as it is referred to in the agricultural industry, began in the 1940s with a focus in the U.S. on improving crop yields through recommendations on fertilizer and lime usage, primarily as a result of the 1930s drought in the southwestern U.S. One of the earliest field-testing programs began at the South Carolina Agricultural Experiment Station (Moore 2011). *However, the methods of sampling were quite crude, many of the reports were simply based on the respondent's judgment about conditions in his locality…*[4]*.* Prior to the 1940s, the focus was on collection of many grab samples, since the agricultural community was aware of the concept of soil variability (Cline 1944, Lyon 1932, Waynick and Sharp 1919, Robinson and Lloyd 1915). Around this time, the concept of a sampling unit was introduced (Yates and Zacopany 1935) and by the 1940s widely adopted (Reed and Rigney 1947, Cline 1944, King and Simpson 1944). The concern was that an unbiased estimate of the mean requires that every constituent in a sampling unit have an equal chance of being sampled (Snedecor 1940). Cline (1944) cites the proceeding reference indicating the agricultural industry was aware and applying statistical tools for soil sampling. Although, research on composite sampling was conducted prior to the 1940s (Cline 1944, Waynick and Sharp 1919, Robinson and Lloyd 1915) its application was not widespread.

---

[4] Stephen, 1949

The date of adoption of composite sampling by the agriculture industry is not exactly clear, but by the end of World War II, the practice was widespread. Adoption of composite sampling was the result of realization that individual grab samples were not reflective of large areas due to the earlier soil variability studies (Cline 1944, Lyon 1932, Waynick and Sharp 1919, Robinson and Lloyd 1915) and the concept of treating a farm as a single sampling unit. Further, collection of enough grab samples to be representative of a sampling unit (e.g., farm) was cost prohibitive.

Typically, with composite sampling 12-20 increments are collected per sampling area. Reed and Rigney (1947) conducted a study looking at the number of increments needed per composite sample and number of subsamples analyzed per sampling unit to yield a mean value with 5% error. The study found the need for a greater number of increments at lower measured nutrient concentrations. Meeting the desired error rate was not possible with the collection of only one or two subsamples. In many situations, it was impossible to obtain a representative estimate of the mean from samples with even 10 increments, especially from a non-uniform sampling unit. Even 30-increments were insufficient for non-uniform sampling units with low nutrient concentrations. Studies by Reed and Rigney (1947) demonstrated the futility of collecting enough grab samples to yield a representative result, and even with composite sampling, there are limitations. In addition to discussion on how to collect a sample, there were some who recognized sample preparation procedures are also important. Cline (1944) summarizes that air drying, sieving, and grinding of the soil sample are necessary practices.

The focus from the 1940's to the present has been on: (1) spatial variation of soil properties (Wildung and Drees 1983, Mausbach *et al.* 1980, Campbell 1979, Drees and Wilding 1973, Nielson *et al.* 1973, Beckett and Webster 1971, Maclean *et al.* 1945), (2) different types of sampling designs (Snedecor and King 1942) including composite sampling (Barbizzi *et al.* 2004, Boswell et al. 1996, Garrett and Sinding-Larsen 1984, Cameron 1971, Duncan 1962, Onate 1953, Reed and Rigney 1947), (3) development of a sample theory (Sampford 1962), and (4) application of classical statistical tools and theory (Gomez and Gomez 1984, Snedecor and Cochran 1967, Cohen 1959, Onate 1953, Snedecor 1940).

1.2.2 Mining Sampling

Documentation on approaches of how to collect a representative sample in the mining industry goes back to the late 1800s and early 1900s (Herzig 1914, Wright 1910, Rickard 1906, Clarkson 1894). Most of the sampling at this time focused on the collection of bulk samples in a systematic fashion with a primary issue being a means to reduce the sample to a manageable size for assaying, while still maintaining representativeness.

> *An accurate sample represents a true cross-section of the ore; it depends therefore, upon the uniformity of size of the groove or furrow; that is to say, an equal amount of ore must be broken across every part of the entire width of the lode. Among the things to be avoided one must mention the so-called 'grab sample'. This is the last resort of inadequacy.*[5]

---

[5] Rickard 1904

Typically, the bulk samples were collected in a transect design, with samples collected over the length of the ore-body then averaged to determine the assay of the ore. This approach falls under the category of probabilistic sample designs discussed further in Section 1.2.3.2.

The book by Rickard (1906) was the first on sampling for the mining industry. Yet, years later Herzig (1914) indicates, mine sampling is still neither understood nor applied and refers to grab sampling as "*pick analysis*", a haphazard approach not recommended for sampling. A common practice of the time was to grab a preliminary pick sample to guide a more comprehensive sampling campaign later. Ironically, the environmental industry uses this same approach over 100 years later, although it is typically referred to as a preliminary or screening sample. Interestingly, the concept of milling the sample prior to analysis to obtain a representative result was recognized as early as Clarkson's (1894) paper.

From the 1910s to 1950s improvements in mining industry sampling were incremental with research focused on the act of field sampling and sample designs based on statistical theory. Beginning in France in the 1950s, Pierre Gy began publishing a series of papers on developing a sample theory to provide the process of sampling a scientific foundation. Gy indicated the Fundamental Error (FE) of a sample relates to particle size and the mass of sample collected as shown in the following simplified equation (Pitard 1993);

$$FE = \frac{\sqrt{20}(d^3\ )}{m} \qquad \text{(eq. 1)}$$

where:

FE = sampling fundamental error

20 = sampling constant

d = maximum particle size (cm)

m = sample mass analyzed (g)

Pitard (1993) simplified Gy's FE equation by deriving a sampling constant, which is based upon (1) f – particle shape factor, (2) g – size range factor of re particles in the population, (3) l – liberation factor of the particles in the population, and (4) c - mineralogical composition factor of the particles in the population. Pitard's (1993) and Gy's (1992, 1982) work show for a highly heterogeneous material that the conventional grab sampling technique requires a mass of material impractical to collect for a single sample. Even when multiple grab samples are collected from a highly heterogeneous material, the number of samples needed to obtain a representative estimate of the mean is impractical (i.e., essentially the entire population must be sampled).

Pierre Gy's sample theory shows the only way to achieve an acceptable level of uncertainty for an estimate of a highly heterogeneous population is through the collection of incremental samples; a concept initially introduced to the mining industry (Clarkson 1894). Publication of Gy's books in English (Gy 1999, 1992, 1982) led to a wider awareness of Gy's sample theory. Gy's theory and its explanation were simplified in Pitard's book (Pitard 1993). However, as noted by Carrasco et al. (2004), even within the copper mining industry of today, Gy's sample theory is not well known. Visman (1971, 1969) also introduced his general sampling theory at about the same time. Visman's theory was also based on the concept of collecting incremental samples in order to minimize bias.

At the same time as Gy was publishing his sample theory, geostatistical analysis tools were being developed to support the mining industry (Matheron 1963). Unlike classical statistical methods, geostatistics involves the use of spatial properties of the data. Classical statistics assumes sample results are independent, whereas geostatistics assumes a spatial dependence or autocorrelation. Geostatistical results are expressed in a variogram that models spatial dependence or a kriged map. Kriging of the data is performed to identify data gaps in sample locations. Geostatistics are often used to analyze limited datasets in order to elucidate patterns not apparent without additional data collection. Isaaks and Srivastava (1989) provide a detailed discussion of the application of geostatistical methods for analysis of environmental data.

1.2.3 Environmental Soil Sampling

The environmental industry is a relatively young field with sampling implemented with the creation of environmental regulations such as Resource Conservation and Recovery Act (RCRA) and Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) in the late 1980s. The USEPA provides numerous documents on guidance on the collection of soil samples (USEPA 2000a, 1995, 1991, 1984). Although, many of the documents focus on the details of soil sampling collection, they often ignore where to collect the soil samples, how many samples to collect, and the type of sample design to use. In addition, a review of the USEPA guidance yields little scientific basis for the soil sampling methods recommended, how the proposed approaches came into being, or who individually was responsible for developing the guidance. A review of the literature suggests much of USEPA's sampling guidance is derived from statistical treatment of data with the work of Cochran (1977) providing the basic framework. USEPA (1995) recommends the judgmental approach with grab samples as being preferred for establishing a threat or identify sources, yet no reference or data provides support for this position. The fundamental problem with conventional grab sampling is the small mass of soil sampled to represent a much larger area and the associated high probability of missing the contaminant of interest (ITRC 2012). One never knows if the high contaminant concentration observed with a grab sample is representative of the entire site. Another sample might yield even higher concentrations or nothing at all. Ironically, the USEPA was aware of these issues; however, adoption of the methodologies identified by Mason (1992, 1983), namely the concept and applications of Gy's sample theory, were not incorporated into USEPA's regulatory guidance.

The preponderance of the sampling guidance provided by USEPA and states focuses on the collection of judgmental grab samples and in many instances; regulators forbid the collection of composite soil samples. The apparent focus on the collection of grab samples is the desire by the regulator to measure the highest possible contaminant level at the site of interest and the belief that composite samples will dilute the results (USEPA 1995, Barth *et al.*

1989). Numerous papers have been published discussing composite sampling in the environmental industry as an alternative to judgmental sampling as documented in Boswell *et al.'s* (1996) annotated bibliography. However, the concept of composite sampling, even to the present, has not gained traction in the environmental industry. Even among environmental consultants, the impetus of collecting a composite sample is the result of a perceived economic benefit due to decreased sample processing, shipping, and analysis (Splitstone 2001). Thus, environmental field sampling has focused primarily on judgmental sampling designs with probabilistic sampling designs being a rarity.

Under the judgmental approach, once the soil sample is collected from the field it is typically placed into an amber 4 oz glass jar and this approximately equates to a mass of 250 g. At the laboratory, the analyst opens the jar and typically scoops off from the top the desired aliquot for digestion/extraction; for metals analysis this amounts to 0.5 to 2-g of material. This is typically the extent of sample preparation unless specific directions are provided to the analytical laboratory. USEPA Method 3050B for metals calls for use of a mortar and pestle, if needed, however discussions with personal associated with commercial analytical laboratories indicates this is rarely done, primarily because this sample preparation step would have detrimental effect on sample throughput. Issues such as sample segregation during transport or ensuring collection of a representative sample from the jar are ignored.

1.2.3.1 Judgmental/Non-Probabilistic Sampling

Collection of soil samples for environmental investigations often employs a judgmental or non-probabilistic sampling approach, which is subjective in nature and not typically based on scientific rationale, although expert judgment is often invoked. The judgmental approach includes a number of designs such as convenience, expert judgment, haphazard, quota, snowball, or volunteer. Table 1-2 presents a summary of the pros and cons of the various designs based on USEPA (2002) and Mason (1992). These designs can work if the correct conceptual model is developed for the site of interest, but can lead to bad decisions if an incorrect conceptual model is developed or not enough samples are collected.

Some of the inherent problems with this approach include: (1) reliance on the sampler’s judgment of where to collect the sample or how much to collect, (2) information only obtained about the sample, (3) inability to statistically extrapolate results from one area to another, (4) inability of estimating or inferring population parameters, as well as testing an hypotheses about the sample population with a defined error rate. However, the single biggest issue typically involves an insufficient number of samples collected, with the limitation being an economic one.

Table 1-2. Comparison of the pros and cons of various sampling designs.

Sample Design	Ease of Use	Conceptual Model Based	Biased	Useful for Decisions	Yields Questionable Results	Applicability to Heterogeneous Sites	Cost Effective
Judgmental or Non-Probabilistic							
Convenience	Yes	No	Yes	No	Yes	No	Yes
Snowball	Yes	No	Yes	No	Yes	No	Yes
Quota	Yes	?	Yes	No	Yes	No	Yes
Volunteer	Yes	No	Yes	No	Yes	No	Yes
Haphazard	Yes	No	Yes	No	Yes	No	No
Probabilistic							
Random	Yes	No	No	Yes	Yes	No	No
Stratified Random	No	Yes	No	Yes	Yes	Yes	Yes
Systematic Grid	Yes	No	Yes	Yes	No	No	No
Systematic Random	Yes	No	No	Yes	No	Yes	No
Search	Yes	No	Yes	No	Yes	No	No
Transect	Yes	No	Yes	No	Yes	No	No
Composite	Yes	Yes	No	Yes	Yes	Yes	Yes
Incremental	No	Yes	No	Yes	No	Yes	Yes
Ranked Set	No	Yes	No	Yes	No	Yes	No

The other issue revolves around proper sample population representation. Judgmental sampling often leads to underestimation of the mean, artificial increase in variability, generation of outliers, non-reproducible results, and non-agreement between replicate and split samples (Ramsey 2009, 2006; Gy 1986). Under the judgmental approach, an arbitrary number of grab samples is often collected from the area of interest. Oftentimes the boundary of the sampling area is ill defined or not established. Additionally, regulatory officials often determine sampling locations based on visual observations and thus the sample designs have a biased component. Finally, the basis for the number of samples collected often uses the simple formula:

$$n = BS / C_s \qquad \text{(eq. 2)}$$

where;

n= number of samples collected,

BS = budget for sampling, and

C_s = cost per sample.

Typically, the budget for sampling is often developed after other project costs have been decided. These other project costs are often given more weight than the budget for sampling such that the number of samples collected is a variable adjusted to achieve the total project budget available.

An USEPA document for Superfund[†] sites recommends that 4 to 5 grab samples be collected for grid cells less than 30 x 30 m and collection of nine grab samples for grid cells greater than 30 x 30 m (USEPA 1995). Additionally, the USEPA has developed guidance on the number of grab samples needed based on a desired coefficient of variation (CV), power, confidence level, and minimum detectable relative percent difference (RPD) using an one-sided, one-sample t-test to characterize an area and compare the results to background (Barth *et al.* 1989). For example, a CV of 20, a power of 95%, a confidence level of 95%, and RPD of 20% indicate the need for 13 samples for both populations compared, or 26 total grab samples. An increase of the CV, power, or confidence level or decrease in RPD results in the need for even more samples. Clearly, increasing the number of samples collected improves the probability of a representative sample, but a question remains on how to balance this with the typical project budget? A common experience is that the number of samples collected is significantly less than the number of samples needed. *Thus, what is desirable is not affordable, and what is affordable is not adequate. How do I deal with this dilemma?*[6]

The results for the judgmental samples are assumed to represent the concentrations within the area of interest and the concentrations of the individual samples are generally assumed Gaussian or normally distributed. However, numerous studies have demonstrated the distribution of results for environmental samples is not typically normal (Reimann and Filzmoser 2000, Singh *et al.* 1997, Moore, 1995, Ahrens 1954). Thus, the application of parametric statistical analysis methods to restore data symmetry can lead to erroneous conclusions

---

[6] Gore and Patil 1994

about the site (Reimann and Filzmoser 2000, Ingamells and Pitard 1986). Results from grab samples are also sometimes used to calculate the mean concentration for an area using geostatistical approaches, such as kriging, where the concentration gradient between sampling points can be expressed by some mathematical concept such as linearity, power curve, or half the distance weighting. The assumption the grab samples are "representative" of the analyte concentrations within the area of interest is generally not tested, although the concentrations determined for grab duplicate samples collected nearby often do not agree. If enough grab samples are collected, anomalous results are often evident during the geospatial analysis of the data and various ad-hoc approaches are used to adjust the data (e.g., calculating the mean of co-located samples results, assigning a value of ½ the detection limit to non-detect samples). Gy (1986) states "*No financial decision or settlement should be made on the basis of analytical results obtained on non-probabilistic samples".*

1.2.3.2 Probabilistic Sampling Designs

Other approaches using the collection of grab samples include probabilistic sample designs such as random, stratified random, systematic grid, systematic random, composite, incremental, rank set, search, and transect sampling (USEPA 2002a, 1995). Each of these designs has its strengths and weaknesses (Table 1-2). For example, a random sample design involves the arbitrary collection of samples within a defined area. This approach leads to biased results and is not suitable for sites with contaminant heterogeneity. Gy

(1986) notes out those probabilistic sampling designs that do not include incremental sampling will lead to incorrect sample results.

In some cases, systematic grid sampling using a grid-node approach is used where the area of interest is divided into a number of individual grids (exposure areas or desired potential remedial removal volume), the size of each being a function of the total area to be assessed and the future land use envisioned. This approach is often used if a statistician or software such as Visual Sampling Plan [VSP] (Matzke *et al.* 2007) is involved in the project. One issue with the use of VSP is that it assumes a normal distribution, but as mentioned earlier, environmental data typically have a skewed distribution. There is an option of using a non-parametric hypothesis test, but it requires the user to provide an estimate of the standard deviation. Given that the degree of heterogeneity is not predictable in a quantifiable sense, one is simply taking a guess. Within each grid, one grab sample is collected and shipped to an offsite laboratory where samples are processed and analyzed. This single sample is supposed to represent the entire area of interest. The sample result is then compared to some action level for the contaminant of interest and if over the limit some action is taken and if below no action is taken. Because of the cost of collecting a large number of samples, use of this approach is rare for non-governmental projects.

1.2.3.3 Recognition of a Problem with Conventional Grab Sampling

In the early 1980s at Eagle River Flats, Alaska, a situation became apparent where ducks were mysteriously dying in the marshes located on the military range. Later studies documented the extent of the duck mortality but not the cause. Grab sediment sampling (81 total samples) conducted by a variety of agencies and contractors from 1985 to 1989 failed to detect any contamination (Racine *et al.* 1992). Racine *et al.* (1992) used a composite transect sampling technique and analyzed the samples for high explosives, but none were detected. However, one of the samples emitted a white vapor when the sample container was opened and stirred, suggesting the presence of white phosphorus. Further studies with the compositing method established the mortality occurred because dabbling ducks were consuming particles of white phosphorus, which are particularly toxic (Nam *et al.* 1994). Later, the question was raised as to why the conventional grab sediment sampling approach failed to detect the white phosphorus contamination, whereas the compositing method yield elevated levels in the sediment. Studies by Walsh *et al.* (1997, 1996) indicated the particulate distribution of white phosphorus was extremely heterogeneous. Consequently, grab sampling resulted in a low probability of encountering individual particles, whereas the compositing technique had an increased probability of including them. Ironically, USEPA guidance was available on sampling approaches to use when particulates are present (Mason 1982). The approaches offered were based on Gy's sample theory. However, Mason (1982) appears to be unfamiliar to many in the environmental industry including both regulators and contractors. *The sampling of particulate materials is an orphan in need of assistance.*[7]

Based on the white phosphorus observations, a similar question was raised about whether military training with munitions containing explosives: 2,4- or 2,6-dinitrotoluene (DNT), octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine (HMX), nitroglycerin (NG), perchlorate, hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX), or 2,4,6-trinitroluene (TNT) results in an environmental impact. The USEPA banned training with artillery and mortar weapon systems containing energetic materials at the Massachusetts Military Reservation (MMR) through an administrative order due to an assumption of an environmental impact (USEPA 1997). Again, the conventional grab soil sampling methodology was employed and results were inconsistent, with infrequent detections of explosives on the range and many non-detections (Clausen *et al.* 2011, 2004; Ogden 1998). Re-sampling of the same area often resulted in significantly different results (e.g., no detections versus percent levels of explosives). Use of the composite sampling approach developed for the white phosphorous situation improved sample quality for soil samples obtained in the impact area of military ranges where explosives residues were introduced (Jenkins *et al.* 1997a, 1996). However, field compositing did not consistently yield reproducible results (Ogden 1998). One reason was the sample preparation procedure promulgated by the USPEA for explosives, Method 8330 (USEPA 1994a), was based on an assumption of aqueous release of a contaminant to the environment. Both the grab and

---

[7] Gy 1986

composite sampling approaches used the sample preparation methods in Method 8330 (USEPA 1994a). Research over the last decade has demonstrated that special field sampling and sample preparation procedures are necessary to obtain a representative and reproducible results when energetic residues are present in soil (Hewitt *et al.* 2009, 2007, 2005; Walsh and Lambert 2006, Walsh *et al.* 2006, 2005, 2003, 2002; Jenkins *et al.* 2006, 2005, 2004a, b, 2001, 1999; Hewitt and Walsh 2003; Thiboutot *et al.* 2002). The decade long focus on sample preparation procedures culminated with a modification to USEPA Method 8330 resulting in Method 8330B (USEPA 2006), which included guidance on surface soil sampling using an incremental approach. This sampling approach is now referred to as ISM, multi-increment sampling (MIS™), or incremental sampling (IS), hereafter referred to as ISM in this dissertation.

1.2.3.4 Incremental Sampling Methodology

The ISM approach includes changes to the conventional field and laboratory sample-preparation procedures (ITRC 2012).

> *Incremental Sampling Methodology (ISM) is a structured composite sampling and processing protocol that reduces data variability and provides a reasonably unbiased estimate of mean contaminant concentrations in a volume of soil targeted for sampling. ISM provides representative samples of specific soil volumes defined as decision units (DUs) by collecting numerous increments of soil (typically 30–100 increments) that are combined, processed, and sub-sampled according to specific protocols[8].*

ISM is largely based on Gy's sample theory (Gy 1999, 1992, and, 1982). One of the differences between composite sampling and ISM is in the number of

---

[8] ITRC 2012

increments collected. Composite samples typically consist of 3 to 20 increments, whereas ISM consists of 30 or more increments (ITRC 2012, Hewitt *et al.* 2009). The purpose of collecting more than 30-increments with ISM is to normalize the distribution of measurements (per the Central Limit Theorem in statistics) and ensure the sample is physically representative of the area sampled (i.e., the environmental population). Another difference is that ISM involves milling or other form of sample reduction rather than homogenization.

In addition, to these methodology changes, ISM involves a different approach to thinking about the sampling problem. Conventional sampling focuses on sampling a specific point and then extrapolating between sampling points to define an area of contamination. The assumption with grab sampling is the result from one sampling point is auto correlated with a nearby sampling point, which may be true for an aqueous contaminant release. However, detonation of a munition results in a random particulate deposition. Thus, there is no co-dependency of concentration between one sampling point and an adjacent point. Further, in many cases, the outer boundary of interest is ill defined or not defined at all with conventional grab sampling. In contrast, with ISM one first defines the area of interest where a decision is needed (the decision unit - DU). A multi-incremental sampling approach is then used to collect increments from within the DU. The final sample is representative of this DU. Thus, ISM involves a direct, area-focused sampling effort whereas the grab method is point-focused.

The term DU (alternately termed area of concern, sampling area, exposure unit, contamination zone, population, or habitat) refers to the area a sample is intended to represent. The increments can be collected in a random fashion or systematically. In the systematic-random pattern, a random starting point is selected and increments are gathered on an even spacing as the sampler walks back and forth from one corner of the DU to the opposite corner (Figure 1-1). Ironically, the concept of the DU appears to be the same as the sampling unit concept developed in the agricultural community back in the 1930s-1940s (Reed and Rigney 1947, Cline 1944, King and Simpson 1944, Yates and Zacopany 1935).

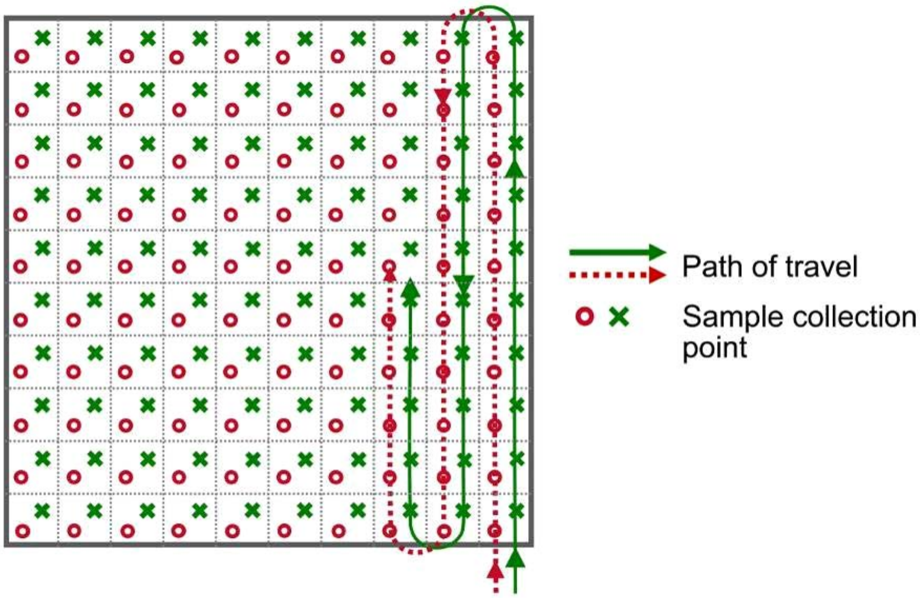


Figure 1-1. Example of multi-increment sampling using a systematic-random sampling design for collecting two separate 100-increment samples.

To obtain representative and reproducible results, the sampling strategy must address the compositional and distribution heterogeneity of the constituents of concern (Pitard 1993). Compositional heterogeneity is due to soil-sized particles within the population not all having the same concentration. This heterogeneity is at a maximum when a portion of the contaminant is present as discrete particles. Error due to compositional heterogeneity is the fundamental error (FE) and inversely relates to the sample mass. Distributional heterogeneity is due to contaminant particles being scattered across the site unevenly, sometimes with a systematic component as well as a short-range random component. Error associated with distributional heterogeneity inversely relates to the number of individual increments used to build the sample. This type of error is at a maximum when a single discrete sample is used to estimate the mean for a larger DU. To reduce the influence of distributional heterogeneity in the estimate of the mean concentration, the collection of 30 or more evenly spaced increments to form an individual soil sample has been recommended (Jenkins *et al.* 2006a, 2005, 2004a,b; Hewitt *et al.* 2012, 2009, 2007, 2005, Walsh *et al.* 2005). The objective of the multi-increment strategy and systematic random design is to obtain a proportional amount of residue particles of every composition and shape. Instead of collecting and analyzing single point samples and integrating the results for an area or assuming a single point is representative of the entire area, ISM samples are built by combining a number of increments of soil from within the DU to obtain a ~ 1 to 3-kg sample representative of the entire area of interest. Contrast this with conventional grab soil sampling where collection of less than 250-g of material occurs from a single point.

Employment of ISM at over 30 military ranges has consistently demonstrated the presence of energetic compounds in surface soils (Clausen *et al.* 2004; Pennington *et al.* 2004; Taylor *et al.* 2004; Ampleman *et al.* 2003a,b, Jenkins *et al.* 2001, 1998, Thiboutot *et al.* 2003, 2000a,b, 1998). Further, these studies have confirmed, under ideal conditions, consumption of a large percentage of the energetic material during detonation. However, these studies also demonstrated field conditions are not always ideal, and consequently, not all munitions undergo a high-order detonation thereby consuming all of the explosive material. A low-order detonation results in the distribution of energetic material into the environment as particulate residues (Taylor *et al.* 2004, Hewitt *et al.* 2003, Jenkins *et al.* 2002, 2000) in an extremely heterogeneous manner (Jenkins *et al.* 1999, 1997a,b, 1996). Research indicated explosive concentrations in grab samples vary substantially even over short distances (Jenkins *et al.* 1999, 1997a, b). Results of samples collected within grids as small as 10 × 10-m using nine-grab samples varied by two orders of magnitude demonstrating short-range heterogeneity (Jenkins *et al.* 2006). Comparison of ISM results with grab samples from soils with energetic residues has repeatedly shown that representative and reproducible results are not possible with conventional grab sampling (Jenkins *et al.* 2005, 2004a,b, 1999, 1997a,b, 1996). Consequently, a field sampling and sample preparation approach was developed, referred to as ISM, to address the conventional grab sampling shortcomings, which is now incorporated into USEPA Method 8330B (USEPA 2006), and program guidance (Clausen *et al.* 2012b, Hewitt *et al.* 2012, 2009, 2007, USACE 2009, Thiboutot *et al.* 2003b, 2002, 1998).

1.3 Philosophy of Soil Sampling and Sample Representativeness

The environmental industry has increasingly questioned whether conventional soil sampling yields a result that is representative of what is present in the field and whether changes to existing procedures are warranted (ITRC 2012). However, before addressing this issue, it is first necessary to define the meaning of a sample and then the concept of representativeness and a representative sample.

A sample is small part of the whole and should provide information overall on the item of interest. Statistics define a sample as one that is a subset of population, but that provides information on the total population. Within the environmental industry, the meaning of a sample is not universal, and it is my impression after 20 years that many in the industry really do not understand the concept of a sample.

Something that is representative serves as an example or type for others of the same classification. The definition of a representative sample is a subset of a statistical population that accurately reflects the members of the entire population such that the sample contains the contaminants of interest in the same proportions as the environmental population (Pitard 1993, Gy 1992, 1982, Barth *et al.* 1989). The two advantages of sampling are lower cost and faster data collection than measuring the entire population (e.g., with a census). Recently, a discussion in the environmental industry has revolved around this concept of representativeness, what it is meant by the use of this term, and how one goes about making a determination of representativeness.

> *From the early age of the industrial revolution miners, geologists, metallurgists, chemists and many others have expressed concern about the representatives of samples, which is essential for any kind of evaluation and quality control program. Many 'experts' developed their own rules of thumb by following their experience, common sense, and intuition.*[9]

When a sample is not representative, the result is a sampling error. A sampling error is a statistical error where the derived results from analysis do not represent the entire population. Gy's sample theory allows for quantification of SE and an estimation of sample accuracy or bias (Gu 1992, 1982, Pitard 1993). The question of representativeness stems from the observation that in a number of cases re-sampling (e.g., for pre-remediation activities) fails to find the contaminant of interest, precision is poor (i.e.. replicate analyses of the same sample area or same sample do not agree), significant outliers exist, or split samples analyzed by different laboratories yield different results for the same material (Bonczek *et al.* 1996, Grant *et al.* 1996, Jenkins *et al.* 1996, Moore 1995). Related to this question is how representative a single grab sample for a population of some item of interest is over a specific area? Is it possible that a given sample is unique (e.g., only one sample from a population has a contaminant particle) or does the sampling methodology, including sample preparation influence the result? Clearly, quantification of the sampling error, a measure of representativeness, is not possible with a single sample. Thus, collection of multiple samples is necessary to gain some insight into the population as well as quantify the sample error. However, the error rate is not

---

[9] Pitard, 1993

static, such that increasing the sample size can change the value determined. Therefore, if collection of multiple samples from the same area occurs and the variance of respective values is large or biased, is it acceptable for the intended purpose to calculate a statistical average as a basis for making a decision? Gy's theory states that the variance of sampling results is a function of the maximum particle size in the sample, which can be reduced through particle size reduction via grinding (Pitard 1993, Gy 1992, 1982)

To ensure sample representativeness a number of sampling methods have been developed (e.g., simple random sampling, systematic random sampling, stratified random sampling, systematic random stratified sampling). In addition, development of elaborate statistical approaches allows for situations where soil property outliers exist in a set of samples from a given population. If some of the property values are below the detection limit of the particular analytical method (i.e., non detects), then treatment of the data with substitution methods (Helsel 2005), Cohen's MLE (Miesch 1967), Delta-Lognormal Method (Atchison 1955), or a variety of new methods such as hypothesis tests, survival analysis, or regression probability plots (Helsel 2005) may be warranted. However, these approaches lead to several questions, such as, are these data analysis approaches correct, do they yield a representative result, and does one even need to use these approaches? Non-detected results or "left censored data" are inherently a form of information loss. Therefore, it is desirable to collect a sample in a manner that reduces the need for statistical procedures to address data quality problems such as large outliers and data censoring.

The objective of a representative sample is to obtain an unbiased estimate of the population property or parameter of interest (e.g., the mean contaminant concentration) and a bound on the error of this estimate. From a broad philosophical perspective, a representative sample is one that should allow the same decision to be made at some level of confidence for each sample collected from the same area of interest every time. Meeting these conditions means the results are reproducible and representative of the conditions in the field.

For soil contamination, it is often also desirable to estimate the spatial distribution or dispersion of the contaminant(s). Although collection of a much smaller mass/volume for the sample occurs from the area of interest and even smaller mass/volume is prepared for analysis; the analyzed sample should still be representative of the larger scale. Since, representation of an area with a single point sample is generally not possible, multiple samples are typically collected.

If multiple samples are representative of the entire population, then the sample results should yield a consistent estimate of the mean. However, in most conventional soil sampling programs, employing judgmental sampling, a determination of representativeness through quantification of reproducibility, accuracy, and precision is generally not performed. Of the three approaches to assess sample error, only duplicate samples are typically collected to test reproducibility. A reproducible result is a repeatable one (i.e., collection of duplicate/replicate field samples or laboratory splits yields the same result every time). However, grab sampling often results in non-agreement with duplicate results. Unlike determining accuracy, assessment of the precision of a replicate sample is quantifiable by estimating the relative RPD for the comparison of two results or coefficient of CV or sometimes referred to as the percent relative standard deviation (RSD) when three or more results are available. Most environmental practitioners have experienced the situation where duplicate sample results differ by a factor of 2 to 10, which implies a lack of reproducibility. When dealing with particulate residues introduced into the environment, the difference in replicate values based on calculation of CV is several hundred percent (Clausen *et al.* 2012a, 2007, Jenkins *et al.* 2004, 1997a,b, 1996). The current solutions to address different duplicate results vary from tossing out the high or low result, averaging the two results together, ignoring the duplicate sample result, or re-sampling (Helsel 2005). Clearly, these approaches should not be acceptable, yet they are the norm for the environmental industry. Accuracy is a measure of sample correctness, which is typically unknown since the amount of contaminant is not known (Mason *et al.* 1992), although analytical accuracy is often quantified through evaluation of spiked and referred samples.

Figure 1-2 illustrates the issues associated with accuracy and precision of a surface soil sample, B8-A, and six duplicate samples, B9 through B14, collected from the berm face of a small-arms range located at Camp Edwards, Massachusetts. Table 1-3 depicts the results for the soil samples collected in wheel pattern with Sample B8-A located at the center and the rest of the collected samples arranged in a circle 15-cm apart (Figure 1-2). If collection of only a single sample in Table 1-3, such as B8-A occurred, how would one know if this represented the highest copper concentration in the area? If a single duplicate sample had been collected, such as B11, which is the correct sample result B8A or B11? Collection of a single duplicate sample within several inches of the original sample is the normal approach for quality assurance/quality control testing for environmental investigation studies. The collection of six duplicate samples reveals that the area sampled has a high degree of heterogeneity. However, had the normal environmental investigative process been followed of collecting a single original sample and one duplicate sample the degree of heterogeneity would not be apparent unless the right sample had been collected (e.g., samples B8-A and B11). If samples B8-A and B14 had been collected, one might draw the conclusion that distribution of the contaminant is rather homogenous.

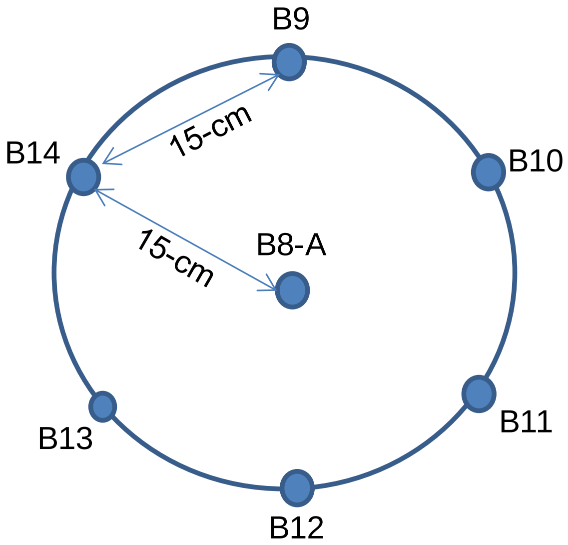


Figure 1-2. Layout of duplicate sample collection from a small-arms range at Camp Edwards.

If one considers a different situation, such that Figure 1-2 represents a larger area, such as a 1-acre residential property, then the question becomes does the estimated mean value accurately reflect the conditions across the entire property. The collection of multiple samples analyzed for copper indicates the estimated mean is 462 mg/kg (Table 1-3). Clearly, the example for copper in Table 1-3 suggests the estimated mean value has a high bias due to the result of the B13 sample. If I ignore the high B13 value of 2,316 mg/kg copper and re-estimate the mean, the new value is 181 mg/kg copper. Conversely, ignoring the low value and re-calculating the mean yields a new value of 645 mg/kg copper. If the regulatory risk value for copper was 400 mg/kg, it is apparent that different environmental investigation outcomes are possible depending on the data analysis used. Even, if one just considered the individual sample results, different outcomes are possible. Several questions raised by this example include: (1) is the original B8-A sample representative of the population, (2) if not, are any of the duplicate samples representative of the population, and (3) if no single sample is representative, are the combination of original and duplicate samples representative or what combination of duplicate samples are representative?

Ironically, the basis for regulatory guidance on how to perform a risk assessment to determine the impact to humans or ecological receptors is on obtaining an estimate of the mean concentration of a contaminant over some defined area. Yet, the common practice is to collect a few grab samples from discrete locations, interpolate between the sampled points, and then derive a mean value. It has been this author’s observation after 20 years working in the environmental industry that the mean contaminant concentration is often ignored and the maximum observed at a single grab sample, a discrete point, is used for the risk calculation. This single result is inferred to be representative of an area.

Table 1-3. Replication of metal results for co-located grab soil samples from a small-arms range at Camp Edwards, Massachusetts[10].

Element	Sample Identification							Mean (mg/kg)	Std. Dev. (mg/kg)	CV (%)
	B8-A	B9	B10	B11	B12	B13	B14			
Al	4323	4202	5390	4983	4339	3605	5938	4676	801	17
Cu	84	257	430	2316	29	24	109	462	830	180
Fe	5691	5630	6811	6646	5628	4866	7774	6116	999	16
Mg	602	793	962	974	733	723	1065	839	162	19
Mn	41	38	54	54	39	42	90	51	19	37
Pb	277	345	590	549	264	720	370	445	175	39
W	429	625	1054	1374	292	142	777	666	439	66
Zn	24.8	41.7	46.6	61.0	<15	16.5	35.4	37.7	15.9	42

The argument made, often by regulatory agencies, is this approach is conservative. However, this approach is contrary to the concept of an exposure unit where the volume of soil to which a receptor is exposed is used to evaluate chronic risk. Hadley and Mueller (2012) make a compelling case that the “hot” samples are inconsistent with the concept of an exposure unit and that extrapolating from a single result to an area is inappropriate. An often-overlooked issue is that, owing to compositional and distribution heterogeneity, a grab sample does not possess a “sample support” (i.e., it does not represent a

---

[10] Soil samples collected on spacing of 6 inches from the center of the berm face on Bravo Range. Highlighted text denotes the known contaminants introduced into the environment.

well-defined volume of “unsampled” uniformly contaminated soil centered at the location from which the sample was collected). As the soil is not homogeneous and it is not possible to re-sample and re-analyze the same soil twice, re-sampling at the same spatial coordinates will generally not provide any useful information.

One of the reasons for this disconnect between what is done and what is needed from sampling, in this author’s experience, is the lack of formal training in sample design and theory (Carrasco *et al.* 2004, Gy 1986). Teaching of these concepts is rare in the fields’ common to the environmental industry such as geology, engineering, chemistry, and toxicology. In addition, statistics and geostatistics are important tools applicable to the environmental industry. However, it is this author’s experience that few individuals are well versed or familiar with geostatistical topics as well as sample design. The other reason for the disconnect between what is done and what is needed for sampling, in the author’s opinion as well as Gy (1986), is the perception that sampling is not a science: because everybody knows how to sample and it is a simple process. *Sampling is an error-generating selection process that must be appraised in terms of accuracy and reproducibility, exactly in the same way as the other parts of the analytical process. Because of lack of information, this point is not adequately perceived by most decision makers.*[11] As an example, the project manager typically does not ask the entry-level environmental professional on their first field sampling effort if they know how to sample. The assumption made

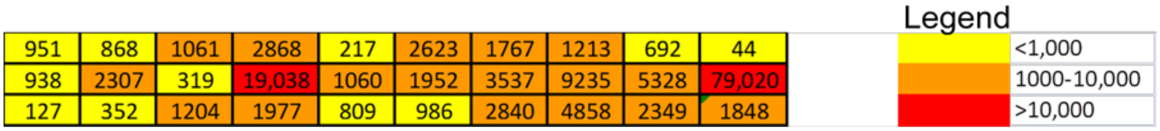
---

[11] Gy 1986

by the program manager is that the individual knows how to collect a sample. If direction is provided, the extent takes the form of a map with locations of where to collect a sample. In many instances, the entry-level environmental professional is simply handed a sampling tool and global positioning satellite unit and told to collect samples.

As mentioned in Section 1.2.3.4 ISM is a new sampling approach introduced to the environmental industry. One of the criticisms of ISM is the belief that information about the contaminant distribution is lost. Further, if one desires to locate “hot spots”, then grab sampling is the only means to accomplish this task. The hot spot term is common in the environmental industry and is supposed to define an area of elevated contaminant concentration. However, regulatory guidance does not state the mass needed to create or the spatial dimensions necessary for delineating a hot spot. Further, Hadley and Mueller (2012) question the concept of a hot spot given that risk is function of the dose (i.e., mass of contaminant over an exposure area).

Discussion of an example will best address this criticism of ISM, but also shed some light on the fallacy of the hot spot concept. Figure 1-3 depicts grab sample results for lead obtained from the berm face of a small-arms range. The dimensions of the berm were 3-meters high by 100-meters long with samples collected over a 2-cm depth. The collection of the 30 grab samples involved using a grid-cell centered sampling approach. The data reveal two high concentration areas colored in red.



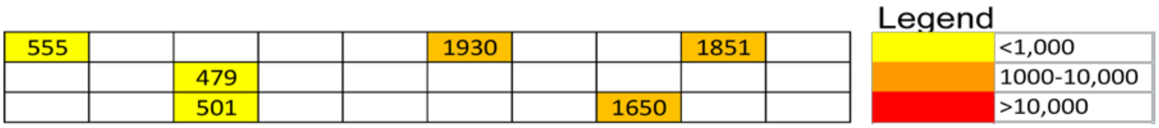
951	868	1061	2868	217	2623	1767	1213	692	44
938	2307	319	19,038	1060	1952	3537	9235	5328	79,020
127	352	1204	1977	809	986	2840	4858	2349	1848

Legend

🟨	<1,000
🟧	1000-10,000
🟥	>10,000

Figure 1-3. Lead results for grab samples collected from the berm face of a military small-arms range using a grid-node centered sampling approach.

Collection of a second set of six samples took place using a random grid-centered grab sampling approach (Figure 1-4). These individual samples were also co-located with those collected in Figure 1-3 with the collection distance being less than 15-cm apart. To characterize the small-arms range berm face depicted, there is a need for 5 to 15 samples based on discussions with colleagues in the environmental industry, USEPA guidance, and sampling design tools (Matzke *et al.* 2010; USEPA 1995, Griffiths 1971). Comparing the results for Figures 1-3 and 1-4 indicates the judgmental grab samples in Figure 1-4 yielded no apparent hot spot. The probability of encountering one of the hot spots with six samples is 44% and the probability of finding both hot spots is 8%. Further, a closer inspection of the data reveals up to a factor of three difference between some of the Figure 1-3 and 1-4 co-located sample results.



555					1930			1851	
		479							
		501					1650		

Legend

🟨	<1,000
🟧	1000-10,000
🟥	>10,000

Figure 1-4. Lead results for grab samples collected from the berm face of a military small-arms range using a judgmental sampling approach.

A question raised from this example is how many grab samples are necessary to characterize this particular site and find the hot spots. If the objective was to find hot spots within an individual grid, then the collection of six samples was clearly inadequate. However, the collection of 30 grab samples as shown in Figure 1-3 may be adequate. Is it reasonable to conclude that a grab sample from a 0.5 cm diameter area is representative of the concentration over the entire 3 x 10 m grid cell? If one has no prior information of the presence of a hot spot, how does one determine how many are likely present or the associated size of such an entity and the appropriate number of samples to collect? Ramsey (2006) illustrates this dilemma using (Figure 1-5) and asks the following questions: (1) are all of these hot spots, (2) what concentration level is important, and (3) what areal dimensions are critical (i.e., hectare versus 1 square centimeter)? Figure 1-5 shows three elevated points on the left, a number of elevated points and small areas in the center, and an area of contamination and absence of contamination on the right.

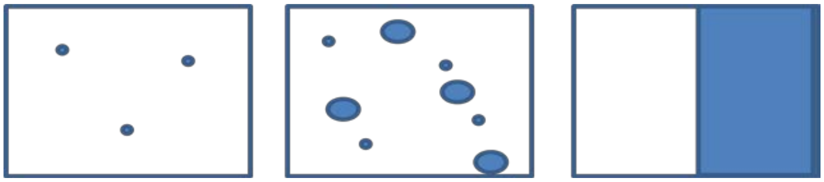


Figure 1-5. Hot spot depiction with colored areas representing contamination and non-contaminated areas white, modified from (Ramsey 2006).

Hadley and Muller (2012) note no common definition of a hot spot exists including how large an area it needs to cover or concentration value exceeded to be considered a hot spot. In the examples above, the hot spot definition was an arbitrary value of lead greater than 10,000 mg/kg. If, a value of more than 5,000 mg/kg lead was the criteria for identifying a hot spot in Figure 1-3, this still yields two hot spots. However, the hot spot on the right side of the figure is now larger and spans three grids. In this example, it is not clear which definition of a hot spot (more than 10,000-mg/kg lead versus 5,000-mg/kg) is right. As Ramsey (2006) notes, neither the entire population nor an individual atom is likely a hot spot. Hot spots are based on some arbitrarily defined area and are a random phenomenon. Typically, the basis for identification of a hot spot is on spatially analyzing a number of sample results and interpolation. The key with interpolation is the assumption that adjacent sample points are correlated. However, since hot spots are random in nature, it is not appropriate to assume some autocorrelation between sample points. This example points to the fallacy of the hot spot concept. Further, the basis for regulatory guidance on assessing risk is the exposure over an area. Thus, the need for identifying hot spots appears unwarranted. After all, the calculations of risk require the average concentration across the exposure area, not the maximum concentration at a point. Thus, there seems to be a misplaced concern and emphasis on a single grab sample result and by connection the identification of hot spots. Appropriate application of the regulatory process does not require knowing the result at a specific point. In addition, once collection of the sample occurs that sample point no longer exists. Therefore, if there is some concern about the validity of the result there is no means to resample that point.

A calculation of the estimate of the (population) mean or a 95% upper confidence (UCL) (of the population mean) is typically used for environmental investigations. Hence, the mean or UCL is used in a risk evaluation calculation and compared to a regulatory action level number, or compared to a background number to determine if a risk is posed by the contaminant in the area of interest. In the example above, the estimated mean for the 30 grid-node centered grab samples is 5,060 mg/kg of lead and the mean for the six random grab samples is 1,161 mg/kg of lead (Table 1-4). The results also indicate the distribution for the grid-node centered data is highly skewed whereas there is less skewness with the judgmental grab sample results. This is apparent by comparing the mean and median values. The lesser bias with the judgmental grab samples is likely the result of chance.

If one subsamples the grid-centered data repeatedly in Figure 1-3 using a bootstrap technique with varying number of subsamples a range of population distributions of the maximum value is possible. It is quite evident that repeatable sub-sampling yields different results even when the number of subsamples is held constant. The frequency of hot spot detection, based on a value less than 10,000 ppm, improves as the number of samples to assess the population increases. Clearly, more samples are preferable, but in the example provided, the pressure on the project manager is to collect the minimum number of samples. In this example, if the population is defined by the 30 samples and yet only 5, 7, 10, 15, samples have been collected, there is no mechanism to know if the number of samples collected is an adequate representation of the total population. This example illustrates that in all likelihood nothing less 30 samples is representative of the total population. Sampling of the area depicted in Figure

1-3 also used the ISM approach involving the collection of 3-increments per grid for a total of approximately 100-increments across the entire DU (all grids combined). Although, three replicate samples are typically collected per DU (ITRC 2012), the sampling was repeated 30 times for statistical comparison purposes. Table 1-4 indicates a high degree of precision between the ISM replicates as evidence by the nearly identical mean and median values. The collection of replicate samples provides information on whether the sampling approach addressed the inherent heterogeneity present in the sample. One of the criticisms of ISM is the related loss of information on the degree of heterogeneity present.

Table 1-4. Comparison of the estimated means for lead for the grid-node centered grab samples, biased grab samples, and incremental collected samples.

	ISM (100-inch)	Grab	Grab
	Systematic Random	Grid-Centered	Biased
n	30	30	6
Mean (mg/kg)	2,717	5,060	1,161
Median (mg/kg)	2,718	1,238	1,103
Minimum (mg/kg)	2,440	44	479
Maximum (mg/kg)	2,936	79,020	1,930
STD (mg/kg)	119	14,438	718
RSD (%)	4	285	62

STD – standard deviation. RSD – percent relative standard deviation

Calculation of a high CV or RSD value for the replicates would indicate a potential problem with field sampling or sample preparation. The replicate samples should be collected in the same manner, processed using the same protocols, and analyzed with the same instrumentation as the original samples. To calculate the RSD, use of a minimum of the three replicates is typical. The RSD provides information on the total SE. Although, this approach does not specifically isolate the source of error, it is generally safe to conclude the majority of the total sample error arises from inadequate accommodation of heterogeneity. If one desires to understand the degree of heterogeneity, then the design of the ISM approach can acquire this information.

Another criticism of the ISM approach is the potential loss of information on the spatial distribution of contaminants. If the objective is locating a hot spot, then the design of an ISM approach can start with that objective in mind. For example, if there is a desire for a more refined contaminant distribution for the area depicted in Figure 1-3, the DU could be further subdivided into smaller sampling units (SUs). One possible approach would be to designate the depicted grids in Figure 1-3 as individual SUs and collect ISM samples from within each. Consequently, this approach would involve the collection of 30 SU samples consisting of 30 to 50 increments each. Thus, the ISM approach is amenable to identifying hot spots, but requires some thought on the hot spot dimensions of importance before sampling. The argument remains that there is a loss of spatial information even from within the smaller SUs, since the result obtained is the average concentration for that area. However, this argument holds ISM to a higher standard than conventional grab samples. The grab sample only represents the physical location sampled, oftentimes this is a 0.5 to

10 cm diameter area. One has to question whether it is appropriate to extrapolate from a single point of several cm in diameter to some much larger area (1 to 100,000 m^2). The hot spot is an inferred physical presence typically based on extrapolation of several points, although in some instances the author has experienced situations where a single elevated grab sample result was inferred to represent a hot spot of much larger spatial dimensions. It does not necessarily mean each location within the hot spot is highly contaminated, although, this is an assumption often made within the environmental industry. Thus, the size of a hot spot is an arbitrary physical construct often interpolated to encompass an area based on as little as a single sample point. There is no consensus on the size of hot spot of importance. Should one be concerned with finding a hot spot the size of a football field, an acre, 10 x 10 m or 1 x1 m? The key component of ISM is defining the size of the hot spot, if this is an objective, before sampling begins.

Figure 1-6 depicts a hypothetical incremental sampling approach for a military range using a single large, 50 x 50 m, DU. A single DU may be appropriate if the final decision involves an action pertaining to the 50 x 50 m range. If the desire is collection of information on the distribution of contaminants on the military range on a scale smaller than 50 x 50-m, then the DU has been selected incorrectly. For example, if one wants to know the contaminants associated with demolition activities or individual detonation craters as shown in Figure 1-6, then a DU with smaller spatial dimensions is in order. One possible approach is to subdivide the area into smaller DUs or subdivide the larger DU into individual SUs. For illustrative purposes, Figure 1-6 depicts 25 individual 10 x 10-m SUs within one larger 50 x 50-m DU. One unique concept of the DU is that there are no right or wrong sized DUs. The size of the DU is solely a function of the objectives of the study and the end use of the data; issues presumably decided upon during the planning stages of a project.

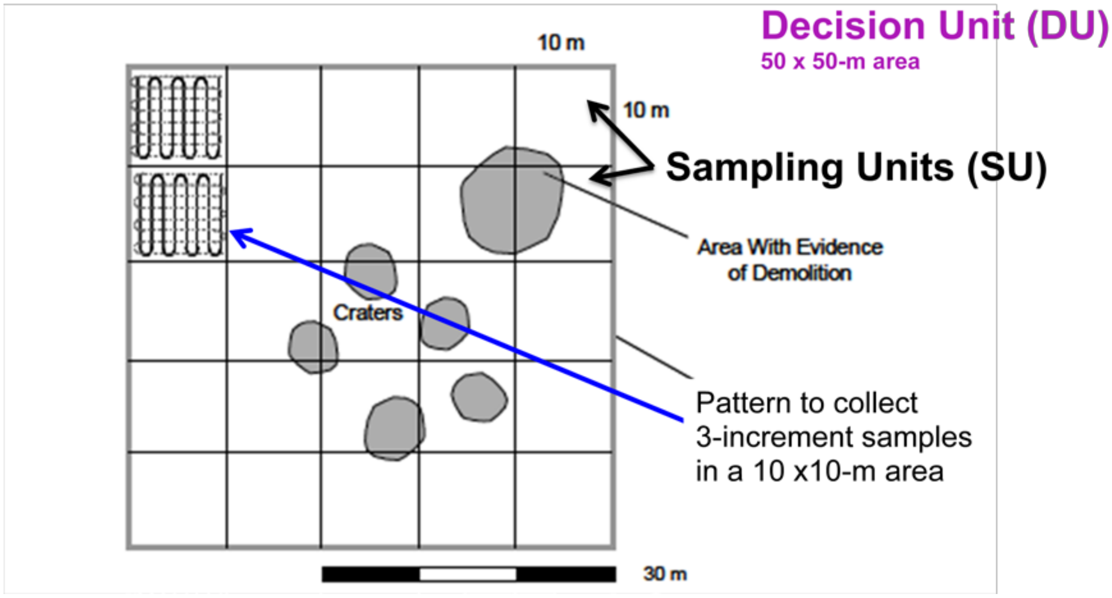


Figure 1-6. Incremental sampling approach with decision units and sampling units.

Given the proceeding discussion on the disadvantages of conventional grab sampling, is there a rationale argument for the use of ISM? The value of ISM appears to encompass a number of variables: (1) it provides a value that is representative of the area of interest from which a direct risk calculation can be performed, (2) the process requires explicit defining of the area of interest (i.e., DU or SU), (3) the approach provides a more reliable and precise estimate of the mean, (4) collection of replicate samples can determine the total sample error, (5) quantification of the error associated with different sampling and processing steps can be determined, (6) fewer number of samples are required as compared to the conventional grab sampling technique, and (7) the fewer number of samples results in lower field sampling, sample preparation, and analysis costs. The following chapters of this dissertation explore these issues and assesses whether the advantages of ISM outweigh the disadvantages of conventional grab sampling. Further, the dissertation evaluates whether there is value in applying ISM to the situation of introduction of solid metallic residues to surface soil as a result of military small-arms training.

1.4 Research Hypothesis

The research hypothesis is that metallic residues (solid particulates) deposited on the berm face of military small-arms ranges occur in an extremely heterogeneous fashion, similar to the distribution of explosive compounds. The heterogeneous distribution of contaminants is not adequately addressed by conventional grab surface-soil sampling, which yields non-representative and non-reproducible sample results when trying to estimate the mean concentration for an area of interest. The secondary hypothesis is an alternative field surface soil sampling approach loosely based on compositing techniques, including modification to laboratory sample preparation procedures, is required to yield representative and reproducible estimates of the population mean.

1.5 Dissertation Research Significance

The findings of the present dissertation have had and will continue to have a significant impact on the environmental industry, DoD, Army, State, and Federal environmental regulatory agencies. There are approximately 3,000 small-arms ranges managed by the DoD (ITRC 2005) and 9,000 non-military ranges in the U.S. (USEPA 2005). The DoDs interest in sampling their ranges stems from directives mandating environmental assessments of the impacts from munitions deposition at all active DoD facilities including training and testing ranges (USDOD 2007, 2005). I approached the USEPA in 2012 to make them aware of my research on military small-arms ranges in light of potential upcoming updates to the USEPA SW-846 Methods manual, i.e. *Test Methods for Evaluating Solid Waste, Physical/Chemical Methods, office of Solid Waste and Emergency Response*. As a consequence of these discussions, the USEPA agreed to convene a SW-846 Inorganic Workgroup to discuss changes to USEPA Method 3050B, *Acid Digestion of Sediments, Sludges, and Soils,* in late 2012 (USEPA 1996a). The last update of Method 3050B occurred in 1996. Since 2012, the Inorganic Workgroup, which consists of representatives from each of the USEPA regions, commercial environmental laboratories, academic institutions, and government research laboratories, has been making modifications to Method 3050B. Research from the present dissertation has been the impetus for the method re-write, as well as the inclusion of an appendix that will provide sampling and sample processing guidance on the specific situation of soils containing metallic residues. The Method 3050B update is anticipated to be issued in 2015 as Method 3050C as part of update VI for SW-846 *Test Methods for Evaluating Solid Waste, Physical/Chemical Methods, office of Solid Waste and Emergency Response.* Method 3050C when issued will have a major impact on how environmental consultants collect samples in the field as well as how commercial environmental analytical laboratories prepare the samples prior to analysis. Further, it is anticipated many state environmental agencies will adopt or incorporate the Method 3050C procedures into their respective rulemaking.

In addition, the US Army Corps of Engineers (USACE) developed guidance for sampling MMRP sites using the incremental sampling approach, including a section focused on military ranges having soils with metallic residues present (USACE 2009). This section on sampling soils with metallic residues provides information based on my input and dissertation findings on the special soil sampling and sample processing techniques needed for ranges where anthropogenic metallic residues are expected. Also, the DoD, with the my input, is presently developing guidance for all services for implementation of the ISM technique for all projects under the MMRP such as Formerly Used Defense Sites, Base Reduction and Closure, Formerly Utilized Sites Remedial Action Program, and active sites in any phase of assessment from preliminary investigations through remedial action. This document is anticipated to be issued by the DoD under the Assistant Chief of Staff for Installation Management in FY15.

Because of my dissertation research and input to the USEPA SW846 Method 3050B changes, contribution to the USACE and DoD guidance documents, and participation on the ITRC ISM team, I have been increasingly sought out within USACE on how to investigate contaminated sites. For example, the Baltimore District of USACE recently sought my expertise on how to sample a former strategic metals storage site, which led to my involvement in developing the sampling design. One significant impact of my dissertation research is the acceptance by the risk assessors with both the USACE and the Maryland Department of Environmental Protection of the idea of using a single incremental sample from an area to calculate risk. As discussed in Section 1.3, the risk assessment calculations require an assumption of exposure over some defined area. However, until recently these calculations involved using grab samples. Individual grab results were often combined through interpolation methods to derive a risk number for an area or in an effort to be conservative; a single grab value of the highest concentration was used to calculate risk and then specified to be representative over some arbitrarily defined area. The acceptance of the incremental sampling approach and application over a defined area for the strategic metals storage site is based on the research associated with my dissertation.

Finally, one of the questions raised by the findings from my dissertation is the impact on milling on metals bioavailability assessments. The Army is concerned that milling of a soil sample increases the metal surface area, and consequently, the metals digestion efficiency, as well as exposing metal from within the soil grain normally not available to the digestion process. Consequently, an increase in soil metal concentrations is expected with milling that is not representative of the conditions in the field. Consequently, the U.S. Army Environmental Command recently funded me to conduct a 2-year study to quantify the metals concentration increase from milling and the associated impact on several common metals bioavailability tests.

1.6 Dissertation Overview

The dissertation research assessed whether the conventional grab surface soil methodologies used by the environmental industry are appropriate for the characterization of metallic residues (solid particulates) introduced into the environment because of military training. The underlying basis for the existing conventional field grab sampling and sample preparation methodologies is the assumption of an aqueous contaminant release to the environment.

The study first involved collection of surface soil samples from four different small-arms ranges located at military installations across the U.S. using the conventional grab sampling methodology (USEPA 2000a, 1995, 1991, 1984) and sample preparation following USEPA Method 3050B (USEPA 1996a). The objective of this phase of research was to determine the degree of heterogeneity associated with metallic residue deposition at military small-arms ranges. Evaluation of the data involved the use of statistical tools such analysis of variance, bias, and t-test. The dissertation findings indicated a high degree of heterogeneity associated with metallic residue deposition at military small-arms ranges.

The second objective of this research was to assess whether conventional grab surface-soil sampling was appropriate for situations where the distribution of the contaminant of interest (e.g., metals, occurs in the environment in a highly heterogeneous distribution). The results indicated conventional judgmental surface-soil grab sampling approaches yields data with uncertainties greater than the measured value for soils with metallic residues introduced into the environment from military small-arms training.

The large grab sample error led to the third objective; an exploration of modifications to the field sampling approach and sample preparation procedures that would improve the precision and accuracy of the results. Chapters 2 - 5 provide a presentation of the various field and laboratory sample preparation methods tested and associated results. Additionally, Chapters 2 - 5 have been submitted for publication in peer-reviewed professional journals.

Chapter 2 of the dissertation discusses an evaluation of conventional grab surface soil sampling techniques from four small-arms range impact berms and assesses the total sample error through statistical analysis and simulation studies. The results indicated conventional grab surface-soil sampling and sample preparation methodologies, used for the past 30 years by the environmental industry, yield results with very large uncertainties for soils where the contaminant has been introduced as a metallic residue. Reduction of the sampling error is possible through collection of an adequate number of grab samples, although project budgets typically limit the number of samples collected. Simulation studies suggested for a typical small-arms range that 100 grab samples are necessary to yield a result with an acceptable degree of uncertainty. However, to date, there has been a general lack of focus on data quality issues other than analytical performance. Therefore, an insufficient number of grab samples are typically collected to satisfy the requirements of using the grab sampling approach to yield quality results representative of the conditions in the field.

Chapter 3 compares the results of conventional grab surface soil sampling with ISM to see if the latter is a reasonable. A related issue evaluated is how many increments are necessary to achieve a representative surface-soil sample from the area of interest using ISM. Again, the primary tools for data analysis are statistical. The results indicated ISM yields a higher data quality with a lower total sampling error concomitant with a higher degree of sample reproducibility and precision than the conventional grab sampling methodology, when the contaminant is a solid metallic residue. In addition, the findings indicated a minimum of 30 increments are necessary per ISM sample to reduce the total sample error and diminishing returns are evident with collection of over 50 increments. Data quality improvements are also possible with fewer ISM samples than conventional grab samples.

Chapter 4 addresses field sampling issues associated with the adoption of ISM. ISM yields soil samples consisting of 0.5 to 5 kg of material versus several hundred grams for conventional grab samples. Because most environmental laboratories are ill suited to handle large volumes of soil cone-and quartering along with a rotary splitter were assessed for their appropriateness to reduce the sample volume/mass. The results indicated field-splitting methods are not appropriate and do not yield equivalent, reproducible, or representative sample splits prior to sample preparation in the laboratory. Although collection of incremental samples improved data quality by reducing the total sample error, the level of reduction by this process alone is not sufficient to meet most data quality objectives.

Chapter 4 addresses whether a larger digestion mass is sufficient to improve data quality in lieu of milling. ISM embodies both incremental sampling in the field as well as additional sample preparation steps in the laboratory (e.g., milling and subsampling). Collection of larger digestion aliquots coupled with field incremental sampling was insufficient to overcome the error associated with sample heterogeneity. Consequently, modifications to the sample preparation procedures employed in the laboratory (e.g., milling and sub-sampling following milling) are necessary to reduce the total sampling error to acceptable levels and that simply collecting a multi-increment is inadequate.

Chapter 5 discusses the use of milling equipment to reduce the particle size of the soil and metallic residues to improve the equivalent distribution of the metallic contaminant throughout the sample. Milling increases the probability a digestion aliquot has the same contaminant to soil proportion as the total sample and the DU. Since a variety of milling equipment and materials is commercially available, the evaluation involved testing several different devices. The results indicated milling is necessary to reduce the total sample error. In addition, the evaluation of the milling time interval assessed the impact on total sample error for the Puck and Roller Mills. Puck, Ring and Puck, and Roller/Ball Mills all yielded satisfactory reductions in the total sample error as compared to an unground soil sample. The mortar and pestle, although yielding an improvement in total SE, was unsatisfactory in comparison with the reductions possible with the milling equipment. The results also indicated for small-arms range metals (antimony, copper, lead, tungsten, and zinc), a milling interval of 5-min is necessary with the Puck or Ring and Puck Mills and 18 hrs for the Roller Mill.

Chapter 6 provides a summary of the dissertation findings presented in Chapters 2 through 5.

CHAPTER 2

POOR PRECISION OF CONVENTIONAL GRAB SAMPLES FOR CHARACTERIZATION OF METAL CONTAMINATION AT SMALL-ARMS RANGES: IMPACT ON UPPER CONFIDENCE LIMIT CALCULATIONS [12]

2.1 Abstract

Metal fragments from routine military training are heterogeneously distributed at small-arms ranges. Environmental characterization of small-arms ranges typically relies on grab sampling despite the fact that it may over or under represent metallic residues in soils. To address this, a study of conventional grab sampling methods for soils containing metallic residues involved the collection of 30 to 48 grab samples from four small-arms-range impact berms (Fort Wainwright, Alaska: Fort Eustis, Virginia: Camp Ethan Allen, Vermont: and Kimama Training Site, Idaho). The sites represent a variety of soil types and physiographic locations. Surface soils were analyzed for 25 metals with only four (antimony, copper, lead, and zinc) likely introduced from the use of small-arms projectiles. Statistical analyses of the data for each of the four sites indicated large variances with positively skewed, non-normal distributions for the four

---

[12] This chapter has been submitted for publication in: Clausen, J., T. Georgian, K. Gardner, and T. Douglas (2014). “Poor Precision of Conventional Grab Samples for Characterization of Metal Contamination at Small-Arms Ranges: Impact on Upper Confidence Limit Calculations.” *Risk Analysis.*

small-arms-related metals. Field duplicates for the anthropogenic metals also exhibited poor agreement. These results are consistent with the presence of metallic nuggets or fragments having a random and highly heterogeneous spatial distribution. In contrast, the 21 native metals (e.g., iron, manganese, etc.) yielded results with near normal distributions and low variances. Field duplicate results suggest better precision and less heterogeneity for native metals than for anthropogenic metals.

Computer simulations (re-sampling methods) using a set of lead results from 48 grab samples (collected systematically from a small-arms impact area) were conducted to assess the reproducibility of conventional grab sampling. The simulations entailed random selections of subsets of m concentration measurements from the set of 48 lead results and calculating sample means or 95% upper confidence limits (UCLs) of the means. For the sample size $m = 7$, the calculated 95% UCLs spanned nearly three orders of magnitude (e.g., 50 to 40,000 mg/kg). The variability of the sample means markedly increased as m decreased from 35 to 7. The simulations suggest small grab sample sizes can produce highly variable and uncertain estimates of population means resulting in potential gross underestimation of the extent of metal contamination. Large numbers of grab samples or a different a sampling approach such as composite or incremental sampling is necessary to characterize soils containing metallic residues.

2.2 Introduction

The DoD conducts environmental sampling to assess the loading of munition constituents (e.g., metals and energetics) at its training ranges. Training activities typically involve small-arms as well as larger projectiles such as artillery and rocket weapon systems. Deposition processes, which are often heterogeneous, occur at a variety of scales, involving impact areas ranging from a few tens to thousands of acres (Jenkins *et al.* 2005a). Consequently, the ability to accurately identify and quantify the loading or spatial deposition of training range materials is a major challenge. Limited physical access to training ranges further restricts the collection of samples representative of background and anthropogenic loadings of metals and energetics. Budgetary constraints are a significant limitation restricting the number of samples collected and analyzed (ITRC 2012). Consequently, environmental practitioners need tools to develop and to apply defensible sampling designs that yield informative and statistically significant results with the fewest number of samples possible.

Until recently, the characterization of surface soils for environmental purposes was primarily accomplished using conventional grab or discrete sampling techniques conducted in a judgmental (e.g., biased) or random manner (USEPA 2002b, 1995, 1991, 1984). However, over the past decade, research has shown that energetic residues released into the environment as particulates, result in extremely spatially heterogeneous distributions (Hewitt *et al.* 2009, 2007, 2005a; Jenkins *et al.* 2005a, b, 2001, 1998, 1997a, b, 1996). This body of research also suggests when particulates are present in soils, grab sampling does not accurately characterize mass loading (Pitard 1993; Gy 1992, 1982). As soils at military small-arms ranges often contain metallic particles (e.g., bullet fragments), the efficacy of conventional grab sampling for characterizing residual metal contamination has been recently questioned (Clausen *et al.* 2012a, b; Hewitt *et al.* 2012). Previously, Clausen and Korte (2009a) showed anthropogenic metals are heterogeneously distributed on small-arms ranges. A variety of metals used in military munitions antimony, copper, lead, and zinc are usually of interest at small-arms ranges (Clausen and Korte 2009a).

The typical goal of sampling for environmental characterization involves estimating the mean, usually by calculating a 95% UCL of the mean (USEPA 2013, 2002a, 2000b, 1992, 1989). The absence of unacceptably high contaminant concentrations warranting remedial action is demonstrated when the 95% UCLs are less than pre-defined decision limits for human or ecological risk. To obtain reproducible and meaningful estimates of the mean, the sampling design and laboratory analytical methods need to account for compositional and distributional heterogeneity (Pitard 1993; Gy 1992, 1986).

There is concern as to whether conventional grab sampling yields a reliable estimate of the mean for study areas when the soils contain metallic residues or metallic particles (ITRC 2012; Hewitt *et al.* 2009). For example, Hadley and Mueller (2012) have argued that a grab sample is representative of only a “sampling point”. This suggests that when a high degree of contaminant heterogeneity is present a sample result is only representative of the aliquot of material analyzed at the laboratory, typically 0.5 to 2 g. Therefore any inferences about the ”true” metal concentration in the remaining portion of the soil in the sample container are not possible unless a sufficient number of replicate subsamples are analyzed to obtain a reliable estimate of the mean of the entire collected sample. However, multiple analyses of material in an individual sample container is typically not practical because of cost constraints; additionally the total number of analyses would increase dramatically if multiple analyses were required for each sample collected. If owing to heterogeneity, a single increment (subsample) is not representative of an entire sample, then the common practice of inferring contaminant concentrations at adjacent locations from the concentration reported for a single laboratory sub-sample is flawed. Similarly, the validity of making inferences about the total sample population based on a few analytical results is questionable (Hadley *et al.* 2011). Yet, the collection of grab samples is the norm within the environmental industry and has been since the inception of environmental soil sampling (USEPA 2000a, 1995, 1991, 1984).

When conducting environmental testing, it is a common practice to evaluate the quality (precision) of grab sampling results using duplicate samples. Usually, duplicate sampling involves collection of one sample for every 10 to 20 environmental samples. Duplicate samples are two aliquots of a parent material (sample) “mixed” in the field prior to being divided for separate laboratory analyses. In instances where samples cannot be split in the field (e.g., for analysis of volatile contaminants) “co-located” samples are collected, which consist of two samples very near one another. The duplicate or co-located samples are used to make inferences about within-sample variability or, less commonly, sample-to-sample variability. However, representative split samples require rigorous sample processing procedures not readily implementable in the field (e.g., particle size reduction using a high-energy mechanical grinding device).

The basis for inferences about within-sample precision using co-located duplicate samples typically relies on the implicit assumption of a strong spatial correlation. However, studies conducted for energetic materials released into the environment have shown that the particulate distribution is random and that autocorrelation between sample points is not a valid assumption (Jenkins *et al.* 1997a, b, 1996). Further, if a grab sample is unique to the physical point from which it is collected, then the gathering of duplicate or co-located samples, spatial mapping using kriging, or variogram data analyses are likely not appropriate. Removal of the sample from the environment makes it impossible to obtain a true duplicate sample because the sample point no longer exists (Hadley and Mueller 2012). An adjacent co-located sample or duplicate containing a heterogeneous distribution of particulates is not related to the original sample as shown by the non-agreement between the original and duplicate sample results (Walsh *et al.* 2005; Jenkins *et al.* 1997a,b, 1996).

Observation of a large duplicate disparity rarely results in an attempt to determine the source of the discrepancy and its affect on the representativeness of the data. Evaluation of precision typically involves using arbitrary acceptance criteria that can affect data usage or decision-making. A common practice entails dismissal of a large disagreement, such as an “outlier” or anomalous result. In some cases, to be “conservative” when conducting a risk analysis, the larger value is accepted and the lower value discarded; in other instances, the results may be averaged.

Grab sampling results from small-arms-ranges are often used to describe the spatial and temporal depositions of metals. However, the small sample sizes usually used (e.g., n < 8) result in highly variable estimates of the mean (ITRC 2012). This paper discusses statistical resampling conducted through computer simulations to explore the problem of sample reproducibility (precision), an approach used because physical resampling is not practical. The number of grab samples required to yield an accurate representation of the conditions in the field is several times larger than what many environmental practitioners routinely collect.

2.3 Methods

2.3.1 Soil Sample Collection

I collected surface soil samples for this study from (1) Range 16 Records Range located at the Fort Wainwright, Alaska Small-arms Range Complex; (2) Range 4-3 located at Camp Ethan Allen, Vermont; (3) Northern Berm located at the Kimama Training Site in Kimama, Idaho; and (4) 1000-inch Range at Fort Eustis, Virginia. These four ranges represent a variety of physiographic locations, environmental conditions, surface soil types, munition usage, and length of time since last use.

The Fort Wainwright Small-Arms Complex is an active range located on surficial sand and gravel deposits mapped as the "Chena Alluvium" by Péwé *et al.* (1976). The target berms consist of unconsolidated sand and rounded river gravel with very little organic material or vegetation. Grain size analysis of the berm soil, following American Society for Testing and Materials D6913-04 (ASTM 2009) yielded a particle size distribution consistent with silty sand with gravel (Clausen *et al.* 2013). Bullet fragments (5.56, 7.62, and 9 mm projectiles) from training are readily visible near the top of the berm.

The Camp Ethan Allen Range 4-3 berm material consists of loamy sand (Clausen *et al.* 2012a). This active range allowed training with 5.56, 7.62, and 9 mm projectiles.

The Western Berm at the Kimama Training Site consists of soils of volcanic origin; the analysis of two samples yielded a determination of poorly graded sand with silt (Clausen *et al.* 2013). The last known use of this particular range occurred in 1993 with training beginning in 1969. The ordnance used on the small-arms range included 7.62 mm and .45, .22, and .50 cal.

The 1000-inch Rifle Range at Fort Eustis is a former small-arms training range used between 1920 and 1941 for target practice with .22-, .30-, and .45-cal munitions. The berm soils consist of silty sand (URS 2010).

I used several approaches to estimate the number of grab samples (sample size) needed to achieve a representative result. At each of the ranges, the berm dimensions were approximately 3 m high by 100 m long. An USEPA (1995) recommendation involved collecting five grab samples for a grid size of 30 × 30 m, that yields 15 samples in our case. Although, another approach is to collect duplicate samples at each grid node, yielding 27 samples (Griffiths 1971). Another process involved the use of a statistical software program, Visual Sampling Plan (VSP), to determine the appropriate number of grab samples (Matzke *et al.* 2010). Depending on the assumption used in development of the sample design (e.g., tolerances for Type I and Type II errors of 5% to detect difference of one to three standard deviations) VSP calculated a sample size of 10 to 30 grab samples.

The range at Fort Wainwright consists of 16 individual berms at varying distances downrange from the firing point (Figure 2-1). We collected grab samples from all 16 berms at 100- m downrange from the firing point. We sampled at the upper left, upper right, and middle center of each berm, yielding 48 samples (Figure 2-1). In contrast, the other installations contained a single contiguous berm (Figure 2-2). At Camp Ethan Allen, I collected 30 grab samples from the berm by using a systematic central aligned grid design (Gilbert 1987). A similar grid design used at the Kimama Training Site and Fort Eustis yielded 30 and 33 samples, respectively. At each of the four sites, I collected grab samples of soil sufficient to fill a 4- oz amber glass container, approximately 225 g, by using a 2-cm diameter steel CMIST corer (Walsh 2009) to a depth of 5 cm.

Figure 2-1. The small-arms Range 16 Records Range berms at Fort Wainwright. The inset shows where the grab samples were collected from at each of the 16 berms (G = grab sample; # = sample number, A, B, C = location on berm face).

Figure 2-2. An example of a single contiguous berm at Fort Eustis with an overlay of a systematic central aligned grid design.

2.3.2 Sample Preparation and Analytical Procedures

Processing of the grab samples followed USEPA Method 3050B (USEPA 1996a), which involved removing a 1- g sub-sample from the top of the sample jar and performing the acid digestion procedure. Method 3050B does not mandate sieving soil samples or present any specific requirements for homogenization or sample processing. The U.S. Army Corps of Engineers, Engineer and Research and Development Center (ERDC), Environmental Laboratory located in Vicksburg, MS followed USEPA Method 6010 (USEPA 2006c) in their sample analysis by using a Perkin Elmer ELAN 6000 quadrupole inductively coupled plasma (ICP) mass spectrometer with the factory-supplied Ryton plastic spray chamber and fixed cross-flow nebulizer. The ERDC, Cold Regions Research and Engineering Laboratory located in Hanover, NH, also analyzed some samples and used a Thermo Scientific iCAP 6000 Series ICP optical emission spectrometer system capable of dual radial and axial plasma views for the 27 metals analyzed, following USEPA Method 200.7 (USEPA 1994b).

Inferences about sample representativeness involved an evaluation of the reproducibility of the concentration measurements. Calculating the sample CV allowed a qualitative evaluation of precision. The CV, which is the ratio of the sample standard deviation to the (sample) mean, is a normalized measure of dispersion, often referred to as the RSD. Because the “true” (population) mean metal concentration of each of the berms was unknown, a determination of bias was not strictly possible. An evaluation of apparent bias assumed the entire set of samples collected from each study area was sufficient to approximate the mean.

2.4 Discussion and Results

2.4.1 Distributions of Metals

A statistical summary of the results for the four small-arms-range berms indicates that the four anthropogenic metals of interest (antimony, copper, lead, and zinc) consistently exhibit the highest variability of the analytes measured and have positively skewed distributions (Table 2-1). For example, the mean and median lead grab sample values (432 and 85.7 mg/kg, respectively) at Fort Wainwright differ by nearly a factor of five, indicating the distribution of grab lead concentrations is very positively skewed, (Figure 2-3). Similar results are evident for the distribution of lead at the three other berms (Table 2-1). In addition, similar variances and skewed distributions are evident as well for antimony, copper, and zinc (Table 2-1), also. The apparent skewness in the data can have many causes, such as biased sampling, multiple populations, or outliers (Singh *et al.* 1997). The data sets used in this study did not have any outliers attributable to noncompliant sampling or analytical procedures.

The metals lead, copper, zinc and antimony are known constituents of small-arms munitions and the positive skewed distributions are consistent with spatially heterogeneous distributions of metallic particles (e.g., bullet fragments) released in soils from small-arms use. Evidence for the assumed native metals includes distributions more Gaussian with much smaller variances (Table 2-1). As an example, the sodium histogram for Fort Wainwright (Figure 2-3) suggests a relatively normal distribution. The distributions of native metal concentrations are the result of natural sedimentary fluvial depositional processes and the observed variability is small relative to anthropogenic metals (Table 2-1).

The skewed distributions for lead, copper, zinc, and antimony likely resulted from a combination of anthropogenic contamination from small-arms use (e.g. producing the extreme values near the right tails of the distributions) and the presence of metals in native soils (e.g., resulting in some of the smaller reported values). The anthropogenic metals exhibited the highest RSDs (up to 350%) at each of the four sites; RSDs for the native metals were much smaller, typically less than 30% (Table 2-1). Consistent with the known anthropogenic metals and their mode of distribution into the environment, the analytical results suggest an extremely heterogeneous and random distribution of metal contamination. Skewness in a dataset is common when the underlying distribution is inherently heterogeneous (Jenkins *et al.* 1997a, b, 1996). Earlier studies have documented this difference in CV with anthropogenic contamination yielding higher values (White and Hakonson 1979) than non-anthropogenic impacted soils (Mausbach *et al.* 1980, Harrison 1979, Mathur and Sanderson 1978).

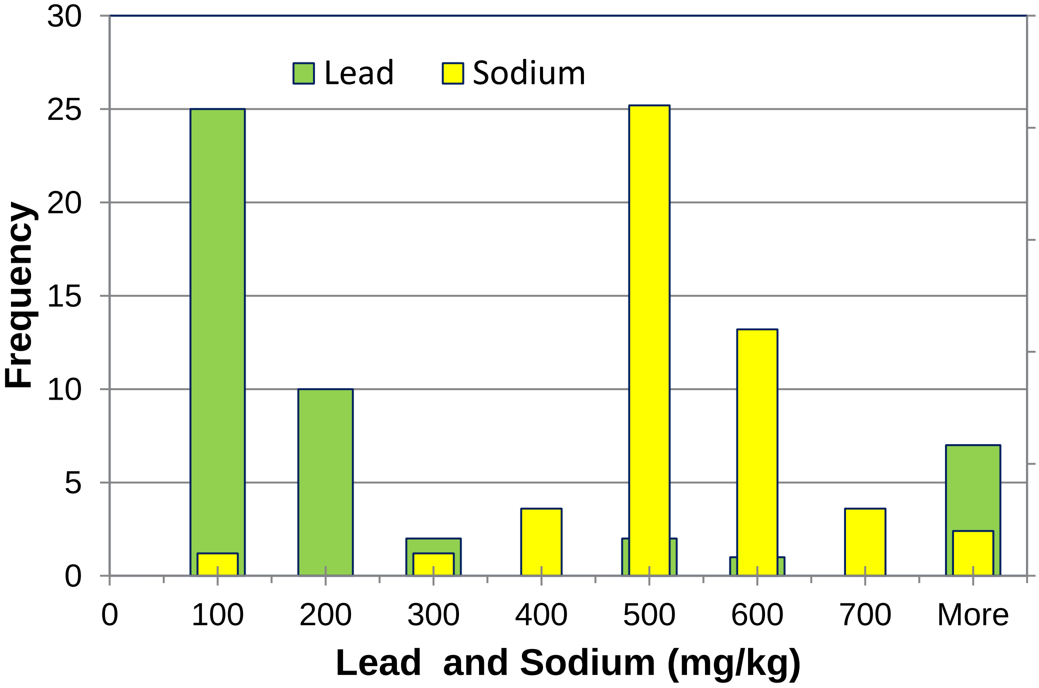


Figure 2-3. Distribution of lead and sodium results for grab samples collected from the berm face of a military small-arms range at Fort Wainwright.

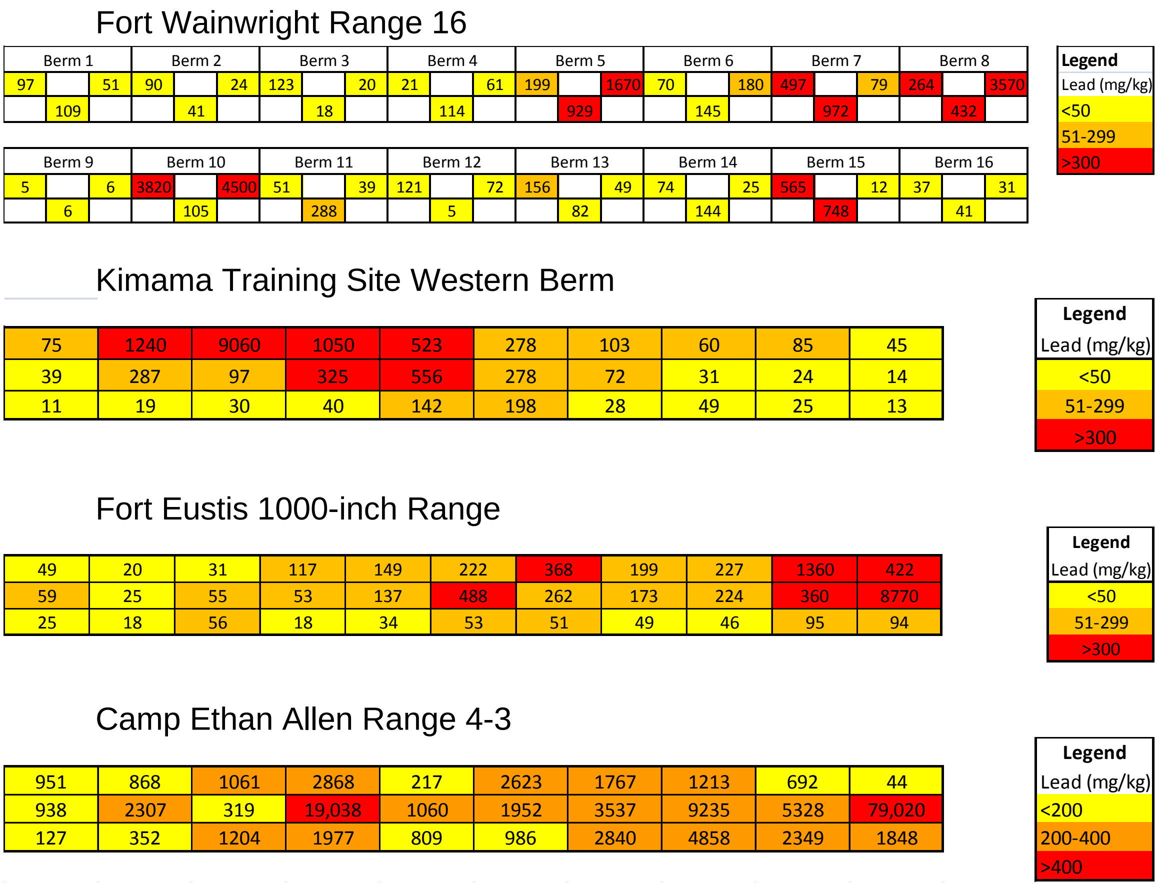
Figure 2-4 shows the spatial distribution of lead (mg/kg) results for the berms (Figure 2-1) at Fort Wainwright, Alaska with berms 1 through 8 representing the left side of the range and berms 9 through 16 the right. The results indicate most of the elevated lead values occur in the center of the berm although there are exceptions such as the elevated levels observed at Berm 15. The random heterogeneous spatial distribution for lead, antimony, copper, and zinc at the other three berms (Fort Eustis, Kimama Training Site, and Camp Ethan Allen) is similar to the distribution at Fort Wainwright (Figure 2-4 and Figures S1, S2, S3, and S4 in the Appendix).

Table 2-1. Summary of grab sample results for the four small-arms range berms sampled.

	Sample Mass (g)	Ag	Al	As	Ba	Be	Ca	Cd	Co	Cr	Cu	Fe	K	Mg	Mn	Na	Ni	P	Pb	Sb	Se	Sr	Si	S	Tl	V	W	Zn
Range 16 Records Range, Fort Wainwright, Alaska																												
n	48	NA	48	47	48	NA	48	NA	47	48	48	48	48	48	48	48	48	48	48	8	0	NA	48	48	NA	48	NA	48
Mean (mg/kg)	159	NA	9980	8.85	104	NA	5566	NA	9.55	18.9	81.0	18143	1005	5235	291	475	21.1	508	432	14.0	<2.00	NA	53.7	114	NA	34.8	NA	52.6
Median (mg/kg)	162	NA	10100	8.58	104	NA	5535	NA	9.47	19.2	27.5	18400	1015	5260	293	481	21.7	509	85.7	7.41	<2.00	NA	49.2	98.6	NA	35.4	NA	47.9
Minimum (mg/kg)*	103	NA	1060	6.23	10.6	NA	529	NA	7.24	2.04	2.58	1960	112	558	29.5	44.7	2.26	50.8	5.01	2.07	<2.00	NA	3.83	9.13	NA	3.59	NA	4.88
Maximum (mg/kg)	197	NA	14200	12.7	170	NA	9010	NA	11.4	22.6	852	22700	1340	6540	384	744	25.3	673	4500	32.6	<2.00	NA	135	581	NA	41	NA	146
STDEV (mg/kg)	19.1	NA	1655	1.36	19.5	NA	1309	NA	0.766	2.78	177	2820	175	817	48.9	117	3.14	84.8	978	13.5	NC	NA	22.5	78.7	NA	5.32	NA	22.3
RSD (%)	12	NA	17	15	19	NA	24	NA	8	15	218	16	17	16	17	25	15	17	226	97	NC	NA	42	69	NA	15	NA	42
Range 4-3, Camp Ethan Allen, Vermont																												
n	30	NA	30	NA	30	NA	NA	30	30	30	30	30	30	30	NA	NA	30	30	30	30	NA	30	NA	NA	NA	30	30	30
Mean (mg/kg)	144	NA	4219	NA	20.0	NA	NA	0.760	7.40	6.22	300	12381	1976	197	NA	NA	9.91	710	5060	87.8	NA	12.9	NA	NA	NA	11.2	0.799	66.1
Median (mg/kg)	146	NA	4231	NA	17.4	NA	NA	0.735	7.44	6.19	270	12380	1967	195	NA	NA	10.0	722	1238	10.0	NA	12.1	NA	NA	NA	11.2	0.745	61.9
Minimum (mg/kg)	122	NA	3784	NA	15.7	NA	NA	0.620	6.31	5.58	69.8	9975	1748	175	NA	NA	8.67	468	43.9	0.898	NA	9.86	NA	NA	NA	9.14	0.400	35.8
Maximum (mg/kg)	159	NA	4577	NA	39.5	NA	NA	1.13	8.38	6.99	598	14700	2325	242	NA	NA	10.7	1043	79020	2072	NA	22.8	NA	NA	NA	15.4	1.37	111
STDEV (mg/kg)	10.1	NA	158	NA	5.61	NA	NA	0.104	0.486	0.409	132	797	124	15.4	NA	NA	0.439	116	14438	375	NA	2.72	NA	NA	NA	1.17	0.266	17.6
RSD (%)	7	NA	4	NA	28	NA	NA	14	7	7	44	6	6	8	NA	NA	4	16	285	427	NA	21	NA	NA	NA	10	33	27
Kimama Training Site western berm, Kimama, Idaho																												
n	30	0	30	29	30	0	30	0	30	30	30	30	30	30	30	30	30	30	30	4	1	NA	30	30	0	30	NA	30
Mean (mg/kg)	100	<2.00	6835	3.51	68.7	<2.00	1776	<2.00	4.50	15.8	23.0	9906	1768	2179	198	65.9	10.5	542	493	19.6	2.21	NA	167	92.0	<2.00	18.0	NA	45.4
Median (mg/kg)	99	<2.00	6955	3.31	69.4	<2.00	1755	<2.00	4.6	15.9	18.1	10050	1835	2185	198	65	10.7	527	73.5	3.02	2.21	NA	177	80.5	<2.00	18.2	NA	45
Minimum (mg/kg)	66.5	<2.00	3030	2.62	30.3	<2.00	892	<2.00	2.16	7.81	9.8	4690	799	1030	94	29	5.15	243	11.1	2.16	2.21	NA	85.3	47.9	<2.00	8.35	NA	22
Maximum (mg/kg)	135	<2.00	8150	7.8	89.2	<2.00	2580	<2.00	5.09	18.5	74	11600	2120	2530	278	89	12.2	727	9060	70.2	2.21	NA	236	195	<2.00	20.8	NA	56
STDEV (mg/kg)	17	NC	908	0.950	10.4	NC	339	NC	0.539	1.88	15.2	1187	244	276	28.8	10.8	1.26	95.0	1645	33.7	NC	NA	47.3	39.8	NC	2.26	NA	6.88
RSD (%)	17	NC	13	27	15	NC	19	NC	12	12	66	12	14	13	15	16	12	18	334	172	NC	NA	28	43	NC	13	NA	15
1000-inch range, Fort Eustis, Virginia																												
n	33	32	33	33	33	33	33	33	33	33	33	33	33	33	33	NA	33	33	33	8	33	NA	0	0	7	33	NA	33
Mean (mg/kg)	78.8	0.707	7129	1.36	40.3	0.242	1547	0.770	2.75	8.11	43.3	6577	572	673	196	NA	9.75	192	434	11.0	2.05	NA	<2.00	<2.00	0.318	19.8	NA	28.6
Median (mg/kg)	22.0	0.450	7300	1.26	38.8	0.23	1160	0.74	2.67	8.53	13	6540	559	647	201	NA	6.72	168	94.3	1.01	1.98	NA	<2.00	<2.00	0.32	18.4	NA	28
Minimum (mg/kg)	1.00	0.11	4860	0.79	28.3	0.13	464	0.5	1.82	4.42	7.2	4460	396	483	35	NA	4.07	125	17.6	0.023	0.79	NA	<2.00	<2.00	0.080	11.3	NA	21.0
Maximum (mg/kg)	116	3.66	10500	2.34	52.7	0.41	6330	1.18	4.14	10.9	755	9440	751	1080	768	NA	92.8	403	8770	69.6	3.26	NA	<2.00	<2.00	0.5	33.6	NA	48
STDEV (mg/kg)	19.9	0.792	1288	0.398	7.22	0.069	1143	0.164	0.614	1.47	129	1258	90	108	137	NA	15.1	64.1	1517	24.0	0.630	NA	NC	NC	0.140	5.38	NA	6.47
RSD (%)	25	112	18	29	18	28	74	21	22	18	298	19	16	16	70	NA	155	33	350	219	31	NA	NC	NC	44	27	NA	23

NA - not analyzed, NC - not caluclated, RSD - percent relative standard deviation, STDEV - standard deviation

Minimum (mg/kg)* is the minimum reportable concentration



Fort Wainwright Range 16

Berm 1			Berm 2			Berm 3			Berm 4			Berm 5			Berm 6			Berm 7			Berm 8		
97		51	90		24	123		20	21		61	199		1670	70		180	497		79	264		3570
	109			41			18			114			929			145			972			432	

Berm 9			Berm 10			Berm 11			Berm 12			Berm 13			Berm 14			Berm 15			Berm 16		
5		6	3820		4500	51		39	121		72	156		49	74		25	565		12	37		31
	6			105			288			5			82			144			748			41	

Legend
Lead (mg/kg)
<50
51-299
>300

Kimama Training Site Western Berm

75	1240	9060	1050	523	278	103	60	85	45
39	287	97	325	556	278	72	31	24	14
11	19	30	40	142	198	28	49	25	13

Legend
Lead (mg/kg)
<50
51-299
>300

Fort Eustis 1000-inch Range

49	20	31	117	149	222	368	199	227	1360	422
59	25	55	53	137	488	262	173	224	360	8770
25	18	56	18	34	53	51	49	46	95	94

Legend
Lead (mg/kg)
<50
51-299
>300

Camp Ethan Allen Range 4-3

951	868	1061	2868	217	2623	1767	1213	692	44
938	2307	319	19,038	1060	1952	3537	9235	5328	79,020
127	352	1204	1977	809	986	2840	4858	2349	1848

Legend
Lead (mg/kg)
<200
200-400
>400

Figure 2-4. Lead results for grab samples collected from the berm face of Range 16 at Fort Wainwright; Kimama Training Site; 1000-inch Range 16 at Fort Eustis; and Range 4-3 at Camp Ethan Allen.

A comparison of co-located samples, collected within 0.5 m of each other at Range 4-3 at Camp Ethan Allen, indicate limited short-range spatial variability for the native metals (Table 2-2). The relative RPDs for the anthropogenic metals are larger than for the native metals by approximately an order of magnitude, suggesting much larger short-range variability. The RPD is the absolute value of the difference between duplicate results divided by the mean and is a more common measure of precision for environmental chemical analyses than the RSD (for duplicates, the RPD = RSD × $2^{1/2}$). Based on an initial evaluation of the metal results, the larger RPDs for the anthropogenic metals may not appear particularly problematic. However, RPDs can vary by several orders of magnitude as suggested by previous studies (Clausen *et al.* 2012a, 2007, Jenkins *et al.* 2004a, 1997a, b, 1996).

Consider the potential consequences of making a monitoring, management, or cleanup decision based on an individual grab sample. In particular, assume application of the USEPA 400 mg/kg lead decision limit for soil to each grid (rectangular cell). In Grid 23, the original lead value is 319 mg/kg but the duplicate result is 479 mg/kg (Table 2-2), yielding two conflicting outcomes for determining whether the concentration of lead is less or greater than 400 mg/kg. The relatively large variances and positively skewed distributions for the anthropogenic metals data from each of the four sites raises a question regarding whether the quality of the data is adequate to obtain reliable estimates of mean concentrations (e.g., calculations of 95% UCLs of the mean).

Table 2-2. Comparison of spatially co-located samples from Range 4-3 at Camp Ethan Allen.

Grid ID	Sample	Mass (g)	Al	Ba	Cd	Co	Cr	Cu	Fe	Mg	Mn	Ni	P	Pb	Sb	Sr	V	W	Zn
Grid 6	Orig.	147	4449	39.5	0.73	7.17	6.28	209	12200	2257	221	10.2	586	692	5.26	16.9	10.2	0.62	60.6
	Dup	147	4206	17.6	0.58	6.84	5.20	392	11070	1922	190	8.92	532	1851	24.1	14.3	8.85	0.69	67.0
	RPD (%)	**0.1**	**1**	**19**	**6**	**1**	**5**	**15**	**2**	**4**	**4**	**3**	**2**	**23**	**32**	**4**	**3**	**3**	**3**
Grid 7	Orig.	152	4117	30.7	1.00	8.38	6.99	248	14700	1877	197	9.93	1043	4858	27.8	11.4	15.4	0.60	60.7
	Dup	154	4065	14.6	0.58	6.83	5.97	280	11090	1805	173	8.87	783	1650	11.9	10.7	9.88	0.65	61.2
	RPD (%)	**0.4**	**0.3**	**18**	**13**	**5**	**4**	**3**	**7**	**1**	**3**	**3**	**7**	**25**	**20**	**2**	**11**	**2**	**0.2**
Grid 15	Orig.	147	4328	18.0	0.85	7.80	6.91	361	13520	1999	197	10.3	832	2623	23.6	11.4	12.9	0.74	74.6
	Dup	156	4412	17.9	0.96	7.47	6.86	270	13180	2062	207	10.5	804	1930	16.4	12.2	12.4	0.80	63.4
	RPD (%)	**2**	**0.5**	**0.2**	**3**	**1**	**0.2**	**7**	**1**	**1**	**1**	**1**	**1**	**8**	**9**	**2**	**1**	**2**	**4**
Grid 22	Orig.	147	4058	16.0	0.73	7.47	6.09	252	12630	1882	178	9.28	734	1204	9.89	12.0	11.8	0.75	60.4
	Dup	150	4131	16.0	0.52	6.56	5.58	224	10800	1934	179	9.29	546	501	3.48	10.1	8.75	0.66	57.6
	RPD (%)	**0.5**	**0.4**	**0.0**	**8**	**3**	**2**	**3**	**4**	**1**	**0.1**	**0.0**	**7**	**21**	**24**	**4**	**7**	**3**	**1**
Grid 23	Orig.	154	4324	17.3	0.71	7.69	6.36	163	12320	1871	193	9.75	813	319	2.69	11.6	11.5	0.49	47.6
	Dup	139	3968	15.2	0.60	6.90	5.56	229	11660	1734	179	8.74	816	479	5.68	12.5	10.4	0.51	52.1
	RPD (%)	**3**	**2**	**3**	**4**	**3**	**3**	**9**	**1**	**2**	**2**	**3**	**0.1**	**10**	**18**	**2**	**3**	**1**	**2**
Grid 30	Orig.	158	4278	17.1	0.71	7.30	5.59	217	12490	1960	196	10.0	623	951	9.26	12.5	11.0	0.92	55.7
	Dup	147	4184	16.0	0.64	7.21	5.95	216	11810	1937	200	9.70	668	555	4.17	11.9	10.0	0.64	55.0
	RPD (%)	**2**	**1**	**2**	**3**	**0**	**2**	**0.2**	**1**	**0.3**	**0.5**	**1**	**2**	**13**	**19**	**1**	**2**	**9**	**0.3**

2.4.2 Upper Confidence Limit Calculations

A typical approach for environmental investigations involves comparing sample maxima or 95% UCLs of the mean (USEPA 1992, 1989) with risk-based thresholds such as the USEPA Regional Screening Levels for soil. Additional remedial action or investigation is necessary when the sample maximum or UCL exceeds the regulatory threshold.

In the current study, I collected 48 grab samples from the berm at Fort Wainwright, a larger number of grab samples than typically collected (ITRC 2012, Hadley *et al*. 2011). Personal observations and discussions with environmental consultants and regulatory officials suggest no consensus on the number of samples needed to characterize a small-arms-range berm. The reason for this non-consensus is the lack of regulatory guidance on the minimum number of grab samples needed (ITRC 2003; USEPA 2000a, 1991, 1984). The required number of samples varies depending on the project objectives and available resources (ITRC 2012; USEPA 1995; Singh et al.1997; van Ee 1990). The various state and federal agencies do offer a variety of statistical approaches to estimate the number of grab samples needed based on the expected variability (e.g., commonly measured by the standard deviation) and the tolerances for (Type I and Type II) decision errors (Barth et al. 1989; Barth and Mason 1984; Gilbert and Doctor 1985; Mason 1983). Unfortunately, statistical methods are not commonly used to determine the number of environmental samples needed. This often occurs for a variety of reasons. Many environmental practitioners lack expertise with statistical methods. In some instances, there is a lack of prior information about the variability of the contaminant concentrations in the study-area soils to estimate the required sample size. In other instances, the variability estimated for the grab samples is so large that it results in a sample size larger than what is affordable to collect. As there is no industry standard, to evaluate the representativeness of smaller numbers of samples than the 48 collected, for our assessments I assumed as few as seven grab samples was adequate to characterize the metal contamination in the small-arms-range berm.

To determine how well seven samples could represent the estimated mean for our dataset, I selected seven lead results at random from the total of 48 grab samples (Table 2-3) by using sampling without replacement (i.e., once I selected a value from the set of 48, I did not select it again) 200 times using the Resampling Stats Version 4.0 for Excel add-on by Resampling Stats, www.resample.com. Using ProUCL Version 5.0, I subsequently calculated a

95% UCL for each of the 200 sets of seven grab samples (USEPA 2013). Table 2-3 shows some descriptive statistics for the calculated 200 UCLs for lead with the median value = 816 mg/kg. Approximately 30% of the UCLs are less than 400 mg/kg (a common decision limit for human exposure to lead in soils), about 20% are less than 300 mg/kg, and 14% are less than 200 mg/kg (see Figure S5 in the Appendix). In contrast, the estimated mean lead concentration of the 48 samples from the Fort Wainwright berm is 432 mg/kg. If I assume 432 mg/kg of lead is approximately equal to the population mean, a sample size of seven grab samples would result in a false negative 64% of the time. The UCLs for lead range from 53 to 35,991 mg/kg, an interval that nearly spans three orders of magnitude. The distribution of UCLs is also positively (right) skewed and exhibits several large outliers (see Figure S6 in the Appendix).

Table 2-3. Summary of descriptive statistics for the 95% upper confidence limit for lead using the Fort Wainwright data set.

Variable	n	Mean	STD[a]	Minimum	Median	Maximum	IQR[b]
95% UCL	200	2915	4238	53	816	35,991	5210

[a] standard deviation, [b]inter quartile range

The 95% UCLs most commonly recommended by ProUCL were the following: Approximate Gamma UCL, Adjusted Gamma UCL, Student's-t UCL, and Chebyshev (Mean, Sd) UCL (Table 2-4) with other types of UCLs selected less than 6% of the time (Figure 2-5). The ProUCL software provides a recommendation on the preferred type of UCL calculation based on the total number of samples, population distribution of the dataset, number of censored samples, etc. (USEPA 2013). For the calculation methods commonly selected, the Adjusted Gamma UCLs and Chebyshev UCLs tended to be the largest UCLs; the median UCL for both of these methods was approximately 6,000 mg/kg lead. The Student's-t method produced the smallest UCLs; the median was roughly 200 mg/kg lead. The Approximate Gamma UCLs tended to be several times larger than the Student's t UCL; the median was about 600 mg/kg.

Table 2-4. Summary of descriptive statistics for the lead 95% upper confidence limits calculated by different methods.

Method	n[a]	Mean	STD[b]	Min[c]	Median	Max[d]
Adjusted Gamma U	48	6169	2611	691	6356	12,983
Approximate Gamma	83	837	812	152	621	5241
Chebyshev Mean Sd	6	1116	856	448	747	2767
H-UCL	1	607	NA[e]	607	607	607
Hall's Bootstrap	5	18,561	12,512	1773	17,805	35,991
Student's t UCL	37	221	147	53	170	642
Chebyshev	20	5455	1718	1613	5928	7268

[a]number of method selections, [b]standard deviation, [c]minimum, [d]maximum, [e]not applicable

The variable UCLs suggest large heterogeneity (e.g., owing to metal fragments in soil) and small sample sizes produce unreliable estimates of the population mean. Although the actual population mean is unknown, a bootstrap method applied to the set of 48 grab samples allowed for calculation of confidence limits of the population mean. I then compared the simulated UCLs with the confidence interval (CI) for the population mean to evaluate bias (i.e., in terms of the number of times the simulated UCLs fall outside the CI, thus, over or under estimating the population mean).

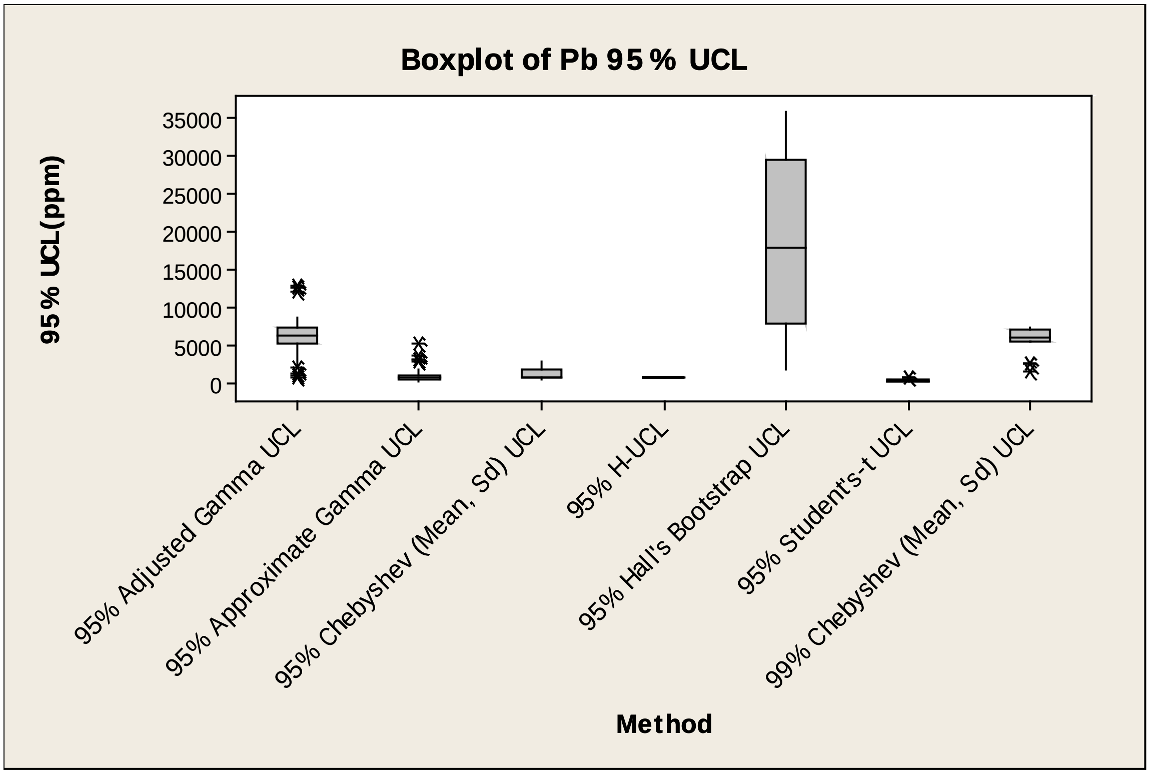


Figure 2-5. Box plot of the lead 95% upper confidence limits by various calculation methods in ProUCL (USEPA 2013).

To calculate confidence limits of the mean, I randomly selected a set of seven results from the set of 48 grab sample results by using sampling with replacement 10,000 times (i.e., I could select each concentration more than once) to calculate a non-parametric bootstrap confidence limit of the population mean. I calculated a one-sided upper 95% confidence limit of the mean (i.e., the 95% of the set of 10,000 means) of 679 mg/kg and a two-sided 95% CI of 193 to 736 mg/kg were calculated. Consequently, the lead population mean is unlikely to be less than 200 or greater than 700 mg/kg, an assumption used to evaluate the simulated UCLs. The UCLs from the simulation over estimate the mean 55% of the time (i.e., 55% of the 200 UCLs exceed 700 mg/kg). About 15% of the

UCLs are at least one order of magnitude larger than 700 mg/kg; and roughly 14% of the UCLs are less than 200 mg/kg, under estimating the population mean. Therefore, 69% of the time, the sets of seven grab samples yield an UCL biased either high or low relative to the population mean; only 31% of the UCLs from the simulation are between 200 and 700 mg/kg. Clearly, seven grab samples are insufficient to provide a representative and reliable population estimate of mean for the area of interest. Therefore, I conducted additional resampling simulations (described below) to estimate how many more samples may be needed.

2.4.3 Resampling Simulations Using Different Sample Sizes

Using the Fort Wainwright lead data, I performed resampling simulations by selecting *m* (sample size) = 5, 7, 10, 15, 20, 25, 30, and 35 randomly from the set of 48 grab results (Table 2-3). For each value of *m*, I repeated the process 300 times. I then calculated the mean of *m* for each of the 300 repetitions and plotted them against *m* (Figure 2-6) to qualitatively assess the variability of the means as a function of the sample size *m*. As shown by the conical pattern for the plotted values in Figure 2-6, the variability is large for small values of *m* and decreases as *m* increases (e.g., as expected from the Central Limit Theorem). The values predominantly fall within the 95% CI of the mean (200 to 700 mg/kg) when the sample size *m* is at least 15 to 30. Also, the frequency of values less than the decision limit of 400 mg/kg increases as *m* decreases. If I assume the population mean is larger than 400 mg/kg (e.g., as the mean of the 48 grabs is

432 mg/kg), it seems likely that sample sizes of less than 15 will produce relatively large false negative rates. An increasing value of *m* resulted in lower error, approximately 50% at $m = 35$, whereas lower values of m had significantly higher errors such as roughly 200% at $m = 10$ (see Figure S7 in the Appendix). Values of $m < 25$ yielded non-normal skewed distributions, making it difficult to represent the mean lead concentrations in surface soils (Figure 2-7). Simulations performed with the anthropogenic metals (copper, antimony, and zinc) yielded similar observations, as did simulations performed with lead and the anthropogenic metals for the other three sites. Consequently, the results from this study suggest the necessity of collecting at least 35 grab samples, which is three times greater than the number of samples often collected, to estimate the mean anthropogenic metal concentrations in soils at small-arms ranges.

2.5 Conclusions

Many environmental characterization studies on small-arms ranges use grab samples to characterize deposition of metals such as lead, copper, zinc, and antimony. However, few studies have used statistical methods to account for the variability of grab sample results when making inferences about the mean metal concentrations in soils from small-arms ranges. This is a function of the number of grab samples for environmental studies commonly driven by cost rather than data quality considerations. The results from this study suggest the necessity of collecting a significantly greater number of grab samples than typically collected to characterize small-arms range berms containing heterogeneous distributions of metallic residues. Conventional grab sampling approaches usually do not result in samples that adequately represent the population mean or yield acceptable precision unless large numbers of samples are collected. The combination of small sample sizes, large variances, and positively skewed distributions for anthropogenic metals collected during this study resulted in large uncertainties for estimates of mean concentrations. The poor reproducibility for duplicate samples also suggests a large degree of uncertainty with spatially co-located grab samples. Either large numbers of grab samples or a different sampling approach (e.g., composite or incremental sampling) will likely be needed for situations where a heterogeneous distribution of metallic residues occurs.

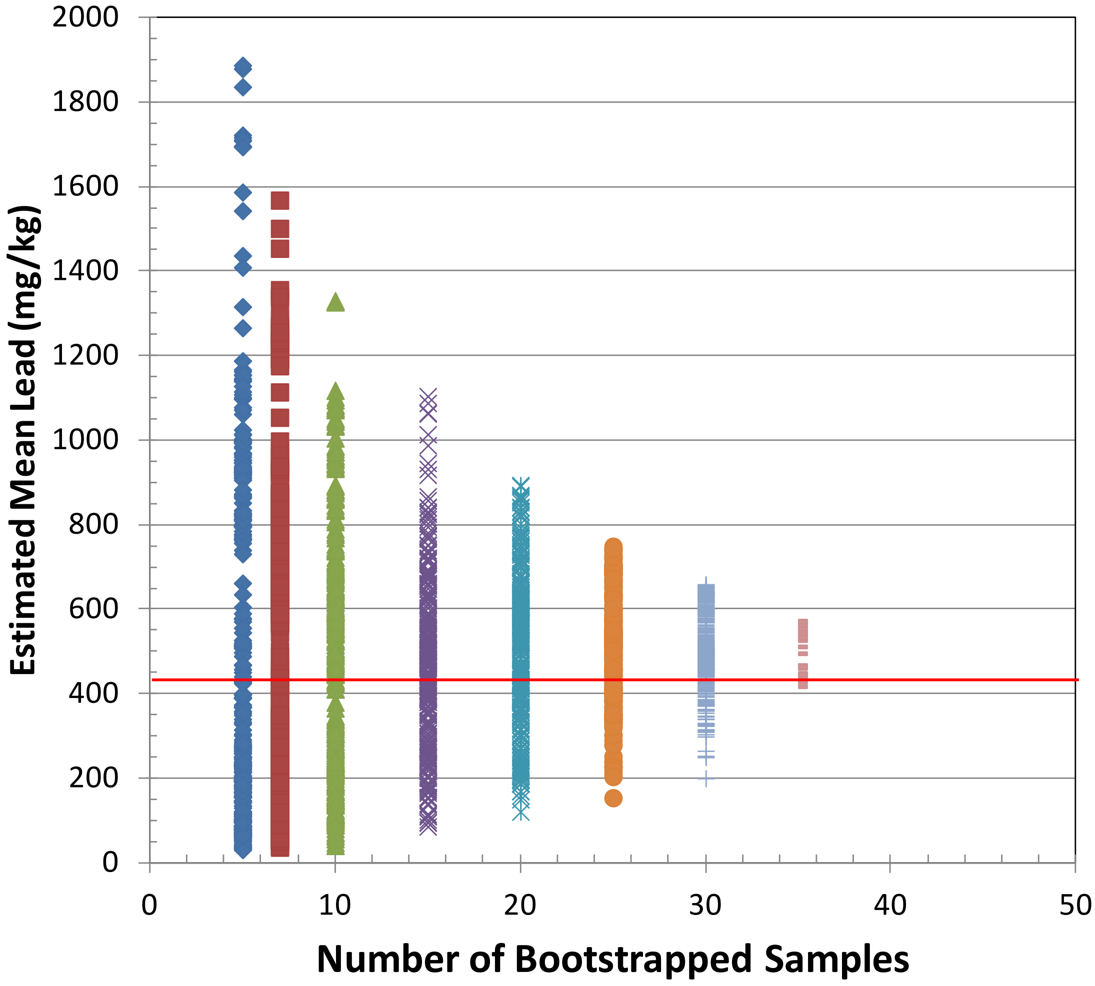


Figure 2-6. Estimated mean lead levels versus the number of bootstrapped samples for Range 16 at Fort Wainwright. I performed 300 hundred simulations for each value of *m* (number of bootstrapped samples).

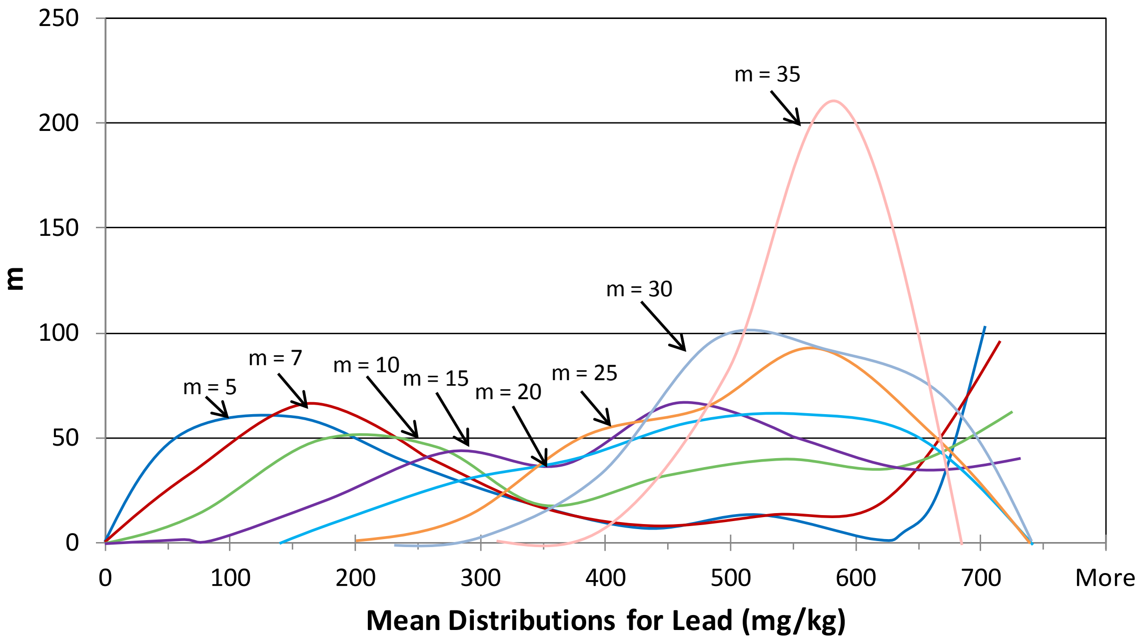


Figure 2-7. Population distribution for lead as a function of *m*, (i.e., number of samples per simulation). I performed 300 hundred simulations for each value of *m* (number of bootstrapped samples).

CHAPTER 3

GRAB AND INCREMENTAL SAMPLE ERROR FOR SOILS CONTAINING METALLIC RESIDUES: IMPACT TO ENVIRONMENT RISK ASSESSMENT [13]

3.1 Abstract

Until recently, the typical technique for collecting surface soil samples to support environmental investigations entailed conventional grab sampling using simple random, systematic, or judgmental (e.g. biased) sampling. All of these approaches have shortcomings when encountering heterogeneously distributed materials. A new sample collection and sample processing approach developed to address the heterogeneous distribution of particulates, principally energetics, led to a technique referred to as the ISM. Increasing acceptance of ISM has led to its application for a variety of contaminated sites and associated contaminants. This study addresses errors associated with the ISM in comparison with conventional grab sampling for soils containing a heterogeneous distribution of metallic residues.

This study used three different sampling designs (systematic random grab, judgmental grab, and multi-increment) to collect surface soil samples from

---

[13] This chapter has been submitted for publication in: Clausen, J., T. Georgian, K. Gardener, and T. Douglas. (2014). "Conventional grab and incremental sample error for soils containing metallic residues: Impact to Environment Risk Assessment." *Soil and Sediment Contamination*.

a small-arms impact berm and analyzed them for 25 metals. Based on the composition of the small-arms projectiles used at the ranges four of the metals were considered anthropogenic in origin: antimony, copper, lead, and zinc. The grab sample results had large variances, positively skewed non-normal population distributions, extreme outliers, and poor agreement between duplicate samples. Good reproducibility for duplicates often was poor even when the grab samples were within several feet of each other. Large extreme values heavily influenced the estimated grab sample means for antimony, copper, lead, and zinc. In contrast, the ISM data sets have small variances, normal population distributions, small variances leading to similar median and mean concentrations, and good agreement between replicate samples.

3.2 Introduction

Characterization of surface soils for environmental purposes typically uses of conventional grab (discrete) sampling techniques conducted in a judgmental (e.g., biased) or random manner (USEPA 2002b, 1995, 1991, 1984). However, at sites with energetic residue deposition, conventional grab sampling methods often yield inconsistent and non-reproducible results (Clausen 2011, 2004; Hewitt *et al.* 2005; Walsh *et al.* 2005; Ogden 1998; Jenkins *et al.* 1996; Racine *et al.* 1992). Research conducted over the last decade reveals that energetic residues released into the environment occur as particulates distributed in an extremely spatially heterogeneous manner (Hewitt *et al.* 2009, 2007, 2005; Jenkins *et al.* 2005a, 2001, 1997a, b, 1996; Walsh *et al.* 1997, 1996; Walsh *et al.* 1997, 1996).

Results from these studies indicate that when particulates are present, grab soil sampling is inappropriate—a finding consistent with Gy's sampling theory (Pitard 1993, Gy 1992, 1982).

One of the outcomes of recent research on the particulate nature of energetic residues were the modifications (Hewitt *et al.* 2009, 2007) to USEPA Method 8330 (USEPA 1996d) for explosives, resulting in the update Method 8330B (USEPA 2006). Collectively, the modifications to the field sampling and sample processing techniques are referred to as the ISM, multi-incremental sampling (MIS)™, or Incremental Sampling (IS) (ITRC 2012; Ramsey 2009; USACE 2009; Hewitt *et al.* 2009, 2007). The DoD, the regulatory community, and environmental consultants are now using ISM for other analytes such as metals (Florida, 2013; Hewitt *et al.* 2012; ITRC 2012, Alaska 2009; Hawaii 2008). However, research and guidance is lacking on the use of ISM versus conventional grab sampling for sites with heterogeneous distributions of metallic residues.

Active DoD facilities are required to assess the environmental impacts of munitions deposition on training and testing ranges (USDoD 2007, 2005). The frequent use on Army training lands of munitions containing metals leads to metal accumulation on surface soils (Clausen and Korte 2009a; Clausen *et al.* 2004). Studies of small-arms ranges indicate that metal deposition is largely spatially heterogeneous, similar to the distribution of explosives (Clausen *et al.* 2014a, *2013,* 2012a; Clausen and Korte 2009a). Clausen *et al.* (2014a), ITRC (2012), and Hewitt *et al.* (2012) questioned the appropriateness of using conventional grab sampling for soils containing metallic residues (particulates).

Therefore, this paper focuses on the use of ISM to estimate mean metal concentrations and compares the results to conventional grab sampling results. Estimating the population mean is a typical goal of environmental studies and often involves calculating a 95% UCL of the mean (USEPA 2013, 2000b, 2002a, 1992, 1989). The 95% UCL of the mean is often compared with background concentrations or fixed decision limits (e.g., cleanup, regulatory and risk-based thresholds). This paper explores the sample reproducibility (precision) and uncertainty for sample collection procedures at sites with heterogeneous distributions of metallic residues by comparing conventional grab and incremental sampling results from the same study area.

3.3 Methods

3.3.1 Soil Sample Collection

This study involved the collection of soil samples from a small-arms range berm at Range 4-3 at Camp Ethan Allen, Vermont. The range dimensions are approximately 3 × 100 m and the soil consists of sand and gravel material containing visible small-arms bullet fragments. In addition, a background location located approximately a half mile from the range was sampled with 50-increment ISM samples collected in triplicate. Clausen *et al.* (2014a, 2013, 2012a) presents additional information on the characteristics of the range.

Clausen *et al.* (2014a) describes the approach for determining the number of grab samples (sample size) needed to achieve a statistically representative result. Using a systematic central aligned grid design 30 grab samples were collected based upon the recommendations of Matzke *et al.* (2010), USEPA (1995), (Gilbert 1987), and Griffiths (1971). Each grab sample was collected to a depth of 5 cm using a 2 cm diameter corer (Walsh 2009) and placed in a 4 oz amber glass container. Additional material was collected as needed to result in a fully packed jar, yielding a sample mass of approximately 0.2 kg.

Collection of ISM surface soil samples followed the methodology outlined in (Clausen *et al.* 2013a, 2012a, ITRC 2012); Figure 3-1 summarizes all of the steps, including sample preparation. To address the compositional and distributional heterogeneity (e.g., to obtain a representative and reproducible estimate of the mean concentration), the sampling strategy requires acquiring an adequate number of particles of the constituents of interest. The particles must be present in the sample in roughly the same proportion as in the DU, or study area (ITRC 2012). The DU is an area of interest about which one plans to make a decision based on the outcome of the soil concentration data (ITRC 2012; Ramsey 2009). To obtain a representative sample, the ISM entails collecting many increments (e.g., 30–100 soil cores) over the entire DU in a systematic random pattern (Figure 1-1) and combining them to prepare a composite sample; the total mass of each ISM sample is usually 0.5 to 2 kg.

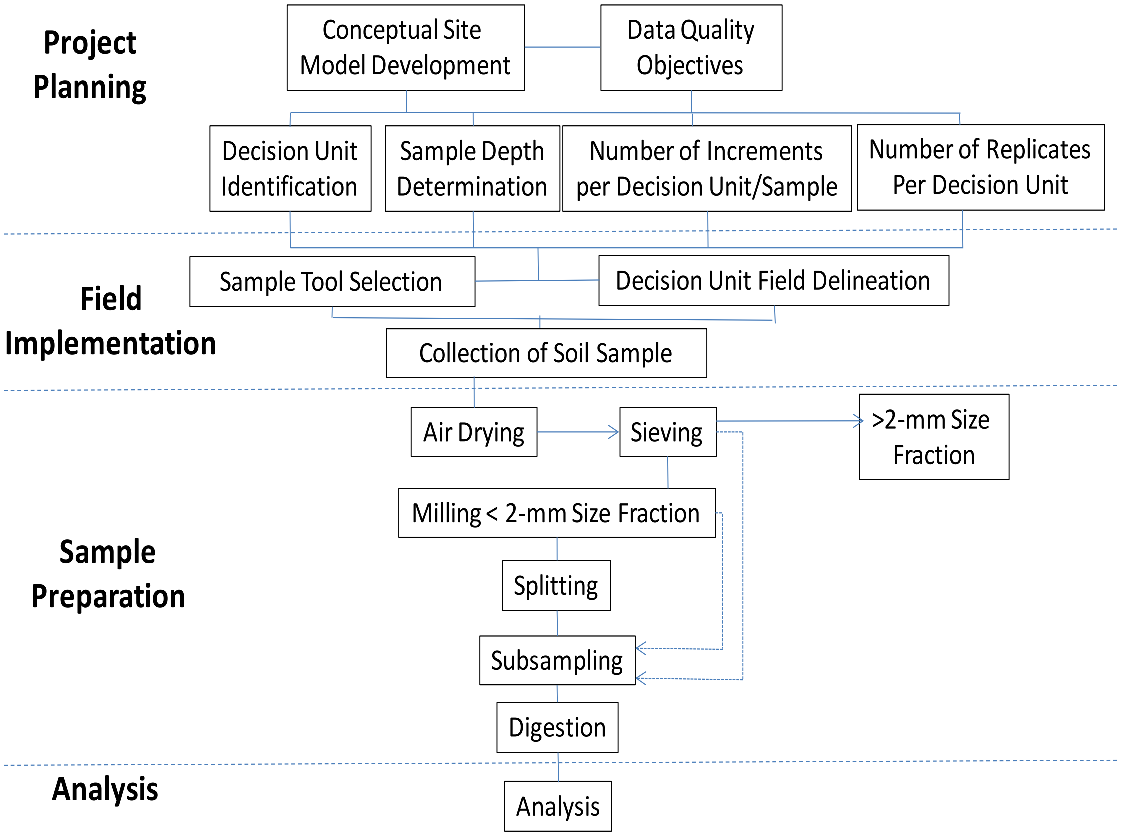


Figure 3-1. Flowchart of the incremental sampling process.

Figure 1-1 illustrates how to collect two independent ISM samples. This project collected seven independent field replicates. To reduce compositional and distributional heterogeneity for each sample, 100 evenly spaced increments were collected and combined from the 3 ×100 m berm face into a composite sample with a total mass of at least 1 kg, consistent with USEPA Method 8330B (USEPA 2006) and recommendations by ITRC (2012), Jenkins *et al.* (2004a,b, 2005a, 2006), Walsh *et al.* (2005), and Hewitt *et al.* (2005, 2007).

3.3.2 Sample Preparation and Analysis Procedures

Grab sample processing followed the general approach outlined in USEPA Method 3050B (1996a), which involved collecting a single 2 g aliquot for digestion from the top of the sample jar. The ISM samples were processed using a modified method described in Clausen *et al.* (2013a). The ISM samples were air-dried, passed through a 10-mesh sieve prior to milling, and then sub-sampled. Sub-sampling involved using the procedure described in USEPA Method 8330B (USEPA 2006a). Briefly, the ground soil was spread on to a sheet of aluminum foil as a thin layer 1 to 2 cm thick and 20 aliquots were selected randomly and combined to yield a 2-g digestion aliquot.

The instrumental analyses of the grab and ISM samples were made with a ICP-optical emission spectrometer (ICP-OES) using a Thermo Scientific ICAP 6000 Series instrument following Method 200.7 (USEPA 1994). The instrument provided results for aluminum, barium, cadmium, chromium, cobalt, copper, iron, magnesium, manganese, nickel, lead, phosphorous, antimony, strontium, vanadium, and zinc.

3.4 Results and Discussion

From our study area DU, we collected 30 grab samples; six sets of seven ISM samples prepared from 5, 10, 20, 30, 50, and 100-increments; and a single 200-increment sample. Table 3-1 presents descriptive statistics for the 30 grab sample results; six sets of ISM results; a set of pooled ISM results from the ISM samples prepared from 30, 50 and 100 increments; and laboratory replicate analyses of the 200-increment ISM sample.

The results in Table 3-1 and Figure 3-2 show less variance and better precision for the anthropogenic ISM results than the grab samples; a finding consistent with Boudreault *et al.* (2012). Table 3-1 suggests that copper, lead, antimony, and zinc are anthropogenic contaminants. The concentrations in the DU are several times larger than the mean background concentrations. The sample means of the grab samples for lead and antimony were 4 to 7 times greater than the medians; and the RSD were 285 and 427%, respectively. The large outliers and large sample means relative to the medians indicate that the lead and antimony distributions are highly positively skewed. The mean and median concentrations for copper were similar as were the values for zinc. The RSD for copper was 44%, which is greater than our target criteria of 30%; the RSD for zinc was slightly lower at 27%.

The smaller variability for copper and zinc (relative to lead and antimony) may be owing to less copper and zinc contamination that is anthropogenic in origin. The mass of anthropogenic copper and zinc may have been insufficient to dominate the natural distribution of these metals. We believe that this explains the small differences between the copper and zinc means and medians, and the smaller RSDs. Conversely, for the metals assumed to be naturally occurring (aluminum, barium, cadmium, cobalt, iron, magnesium, manganese, nickel, phosphorous, strontium, and vanadium) the grab means and medians were similar and the RSDs less than 30%.

In contrast to the anthropogenic grab metal results, the pooled 21 ISM results (consisting of 30, 50, and 100 increments) had estimated means similar to the medians and the RSDs were less than 25% for all metals (anthropogenic and native). A comparison of the estimated means by number of increments collected shows for those ISM samples with more than 30 increments, the difference in values with the pooled result is generally less than 21% (Table 3-1). The one exception was the chromium results for the 30- and 50-increment samples, where the percent differences from the pooled ISM mean were 51% and 30%, respectively. The larger differences were a result of chromium cross-contamination during milling. The bowl and puck are composed of chromium steel. During the milling process, chromium (as well as iron, manganese, and nickel to some degree) is lost from the bowl and puck and contaminates the ISM samples (Clausen *et al.* 2012a, 2010; El Khoukhi *et al.* 2005).

The differences (e.g., as measured by central tendency and dispersion) between the grab and incremental data sets are notable for lead and antimony, to a lesser degree for copper, but not for zinc. The concentrations of lead reported from the grab samples are extremely variable. The estimated mean grab sample lead concentration of 5,060 mg/kg is strongly influenced by a single large value, the maximum detected value of 79,020 mg/kg (Figure 3-2). If a single grab sample was used to characterize this DU, the majority of the time the result would be lower than the mean. Exclusion of the 79,020 mg/kg grab sample outlier yields a change in the mean of a nearly a factor of two, from 5,060 to 2,510 mg/kg, and a change in the RSD of a nearly a factor of four, from 165% to 43%.

Table 3-1. Statistical summary of grab and incremental sampling methodology samples collected at Camp Ethan Allen.

Berm Face Decision Unit Grab Sample Results																
	Al	Ba	Cd	Co	Cr	Cu	Fe	Mg	Mn	Ni	P	Pb	Sb	Sr	V	Zn
n	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30
Mean	4219	20.0	0.760	7.40	6.22	300	12381	1976	197	9.91	710	5060	87.8	12.9	11.2	66.1
Median	4231	17.4	0.735	7.44	6.19	270	12380	1967	195	10.0	722	1238	10.0	12.1	11.2	61.9
Minimum	3784	15.7	0.620	6.31	5.58	69.8	9975	1748	175	8.67	468	43.9	0.898	9.86	9.14	35.8
Maximum	4577	39.5	1.13	8.38	6.99	598	14700	2325	242	10.7	1043	79020	2072	22.8	15.4	111
STD	158	5.61	0.104	0.486	0.409	132	797	124	15.4	0.439	116	14438	375	2.72	1.17	17.6
RSD (%)	4	28	14	7	7	44	6	6	8	4	16	285	427	21	10	27
Berm Face Decision Unit Incremental Sampling Methodology Results																
n	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
Mean	6760	30.9	1.15	9.37	344	559	17208	2068	232	13.1	762	2583	21.0	24.9	17.4	72.2
Median	6730	31.0	1.18	9.31	358	553	17650	2030	227	13.0	758	2539	20.2	25.3	17.1	70.0
Minimum	6130	26.0	0.94	8.73	228	349	14465	1940	216	12.1	689	1835	15.0	21.5	15.4	58.2
Maximum	7510	33.7	1.41	9.94	442	882	20405	2243	254	14.0	847	3595	28.9	28.2	20.2	95.9
STD	399	1.53	0.155	0.320	53.9	141	1658	94.1	11.5	0.52	38.7	488	3.85	1.82	1.25	10.5
RSD (%)	6	5	13	3	16	25	10	5	5	4	5	19	18	7	7	15
^{1}Mean 5-inc.	6230	29.1	1.16	8.85	262	539	17084	2088	220	12.4	727	2989	23.5	24.2	15.9	79.7
^{1}Mean 10-inc.	4240	16.3	0.749	6.64	6.08	1277	12382	1987	192	9.30	721	2132	18.5	12.5	10.1	193
^{1}Mean 20-inc.	5824	29.9	0.875	8.12	220	473	15005	2270	235	13.2	712	2689	23.1	17.9	14.3	80.7
^{1}Mean 30-inc.	7224	31.2	1.30	9.56	395	573	17435	1992	228	13.5	792	2664	22.7	26.4	18.6	67.6
^{1}Mean 50-inc.	6604	30.2	1.19	9.15	341	457	15946	2021	222	13.1	737	2156	17.6	23.1	16.9	67.2
^{1}Mean 100-inc.	6453	31.3	0.963	9.39	296	648	18242	2191	245	12.7	757	2929	22.8	25.3	16.6	81.8
^{2}Mean 200-inc.	5816	30.0	1.83	8.58	223	550	16752	2125	225	12.2	611	2717	22.6	22.0	15.2	77.2
^{3}Mean$_{\text{Weighted}}$	6717	30.9	1.18	9.33	338	559	17187	2071	232	13.1	755	2589	21.1	24.8	17.3	72.4
^{4}RPD	3.6	0.7	10.7	2.1	10.3	0.4	0.6	0.6	0.7	1.7	5.3	1.2	1.7	3.0	3.2	1.6
Background Decision Unit Incremental Sampling Methodology Results																
n	3	3	3	3	3	3	3	3	3	3	0	3	3	0	3	3
Mean	11300	47.9	<0.002	7.80	368	15.8	21300	2880	306	13.1	NA	37.7	4.66	NA	23.5	42.1
Median	11300	47.6	<0.002	7.90	384	10.8	21300	2880	305	13.3	NA	38.0	4.85	NA	23.6	41.9
Minimum	11200	47.4	<0.002	7.49	319	10.6	21200	2860	300	12.1	NA	36.8	4.17	NA	23.1	41.5
Maximum	11400	48.6	<0.002	8.01	402	26.1	21400	2900	313	14.0	NA	38.3	4.97	NA	23.8	42.9
STD	100	0.643	<0.002	0.274	43.7	8.89	100	20.0	6.56	0.961	NA	0.794	0.431	NA	0.361	0.721
RSD (%)	1	1	NA	4	12	56	0	1	2	7	NA	2	9	NA	2	2

Units of mg/kg unless noted.
Highlighted metals represent those of anthropogenic origin.
STD - standard deviation, RPD - relative % difference, RSD - % relative standard deviation
^{1}Mean of field replicates n=7
^{2}Single sample, mean of 30 laboratory replicate analyses
^{3}Weighted DU Mean calculated from 22 incremental sampling methodology (ISM) field samples and 100 analyses
^{4}RPD – relative % difference between mean of the weighted mean and mean for the 200-increment field sample, which was analyzed 30 times.

The large variability of the lead grab sample concentrations (Figure 3-2) suggests that estimates of the DU mean will possess large uncertainty. As shown in Figure 3-3, the distribution of lead concentrations is positively skewed;

the median lead concentration (1,238 mg/kg) is several times smaller than the mean. As a result, a small number of grab samples will likely under-estimate the DU mean. This observation is consistent with the findings for impact areas containing energetic particulates from the detonation of military munitions (Hewitt et al. 2005; Jenkins *et al.* 2005b, 2004a).

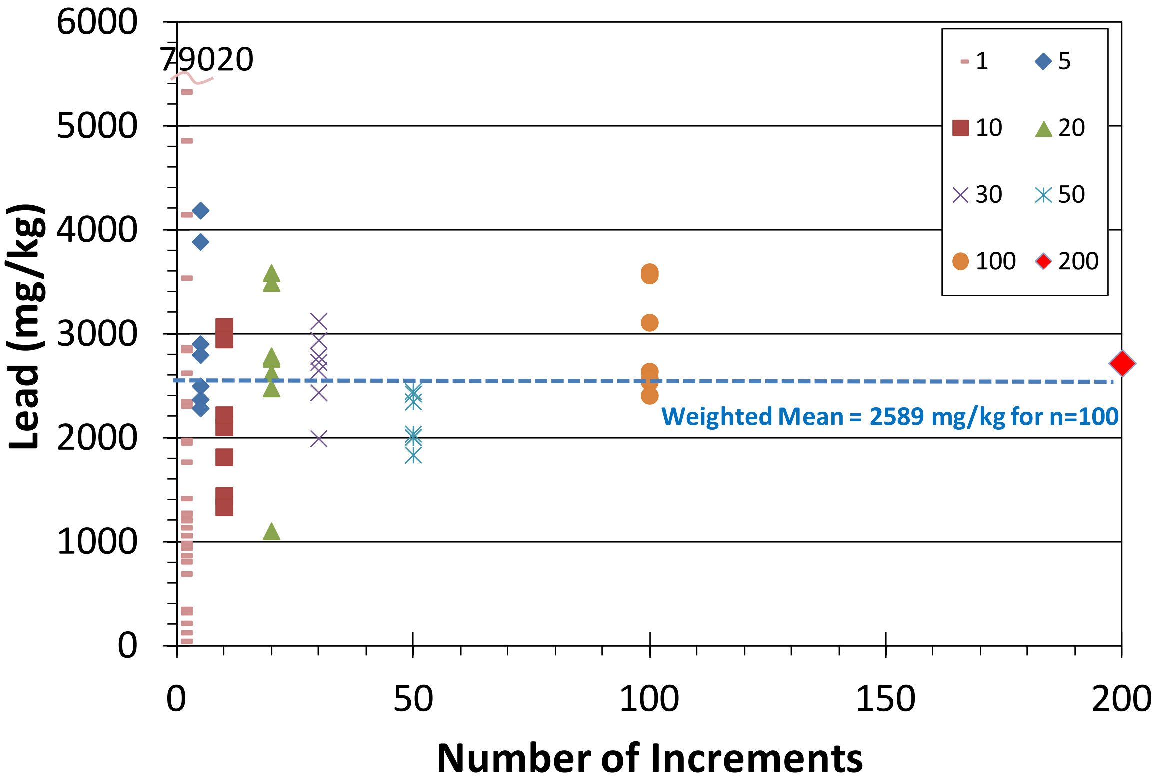


Figure 3-2. Distribution of lead results by number of increments. An increment of one is a grab result whereas increments of 5, 10, 20, 30, 50, 100, and 200 are incremental sampling method results. A weighted incremental sampling methodology mean was calculated for those samples with more than 30 increments.

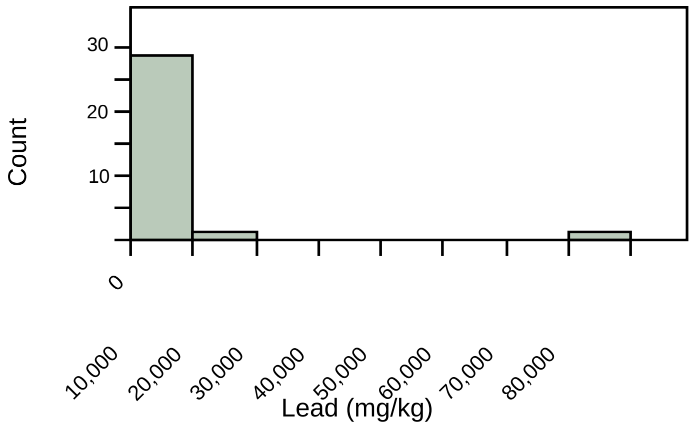


Figure 3-3. Distribution of lead results for the 30 grab samples.

As discussed in Clausen et al. (2014a) if the set of grab samples is repeatedly resampled using a bootstrap method, the simulation produces a wide range of different estimates of the mean lead concentration, which raises several questions. Which grab sample estimate of the mean provides a reasonable estimate of the "true" lead concentration within the DU? How many grab samples are needed for a representative estimate of the mean? Would an alternative sampling approach such as ISM provide an estimate of the mean with less uncertainty?

Clausen *et al.* (2014a) investigated the number of grab samples necessary to yield results representative of the mean concentration and observed a large uncertainty that suggested a minimum of n > 35 and more likely several hundred samples are likely needed to obtain a reasonable estimate of the mean. Even 30 to 50 grab samples for an area as small as 3 × 100 m may be insufficient to overcome the high degree of small-scale heterogeneity for metallic residues on small-arms ranges. This was further illustrated where a significant difference in lead levels for co-located grab samples existed with a separation distance as little as 0.5-m.

For this study, a grab sample from Grid 1, discussed in Clausen *et al.* 2014a, was split into fifths in the laboratory using a sectorial rotary splitter (Table 3-2). Large variability was evident in concentrations of lead (which ranged from 1,136 to 4,147 mg/kg) and antimony (which ranged from 7.9 to 47 mg/kg) for samples with a mass of approximately 250 g.

Table 3-2. Descriptive statistics for five splits prepared from a grab sample from Grid 1 on Range 4-3 at Camp Ethan Allen.

	Al	Ba	Cd	Co	Cr	Cu	Fe	Mg	Mn	Ni	P	Pb	Sb	Sr	V	Zn
n	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Mean	4358	18.4	0.72	6.61	5.83	225	10777	1969	188	9.21	568	1848	16.6	9.29	9.92	59.5
Median	4373	18.5	0.64	6.66	5.83	223	11340	1969	189	9.13	559	1276	9.36	9.13	10.0	60.2
Min	4260	16.7	0.63	6.34	5.74	211	9880	1910	184	9.06	521	1136	7.87	8.76	9.74	57.4
Max	4418	21.1	0.86	6.86	5.96	248	11350	2011	193	9.59	636	4147	47.0	10.0	10.1	60.5
STD	59.3	1.83	0.121	0.208	0.080	14	776	37.8	4.47	0.217	43.6	1289	17.0	0.512	0.152	1.29
RSD	1	10	17	3	1	6	7	2	2	2	8	70	102	6	2	2

STD - standard deviation, RSD - percent relative standard deviation

Similar to determining the number of grab samples needed for estimating the DU mean, the number of increments required for ISM samples needs to be determined prior to sampling. Consequently, I assessed the distribution and mean of the ISM results as a function of *m,* the number of individual increments used to build each incremental sample. Six sets of seven independent ISM samples with values of *m* equal to 5, 10, 20, 30, 50, and 100 were collected. Increasing the number of increments decreased the spread in individual lead results, resulting in distributions more Gaussian shaped, and presumably improved the estimates of the DU mean (Figure 3-4). The results are fully consistent with the central limit theorem, which states that even when the distribution of individual results is non-normal, the distribution of means will approach a normal distribution as the sample size *n* increases. As was observed for small numbers of grab samples, a small number of increments tended to under estimate the mean for lead, as well as copper, antimony, and zinc (Figure 3-5). This observation is consistent with the findings for impact areas containing energetic particulates from the detonation of military munitions (Jenkins *et al.* 2004a, 2005b).

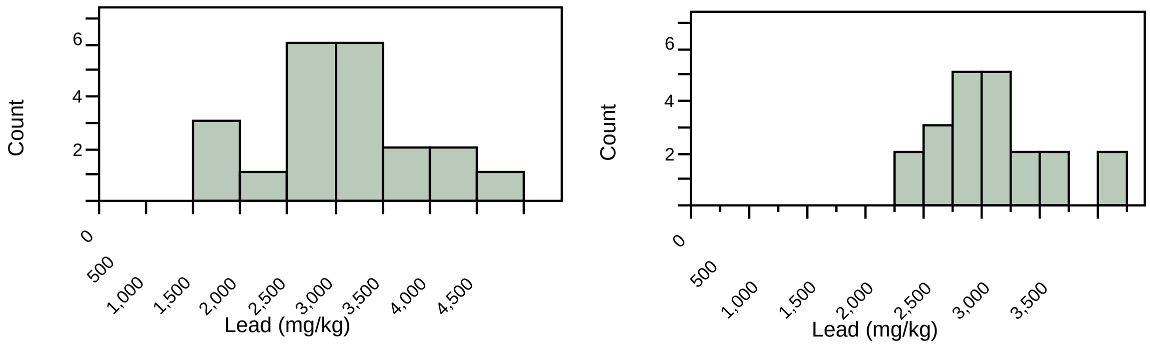


Figure 3-4. Distribution of lead results for the incremental samples of less than 25-increments and more than 25-increments.

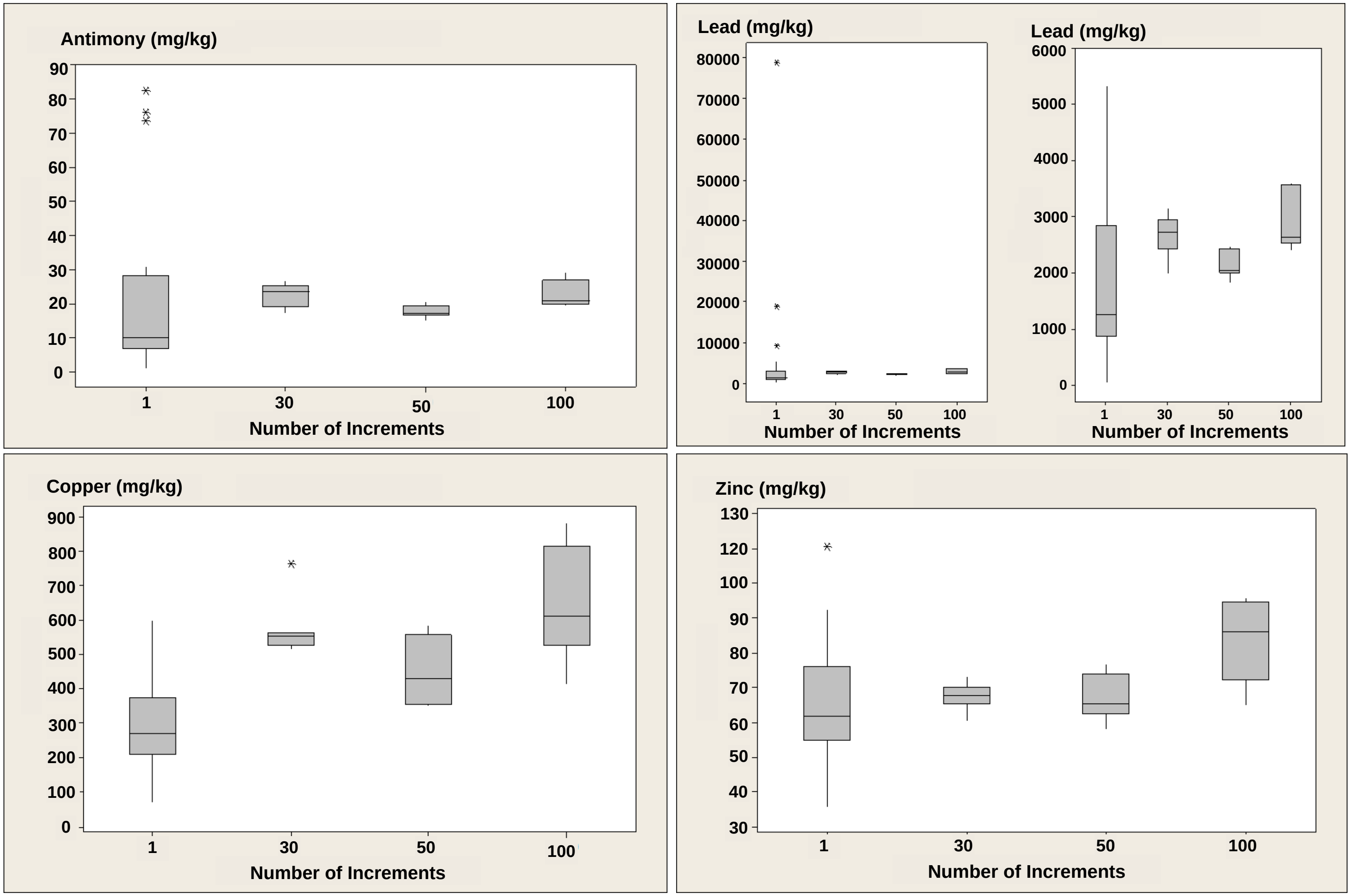


Figure 3-5. Box plots depicting distributions for antimony, copper, lead, and zinc results for the grab and incremental samples of more than 25-increments.

One of the concerns with the ISM approach relates to a fear of “diluting” the analyte of interest by mixing low concentration increments with higher concentration increments. This concern is taken into account during project planning (as illustrated in Figure 3-1) when determining the size of the DU; by definition, the DU is the smallest area or volume of interest for decision making. In this study, the incremental samples resulted in a higher mean concentration than 70% of the grab samples collected from the same DU. Similar observations were made for antimony (Figure 3-5 and Figure S8 in the Appendix), which is not surprising as the core of a small-arms projectile consists of a lead/antimony alloy. The box plots in Figure 3-5 suggest that ISM results in larger mean concentrations for anthropogenic metals. These findings are also consistent with studies comparing ISM and grab sampling for sites with energetic residue contamination (Jenkins *et al.* 2005a, 2004a,b, 1999, 1997a,b, 1996). The characterization of ISM as a sampling methodology that dilutes contamination is largely a misconception. If the objective is to estimate the mean of the entire DU, physically compositing many increments of ISM samples is comparable to numerically averaging a large number of grab samples. The greater number of grab samples or increments the greater the probability of a physically representative sample containing the analyte in the same proportion as the population (DU).

The set of 30 grab samples and three sets of seven ISM samples (30-, 50-, and 100-increments) were compared using the non-parametric two-tailed Kruskal Wallis test for the medians and Levene’s test for the variances. Table 3-

3 summarizes the results of these statistical evaluations for lead, copper, antimony, and zinc. The ISM approach produced much smaller variances for the anthropogenic metals lead and antimony than conventional grab sampling and tended to result in significantly larger median concentrations for lead and copper Boudreault *et al.* (2012) and Jenkins *et al.* (2005a) work.

Table 3-3. Statistical analysis summary for antimony, copper, lead, and zinc.

Metal	k^1	n^2	**Normal**[3]	**KW**[4]	**Levene's**[5]
copper	1	30	Y	p= 0.000 m_{100}, m_{50}, m_{30}> m_1	p= 0.172 No difference
	30	7	N		
	50	7	Y		
	100	7	Y		
lead	1	30	N	p = 0.024 m_{100}, m_{50}, m_{30}> m_1	p= 0.000 $s_1 > s_{30}$, s_{50}, s_{100}
	30	7	Y		
	50	7	Y		
	100	7	Y		
antimony	1	30	N	p = 0.121 m_{100}, m_{50}, m_{30}> m_1	p= 0.014 $s_1 > s_{30}$, s_{50}, s_{100}
	30	7	Y		
	50	7	Y		
	100	7	Y		
zinc	1	30	Y	p = 0.047 $m_{100} > m_1$, m_{30}, m_{50}	p = 0.026 $s_1 > s_{30}$, s_{50}, s_{100} but $s_1 \approx s_{100}$
	30	7	Y		
	50	7	Y		
	100	7	Y		

[1]. *k* = number of increments per sample; for grabs *k* = 1.

[2]. *n* = number of replicates per DU.

[3]. N = not normal at 95% confidence level of confidence; Y = normality assumption not rejected.

[4]. KW= two-tailed Kruskal-Wallis test p-value. m_k = median of data set with *k* increments. At least one median different with 95% confidence if $p \leq 0.05$.

[5]. Levene's test for variances; p-values. At least one variance different with 95% confidence if $p \leq 0.05$. s_k denotes standard deviation of data set with *k* increments.

I identified a statistical difference between the grab and ISM zinc medians but it is not clear if the difference is of practical significance; m_1 (the median for the set of grab samples) is only somewhat smaller than m_{100} (the median of the set of ISM samples prepared from 100 increments each), but $m_1 \approx m_{30}$ and m_{50} (Table 3-3). Similarly, a statistical difference was identified for the variances; the standard deviation s_1 (for the grab samples) is numerically larger than the standard deviations for the ISM data sets. Based on these results, it appears that the ISM approach did not significantly improve the quality of the zinc results in all cases because the soil samples contained little or no metallic particles. Intact projectiles with their copper/zinc alloy jacket (Figure 3-6) were infrequently observed on the berm face.

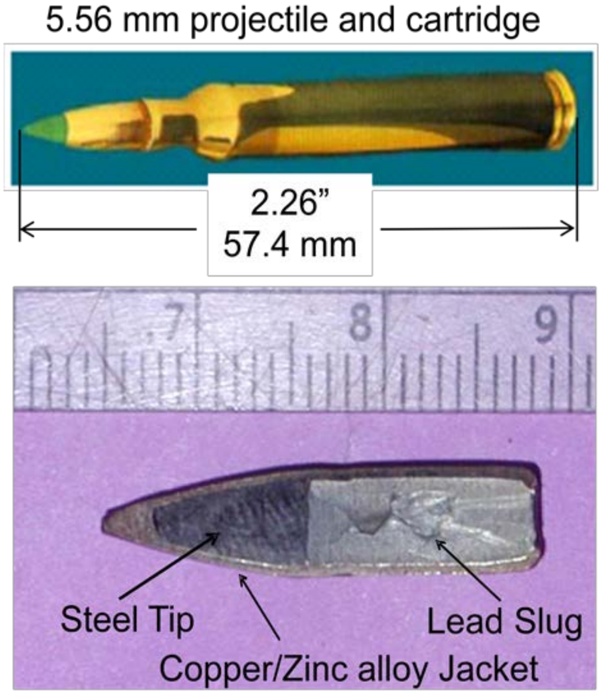


Figure 3-6. Photograph of a 5.56 mm projectile (scale bar is in inches) and cartridge (upper) and projectile cut in half-lengthwise (lower) used in the M-16 Rifle.

The ISM approach did not seem to normalize the data or decrease the variability for copper with 30-increments. A significant difference was not observed for the variances at the 95% level of confidence using Levene's test. I believe this is due to the formation of flat copper plates during milling, thus confounding the results. However, the median copper concentrations for the ISM data sets were about 50 to 100% larger than the median for the grab data set.

A question often raised relates to the accuracy of grab sample data relative to ISM. A direct determination of accuracy is not practical because the total mass of metal (native and anthropogenic) in the soil would need to be known or the entire soil mass in the DU would need to be digested and analyzed with negligible analytical error. The former is generally unknown and the latter would be impractical. To evaluate relative bias for ISM, we pooled results from the single 200-increment, and seven 30-increment, seven 50-increment, and seven 100-increment ISM samples. This yielded a dataset of 22 ISM samples consisting of 1,460 increments for an area of 3,000 m^2 and a soil volume of 75 m^3. All of the ISM samples were prepared, sub-sampled (digestion masses of at least 2 g) and analyzed in the same manner. When laboratory replicates were analyzed for an ISM sample, the results were arithmetically averaged prior to pooling the results. Including the laboratory replicates, I performed 100 analyses for the 22 ISM samples. The pooled means are listed in Table 3-1 ("$\text{Mean}_{\text{Weighted}}$").

Table 3-4 provides RPDs between the pooled ISM mean and ISM samples with 20 or fewer increments; the RPDs range from <1% to 130%. Generally, the differences were much less for the non-anthropogenic metals than the anthropogenic metals. A comparison of the grab sample mean (increment of one) with the pooled ISM mean indicates differences in excess of 41% for all metals analyzed with most differences near 100%. The results suggest a reliable estimate of the mean for this study site is not possible with 30 grab samples.

Table 3-4. Relative percent differences of grab sample and incremental sampling methodology means relative to the pooled incremental sampling methodology mean.

Number of Increments	Al	Ba	Cd	Co	Cr	Cu	Fe	Mg	Mn	Ni	P	Pb	Sb	Sr	V	Zn
1	100	84	41	76	97	95	100	100	99	80	100	99	20	84	80	81
5	5	2.2	2.0	3.8	0.2	1	1	1	2.8	0.5	11.4	13	13.6	7.6	1.8	4.3
10	29	45	37	22	98	134	27	6	15	25	10	20	11	45	36	152
20	2	0.3	26	4.8	16	13	11	8	3.7	6.3	9.2	1	11.3	21	8.7	5.7
30	21	4.8	10	12	51	5	3	5	1.0	9.0	21	1	9.3	17	19	12
50	11	1.3	1.0	7.3	30	16	5	4	1.9	5.1	13	19	15	2.5	8.2	12
100	8	5.1	18	10	13	19	8	4	8.3	2.4	16	11	10	13	5.9	7.1
200	2	0.7	55	0.7	15	1	1	1	0.5	1.9	6.4	2	8.9	2.2	3.1	1.0

Yellow highlighted metals represent those of anthropogenic origin.
Green highlighted increment of one is the grab sample.

Table 3-5 presents descriptive statistics for the 22-pooled ISM sample results and the 30 grab sample results. As a small number of large extreme values heavily influenced the means of the grab samples and the ISM means and median values were similar to each other, I evaluated relative bias qualitatively by comparing the median grab and ISM concentrations. As shown in Table 3-5, the grab sample median values tended to be negatively biased relative to the ISM median values.

Table 3-5. Determination of accuracy for the estimated mean grab sample concentrations for antimony, lead, copper, and zinc relative to the pooled Incremental Sampling Methodology mean.

		Units	Sb	Pb	Cu	Zn
Pooled ISM	n		22	22	22	22
	Weighted Mean*	mg/kg	21.1	2589	559	72.4
	Median*	mg/kg	21.3	2755	541	77.4
	RSD*	%	11	10	9	6
Grab	n		30	30	30	30
	Mean	mg/kg	87.8	5060	300	66.1
	Median	mg/kg	10.0	1238	270	61.9
	RSD	%	427	285	44	27
	Relative Bias	%	53	55	50	20

* represents 22 field samples and 100 replicate analysis
Relative bias – grab median minus Pooled Incremental Sampling Methodology median divided by pooled Incremental Sampling Methodology median times 100.

The grab sample results exhibit significant differences between estimates of the mean for antimony and lead with less difference for copper and zinc compared to the pooled ISM means. The antimony/lead and copper/zinc differences are likely related to the physical depositional process of projectile impact and the mass of material in the projectile. The copper/zinc alloy jacket often separates from the steel tip and lead/antimony slug (Figure 3-6), falling to the range floor, as it passes through the target whereas the steel tip and lead slug proceeds through the target and reaches the impact berm where sampling occurred. Note that despite the heterogeneous spatial distribution of metallic contamination at the berm (e.g., owing to bullet fragments), the ISM RSDs are much smaller than the RSDs for the grab samples. The differences between the grab and ISM results with respect to central tendency and dispersion are largest for the lead and antimony results. If I assume the pooled ISM mean represents the nearly “true” mean of the DU, the accuracy of the mean for the grab sample results of antimony, lead, and copper is poor. Clearly, many more grab samples are required to arrive at an accurate estimate of the mean for antimony, lead, and copper at this DU with Clausen *et al.* (2014a) suggesting that more than 30 grab samples are needed.

Existing regulatory, statistical, and software guidance suggests 30 samples should be more than adequate to calculate the mean for the sampled DU (Matzke *et al.* 2010; USEPA 1995; Mason, 1992; Barth and Mason 1984). Clausen *et al.* (2014a) showed that the estimate of the mean using 30 grab samples for a small-arms range berm was inadequate to obtain a RSD ≤ 30%. Based on the empirical performance of the ISM method at a number of study areas, a RSD of 30% is often selected as a target for total precision to obtain reliable (confident) estimates of the DU mean. Computer simulations showed each independent sampling event yielded significantly different estimates of the mean based on two-tailed Kruskal Wallis tests for the medians and Levene’s test for the variances (Clausen *et al.* 2014a). My findings are of concern because environmental practitioner’s are often pressured to collect the fewest number of samples as possible (ITRC 2012); far fewer than 30 grab samples would normally be collected ((Hadley *et al.* 2011; Jenkins *et al.* 2005b) for this particular DU. It is a common practice in the environmental industry to establish sample sizes based on budgetary considerations alone rather than decision error tolerances or the end use of the data. Clausen *et al.* (2014a) suggests that this currently accepted practice of collecting the fewest number of grab samples as possible yields unreliable estimates of the mean when metallic residues are present.

To estimate how many grab samples are necessary to achieve the same level of data quality obtained with ISM, we used a bootstrap (resampling) method for different numbers of grab samples *m.* Clausen e*t al.* (2014a) discusses the details of the computer simulations. The focus of our earlier paper was to ascertain whether a reliable estimate of the mean is possible using grab sampling.

I performed three hundred bootstrap simulations for each value of "*m*". Figure 3-7 is a plot of the standard error of the mean (SE) (from the simulations) versus the number of grab samples (*m*). The simulations indicate that the SE declines with an increasing number of grab samples, and the trend of the data can be fitted with a power curve with a coefficient of determination $R^2 = 0.9851$. If the fitted curve for the grab samples is extrapolated to the SEs for the ISM samples prepared from 30 to 100 increments, the plot suggests that approximately 100 grab samples would be needed to achieve the same data quality as one ISM sample. My earlier work evaluating the number of grab samples needed to yield a reliable estimate of the mean for the anthropogenic metals at a small-arms range at Fort Wainwright, Alaska yielded similar findings (Clausen *et al.* 2014a). As discussed earlier, there is pressure to minimize the number of samples collected on environmental characterization studies. For many projects, the necessity of collecting 100 grab samples would be cost prohibitive. In contrast, collecting three to seven 100-increment samples from a DU has been shown to be sufficient for most situations (ITRC 2012). The time and cost to collect three ISM samples is 5% to 50% lower than that for 7 to 15 grab samples (Clausen *et al.* 2013b). For the present study, seven field replicate ISM samples yielded results with a higher data quality than with the collection of 30 grab samples.

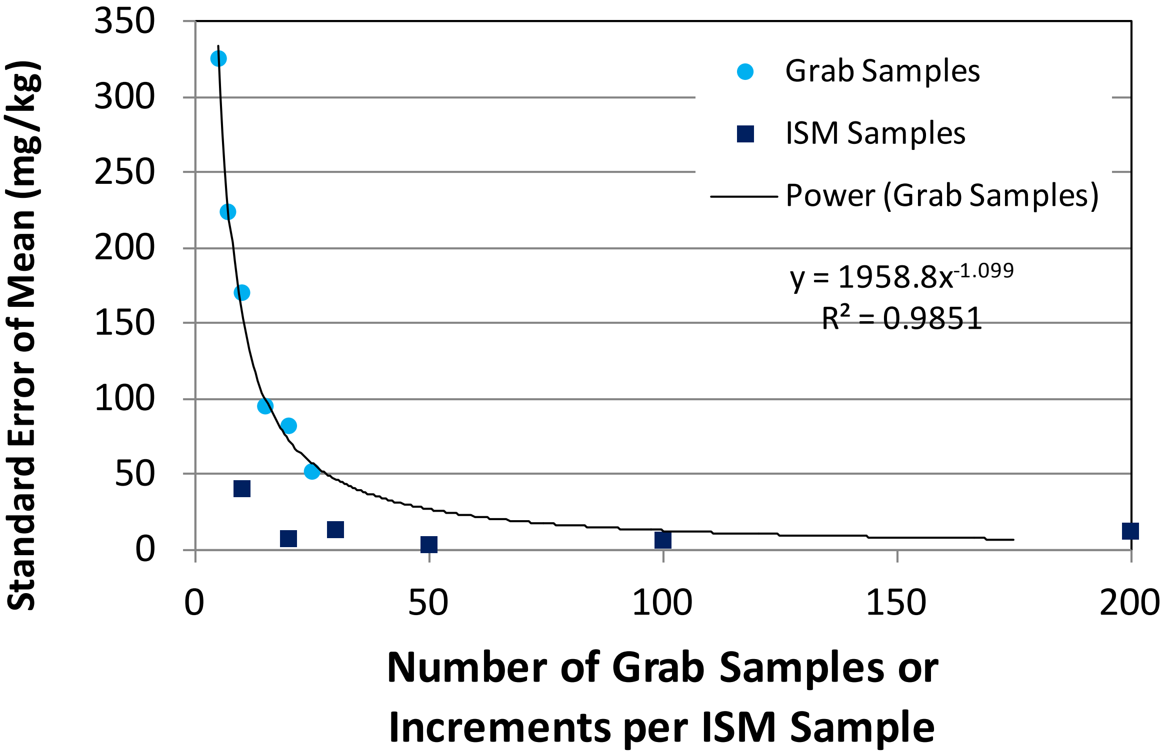


Figure 3-7. Standard error of the mean versus number of grab samples or increments per sample.

The same situation of too few grab samples also applies to the number of increments per ISM sample. In Figure 3-7, it is apparent that an ISM sample consisting of 10 increments yields a larger sample error (SE) than ISM samples with more increments. Similarly, Jenkins *et al.* (2005b) found that the calculated

95% tolerance limits of the mean decreased as the number of increments increased. For this reason earlier studies on the use of incremental sampling for soils with energetic residues recommends the collection of a minimum of 30 increments per sample (ITRC 2012) and preferably 50 to 100 (Jenkins *et al.* 2006a, 2005b, 2004a,b; Hewitt *et al.* 2012, 2009, 2007, 2005, Walsh et al. 2005).

It is clear that 30 grab samples are sufficient, for the most part, to represent the mean for the non-anthropogenic metals suggesting a more homogeneous distribution in the environment (Table 3-6 and Figures SI9–SI11 in the Appendix). The acceptance criterion for accuracy for laboratory analyses of metals is commonly no greater than ±30%. With the exception of barium, cadmium, and strontium all of the non-anthropogenic metals met this criterion. The number of grab samples that need to be collected depends on the heterogeneity (variability) expected for the analyte of interest. In the case of natural background distribution of a metal or an aqueous contaminant release, the collection of a reasonable number of grab samples to yield an estimate of the mean with acceptable uncertainty may be possible. However, for contaminants released as particulates at a small-arms range more than 100 grab samples will likely be required to account for heterogeneity and yield a reliable estimate of the mean.

Table 3-6. Determination of accuracy for the estimated mean grab sample concentrations for native metals relative to the pooled incremental sampling methodology mean.

		Units	Al	Ba	Cd	Co	Fe	Mg	Mn	Ni	P	Sr	V
Pooled ISM	n		100	100	100	100	100	100	100	100	100	100	100
	Mean	mg/kg	5956	29.9	1.16	8.54	16837	2114	227	12.3	656	22.7	15.7
	CI Mean	mg/kg	1.41	1.50	1.53	1.15	7.04	0.91	0.11	0.01	0.47	0.01	0.01
	Median	mg/kg	5870	29.6	1.12	8.66	16955	2098	224	12.1	619	22.8	15.3
	RSD	%	12	14	46	9	7	7	8	9	12	10	10
Grab	n		30	30	30	30	30	30	30	30	30	30	30
	Mean	mg/kg	4219	20.0	0.760	7.40	12381	1976	197	9.91	710	12.9	11.2
	CI Mean	mg/kg	0.12	1.81	0.06	0.001	9.12	1.42	0.18	0.01	1.33	0.03	0.01
	Median	mg/kg	4231	17.4	0.735	7.44	12380	1967	195	10.0	722	12.1	11.2
	RSD	%	4	28	14	7	6	6	8	4	16	21	10
	Accuracy	%	71	67	65	87	74	94	87	81	92	57	72

CI – 95% confidence interval of mean
Accuracy – as percent difference of the mean grab sample result relative to the pooled Incremental Sampling Methodology mean value.

There is a tendency for grab samples to yield a negative bias when estimating the mean. The situation is clearly illustrated when calculating the 95% UCLs (USEPA 1992, 1989), which are often used in environmental investigations to estimate the risk and to determine whether a remedial action is necessary. Table 3-7 presents 95% UCLs calculated using USEPA's ProUCL Version 5.0 software program (USEPA 2013). The calculated 95% UCLs for the grab samples are smaller than the ISM 95% UCLs for Cu whereas, larger 95% UCLs are evident for the Pb and Sb grab data as compared to the ISM data sets. The differences between the UCLs calculated for the grab and ISM samples are predominantly due to the large variability and positively skewed distributions of the grab sample concentrations (Figure 3-3 and Figure SI11 in the Appendix) that ProUCL identified as consistent with lognormal distributions. Consequently, ProUCL selected the (H-UCL) Land's H-statistic (Land 1971, 1975) as the most appropriate 95% UCL for lead and the 95% Chebyshev lognormal UCL for antimony. ProUCL selections are based on simulation studies summarized in Singh *et al.* (2002) and Singh and Singh (2003). USEPA (2013) suggests that the H-UCL can be unreliable. The Student's t-test modified for skewness by Chen (1995) and Johnson (1978) and a Gamma Distribution UCL and Chebyshev, which accommodate lognormal distributions, were also calculated using the grab samples for comparison. The Chebyshev and Student's t-test UCLs were calculated for the ISM samples; the Student's t-test UCLs were selected in every case as more appropriate (Table 3-7). However, ITRC (2012) suggests that the Chebyshev results in better coverage of the population mean when n is less than 7.

As all the calculated 95% UCLs are greater than the USEPA recommended screening level for residential soil, both grab and ISM data would result in the same outcome: contaminant concentrations exceeding the RSL require remediation. However, if bias is assessed as a RPD by using Equation 2, the magnitude of the differences between the UCLs in Table 3-6 and the "true" (population) mean (approximated as the weighted mean in Table 3-5) suggests that the UCLs for the grab copper, lead and antimony results are positively biased relative to the UCLs for the ISM samples (Table 3-8).

2}   RPD = ((95% UCL – "true" mean)/ "true" mean) * 100

Table 3-7. Summary of ProUCL 95% upper confidence limits.

Test	Sample Type	Cu (ppm)	Pb (ppm)	Sb (ppm)	Zn (ppm)
Distribution	Grab	N, S, G, L	L	L	N, S, G, L
Distribution	ISM 30		N, G, L	N, G, L	N, G, L
Distribution	ISM 50	N, G, L	N, G, L	N, G, L	N, G, L
Distribution	ISM 100	N, G, L	N, G, L	N, G, L	N, G, L
Student's-t UCL	Grab	341*	NA	NA	71.6*
H-UCL (lognormal)	Grab	365	9429*	92	72.5
Chebyshev UCL (lognormal)	Grab	429	9457	387*	80.5
Student's-t UCL (Chen Mod)	Grab	341	NA	NA	89.8
Student's-t UCL(Johnson Mod)	Grab	341	NA	NA	91.9
Gamma Distribution UCL	Grab	351	NA	NA	72.3
Student's-t UCL	ISM 30	636*	2933*	25.1*	70.5*
Chebyshev UCL	ISM 30	714	3268	28.1	74.1
Student's-t UCL	ISM 50	527*	2335*	19.0*	72.0*
Chebyshev UCL	ISM 50	615	2557	20.7	78.0
Student's-t UCL	ISM 100	772*	3287*	25.6*	92.9*
Chebyshev UCL	ISM 100	936	3746	29.1	103
USEPA Recommended Screening Level for Residential Soil	NA	310	400	3.1	2300

N – normal distribution, S – skewed distribution, G – gamma distribution, L – lognormal distribution, NA – not applicable, H-UCL – UCL based on Land's H-statistic

* UCL selected as most representative by ProUCL

Table 3-8. Relative percent differences between 95% upper confidence limit and weighted means for copper, lead, antimony, and zinc.

Test	Sample Type	Cu (%)	Pb (%)	Sb (%)	Zn (%)
Student's-t UCL	Grab	39	NA	NA	1
H-UCL (lognormal)	Grab	35	264	337	0
Chebyshev UCL (lognormal)	Grab	23	265	1734	11
Student's-t UCL (Chen Mod)	Grab	39	NA	NA	24
Student's-t UCL(Johnson Mod)	Grab	39	NA	NA	27
Gamma Distribution UCL	Grab	37	NA	NA	0
Student's-t UCL	ISM 30	14	13	19	3
Chebyshev UCL	ISM 30	28	26	33	2
Student's-t UCL	ISM 50	6	10	10	1
Chebyshev UCL	ISM 50	10	1	2	8
Student's-t UCL	ISM 100	38	27	21	28
Chebyshev UCL	ISM 100	67	45	38	42

NA – not applicable

3.5 Conclusions

Unless a large number (e.g., greater than 100) of samples are collected, grab samples yield unreliable estimates of the mean for soils when the contaminant of interest is a solid distributed heterogeneously. This is of particular concern for energetic compound residues and metallic fragments deposited on DoD training lands as individual grab sample results for anthropogenic metals are highly variable. The primary advantage of ISM over conventional grab sampling is its high precision for estimating the DU mean. This is critical for environmental work because inferences about human and ecological risk usually depend on the means of the DUs, which are typically estimated using 95% UCLs. Reliable estimates of DU means are very problematic for grab sampling, when the sample sizes are small, especially when there is large variability and the distributions are positively skewed.

ISM effectively quantifies data quality, principally precision, through the collection of replicate ISM samples. Data quality assessments are also possible with grab samples but require a large number to provide a reliable estimate of the variance. Consequently, a set of replicate ISM samples provides a comparable or better estimate of the DU mean than a much larger number of grab samples. Further, this can be accomplished at a lower project with ISM samples.

CHAPTER 4

SAMPLE PROCESSING OF SOILS CONTAINING METALLIC RESIDUES WHEN USING THE INCREMENTAL SAMPLING METHODOLOGY (ISM)[14]

4.1 Abstract

Results from numerous studies of military training ranges across a variety of soil geochemical conditions indicate energetic and metallic residues in surface soils possess spatially heterogeneous distributions that cannot be readily characterized using conventional grab sampling. Adequate characterization of military unique chemical constituents at training ranges requires an alternative sampling strategy such as the ISM. We implemented a controlled study to explore two aspects of sample processing often considered with ISM: (1) the utility of splitting samples in the field to reduce the sample mass shipped to the analytical laboratory and (2) processing a larger sub-sample mass for digestion in the laboratory in lieu of milling. Statistical evaluations were conducted to determine if these two strategies are viable.

For the first part of the study, we investigated cone-and-quartering and a sectorial rotary splitter. Both sub-sampling techniques resulted in poor precision

---

[14] This chapter has been submitted for publication in: Clausen, J., T. Georgian, K. Gardner, and T. Douglas. (2014). “Sample processing of soils containing metallic residues when using the Incremental Sampling Method (ISM)”. Chemosphere.

for replicate analyses and positively skewed distributions of replicate concentration measurements. Cone-and-quartering mean and median values for a parent soil split into fourths with 10 replicate samples yielded statistically significant differences. Similarly, evaluation of sample splitting using a sectorial rotary splitter for ISM and grab soil samples yielded unacceptable levels of uncertainty. Even with the collection of incremental samples designed to minimize heterogeneity, sectorial rotary splits of the same samples yielded highly variable results with concentrations of copper and lead varying over an order of magnitude.

Increasing the digestion mass by a factor of five, from 2 to 10 g, for unmilled ISM samples resulted in highly variable concentrations relative to milled sub-samples from the same material. In contrast, the milled incremental samples exhibited smaller variability inversely proportional to the sub-sample mass. Although incremental field sampling improves data quality, soils with heterogeneously distributed metallic residues require additional sample processing in the laboratory (e.g., particle size reduction by milling) to achieve good precision.

4.2 Introduction

Combinations of field sampling and laboratory sample processing techniques referred to as the ISM, MIS™, or IS are being increasingly used for characterization studies (Florida 2013; Hewitt *et al.* 2012, 2009, 2007; ITRC 2012; Alaska 2009; Ramsey 2009; USACE 2009; Hawaii 2008). The ISM approach allows for an estimate of the mean contaminant concentrations when particulates are present in environmental media (e.g., soils and sediments) possessing large compositional and distributional heterogeneity (ITRC 2012). The underpinning of ISM is Gy's sampling theory, which was originally developed for the mining industry and suggests total sample error depends on particle size and sample mass (Pitard 1993, Gy 1992, 1982). Initial research on environmental applications of Gy's theory focused on explosive residues (Hewitt *et al.* 2009, 2007, 2005; Jenkins *et al.* 2005a, b, 2001, 1998, 1997a, b, 1996). Particles of explosives released into the environment (e.g., soils at military training ranges) of variable sizes, shapes and compositions result in spatially heterogeneous distributions of analyte concentrations. Similarly, soils at military small-arms ranges often contain metallic residues (e.g., bullet fragments) (Clausen and Korte 2009a) distributed in a highly heterogeneous fashion (Clausen *et al.* 2014a, *2013*a*,* 2012a). The utility of conventional grab sampling for characterizing heterogeneously distributed metallic residues at small arms ranges was questioned by Clausen *et al.* (2012a, b) and Hewitt *et al.* (2012). However, research and guidance on the use of ISM versus for sites with heterogeneous distributions of metallic residues is lacking.

ISM relies on initially identifying the spatial boundaries of the DU, the environmental population of interest (e.g., volume of soil requiring characterization). A DU is “*the smallest volume of soil (or other media) for which a decision will be made based upon ISM sampling*” (ITRC 2012, Ramsey 2009, 2006). For example, a DU may be sampled to determine whether a risk-based threshold is exceeded or remedial activities are needed. The ISM typically entails collecting many increments (e.g., 30 – 100) over the entire DU, usually using simple random or systematic random sampling (as shown in Figure 1-1). The increments are combined to prepare a single composite sample (ITRC 2012) of 0.5 to 5 kg in total mass. In addition to incremental sampling in the field, ISM involves modifications to the sample processing procedures normally used by analytical laboratories. These modifications often include air-drying, sieving, milling, and incremental sub-sampling (Figure 3-1).

Interest in ISM by the DoD stems from directives mandating environmental assessments of impacts from munitions deposition at active DoD facilities including training and testing ranges (USDOD 2007, 2005). The frequent use of munitions containing metals on Army training lands results in accumulation of metal on surface soils (Clausen and Korte 2009a, Clausen *et al.* 2004). Recently, Clausen *et al.* (2014a, b, 2013a, 2012a) demonstrated the inappropriateness of using conventional grab sampling of soils with metallic residues at military small-arms ranges. Use of ISM resulted in better precision and higher estimates of the mean concentrations compared with conventional grab sampling methods (Clausen *et al.* 2014b). Although ISM entails longer sample collection times and larger sample volumes than conventional grab sampling, ISM greatly reduces field and laboratory costs overall, because fewer numbers of samples need to be collected, processed, analyzed, and validated (Clausen *et al*. 2013b). Despite these findings, there is resistance in the environmental assessment community to implement ISM, in part, because environmental analytical laboratories cannot readily process the larger sample mass resulting from multi-increment sampling or a lack of milling equipment. Consequently, there have been attempts to omit or modify some of the ISM sample processing steps shown in Figure 3-1. This paper assesses the value of (1) field splitting to reduce the sample mass sent to the laboratory for analysis and (2) collecting larger sub-sample masses in lieu of milling.

4.3 Methods

4.3.1 Soil Sample Collection

A single 200-increment surface soil sample was collected to a depth of 5 cm using a 2-cm diameter corer (Walsh 2009) yielding a 20-kg sample from the small-arms range berm face DU on Range 4-3 at Camp Ethan Allen in Jericho, Vermont. Camp Ethan Allen has a humid continental climate with warm, humid summers and cold winters with monthly daily average temperatures ranging from 70.6^{o}F (21.4^{o}C) in July to 18.7^{o}F (-7.4^{o}C) in January and annual precipitation of 36.8 in (935 mm). Range 4-3 is used by Army National Guard units for training with small arms such as the M16 Rifle. The berm surface is sparsely vegetated; the soil is a loamy sand containing gravel and visible small-arms bullet fragments

(Clausen *et al.* 2012a). ISM sample collection followed the procedures outlined in (Clausen *et al.* 2013a, 2012a, ITRC 2012). The berm face dimensions are approximately 3 x 100 m. Clausen *et al.* (2014a, 2013a, 2012a) presents additional information on the characteristics of the range.

4.3.2 Sample Preparation and Analytical Procedures

The initial 20-kg sample was sent to the CRREL geochemistry laboratory in Hanover, NH, air-dried, and processed through a Lab Tech Essa sectorial rotary splitter (Model RSD 5/8, Belmont, Australia) operated at 100 rpm to create 12 splits of approximately 0.5 kg each. One of the 12 rotary sectorial splits were selected for the unmilled digestion experiments described in Section 4.4.2. Five of the splits were used for experiments not reported in this paper. The remaining six splits were recombined for the cone-and quartering experiment to prepare a sample of approximately 10-kg, which was subsequently split into quarters of approximately 2.5 kg each. Cone-and quartering entails manually mixing soil prior to manually splitting. Each of the four cone-and-quartered splits was passed through a no. 10 mesh sieve to remove material larger than 2-mm. The sieved (less than 2-mm) fraction of each split were milled separately in a Lab Tech Essa chrome steel ring mill grinder (Model LM2, Belmont, Australia) for five 60-second intervals with 60 seconds of cooling between each interval. The metallic composition of the ring mill and bowl is primarily iron and chromium with lesser amounts of manganese and nickel and other ancillary metals (Clausen *et al.* 2012a).

The procedure in USEPA Method 8330B (USEPA 2006a) was used to sub-sample the four cone-and-quartered splits and the unmilled sectorial split. Briefly, each soil split was spread onto a sheet of aluminum foil to produce a layer 1 to 2 cm thick. Twenty increments were collected in a systematic random fashion (similar to the field sampling approach) and combined to prepare a sub-sample for analysis (i.e., acid digestion and instrumental analysis by inductively coupled optical emission spectrometer (ICP-OES)). Ten 2-g replicate aliquots were analyzed for each cone-and-quartered split (after milling). Fifteen replicate aliquots of mass 0.5, 1, 2, 5, and 10-g were analyzed for the unground split. The remaining portion of the unground split was subsequently milled and 15 replicate subsamples of masses 2, 5, and 10-g were analyzed.

Instrumental analysis was performed at CRREL using a Thermo Fischer iCAP 6300 Duo view instrument equipped with a CETAC ASX-520 auto sampler following USEPA Method 6010 (USEPA 2006c). The ICP-OES operating conditions included setting the RF power torch at 1,150 W, auxiliary gas flow rate at 0.5L/min, nebulizer gas flow rate at 0.7 L/min, and pump flow rate of 50 rpm. The ICP-OES provided results for aluminum, antimony, barium, cobalt, copper, magnesium, manganese, nickel, lead, strontium, vanadium, and zinc. In general, the quantification limit was approximately 1 mg/kg, although most sample results were well above this value.

4.4 Results and Discussion

4.4.1 Splitting

4.4.1.1 Cone-and-Quartering

To assess whether cone-and-quartering in the field is an appropriate technique to reduce sample volume the average metal concentrations of the four splits were statistically compared. As all of the data sets exhibited non-normal distributions, the non-parametric Kruskal-Wallis test was used to compare the median metal concentrations. Statistically significant differences at well over the 99% level of confidence were observed for the anthropogenic small-arms metals antimony, copper, lead, and zinc (Table 4-1).

Table 4-1. Statistical evaluation for the cone-and-quartering experiment using the Kruskal-Wallis test.

Metal	Group*	Median (mg/kg)	p-value for KW test
copper	1	421	0.000
	2	340	
	3	443	
	4	547	
lead	1	3220	0.000
	2	1420	
	3	1260	
	4	3120	
antimony	1	25.8	0.000
	2	9.55	
	3	8.71	
	4	26.1	
zinc	1	73.4	0.004
	2	59.2	
	3	60.9	
	4	65.7	

*Each group consists of n=10 laboratory replicates.

The lead and antimony medians and means of the different splits differ by a factor of two whereas less difference is evident between copper and zinc (Figure 4-1 and Table 4-2). For example, the lead mean values for splits 1, 2, 3, and 4 are 3295, 1447, 1281, and 1337 mg/kg, respectively. These results call into question the appropriateness of field splitting of soils containing contaminants in particulate form even when incremental field sample collection methods are used. My findings are consistent with the work of Walsh *et al.* 2005 and Gerlach and Nocerino (2003) that established cone-and-quartering yields biased results and should be avoided (Gerlach *et al.* 2002; Pitard 1993). In contrast, a smaller difference is evident between the sample means and medians for non-anthropogenic metals aluminum, barium, cobalt, magnesium, manganese, nickel, strontium, and vanadium (Table 4-2) and as illustrated for barium, cobalt, nickel, and vanadium (Figure 4-2).

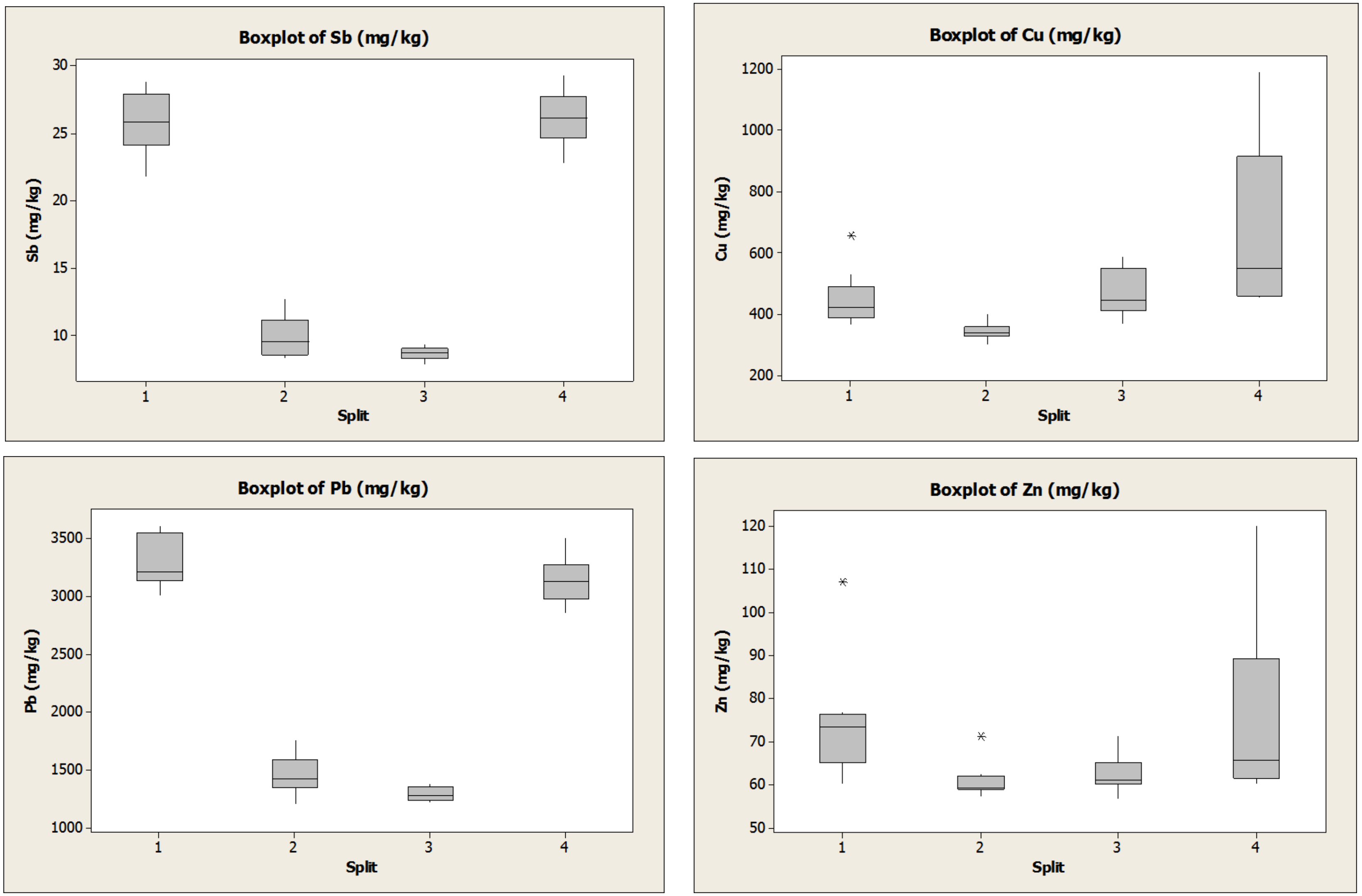


Figure 4-1. Box plots comparing cone-and-quartered splits for antimony, copper, lead, and zinc.

Table 4-2. Summary of metals results for the cone-and-quartering experiment.

	Concentration (mg/kg)												
	Al	Ba	Co	Cr	Cu	Mg	Mn	Ni	Pb	Sb	Sr	V	Zn
Split 1													
n	10	10	10	10	10	10	10	10	10	10	10	10	10
Mean	5863	29.0	7.87	194	450	2144	223	10.9	3295	25.7	19.9	15.8	73.9
Median	5698	29.3	7.81	193	421	2184	227	10.8	3216	25.8	19.1	15.8	73.4
Min	5480	26.7	7.66	187	366	1946	203	10.5	3011	21.8	18.1	15.2	60.2
Max	6725	30.6	8.64	206	658	2262	235	11.6	3614	28.8	23.5	16.5	107
STD	435	1.31	0.286	6.00	87.0	110	11.3	0.365	213	2.17	1.87	0.380	13.2
RSD (%)	7	5	4	3	19	5	5	3	6	8	9	2	18
Split 2													
n	10	10	10	10	10	10	10	10	10	10	10	10	10
Mean	6111	27.9	7.71	172	345	2224	224	10.5	1447	10.0	20.4	14.8	60.8
Median	6078	26.8	7.74	172	340	2079	211	10.3	1420	9.50	20.6	14.7	59.2
Min	5500	24.8	7.29	161	303	2036	201	9.83	1198	8.37	18.1	14.1	57.3
Max	6580	41.4	8.41	188	398	3492	351	12.1	1756	12.7	22.1	15.9	71.2
STD	305	4.90	0.329	6.81	26.0	447	45.0	0.615	163	1.49	1.07	0.501	4.00
RSD (%)	5	18	4	4	7	20	20	6	11	15	5	3	7
Split 3													
n	10	10	10	10	10	10	10	10	10	10	10	10	10
Mean	6687	26.1	7.92	177	467	2080	207	10.6	1281	8.67	22.3	14.8	62.6
Median	6750	26.0	7.91	177	443	2078	206	10.6	1264	8.71	22.6	14.9	60.9
Min	6440	24.8	7.82	174	368	2023	201	10.4	1216	7.89	21.3	14.1	56.7
Max	6860	28.0	8.10	180	587	2145	218	10.8	1369	9.34	23.0	15.1	71.3
STD	160	0.924	0.083	2.08	73.5	31.6	4.46	0.090	58.5	0.461	0.623	0.339	4.54
RSD (%)	2	4	1	1	16	2	2	1	5	5	3	2	7
Split 4													
n	10	10	10	10	10	10	10	10	10	10	10	10	10
Mean	6826	23.9	7.97	175	674	2054	203	10.8	3137	26.1	22.8	15.2	77.0
Median	6798	23.7	7.96	175	547	2057	203	10.9	3123	26.1	22.7	15.1	65.7
Min	6605	22.6	7.84	167	454	2011	199	10.6	2856	22.7	22.0	14.6	60.1
Max	7075	25.2	8.10	180	1190	2085	206	11.0	3505	29.2	23.8	15.8	120
STD	154	0.763	0.088	3.67	266	19.9	2.31	0.106	195	2.09	0.620	0.433	21.9
RSD (%)	2	3	1	2	39	1	1	1	6	8	3	3	29

n - number, Min - minimum, Max - maximum, ND - not detected, RSD - percent relative standard

Highlighted text indicates known contaminants of interest present.

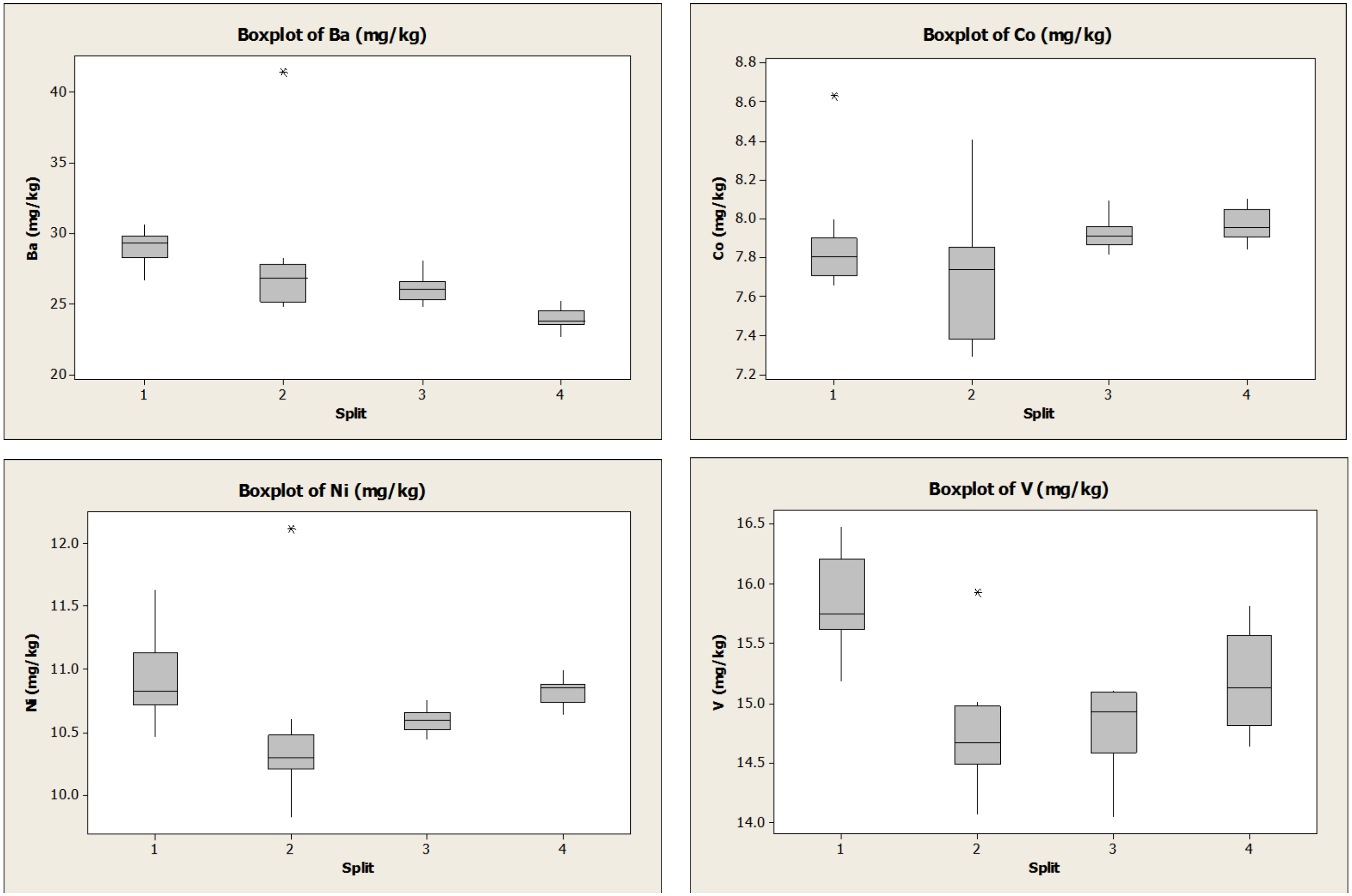


Figure 4-2. Box plots comparing cone-and-quartered splits for barium, cobalt, nickel, and vanadium.

For each split, lead met our target criteria for within-sample (subsampling) variability of less than 15% as measured by RSD (Figure 4-3). The value of 15% was selected for intra sample comparisons and a value of 30% used for inter sample comparisons (Clausen *et al.* 2013a; Taylor *et al.* 2011). Our precision target of 15% was exceeded for zinc splits 1 and 4 and copper splits 1, 3, and 4. Individual replicate concentration results for copper indicate the range in values for split 4 is quite large ranging from 454 to 1,190 mg/kg (Figure 4-4). This particular example suggests particulates of copper are present in the sample. Consequently, the collection of a single or a few samples from a field split has the potential to yield misleading or biased results. Without collection of many replicate samples, the degree of heterogeneity and uncertainty in any individual reported result would not be apparent.

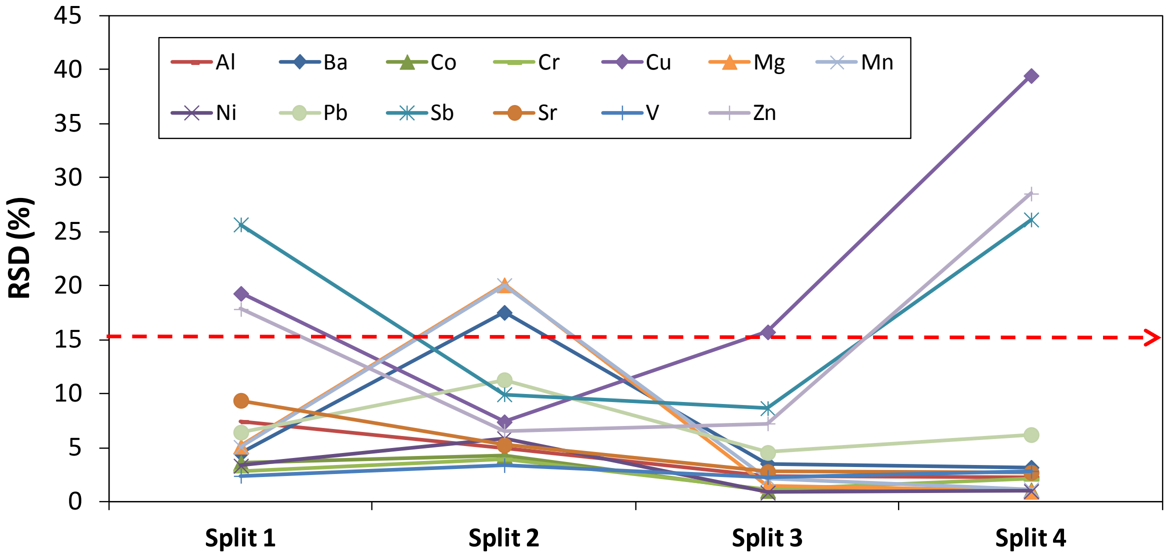


Figure 4-3. Intra split sample variability measured by the percent relative standard deviation.

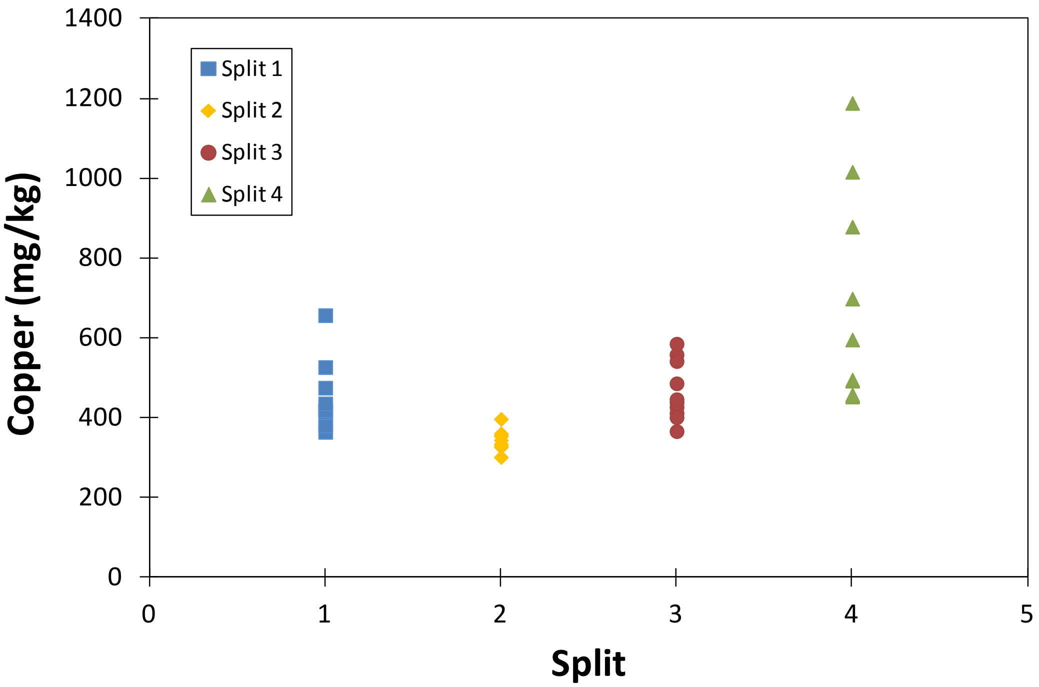


Figure 4-4. Individual value plots of copper replicates for the four splits.

In contrast, the non-anthropogenic metals all had RSDs of less than 10% within splits (Figure 4-3) and between splits (Table 4-3). These results indicate the distribution of the non-anthropogenic metals is relatively homogenous and sample reduction using cone-and-quartering splitting techniques appears acceptable.

Table 4-3. Between split metal variability.

	Mean (mg/kg)												
	Al	Ba	Co	Cr	Cu	Mg	Mn	Ni	Pb	Sb	Sr	V	Zn
Split 1	5863	29.0	7.87	194	450	2144	223	10.9	3295	25.7	19.9	15.8	73.9
Split 2	6111	27.9	7.71	172	345	2224	224	10.5	1447	10.0	20.4	14.8	60.8
Split 3	6687	26.1	7.92	177	467	2080	207	10.6	1281	8.67	22.3	14.8	62.6
Split 4	6826	23.9	7.97	175	674	2054	203	10.8	3137	26.1	22.8	15.2	76.7
RSD%	7	8	1	6	28	4	5	2	47	54	7	3	12

However, as shown in Figure 4-2 outliers are evident for at least one of the samples splits for each metal. Recall that our parent sample consists of 200-increments, yielding a sample mass of approximately 20-kg, which is factor 10 to 40 larger than the typical ISM sample. Pierre Gy's equation, see Equation 1, for the calculation of the fundamental error of a sample suggests increasing error as the mass of the sample decreases with all other variable held constant (Pitard 1993).

4.4.1.2 Sectorial Rotary Splitting

To address whether the specific splitting technique, cone-and-quartering, contributed to the differences in mean anthropogenic metal values a sectorial rotary splitter operated at 60 rpm was evaluated. Rotary splitters tend to produce results more representative than other splitting techniques such as cone-and-quartering because many increments are collected for each sample split (Gerlach and Nocerino 2003).

Two soil samples consisting of 100-increments, approximately 2-kg each, were collected from a military small-arms berm located in the Delta River downrange from the Lampkin Firing Point at Donnelly Training Area located 10 miles east of Delta Junction, Alaska. The Donnelly Training Area has a subarctic climate with short, warm summers, cold, dry winters, and annual precipitation of 11.1 in (282 mm). The berm is constructed of river gravel from the Delta River and was primarily used as a target for firing small arms (5.56-mm), machine guns (.50 cal), and 40-mm grenades. The soil samples were air-dried and then sieved using a no. 10 mesh (2-mm) sieve (but were not milled). Splitting of the less than

2-mm material involved use of a sectorial rotary splitter to make 15 sample splits. Subsamples of 2-g were collected from the 15 splits and digested using USEPA Method 3050B (USEPA 2006a).

Large differences, up to an order of magnitude, are evident between some individual splits (Table 4-4). For example, the difference between the minimum and maximum lead value for Sample #1 is 290 and 2,800 mg/kg, which occurs in splits 13 and 6. Further, if multiple analytes are of interest there is no guarantee the maximum results for all of the analytes will occur in the same split. For example, the maximum lead concentration in Sample #1 occurs in split 6, but the maximum antimony value in Sample #1 is observed for split 1. The RSD of copper for Sample #2 is only 13%, which is below our target of 15% error for intra sample comparison. Given the large RSD for copper for Sample #1 it is suspected the small RSD for Sample #2 is biased low because the sample size (n = 15) was too small to reliably measure the variability. However, it is not possible to demonstrate this without analyzing a larger number of replicates. Although Samples #1 and #2 were collected from the same DU, within-sample variability and the mean metal concentrations differ significantly (Table 4-4).

Table 4-4. Difference in split results using a sectorial rotary splitter for two unmilled soil samples from a small-arms range berm in Alaska.

	Sample #1		Sample #2	
Sub-sample Split Number	Lead (mg/kg)	Copper (mg/kg)	Antimony (mg/kg)	Copper (mg/kg)
1	360	2600	5.5	99
2	330	110	5.0	90
3	920	300	7.6	87
4	300	110	4.3	99
5	280	130	4.3	130
6	2800	140	16	90
7	1600	860	12	88
8	330	540	4.6	99
9	850	1200	4.2	83
10	1500	130	4.5	98
11	380	1900	4.9	99
12	330	120	4.3	110
13	290	130	3.7	80
14	300	120	4.1	87
15	820	110	8.2	84
Median	360	130	4.6	90
Mean	759	567	6.2	95
Std Dev	699	750	3.4	12
RSD (%)	92	132	55	13

RSD – relative standard deviation
Std Dev – standard deviation

4.4.1.3 Splitting Assessment

As noted by Gerlach and Nocerino (2003), Gerlach *et al.* 2002; Pitard 1993 sampling can be a major source of error; therefore, field splitting is not recommended. Our findings support this conclusion when the contaminant of interest is in particulate form, even when field sampling involves incremental collection techniques. Incremental field sampling addresses distributional heterogeneity within the DU but field sample-splitting techniques do not adequately take into account the within-sample compositional heterogeneity when the contaminants of interest are present as relatively pure particles or “nuggets” of material (e.g., fragments of explosives or bullets). This is evident when examining the replicate sample results from each split (Figure 4-4). If sufficient replicate samples are collected the mean analyte concentrations of various splits will differ significantly (Walsh *et al.* 2004). However, this within sample variability is typically not apparent when collecting three replicates as suggested by the ITRC (2012). Our findings indicate sample-splitting techniques do not control distributional heterogeneity when particulates are present in samples (e.g., resulting in splits with significantly different mean concentrations). When the splits are not milled, the fundamental error associated with the parent material is similar to the split samples (Walsh *et al.* 2005). Splitting to reduce sample mass or volume may be acceptable in situations where the analyte of interest is of non-anthropogenic origin and a number of replicates are analyzed (e.g., to estimate the mean and assess the variability). However, when samples contain metal fragments, sample mass reduction using splitting techniques needs to occur after particle size reduction in order to obtain reproducible results.

4.4.2 Digestion Aliquot Mass

To account for heterogeneity and obtain adequate laboratory sub-sampling precision, analysis of larger aliquot masses (in lieu of milling) has been proposed in combination with incremental sampling. Method 3050B calls for the digestion of 2-g aliquots. To evaluate the effect of aliquot mass on precision, 15 replicate 2-, 5-, and 10-g grab subsamples of the same unground parent Camp

Ethan Allen, 200-increment, 20-kg sample were analyzed. Replicate 0.5-, 1-, 2-, 5-, and 10-g aliquots were subsequently analyzed after the parent material was milled.

4.4.2.1 Unmilled Soil

Figure 4-5 shows the within sample variability of lead for the unmilled 2-, 5-, and 10-g samples as well as the milled 0.5-, 1-, 2-, 5-, and 10-g samples. The unmilled samples exhibit a high variability in lead values with median values increasing with the mass of digested material. This is related to the observation that the number of particles present increases with digestion aliquot mass (Figure 4-6) as well as with the lead concentration assuming a uniform particle size of 2 mm. For a soil with 1 mg/kg lead a digestion mass of at least 25-g is needed before one particle of lead on average is encountered (see Appendix Table S12). At a lead soil concentration of 10 mg/kg there are no lead particles present on average until a 5-g aliquot mass is reached. If the probability for encountering a single lead particle at a soil concentration of 1 mg/kg is calculated (see Appendix Table SI13), the value never reaches more than 0.2 even with an aliquot mass of 100-g (Figure 4-9). For soils with a concentration of lead greater than 10 mg/kg a minimum aliquot mass of 100-g is needed to ensure at least one particle is present every time a replicate sample is collected.

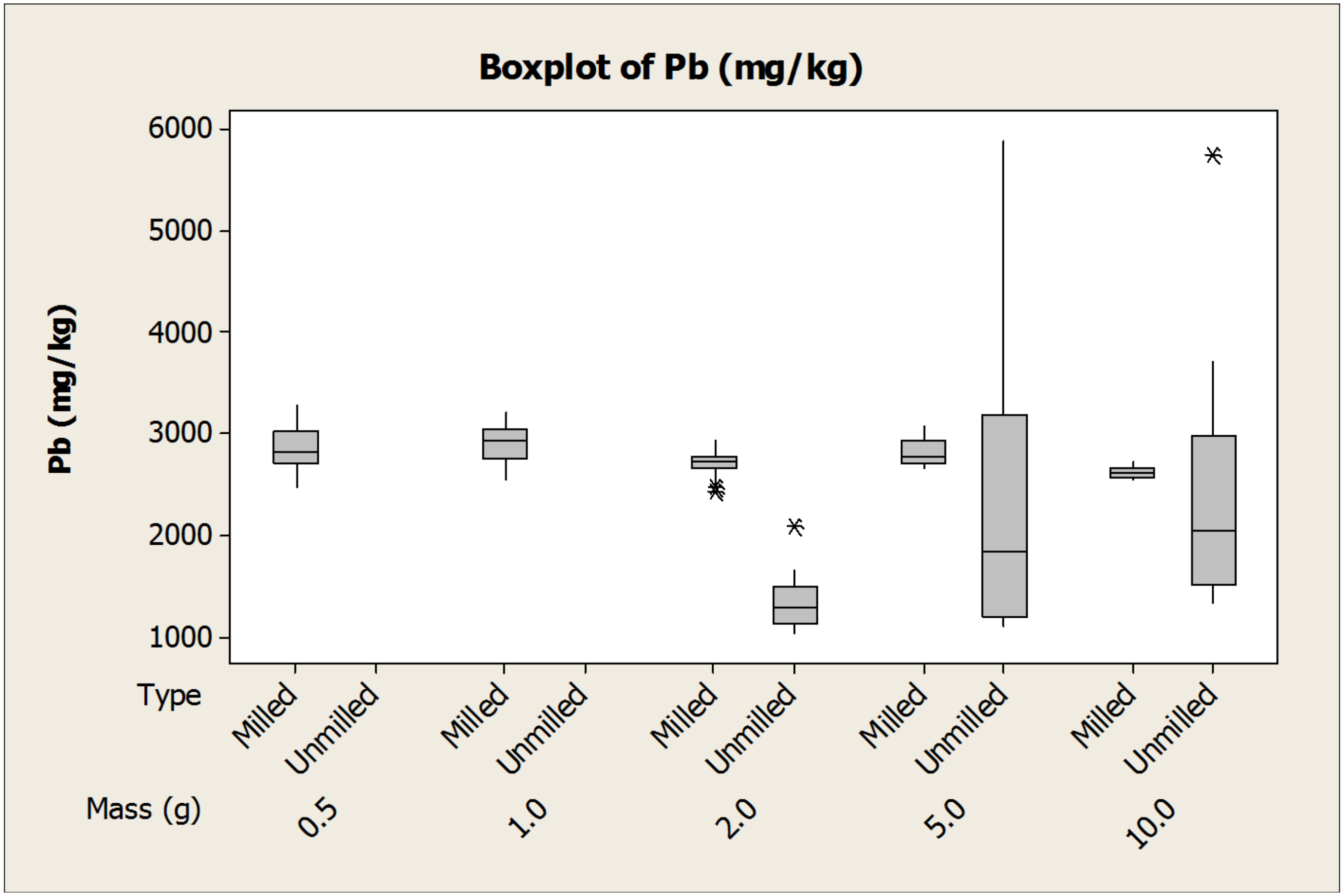


Figure 4-5. Box plot of milled and unmilled soil lead results by digestion mass.

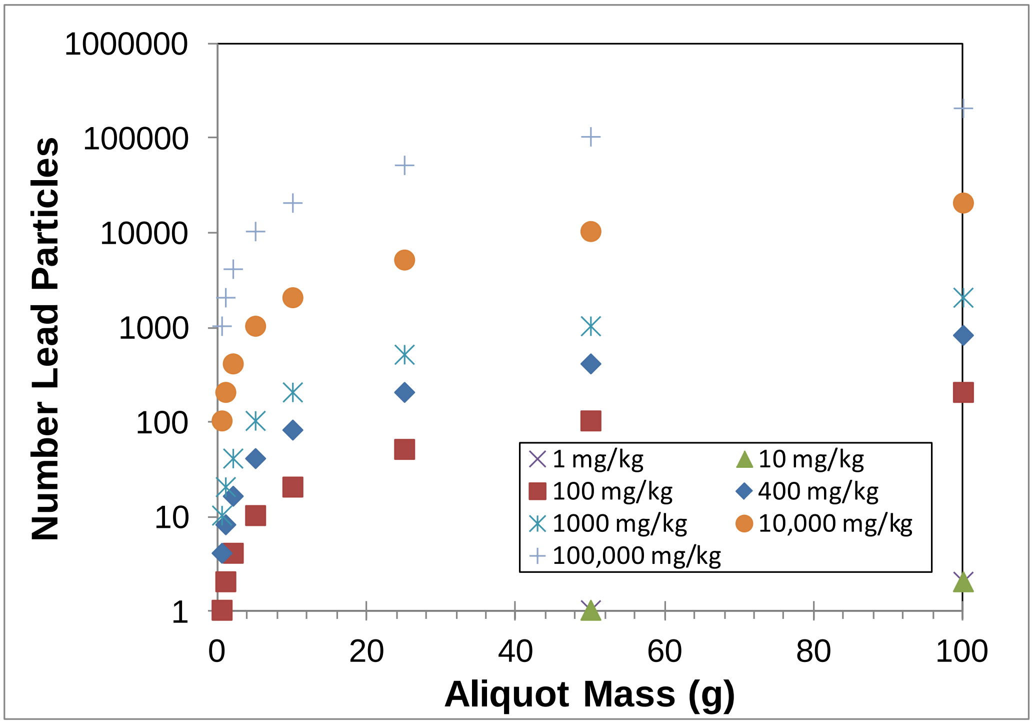


Figure 4-6. Number of lead particles by digestion mass for an unmilled soil sample.

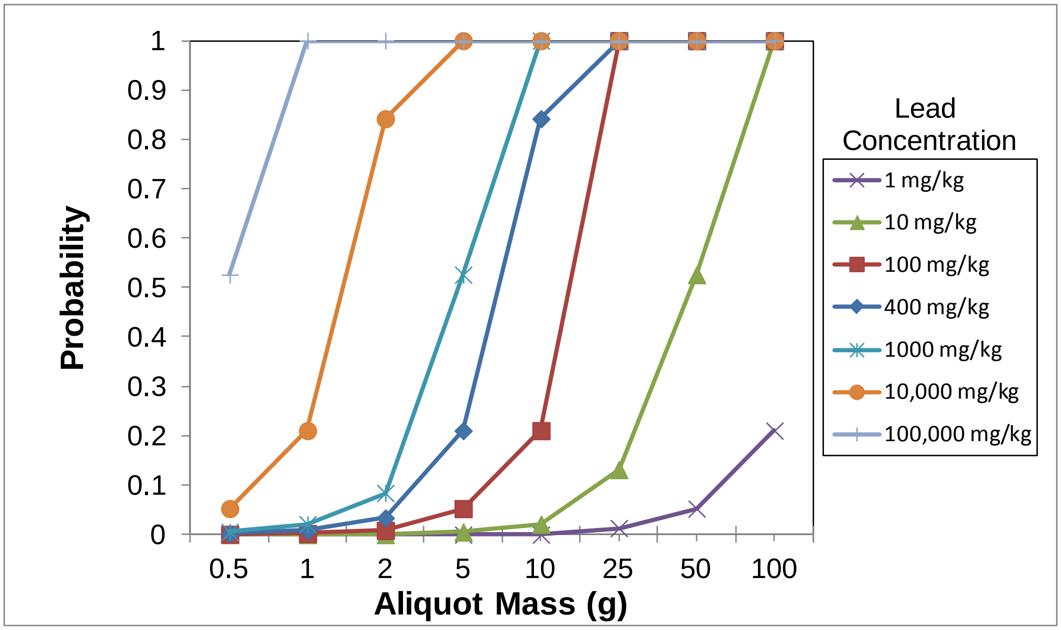


Figure 4-7. Probability of encountering a single lead particle by digestion mass and soil lead concentration for an unmilled soil sample.

Consequently, the increase in variability associated with an increased unmilled aliquot mass reflects the increased probability of having a lead particle in any given replicate sample. However, the number of particles present even at 3,000 mg/kg lead is still only a few hundred assuming uniform particle size (see Appendix Table SI12). Since, the unmilled soil has both soil and lead particles ranging in size from the nanoscale to 2-mm the number of particles from sample to sample will vary considerably in an unpredictable manner as will the concentration. Even if the entire 1-kg soil sample at a concentration of 100,000 mg/kg lead was digested the number of particles theoretically present is still a factor of 5 less than a milled soil sample aliquot of 2-g having a lead concentration of 1 mg/kg. Simply increasing the digested mass of an unmilled sample does not address the inherent heterogeneity of the parent material. A finding consistent with the work of Walsh *et al.* (2002) where increasing the aliquot mass did not reduce the total sample error for energetics in soil.

Therefore, we do not recommend grab sub-sampling of an unmilled soil a finding consistent with Petersen *et al.* (2004); Gerlach *et al.* (2002); Mullins and Hutchison (1982); and Allen and Khan (1971). Furthermore, we suggest the increased variability in lead values for the unmilled samples is due to the increase probability of encountering a greater number of lead particles with the larger aliquot mass. Because, these particles have varying size/mass there is also an increased probability that the variance will increase. This variability in individual results may not be apparent with analysis of three replicate samples as suggested by ITRC (2012). The increased variance for the anthropogenic metals in the unmilled aliquots results in large differences in the sample means (Table 4-5). For example, the mean lead values increase with aliquot mass and vary from 1,600 to 2,395 mg/kg for the 2 and 10-g aliquots, respectively. This phenomenon is also observed for the other anthropogenic metals copper, antimony, and zinc (Figure 4-8).

Table 4-5. Mean metal concentrations of unmilled and milled samples for each sub-sample mass.

		Concentration (mg/kg)											
Sample T	Digested Mass (g)	Al	Ba	Co	Cr	Cu	Mg	Mn	Ni	Pb	Sb	V	Zn
Unmilled	2	4041	16.2	5.79	2.09	337	2180	195	8.39	1344	3.8	8.74	55.8
Unmilled	5	3547	12.6	5.19	1.53	370	2005	180	7.76	2328	21.6	7.28	51.7
Unmilled	10	3676	13.3	5.28	1.73	548	2071	187	7.93	2395	25.7	7.66	70.6
Milled	0.5	5116	28.6	8.55	219	466	2120	229	11.2	2848	24.6	13.9	79.0
Milled	1	5550	28.6	7.74	225	610	2160	227	11.9	2913	23.3	14.6	83.4
Milled	2	5509	31.6	8.75	229	496	2242	235	12.4	2760	21.8	15.2	81.0
Milled	5	5311	28.3	7.98	222	526	2140	227	12.3	2815	21.7	14.8	81.2
Milled	10	5538	27.4	7.41	214	555	2014	211	11.6	2622	16.7	14.5	75.3

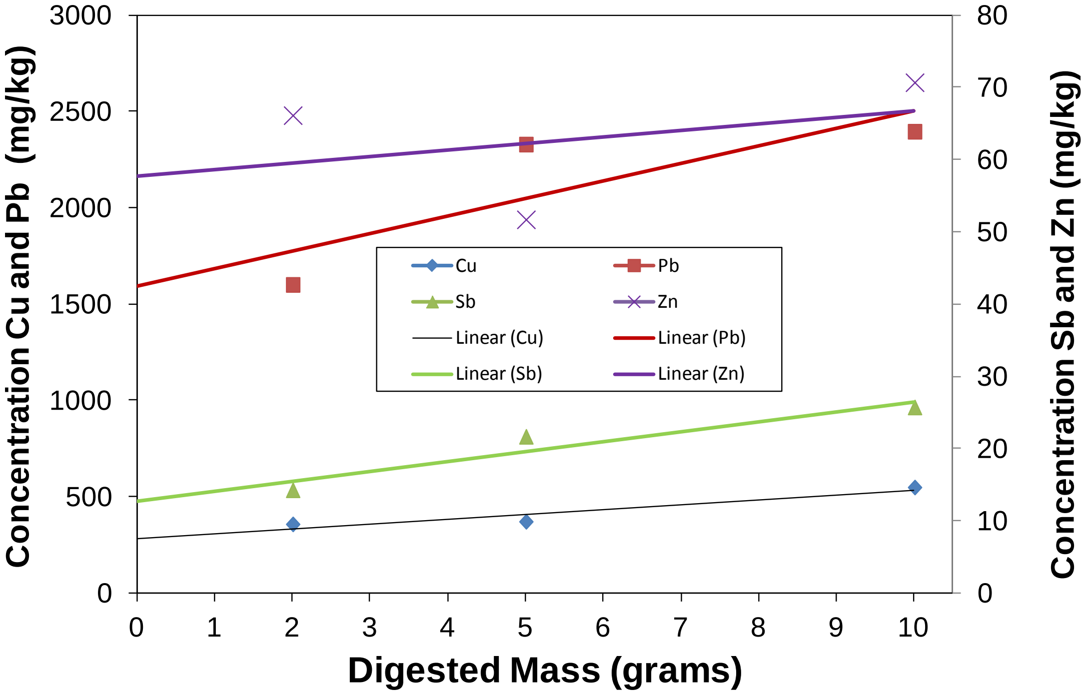


Figure 4-8. Mean anthropogenic metal concentrations of unmilled sub-samples versus digestion mass.

4.4.2.1 Milled Soil

In contrast, the number of small particulates present in the milled sample is large (see Appendix Table SI12), increasing the digestion mass does not greatly change the number of particles per unit mass. All of the digestion mass values evaluated resulted in low sample variability (Figure 4-9) and improved estimates of the mean as compared to the unmilled soil (Table 4-5). For the milled soil sample, assuming a particle size of 75 µm, there are close to 200,000 lead particles in a 0.5-g aliquot having a soil concentration of 1 mg/kg lead. Therefore, the probability of encountering at least several thousand particles of lead is 100% for all aliquot masses and soil lead concentration ranges explored. Consequently, there is a high degree of confidence of obtaining consistent soil lead concentrations for replicate samples across a wide range of lead soil concentrations and aliquot masses. Although, 2-g is the minimum aliquot mass recommended for the milled sample there is an order of magnitude increase in the number of particles with an aliquot mass of 10-g.

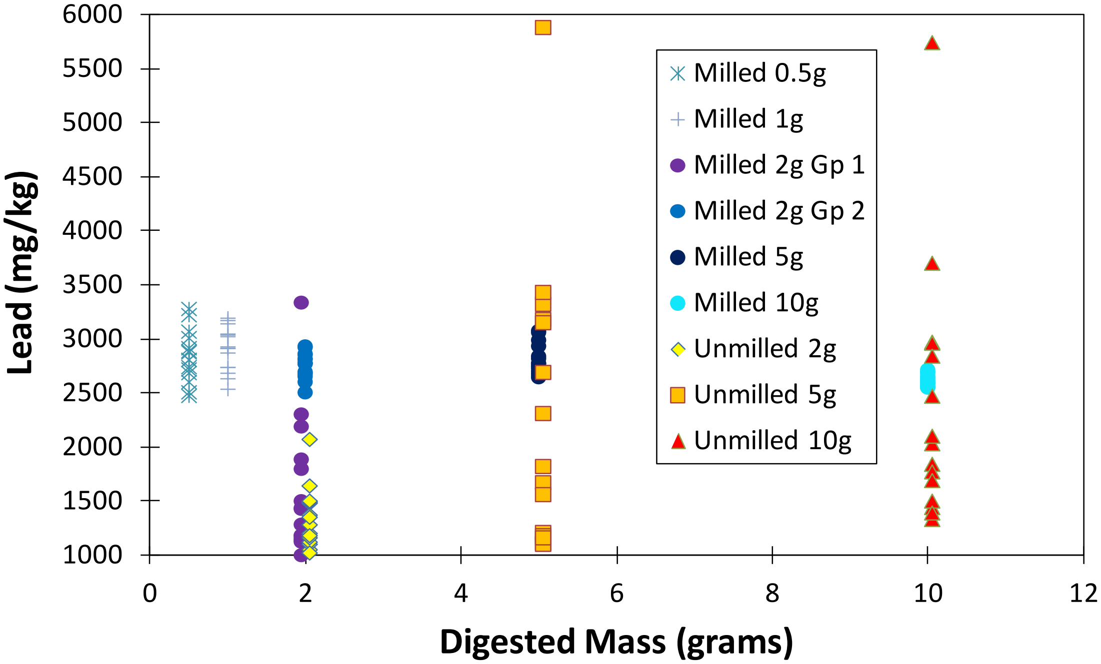


Figure 4-9. Milled and unmilled soil lead results by digestion mass.

As shown in Table 4-5 there is some marginal improvement in precision with increasing aliquot mass due to the increased number of lead particles. For example, all of the milled sample populations for lead have a decreasing variance with increasing aliquot mass (Figure 4-11). Further, the RSDs for lead and antimony of the milled results are consistently smaller than for the unmilled results (Table 4-5). The RSDs of the unmilled zinc and copper replicates of mass less than 10 g are similar to the RSDs of the milled replicates but the means of the unmilled replicates are negatively biased (relative to the means of the milled replicates). The relatively small RSDs of the unmilled replicates of masses less than 10 g are likely an artifact of small sample sizes (i.e., number of replicates).

Table 4-5. Percent relative standard deviations of the unmilled and milled samples by sub-sample mass digested.

		Percent Relative Standard Deviation											
Sample Type	Digested Mass (g)	Al	Ba	Co	Cr	Cu	Mg	Mn	Ni	Pb	Sb	V	Zn
Unmilled	2	2	15	2	5	25	1	1	2	39	69	4	17
Unmilled	5	2	3	3	15	13	2	2	2	57	116	3	4
Unmilled	10	1	4	2	12	90	1	2	1	49	106	6	76
Milled	0.5	3	6	2	3	20	2	2	3	8	12	4	10
Milled	1	4	4	2	1	36	2	1	2	7	12	4	15
Milled	2	1	4	1	2	15	2	2	1	4	7	2	10
Milled	5	3	6	5	2	23	3	3	3	5	9	2	15
Milled	10	4	4	2	1	11	4	3	1	2	5	2	7

It is likely a larger number of replicate samples for the unmilled group would have resulted in larger RSDs and sample means based on the results in Figure 4-6. When distributions are positively skewed small sample sizes tend to underestimate the variance and the mean (e.g., because the likelihood of obtaining an extreme value that lies in the right tail is small). The large variability of the unmilled masses suggests a single sample or even three replicate subsamples will likely be inadequate to obtain an accurate estimate of the mean of the parent sample.

Unlike the unmilled subsamples, increasing the aliquot mass of the milled subsamples had a marginal affect on the mean concentration (Table 4-4), though statistically significant differences between the median concentrations of the milled results were detected at the 99% level of confidence by the non-parametric Kruskal-Wallis test. Increasing the sample mass from 0.5- to 10-g results in smaller sub-sample variability (see Appendix Figures SI14 – SI16) a finding consistent with Hewitt et al (2009), Walsh et al. (2006, 2002), and Gilbert and Doctor (1985).

Consequently, our recommendation is that milling is required for soils expected to have particulate contamination and a minimum aliquot of 2-g digested material. In situations where a large variance is expected (e.g., owing to a heterogeneously distributed contaminant) or a large degree of uncertainty is unacceptable, 5- to 10-g aliquots should be digested. Additionally, as shown by Gilbert and Doctor (1985) if a smaller aliquot mass is used then a greater number of replicate samples are necessary to yield an estimate of the mean with low uncertainty.

4.5 Conclusions

I investigated two sampling processing approaches considered with the application of ISM at sites where trace metals were heterogeneously distributed in soils owing to military small-arms training activities:. The two most common method changes include using (1) Field splitting to reduce the sample mass shipped to analytical laboratories, and (2) preparation of larger digestion aliquots to eliminate the need for milling. I implemented a controlled study using statistical techniques to explore these two aspects of sample processing. . Two important results emerged. First, reduction of sample mass for ISM samples using splitting techniques or devices is not viable when contaminants are in the form of particulates. The splitting techniques fail to control heterogeneity, resulting in splits with significantly different mean concentrations. Even when incremental samples are collected in the field, compositing the increments addresses only the distributional heterogeneity, not the compositional heterogeneity. Second, additional laboratory sample processing, namely milling, is required to account for compositional heterogeneity. Analysis of sub-sample masses as large as 10 g for unmilled samples did not adequately address the compositional heterogeneity. Processing larger unmilled sub-sample masses still resulted in large variability. Unless field sample collection and laboratory sample processing procedures account for both the distribution and compositional heterogeneity, no single result will likely adequately reflect the true concentration at the site. Further, the variability owing to distributional and compositional heterogeneity will likely be apparent only when a large number of replicates is analyzed.

Field splitting using cone-and-quartering and a rotary sectorial splitter prior to laboratory sample preparation resulted in poor sample precision and highly variable results that were not reproducible. Cone-and-quartering mean and median values for four lots of 15 replicate samples yielded statistically significant differences. Furthermore, the data was highly skewed with a non-normal distribution. Sample splitting in the field using a rotary sectorial splitter was also evaluated for incrementally collected soil samples. Even with the collection of incremental samples, field splits of the same samples yielded highly variable results. Copper and lead values for the sample splits varied over an order of magnitude for the same sample. Increasing the digestion aliquot mass without implementing ISM processing steps such as milling and sub-sampling was insufficient to overcome sample heterogeneity issues. Increased sample error was associated with an increase in the digestion aliquot mass. Although incremental sampling in the field improves data quality, soils with heterogeneously distributed metal residues need modifications to the existing sample preparation procedures to achieve acceptable levels of total sample error.

CHAPTER 5

MILLING SOIL SAMPLES WITH METALLIC RESIDUES TO ADDRESS SAMPLE HETEROGENEITY AND REDUCE ANALYTICAL UNCERTAINITY[15]

5.1 Abstract

Metallic residues are distributed heterogeneously at small-arms ranges because of projectile fragmentation upon impact with the target or berm backstop material. As a consequence, soil samples collected from small-arms ranges can include a range of metallic residue particle sizes. This results in lack of reproducibility and unreliable estimates of mean metals concentrations. A new sample collection and processing approach, referred to as the ISM, is becoming more prevalent to address spatially heterogeneous contamination of surface soils. One aspect of the ISM process involves milling of the soil sample prior to analysis. However, the adequacy of milling equipment and length of milling interval necessary for adequate pulverization and mixing of soils containing metallic residues to meet desired levels of precision have not been established. The present study evaluated the use of a puck mill, ring and puck mill, ball mill, and mortar and pestle as well as the milling time necessary to achieve

---

[15] This chapter has been submitted for publication in: Clausen, J., T. Georgian, K. Gardner, and T. Douglas. (2014). “Milling soil samples with metallic residues to address sample heterogeneity and reduce analytical uncertainty”. Talanta.

acceptable levels of precision. The puck mill, ring and puck mill, or ball mill all yield acceptable levels of precision given a sufficient interval of milling. For soils from small-arms ranges containing particulates of antimony, lead, copper, and zinc a milling interval of five minutes with the puck mill is necessary. In contrast, a milling interval of at least 18 hours appears necessary when milling with the less aggressive ball mill. Metal cross-contamination from the puck mill is evident but the levels are not statistically significant, except for chromium. The results from the present study have applicability beyond small-arms ranges to include any environment where introduction of metallic residues occurs, e.g. military impact areas; mine tailing waste; coal gasification plants; sewage sludge; metal refining, production, and finishing, etc.

5.2 Introduction

Small-arms ranges worldwide are used for training with weapon systems that result in the introduction of projectile fragments into the environment containing antimony, copper, lead, and zinc (Clausen and Korte 2009a, Clausen *et al.* 2004). A significant amount of metal, principally lead, is deposited to surface soils, approximately 2 million tons/yr in the U.S. (USEPA 2005). This results in surface soil concentrations up to the percent level (Clausen *et al.* 2009; Cao *et al.* 2003) and metal contamination is thus a problem of international concern (Knechtenhofer *et al.* 2002, Scheinhost *et al.* 2006, Sorvari *et al.* 2006). Conventional grab or discrete sampling methods to address metal contamination in the environment have been questioned (Clausen *et al.* 2014a, b, 2012a, 2013a; Hewitt *et al.* 2012) due to studies with energetic contaminants. Numerous studies of energetic deposition indicate conventional grab sampling does not address heterogeneous distributions of particulates (Hewitt *et al.* 2009, 2007, 2005; Jenkins *et al.* 2005a, b, 2004a, b, 2001, 1998, 1997a, b, 1996). These studies found mean contaminant concentration estimates have a large uncertainty and are non-reproducible. Similarly, Clausen *et al.* (2014a, b, 2013a, 2012a) demonstrated with statistical analyses the inappropriateness of using conventional grab sampling procedures for soils with metallic residue contamination at small-arms ranges. These studies showed the necessity of collecting a much larger number of grab samples, (n > 35 at a minimum and preferably as many as a hundred), than are typically collected to address the large degree of spatial variability. Results from conventional grab sampling did not adequately represent the soil populations or result in acceptable precision for estimates of the population means. Calculated exposure point concentrations used in risk assessments, which are typically 95% UCLs of population means, were highly variable.

Research on the distribution and sampling of energetic residues at firing points and on impact areas led to recommendations (Hewitt *et al.* 2009, 2007, Thiboutot *et al.* 2002, 1998b) and modifications to USEPA Method 8330 (USEPA 1996a) for the sampling and processing of soil, resulting in an update to the Method (8330B; USEPA 2006a). Collectively, the modifications to the field sampling and sample processing methods are referred to as the Incremental ISM, MIS™, or IS (ITRC 2012; Ramsey 2009; USACE 2009; Hewitt *et al.* 2009, 2007). ISM involves the collection of tens to hundreds of discrete field increments from a defined area of interest (typically referred to as the DU) composited into a single sample and modifications to the sample processing procedures (ITRC 2012). The sample processing modifications include air drying, sieving, milling, changes to the sample aliquot mass, and sub-sampling.

The ISM approach allows for an estimate of the mean contaminant concentration when particulates are present in environmental media (e.g., soils and sediments) having a large compositional and distributional heterogeneity (ITRC 2012). The DoD, the regulatory community, and environmental consultants are now using ISM for analytes other than energetics, including metals (Florida 2013, Hewitt *et al.* 2012, ITRC 2012, Alaska 2009, USACE 2009, Hawaii 2008).

Despite the desire for and the use of ISM, information on appropriate milling devices or the milling interval necessary for soils containing metallic residues is non-existent. Milling is one of the principal steps to address contaminant heterogeneity (Clausen *et al.* 2014c) and is one-step of the ISM process (ITRC 2012). Milling is not an atypical process for some agencies (CSSS 2007; Peacock 2002; Mason 1992), but is not specified in USEPA Method 3050B, 3051, or 3052 for metals digestion (USEPA 1996a,b,c), which are commonly used to prepare solid samples. Clausen *et al.* (2014b) showed that simply collecting multiple increments in the field to form a composite sample only addresses distributional heterogeneity in the soil substrate, a finding consistent with Gy's sample theory (Pitard 1993; Gy 1992). Milling was necessary to deal with the compositional heterogeneity for soils containing metallic residues (Clausen *et al.* 2014c) as well as energetics (Hewitt *et al.* 2009; Walsh *et al.* 2007, 2002) to reduce the sub-sampling uncertainty prior to elemental analysis. However, most milling equipment has metallic surfaces exposed to the sample during the milling process. This results in metal cross-contamination of the sample during processing (Hartman 1992). A recent study by Clausen *et al.* (2012a) with glass beads and a contaminated soil from a small-arms range suggests significant metal cross-contamination from a puck mill for chromium, vanadium, and nickel. However, since the predominant metals of interest at small-arms ranges include lead, antimony, copper, and zinc (Clausen *et al.* 2013; Clausen and Korte 2009a; Duggan and Dhawan 2008; Sorvari *et al.* 2006; Vantelon *et al.* 2005; Cao *et al.* 2003k; Craig *et al.* 2002) there may be little to no risk of cross-contamination for these metals.

A related issue is the degree to which particle size reduction per se liberates metals from milled relative to unmilled soil samples. Soil samples are commonly digested in acid to transfer the metals into solution prior to analysis. A reduction in particle size results in increased surface area. The larger surface areas of the milled samples potentially results in larger metal concentrations relative to unmilled samples. Pulverization of a soil with contaminants having an insoluble oxide layer also allows more accessibility to this inner material. In either situation, milling may promote greater digestion efficiency. However, studies conducted by Felt *et al.* (2008) using “clean” (native) soils milled with equipment containing non-metallic milling surfaces (mortar and pestle, ball mill, and pulvisette) suggest changes in metal concentrations owing to improved digestion efficiencies is minimal. Similarly, preliminary studies by Clausen *et al.* (2010b) using a contaminated soil, Ottawa sand, and glass beads found a puck mill with metallic components combined with a modified digestion procedure only yielded slightly elevated concentrations of most metals, (10% to 30%) compared with an unmilled split sample following the standard Method 3050B (USEPA 1996a). However, recent small-arms range studies with a milled contaminated soil suggest significant increases in the estimated mean for antimony, copper, and lead (Clausen *et al.* (2012a). However, the larger metal concentrations for the milled samples were attributed to larger probabilities of capturing relatively pure metal fragments in the laboratory sub-samples analyzed rather than improved digestion efficiency.

Other issues related to milling include the type of milling equipment used and the length of milling to achieve a representative sample. A study of milling equipment by Felt *et al.* (2008) suggested the ball mill was not as effective as a puck mill for metals. However, Clausen *et al.* (2012a) obtained acceptable precision (i.e., percent RSDs less than 15%) with a ball mill. Although, the studies by Felt *et al.* (2008) used a milling interval of six hours, whereas Clausen *et al.* (2012a) milled the sample for 18 hrs, which may explain the differences.

5.3 Methods

The soil used in the following experiments was obtained from the military small-arms berm face at Range 4-3 located on Camp Ethan Allen in Jericho, Vermont. The berm face is approximately 3 x 100 m, sparsely vegetated, and the soil is a loamy sand containing gravel and visible small-arms bullet fragments (Clausen *et al.* 2012a). Range 4-3 is used by Army National Guard units for training with small-arms weapons systems such as the M16 Rifle that fires a 4.5 cm long by 5.56 mm diameter (caliber) projectile round. Clausen *et al.* (2014a, 2013a, 2012a) presents additional information on the characteristics of the range. Camp Ethan Allen has a continental climate with warm, humid summers and cold winters with monthly daily average temperatures ranging from 70.6^{o}F (21.4^{o}C) in July to 18.7^{o}F (-7.4^{o}C) in January and annual precipitation of 36.8 in (935 mm).

5.3.1 Soil Sample Collection

A single 200-increment surface soil sample was collected to a depth of 5 cm using a 2-cm diameter metal corer (Walsh 2009) yielding a 20-kg sample. Equipment rinseate samples previously demonstrated no cross-contamination from the soil corer (Clausen *et al.* 2012a). The 200-increment ISM sample collection followed procedures outlined in (Clausen *et al.* 2013a, 2012a, ITRC 2012).

5.3.2 Sample Preparation and Analytical Procedures

The 20-kg sample was sent to the CRREL geochemistry laboratory in Hanover, NH, air-dried, and sieved through a no. 10 mesh sieve to remove material larger than 2-mm (0.43 kg of which was set aside). A sectorial rotary splitter operated at 100 rpm was used to split the less than 2-mm material in half. Half of the material, 10-kg, was used in earlier experiments described in Clausen *et al.* (2014a,b,c). The other half of unmilled soil was spread out onto a flat surface and 15 replicates collected. The remaining 10 kg of less than 2-mm material was processed through a sectorial rotary splitter operated at 100 rpm to create 12 splits of approximately 0.8 kg each. Five splits were milled with the following equipment: (1) a Lab Tech Essa chrome steel puck mill grinder (Model LM2, Belmont, Australia) at CRREL, (2) a Lab Tech Essa chrome steel puck mill grinder (Model LM2, Belmont, Australia) at APPL Inc. (Clovis, CA), (3) a chrome steel TM Engineering ring and puck mill grinder (Model TM/S) at Test America (Denver, CO), (4) a Stoneware roller mill (Model 803FVM) with polyurethane coated steel jars and agate milling stones at Test America (Denver, CO), and (5) ceramic mortar and pestle at CRREL.

The splits destined for the two puck mills, puck and ring mill, and mortar and pestle were milled for five 60-second intervals with 60 seconds of cooling between each interval. The ball mill sample split was milled for 18 hrs.

The composition of the Lab Tech Essa bowl and puck principally consists of chromium, iron, manganese, and nickel with lesser amounts of additional metals (Clausen *et al.* 2012a). The Lab Tech Essa puck mills can accommodate up to 800 g of material and grind a sample to a 75µm particle size. The TM Engineering ring and puck mill can grind up to 500 g of material in a batch and has a similar bowl, ring, and puck composition as the Lab Tech Essa components. The Stoneware ball mill can accommodate up to six jars and is operated at 290 rpm. The mortar and pestle is composed of glazed, hard, chemical-resistant porcelain and can accommodate up to 70 g of material per batch. Manual disaggregation with the mortar and pestle involved multiple batches combined to form a single processed sample.

A separate split of the parent material was used for a puck mill experiment where 15 replicate samples were collected after intervals of 30, 60, 90, 120, and 300 seconds. A ball mill consisting of polyethylene lined steel cans was also evaluated over 8, 12, 16, and 20 hours using another split of the parent material. Fifteen 2-g replicate samples were collected after each time interval and then milling was resumed.

The procedure in USEPA Method 8330B (USEPA 2006a) was used to sub-sample all of the replicate samples. Briefly, each soil split was spread onto a sheet of aluminum foil to produce a layer 1 to 2 cm thick. Twenty increments were collected in a systematic random fashion (similar to the field sample collection approach) and combined to prepare a 2-g sub-sample for analysis (i.e., acid digestion and instrumental analysis by inductively coupled mass spectrometer (ICP-MS). Ultrapure reagent grade acids were used for the digestions as well as deionized water with a resistivity of less than 10MΩ cm. Fifteen 2-g replicate subsamples were collected, digested, and analyzed for each split (experimental condition).

The split samples shipped to APPL and Test America for milling were sub-sampled, digested, and analyzed at their respective laboratories. The CRREL experimental samples were sub-sampled at CRREL and then shipped to the USACE Environmental Laboratory (EL) in Vicksburg, MS for digestion and analysis. The digestion protocol followed at all three laboratories was Method 3050B (USEPA 1996a). Each of the three laboratories used an ICP-MS following Method 6020 (USEPA 2006b). At APPL Inc., an Agilent 7500cx ICP-MS equipped with an Agilent ASX-500 series sampler was used. The operating conditions were set as follows: RF power at torch = 1600 W, carrier gas flow = 1 L/min., auxiliary gas flow = 0.1 L/min., nebulizer flow = 0.4 mL/min., and pump flow rate = 0.1 rpm. The analysis of samples performed at Test America used an Agilent 7500 Series operated in the collision-cell mode, with helium as the collision gas. Samples at EL were analyzed with a Perkin Elmer Sciex ELAN

6000. The ICP-MS provided results for aluminum, antimony, barium, cadmium, cobalt, chromium, copper, iron, magnesium, manganese, nickel, phosphorous lead, strontium, vanadium, and zinc. In general, the quantification limit was approximately 1 mg/kg for all instruments, although most sample results were well above this value.

5.4 Results and Discussion

Our initial experiments involved a statistical comparison of metal results for the different milling apparatus as well as for unmilled parent material. This was followed with experiments with the puck mill and ball mill to determine the optimum-milling interval for soils with metallic residues. In the following discussion, Puck Mill #1 and #2 and Unground #1 and #2 represent two independent splits of the parent material. The Puck Mill #1 and #2 samples were ground at CRREL.

5.4.1 Comparison of Milling Equipment

To assess the precision of the analytical results the RSD was calculated for the 15 replicate samples prepared from pulverizing with the four different milling procedures/devices. I established a RSD target of 15% as the desired performance criterion (Clausen *et al.* 2014c, 2013a; Taylor et al. 2011).

The unmilled replicate sub-samples yielded RSDs varying from 17% to 257% for copper, lead, antimony, and zinc (Figure 5-1, Table 5-1). These high RSDs are a result of the unpredictability in the number of particulates present in any given sub-sample (Table 5-1). Recall that we employed sub-sampling with 20-increments to obtain a 2-g mass of material for digestion; where each individual increment is approximately 0.01-g. Therefore, the presence (or absence) of a single large particle in any one increment could have a significant impact on the overall metal concentration. However, as the likelihood of an individual increment containing a large anthropogenic metallic particle is usually relatively small, the masses and concentrations of the anthropogenic metals in the sub-samples are typically highly variable. For example, the range in copper concentrations for the Unground #1 varies from 299 to 10,900 mg/kg (see Appendix Table SI17). The RSD results for the mortar and pestle were also above our RSD target of 15%, varying from 28% to 55% indicating this device does not adequately pulverize the sample but, rather, disaggregates it. Pulverization involves reducing the particle size of the individual soil grains and any metallic residues whereas disaggregation involves physical separation of cohesive soil grains.

The puck mill, ring and puck mill, and ball mill each yielded results meeting our performance criteria for the anthropogenic metals of interest copper, lead, antimony, and zinc (Table 5-1). This indicates the reduction in particle size reduced the total sample error due to heterogeneity by increasing the number of metal particles in the sample and improving the probability of encountering a consistent number of particles in the sub-sample, consistent with Gy's theory (Gy 1992).

In contrast, the RSDs for the milled non-anthropogenic metals exhibited little difference; all of the RSDs were less than our target of 15% (see Appendix Figure SI18). We attribute this finding due to the natural geologic depositional processes, which resulted in a relatively homogenous distribution of metals in the loamy sand. The metals present in the soil are inherent to its chemical makeup. This suggests for the material we analyzed the necessity of milling is applicable to those situations where control of the compositional heterogeneity is necessary, such as metal deposition at small-arms ranges. Soils with a non-anthropogenic component or metals contamination from an aqueous source have minimal compositional heterogeneity and therefore milling is not necessary.

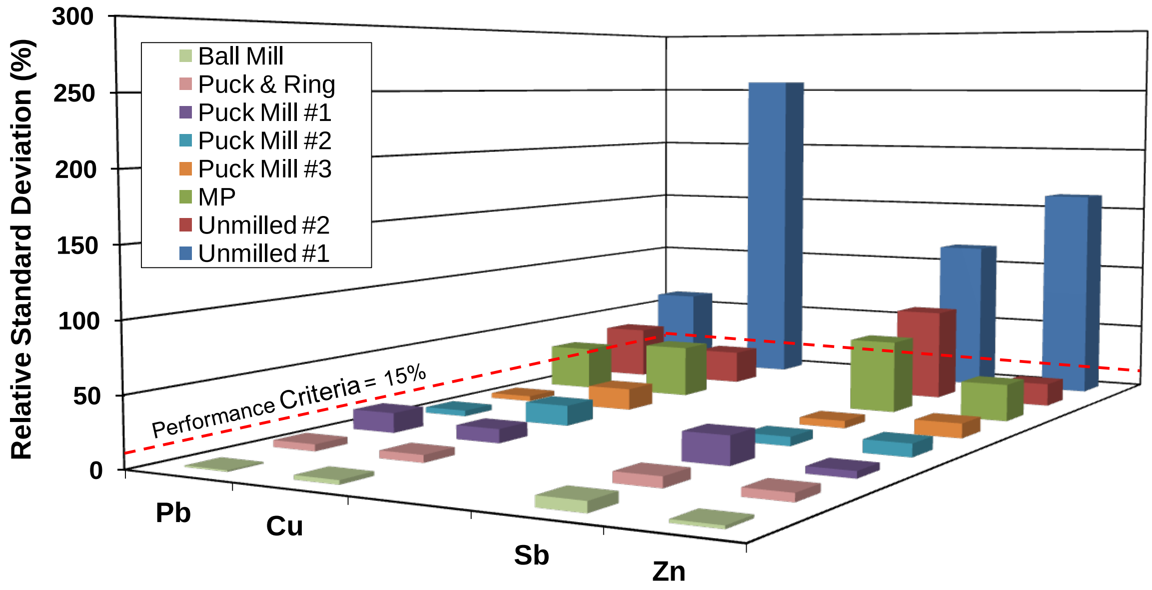


Figure 5-1. Comparison of percent relative standard deviation by milling device for antimony, copper, lead, and zinc.

As mentioned in Section 5.2, Clausen *et al.* (2012a, 2010b) and Felt *et al.* (2008) found differences in mean metal concentrations by milling device and analyte for several uncontaminated reference soils. The current study found the estimates of the mean, medians, and variances varied by metal and milling device as well for a contaminated soil (Table 5-1). The variances of the anthropogenic metals (copper, lead, antimony, and zinc) for the different milled populations differ from one another based on Levene's test of variances at the 95% confidence level.

Table 5-1. Summary of means and percent relative standard deviations by process for the metals studied.

		Al	Ba	Cd	Co	Cr	Cu	Fe	Mg	Mn	Ni	P	Pb	Sb	Sr	V	Zn	BET
Process	Lab	Mean (mg/kg)																(m2/g)
Unmilled #1	TA	3538	11.9	ND	4.37	4.47	1061	9173	NR	179	7.45	NR	2043	15.9	NA	6.19	83.7	NR
Unmilled #2	CRREL	4124	17.1	1.08	6.48	6.12	357	9307	2018	201	9.09	559	1600	14.2	12.6	NR	66.1	0.02
Mortar & Pestle	CRREL	4173	17.3	ND	6.31	6.26	372	13818	2025	184	8.37	596	1359	10.5	13.5	10.3	71.5	0.56
Puck Mill #1	APPL	2930	14.6	ND	4.25	108	279	7702	1520	142	6.46	335	3041	11.2	8.80	5.72	49.9	NR
Puck Mill #2	CRREL	5509	31.6	1.82	8.75	229	496	16676	2242	235	12.4	590	2760	21.8	20.6	15.2	81.0	1.36
Puck Mill #3	CRREL	6123	28.4	1.84	8.42	217	605	16828	2008	215	11.9	631	2673	23.4	23.4	15.2	73.3	NR
Puck & Ring Mill	TA	5707	40.6	ND	6.92	353	346	14200	2144	221	13.2	384	2349	5.20	NR	14.3	50.5	2.32
Ball Mill	TA	4983	31.9	ND	6.00	9.37	319	17240	2212	226	9.38	427	1687	3.47	NR	10.8	51.6	1.75
		RSD (%)																
Unmilled #1	TA	4	5	ND	5	5	257	4	NR	4	7	NR	61	116	NA	4	162	NA
Unmilled #2	CRREL	2	15	2	2	5	25	1	1	1	2	1	39	69	4	NR	17	NA
MP	CRREL	5	5	ND	4	4	39	4	4	3	3	5	32	55	11	4	28	NA
Puck Mill #1	APPL	5	6	ND	4	4	10	4	4	4	4	3	15	21	4	5	5	NA
Puck Mill #2	CRREL	1	4	1	1	2	15	4	2	2	1	2	4	7	2	2	10	NA
Puck Mill #3	CRREL	5	5	2	1	1	16	3	3	2	2	2	4	5	5	2	11	NA
Puck & Ring Mill	TA	6	6	ND	5	5	5	4	5	5	5	5	5	8	NR	5	6	NA
Ball Mill	TA	1	1	ND	1	1	3	1	1	1	1	2	1	8	NR	1	2	NA
NA = not applicable, ND = not detected, NR = no result																		

Lead is the most common metal found in target berms because lead makes up 90% or more of most small arms projectiles. In our study, lead median concentrations were generally the lowest for the unmilled samples and for the soil disaggregated with the mortar and pestle (Figure 5-2). Statistically significant differences between the variances of the two unground data sets were not detected by the Levene's test at the 95% level of confidence for lead or for the other metals copper, antimony, or zinc. However, the unmilled and mortar and pestle populations exhibited the highest variances with large interquartile ranges (IQRs) with notable outliers present. The differences in variances between the unmilled soil and mortar and pestle processed soil are significant by Levene's test at the 95% level of confidence. The high variances and outliers are the result of particulates in the sample and thus any estimate of the mean has a high degree of uncertainty.

The samples processed with the puck mills have higher median values and lower variances (Figure 5-2), which are statistically different from the unmilled samples at a 99% level of confidence using the non-parametric Kruskal-Wallis test for medians and Levene's test for variances. This finding is consistent with earlier milling studies with energetic materials (Walsh *et al.* 2002). The higher metal concentrations for the puck mills are consistent with their more effective pulverization capability as shown by the average surface area of a sample milled with the puck mill yielding mean values of 0.28 to 1.65 m^2/g depending upon the milling interval (Figure 5-3). Surface area was determined using the BET method with a Micromeritics TriStar 3000 (Norcross, GA) instrument. The surface area of the soil sample increased with increased milling time and appeared to stabilize after 90 sec (Figure 5-3). In comparison, the average surface area of the sample milled with the puck and ring mill for 300 sec was 2.32 m^2/g and 1.75 m^2/g with the ball mill after 20 hrs of milling. In contrast, the surface area of the soil disaggregated with the mortar and pestle for 300 sec was 0.56 m^2/g. The surface area for the unmilled soil was not measured but the largest particle size is 2-mm, which corresponds to a surface area of 0.01 m^2/g. (see Appendix Table SI19). We believe the increased particle surface area may be a second mechanism, which results in increased metal recovery, due to increased metal exposure to the digestion acids. The mean concentrations for a number of the non-anthropogenic metals (aluminum, barium, and strontium) are generally higher following milling than the unmilled sample. As mentioned earlier the primary mechanism for higher metals concentrations is the increased number and probability of metal particles in the sample following milling.

Cross-contamination from metallic milling equipment is another possible source of the increased metal concentrations. However, our earlier work with glass beads indicates the puck mill contributed little metal to the sample, with the exception of chromium and iron (Clausen *et al.* 2012a). Table 5-1 clearly shows the cross-contamination contribution of chromium and iron with the use of the puck mill and puck and ring mill in comparison with the unmilled sample.

The ball mill samples exhibited higher median values than the unmilled sample, which is statistically significant but the values are lower than the puck mills, which is consistent with the ball mill being a less efficient milling device. The variances for the ball mill population are much smaller than for the puck mills and statistically significant for lead (Figure 5-2) as well as copper, antimony, and zinc (see Appendix Figures SI20, SI21, and SI22).

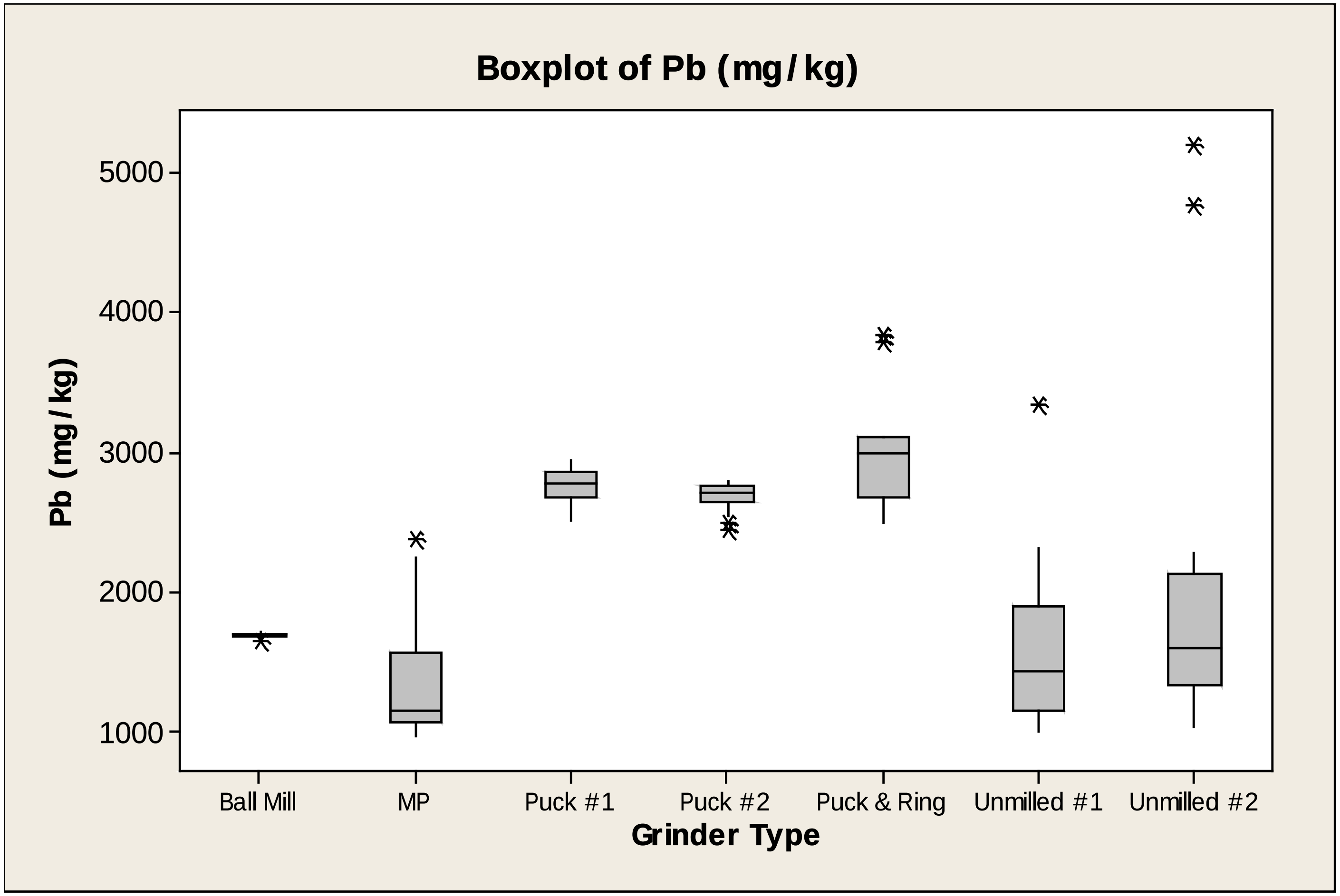


Figure 5-2. Box plot of lead by milled and unmilled splits.

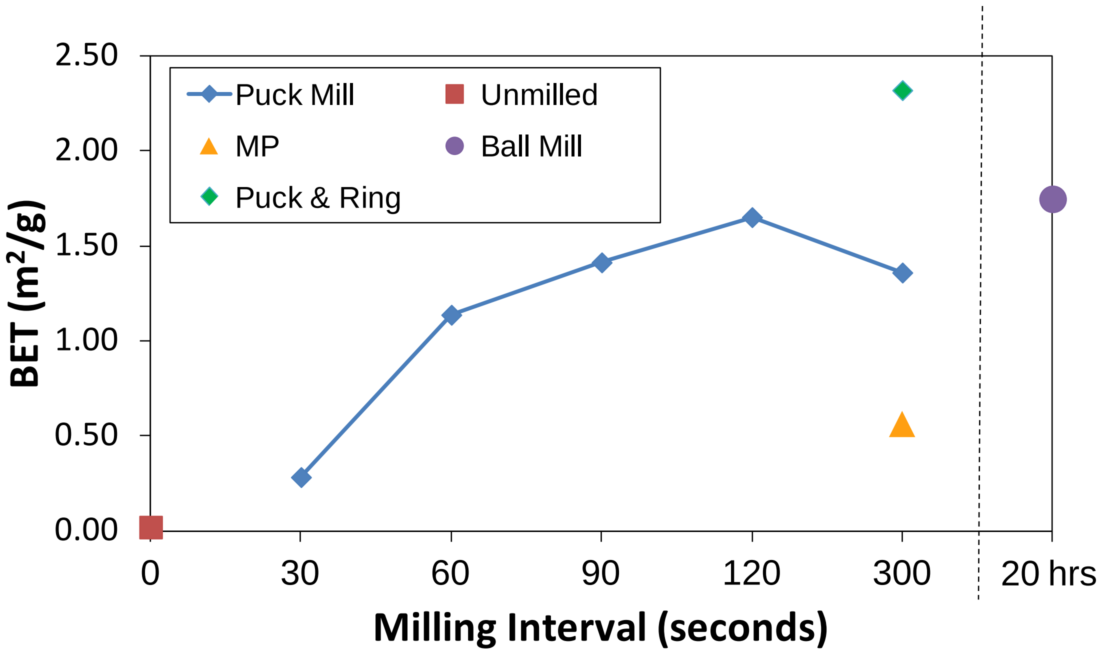


Figure 5-3. Average particle surface area by milling interval for the puck mill, puck and ring mill, unmilled sample, mortar and pestle, and ball mill.

The puck mills tended to produce better precision than the mortar and pestle but this was not consistent for all four anthropogenic metals. Generalizations about the mortar and pestle and the puck mill were problematic because the two puck mills did not consistently produce comparable results. When Levene's test was used to compare the puck mill variances at the 95% confidence level, Puck Mill #1 copper and zinc variances were larger than Puck Mill #2 copper and zinc variances, while the Puck and Ring Mill variances were smaller than the Puck Mill #2 variances. In the case of copper, the results are complicated by the observation from this study that when using the puck mill this metal tends to form small flat plates larger than the nominal particle size of the soil material due its malleability. The concentration of antimony and zinc are much lower in the soil than lead making differences between the milling devices less evident. However, in terms of reducing sub-sampling variability and improving precision, the overall performance for the various milling equipment seemed to be as follows: ball mill > ring and puck mill > puck mill > mortar pestle > unmilled sample..

In contrast, for most of the non-anthropogenic metals little difference in the means/medians is evident by milling device (see Appendix Figure SI23). However, in the case of chromium the puck mill samples have statistically significantly higher means than the unmilled soil or mortar and pestle and ball mill processed soils (Table 5-1). This is likely due to chromium eroding from the puck mill and ring and puck mill equipment surfaces and cross-contaminating the sample. Iron and vanadium concentrations also appear to nearly double as well as compared to the unmilled samples. These findings are not surprising as chromium, iron, and vanadium are major constituents used in the manufacturer of the puck mill components and steel is the material type for the ball mill. Consequently, if chromium is a target analyte the ball mill is the preferred option or possibly a puck mill with non-metallic surfaces. If iron or vanadium are target analytes a non-metallic milling device is recommended. Puck mills are available with bowls, rings, and pucks constructed from agate (SiO_2) from a variety of manufacturers. Ball mills with polyurethane coated alumina jars are also available.

The puck mills and ball mill were identified as the best tools for reducing the total sample error and uncertainty, however, a question remains about the optimum-milling interval to accurately represent the average metal concentration in the soil samples. My expectation is that at some interval of time for milling, significant particle size reduction plateaus in terms of its effect on sub-sampling precision.

5.4.2 Optimum Milling Interval

The results presented in Section 5.4.1 indicate the puck mills and ball mill are appropriate milling devices for reducing the total sampling error to acceptable levels. A study with the puck mill evaluated milling intervals of 30, 60, 90, 120, and 360 sec. The milling intervals evaluated with the ball mill were 8, 12, 16, and 20 hrs.

The same 200-increment parent soil previously discussed was used for all tests. The approach for each piece of milling equipment was as follows: mill the soil for the shortest time interval, collect 15 laboratory replicate sub-samples, and then mill the remaining soil for the next time interval, repeating the procedure until the last interval. For all samples, the digestion and analysis procedures were identical to those previously discussed.

5.4.2.1 Puck Mill

The RSD was calculated for the set of 15 replicate samples for each time interval to assess precision as a function of milling time for the anthropogenic metals (Figure 5-4). The time interval of 0 sec represents the unmilled control sample. The results suggest a milling interval of 60 sec might be sufficient for the anthropogenic metals; however, copper had a RSD of 31% at 90 seconds, exceeding our target for inter-sample precision of 15%. It seems that 120 sec of milling is adequate to achieve RSDs of less than 15% for lead and zinc, but not copper and antimony. At 300 sec the RSD was less than 15% for all analytes accept copper, which had a RSD of 20%. However, the copper RSD at 300-sec of milling is still three times smaller than the copper RSD for the unmilled split. As discussed earlier, the highly malleable nature of copper may be inhibiting particle size reduction. It is possible a milling interval longer than 300 sec would reduce the total sample error to less than 15% but we did not explore this further.

As shown in the Appendix Figure SI24 all of the non-anthropogenic metals had a RSD of less than 15% for the unmilled populations (time zero). Milling did not appreciable improve sample precision for the non-anthropogenic metals. The decreased precision at a milling interval at 120 sec for most non-anthropogenic metals suggests a systematic error introduced during this processing event.

Consequently, milling would not be necessary if the focus of the sampling effort was on the background level of metals at a particular site. A study on the aqueous application of pesticide containing arsenic also found milling was unnecessary (ITRC 2012). Although further studies are warranted, it seems likely that metal contaminants introduced in aqueous form or background studies of native materials, with the exception of ore bodies, are sufficiently homogenous making milling unnecessary.

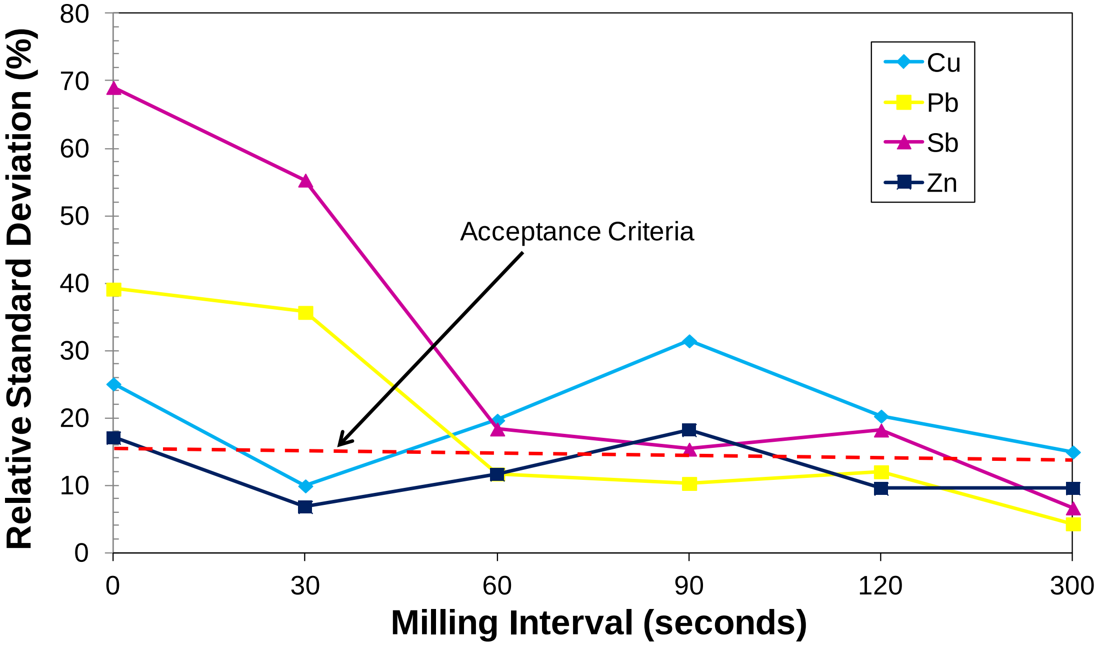


Figure 5-4. Percent relative standard deviations by milling interval for copper, lead, antimony, and zinc.

One of the unanswered questions of using a puck mill is how much the process of milling changes the metals concentration results. It is apparent milling with a puck mill composed of metallic components results in an increase in some metal concentrations, although there is enough variability that clear positive trends are not apparent for all metals (Table 5-2). For the anthropogenic metals lead and antimony, the means appear to increase with milling time and subsequently plateau by 300 sec (Figure 5-5). Overall, the variance for lead and antimony decreased as milling time increased, suggesting particle size reduction of bullet fragments resulted in a corresponding decrease in total sampling error. Antimony is added as a hardening agent in the lead alloy mixture for small-arms ammunition. In contrast, there is a lack of clear relationship between milling time and copper and zinc concentrations. I hypothesize this lack of relationship is owing to the malleable nature of copper and the resulting difficulty milling it. Zinc is a component of the copper alloy used for the casing material of small-arms ammunition. However, milling did improve the data quality of the copper and zinc results. The distributions for both analytes are nearly normal for the 120 and 300-sec intervals and the variances of the replicates analyzed for the two milling intervals are not significantly different at a 95% level of confidence by Levene's test.

As expected, metals that are major constituents of the puck mill surfaces (e.g., chromium and iron) as well as minor constituents (nickel and vanadium) tended to increase as the milling time increased (see Appendix Figures SI25, 26, and 27). However, the chromium and iron levels appear to stabilize after 120 sec of milling (Table 5-2, see Appendix Figures SI25). Manganese, which is a major constituent of the chrome steel puck mill, did not exhibit increased concentrations after milling. Barium, which is not listed as a constituent off the puck mill, increased slightly with milling time.

Table 5-2. Mean metal concentrations by puck mill milling interval.

Time	Mean (mg/kg)																BET	Particle Size
(seconds)	Al	Ba	Cd	Co	Cr	Cu	Fe	Mg	Mn	Ni	P	Pb	Sb	Sr	V	Zn	(m^2/g)	(cm)
0	4124	17.1	1.08	6.48	6.12	357	9307	2018	201	9.09	559	1600	14.2	12.6	10.0	66.1	0.0200	0.0100
30	4754	19.6	0.813	6.71	74.7	357	14,344	1848	183	9.09	739	2348	24.2	18.4	12.0	58.6	0.2832	0.0008
60	5378	29.3	0.928	8.04	135	430	15,597	2187	224	10.2	576	2187	18.6	18.7	14.1	75.1	1.1376	0.0002
90	6518	24.7	1.05	7.69	162	646	16,933	1824	196	11.0	570	2818	24.8	19.9	15.0	66.7	1.4135	0.0002
120	7156	26.8	0.889	8.01	233	538	16,653	1873	200	12.5	629	2562	21.6	27.2	15.6	60.8	1.6515	0.0001
300	5509	31.6	1.82	8.75	229	496	16,676	2242	235	12.4	590	2760	21.8	20.6	15.2	81.0	1.3575	0.0002
Change (%)	34	85	69	35	3641	39	79	11	17	37	5	73	53	64	52	23	6688	

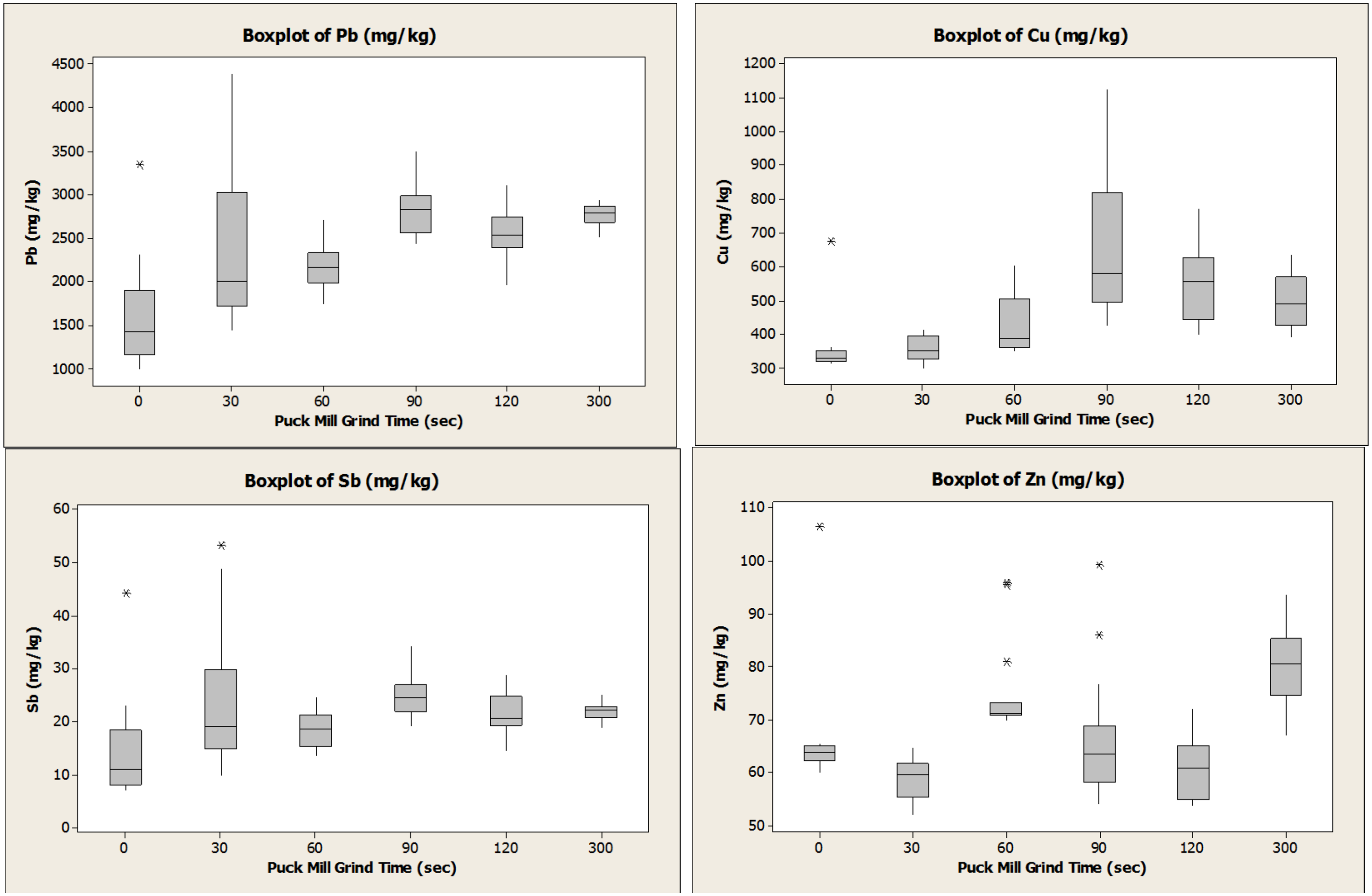


Figure 5-5. Box plots for lead, copper, antimony and zinc concentrations by puck mill milling interval.

5.4.2.2 Ball mill

A summary of the RSDs by milling interval with the ball mill for the metals are shown in Table 5.3 and suggests 8 hrs is sufficient in most cases using a target criterion of 15%. Antimony was the only analyte exceeding our target with a value of 23% at 8 hrs of milling. There is minimal improvement in the RSDs with increased milling time, with all analytes meeting our target at 12 hrs. As a comparison the RSD for the unmilled samples are also shown (Table 5-3) and indicate the ball mill is successful in addressing the heterogeneity present in the sample.

Table 5-3. Relative standard deviations by ball mill milling times.

	Relative Standard Deviation (%)											
	Al	Ba	Co	Cr	Cu	Fe	Mn	Ni	Pb	Sb	V	Zn
Unground	5	NA	5	5	257	4	4	7	61	116	NA	162
Ball Mill - 8hr	4	5	3	9	4	8	2	4	13	23	8	8
Ball Mill - 12hr	5	10	3	11	8	4	4	3	6	7	8	7
Ball Mill - 16hr	4	6	1	6	6	2	2	1	6	14	5	3
Ball Mill - 20hr	4	6	1	6	6	NA	2	NA	6	14	5	3

Similar to the observations with the puck mill, an increase in the milling interval with the ball mill resulted in an increase in metal concentrations (Table 5-4). As the ball mill consists of polyurethane coated steel jars and ceramic balls, cross-contamination of metals during milling is expected to be minimal. Consequently, the observed increase in anthropogenic metal concentrations with milling time is hypothesized to be the result of particle size reduction and an increased number of particles. Particle size reduction results in an increased surface area and during digestion yields higher metal recovery. The increased number of particles results in higher probability that a consistent number of particles will be present in a sub-sample. Our earlier paper, Clausen *et al.* (2014c) indicated that an unmilled sample may have as few as single metallic particle whereas a milled sample has in excess of million particles. It should be noted in this experiment the soil sample consisted of an unsieved sample, whereas all previous experiments utilized the sieved less than 2-mm portion. Milling for 20 hrs with the ball mill is nearly comparable in terms of particle size reduction as milling for 300 sec with the puck mill (Figure 5-3). The lack of significant non-anthropogenic metal concentration increase suggests improved metal digestion efficiency due to increased particle surface area is not apparent.

Table 5-4. Mean metal concentrations by ball mill milling interval.

Time	Concentration (mg/kg)											
(hours)	Al	Ba	Co	Cr	Cu	Fe	Mn	Ni	Pb	Sb	V	Zn
Unmilled	3538	11.9	4.37	4.47	1061	9173	179	7.45	2043	15.9	NA	83.7
8	4275	20.1	4.74	6.47	1013	11553	199	8.15	4787	34.9	7.55	69.2
12	4588	22.4	4.92	7.68	1039	13333	194	8.78	5800	45.9	8.64	80.0
16	4761	26.4	4.80	9.32	1183	13847	202	8.63	6380	45.1	7.75	91.3
12	5058	29.2	4.84	10.4	1315	NA	209	NA	7153	55.2	8.48	93.4

NA = not analyzed

Concentrations of antimony and zinc appear stabilized after 16 hrs of milling (Figure 5-6). The magnitude of lead and copper concentration increases appear to decline with milling time although it is has not stabilized after 20 hrs of milling. This suggests larger particles of lead and copper continue to persist in the sample.

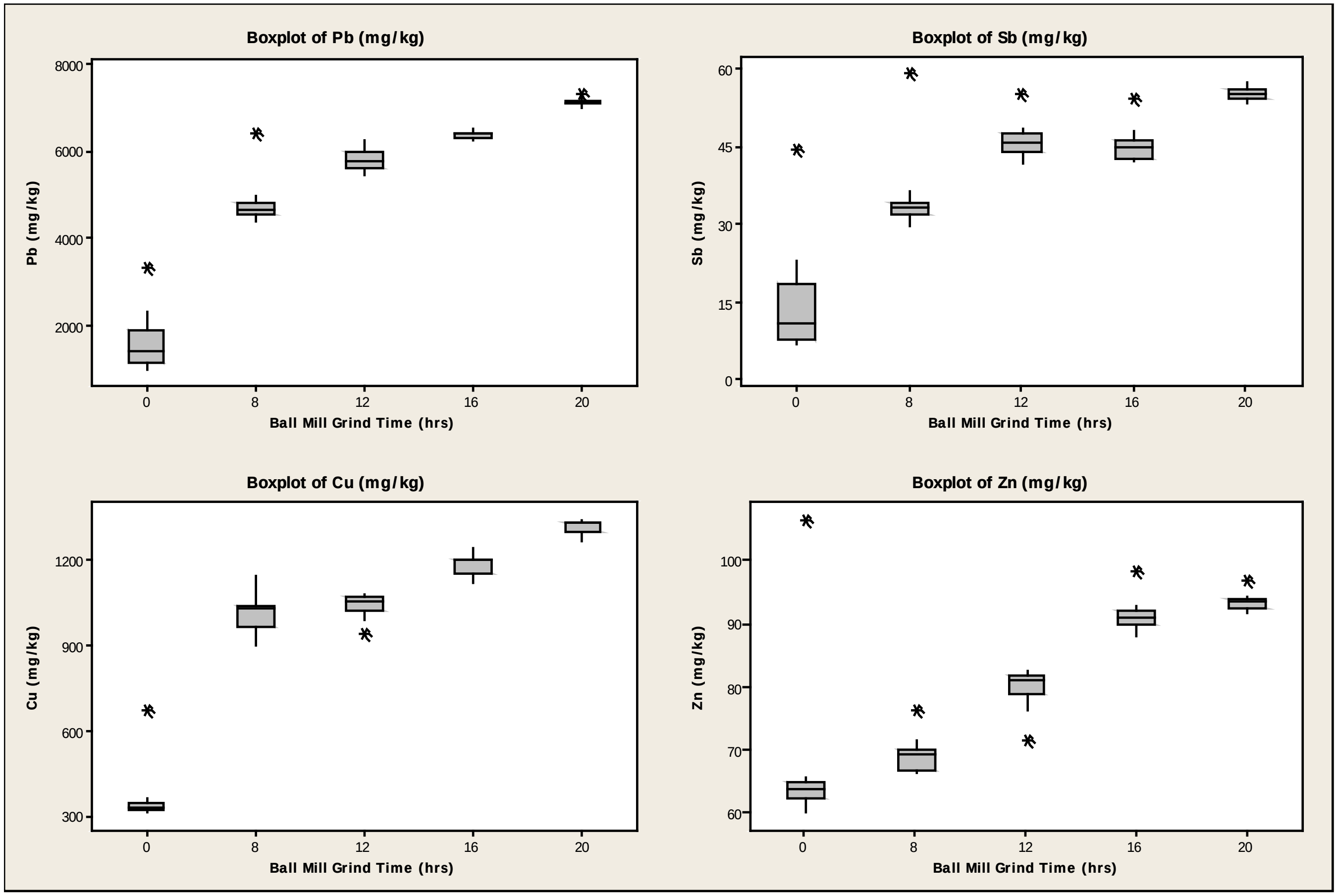


Figure 5-6. Box plots for lead, copper, antimony, and zinc concentrations by ball mill milling interval.

5.5 Conclusions

An evaluation of the necessity of milling for soil from a small-arms range containing a heterogeneous distribution of metal particulates indicates milling is the most optimal sample processing method for obtaining reliable estimates of the mean. The improved estimates of the mean are due primarily to a reduction in variability resulting in increased precision as a result of increasing the number of particulates and the concomitant improved probability of encountering an increased number of particulates in the sample. Improved metal recovery as a result of a reduction in particle size and a corresponding increase in surface area of metal during digestion is a secondary minor mechanism. These observations are consistent with earlier work focused on improved field sampling and sample preparation methodologies for energetic constituents introduced into the environment in particulate form, which led to modifications in the USEPA extraction Method 8330B for energetics. Similarly, modifications to USPA Method 3050B for the digestion of metals are underway to include milling as part of the sample preparation process when metallic residues are present. These modifications will be included in update VI of SW-846.

The disaggregation of soil containing metallic residues using a mortar and pestle, which is an option in Method 3050B, does not adequately address heterogeneity of the sample. A puck mill or ball mill is necessary to reduce the compositional heterogeneity to acceptable levels. A milling interval of 300 sec with a puck mill, puck and ring mill, or at least 16 hrs with a ball mill appears adequate for most small-arms range contaminated soils and contaminants of interest. In the case of copper, a longer milling interval than 300 sec may be beneficial to improve sample precision.

To avoid cross contamination, the grinding surfaces of milling devices should not contain significant quantities of the metals of interest. In particular, if the contaminants of concern include chromium or iron the samples should not be ground in a chrome-steel puck mill. Alternative non-metallic milling devices include a puck mill with agate grinding surfaces or a ball mill with polyurethane coated steel jars and ceramic balls.

The milling methods discussed have applicability beyond small-arms ranges and would be appropriate in any situation where the form of the metal contaminant is introduced into the environment in solid form, e.g. military impact areas; mine tailing waste; coal gasification plants; sewage sludge; metal refining, production, and finishing, etc. It seems likely that the same methods discussed in this paper would also be appropriate for other contaminants introduced into the environment in particulate form such as polyaromatic hydrocarbons at Skeet Ranges, perchlorate from fireworks, or polychlorinated biphenyls in sediments.

In contrast, milling appears unnecessary for evaluating the background metal levels of a soil or a soil contaminated with release of a metal into the environment in aqueous form. Although, improvements in precision are evident for aqueous released metal contaminants the enhancement is only marginal.

CHAPTER 6

SUMMARY OF DISSERTATION FINDINGS

6.1 Conclusions

The research hypothesis tested was that metallic residues (solid particulates) deposited on the berm face of military small-arms ranges is sufficiently heterogeneous that conventional grab sampling is appropriate. If the aforementioned hypothesis was not supported, then the goal of the research was to determine what alternative field-laboratory sample-preparation procedures are required to yield representative and reproducible estimates of the population mean?

The results in Chapter 2 showed that distribution of metallic residues as result of small-arms training yields a heterogeneous distribution. This type of non-Gaussian contaminant distribution was only adequately addressed by collecting a large number of conventional grab samples. However, environmental studies using grab sampling are predominantly driven by cost rather than data quality considerations. Consequently, in most environmental investigations, the limited number of samples results in poor population precision when estimating the mean or calculating an UCL of the mean. The large uncertainties of the mean associated with grab samples can only be overcome by either the collection of a large numbers of grab samples (> 100) or a different sampling approach (e.g., composite, ISM).

Chapter 3 compared and contrasted the ISM method with conventional grab sampling to determine whether this approach yielded improved estimates of the population mean and sample reproducibility. The results indicated less variance and skewness of the anthropogenic sample population and improved reproducibility and precision with ISM as compared to conventional grab sampling. The other major observation was that a higher quality of data is possible using ISM with fewer samples than what normally is collected with the conventional grab sampling method.

Chapter 4 evaluated two sampling processing limitations with the application of ISM for sites where trace metals are heterogeneously loaded to soils from military training with small-arms. The two most common methods changes include using field splitting to reduce the sample mass shipped to analytical laboratories and preparation of larger digestion aliquots to eliminate the need for milling. A controlled study found that a reduction of sample mass for ISM samples using splitting techniques or devices is not recommended when contaminants associated with particulates. Even when incremental samples are collected in the field, compositing the increments addresses only the distributional heterogeneity, not the compositional heterogeneity. Second, additional laboratory sample processing, namely milling, is required to account for compositional heterogeneity. Analysis of sub-sample masses as large as

10 g for unmilled samples did not adequately address the compositional heterogeneity. Unless field sample collection and laboratory sample processing procedures account for the distribution and compositional heterogeneity, no single result will adequately reflect the true concentration at the site. The variability owing to distributional and compositional heterogeneity will be apparent only when a large number of replicates are analyzed.

Chapter 5 assessed whether particle size reduction through milling is necessary to increase the number of particulates in the sample to improve sample homogeneity, to improve sample precision. Multi-increment field sampling methods were insufficient to reduce total SE for soils containing metallic residues. The considerable improvement in sample uncertainty offset the slight increase in metal concentrations due to milling. An assessment of different milling devices found that the ring mill and roller mill devices yielded acceptable results based on pre-selected quantitative error acceptance criteria. Unmilled soil samples consistently yielded a non-normal sample population, with a high degree of variance, and a number of outliers. In addition, the population median value for unmilled samples had a statistically significant low bias as compared to the milled samples using the puck and roller mills at the 99% level. Standard deviations of the unmilled results were five times larger than those of the milled samples using the puck and roller mills. Mortar and pestle results were similar to the unmilled sample results suggesting these do not improve sample homogeneity. In addition to evaluating different types of milling equipment, the impact of milling interval on meeting quantitative performance objectives was assessed. A milling interval of 5-min with the ring mill was sufficient in most situations to adequately reduce the metallic particles to a consistent size, yielding reproducible sample results. Similarly, acceptable results were obtained with the Roller mill when the milling interval exceeded 18 hrs.

Although not evaluated in this document, a cost-benefit analysis performed by Clausen *et al.* (2013) found a potential total project cost savings of 5 to 50% utilizing the approach outlined in this dissertation. Field sampling using ISM is \$20–\$40 higher per sample than conventional grab sampling, which is largely a function of the greater amount of time needed to collect the ISM sample (i.e., the collection of multiple increments). Similarly, laboratory preparation costs run \$40–\$60 higher with ISM due to the additional sample processing steps. Therefore, on a per sample basis, the cost of ISM is approximately 55% to 65% higher than conventional grab sampling. However, the total project cost with ISM is lower due to fewer samples collected. Three replicate ISM samples would be sufficient for a typical small-arms range versus 7 to 15 conventional grab samples for the same DU. The cost savings become greater as the number of grab samples for comparison increases. The reduction of total costs with ISM is primarily a function of the fewer number of samples collected, processed, and analyzed. If this methodology were widely implemented across the DoD the potential impact on DoD annual expenditures, of approximately \$2 billion dollars (GAO 2003), for site characterization and investigation activities could be substantial. To facilitate introduction of these techniques to the DoD and their contractors I am involved with making modifications to USEPA SW-846 Method 3050B. The Method 3050B update is anticipated to be issued in 2015 as Method 3050C as part of update VI for SW-846 *Test Methods for Evaluating Solid Waste, Physical/Chemical Methods, office of Solid Waste and Emergency Response.*

6.2 Future Research

One of the issues raised about the implementation of ISM is the concern related to cross contamination from the metallic milling apparatus used to process the soil. It is not clear if the resulting elevation of metal concentrations is a result of cross-contamination, increased digestion efficiency as result of an increased surface area from milling, or a combination of both. The only means to address this issue is by conducting experiments on soil using the same milling equipment but with metallic and non-metallic components (e.g., a Puck mill puck and bowl, Ball mill balls and roller cans).

A second issue raised was the change of metal soil concentrations and subsequent bioavailability test results due to the milling of the soil. The Army is concerned that milling of a soil sample yields a result not reflective of the true conditions in the field. Milling increases the number of particles, as well as surface area of the soil particle, and exposes the inner portion of the soil grain to the digestion process. Under conventional sample preparation procedures, the inner portion of the soil grain is not digested. Finally, the reduced particle size from milling and greater surface area results in improved digestion efficiencies as compared to the conventional sample preparation. All of these factors potentially contribute to measurements yielding larger soil concentrations as compared to the conventional approaches. The question is whether this concentration increase occurs and, if it is statistically significant. A secondary question is which approach (the conventional method or ISM) yields a more representative assessment of metals bioavailability.

REFERENCES

Alaska. (2009). *Draft Guidance on MULTI INCREMENT Soil Sampling*. Alaska Department of Environmental Conservation, Division of Spill Preventions and Response, Contaminated Sites Program.

Allen, T. and A.A. Khan. (1970). Critical evaluation of powder sampling procedures. Chemical Engineering. **238**:CE108-CE112.

Ahrens, L.H. (1954). The lognormal distribution of the elements. Geochimica et Cosmochimica Acta. **5**:49-73.

Atchison, J. (1955). On the distribution of a positive random variable having a discrete probability mass at the origin. Journal of the American Statistical Association. **50**:901–908.

Ampleman, G., S. Thiboutot, J. Lewis, A. Marois, S. Jean, A. Gagnon, M. Brouchard, R. Martel, R. Lefebvre, C. Gauthier, J.M. Ballard, T.A. Ranney, and T.F. Jenkins. (2003a). *Evaluation of the Impacts of Live Fire Training at CFB Shilo.* TR-2003-066. Defence Research Establishment – Valcartier. Valcartier, Quebec, Canada.

Ampleman, G., S. Thiboutot, J. Lewis, A. Marois, S. Jean, A. Gagnon, M. Brouchard, T. Jenkins, A. Hewitt, J.C. Pennington, and T.A. Ranney. (2003b). *Evaluation of the Contamination by Explosives in Soils, Biomass and Surface Water at Cold Lake Air Weapons Range (CLAWR), Alberta, Phase 1 Report.* DRDC-Valcartier-TR-2003-208-Annex. Defence Research Establishment – Valcartier. Valcartier, Quebec, Canada.

ASTM. (2009). *Standard Test Methods for Particle-Size Distribution (Gradation) of Soils Using Sieve Analysis.* ASTM D6913-04. American Society for Testing and Materials. West Conshohocken, PA.

Barbizzi, S., P. Zorzi, M. Belli, A. Pati, U. Sansone, L. Stellato, M. Barbina, A. Deluisa, S. Menegon, V. Coletti. (2004). Characterization of a reference site for quantifying uncertainties related to soil sampling. Environmental Pollution. **127**:131–135

Barth, D.S. and B.J. Mason. (1984). *Soil Sampling Quality Assurance User's Guide.* EPA/600/4-84-043. U.S. Environmental Protection Agency. Las Vegas, NV.

Barth, D.S., B.J. Mason, T.H. Starks, and K.W. Brown. (1989). *Soil Sampling Quality Assurance User's Guide.* EPA/600-8-69/046. U.S. Environmental Protection Agency. Las Vegas, NV.

Beckett, P.H.T. and R. Webster. (1971). Soil variability; a review. Soils and Fertilizers. **34**(1): 1-15.

Bicking, C.A. (1967). The sampling of bulk materials. Materials Research and Standards. **March**:95-116.

Blaut, J. M. (1959). Microgeographic sampling a quantitative approach to regional agricultural. Economic Geography. **35**(1):79-88

Bonczek, R.R., J. Davidson, A.F. Diefendorf, F.G. Dolislager, N.E. Korte, C.J. Marshall, G.R. Miller, and W.L. Richards. (1996). *Background Concentrations and Human Health Risk-Based Screening Criteria for Metals in Soil at the Paducah Gaseous Diffusion Plant, Paducah, Kentucky.* DOE/OR/07-1417&D2. U.S. Department of Energy. Paducah, KY.

Boswell, M.T., S.D. Gore, L. Lovison, and G.P. Patil. (1996). Annotated bibliography of composite sampling Part A: 1936-92. Environmental and Ecological Statistics. **3**:1-50.

Boudreault, J., J. Dube, M. Sona, and E. Hardy. (2012). Analysis of procedures for sampling contaminated soil using Gy's sampling theory and practice. Science of the Total Environment. **425**:199-207.

Bowley, A.L. (1906). Address to the Economic Science and Statistics Section, British Association for the Advancement of Science, Journal of the Royal Statistical Society. **69**:540-557.

Brumbaugh, M.A. (1954). Principles of sampling in the chemical field. Industrial Quality Control. **19**(4):6-14.

Cameron, D.R., M. Nyborg, J.A. Toogood and D.H. Laverty. (1971). Accuracy of field sampling for soil tests. Canadian Journal of Soil Science. **51**: 65-175.

Campbell, J.B. (1979). Spatial variability of soils. Annals of the Association of American Geographers. **69**(4):544-556.

Cao, X., L.Q. Ma, M. Chen, D.W. Hardison, and W.G. Harris. (2003). Lead transformations and distribution in the soils of shooting ranges in Florida, USA. The Science of the Total Environment. **307**:179-189.

Carrasco, P., P. Carrasco, and E. Jara. (2004). The economic impact of correct sampling and analysis practices in the copper mining industry. Chemometrics and Intelligent Laboratory Systems. **74**:209-213.

Chen, L. (1995). Testing the mean of skewed distributions. Journal of the American Statistical Association. **90**:767-772.

Clarkson, T. (1894). The sampling of materials. The Journal of the Society of Chemical Industry. **13**:214-217.

Clausen, J.L., J. Robb, D. Curry, and N. Korte. (2004). A case study of contaminants on military ranges: Camp Edwards, Massachusetts, USA. Environmental Pollution. **129**:13–21.

Clausen, J., S. Taylor, S. Larson, A. Bednar, M. Ketterer, C. Griggs, D. Lambert, A. Hewitt, C. Ramsey, S. Bigl, R. Bailey, and N. Perron. (2007). *Fate and Transport of Tungsten at Camp Edwards Small-arms Ranges.* ERDC-CRREL TR-07-05, US Army Corps of Engineers, Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH.

Clausen, J. and N. Korte. (2009a). The distribution of metals in soils and pore water at three U.S. military training facilities. Soil and Sediment Contamination Journal: An International Journal. **18**(5):546-563.

Clausen, J. L. and N. Korte. (2009b). Environmental fate of tungsten from military use. The Science of the Total Environment. **407**(8):2887-2893.

Clausen, J. L., A. Bednar, D. Lambert, R. Bailey, M. Kuhlbrush, S. Taylor, and S. Bigl. (2010a). *Phase II Tungsten Fate-and-Transport Study for Camp Edwards.* ERDC-CRREL TR-10-3. US Army Corps of Engineers, Environmental Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH.

Clausen, J.L., J. Kaste, M. Ketterer, and N. Korte. (2010b). Sample preparation and digestion considerations for determining metal deposition at small-arms ranges. International Journal of Environmental Analytical Chemistry. **90**(12):903-921. http://www.tandfonline.com/doi/abs/10.1080/03067310903353495

Clausen, J.L. (2011). Energetic Residue Observations for Operational Ranges: in *Environmental Chemistry of Explosives and Propellant Compounds in Soils and Marine Systems: Distributed Source Characterization and Remedial Technologies.* Ed. M. Chappelle, C. Price, and R.D. George. American Chemical Society Symposium Series 1069, Washington, DC. pp.107-136.

Clausen, J.L., T. Georgian, J. Richardson, A. Bednar, N. Perron, L. Penfold, D. Anderson, G. Gooch, T. Hall, and E. Butterfield. (2012a). *Evaluation of Sampling and Sample Preparation Modifications for Soil Containing Metal Residues.* ERDC TR–12-1. US Army Corps of Engineers, Environmental Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH. http://acwc.sdp.sirsi.net/client/search/asset:asset?t:ac=$N/1006020

Clausen, J.L., J. Richardson, N. Korte, G. Gooch, T. Hall, N. Perron, E. Butterfield, M.E. Walsh, and S. Taylor. (2012b). *Metal Residue Deposition from Military Pyrotechnic Devices and Field Sampling Guidance.* ADA562327. Prepared for U.S. Army Environmental Command, Fort Sam Houston, TX by US Army Corps of Engineers, Environmental Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH. http://handle.dtic.mil/100.2/ADA562327

Clausen, J.L., T. Georgian, A. Bednar, N. Perron, A. Bray, P. Tuminello, G. Gooch, N. Mulherin, A. Gelvin, M. Beede, S. Saari, W. Jones, and S. Tazik. (2013). *Demonstration of Incremental Sampling Methodology for Soil Containing Metallic Residues.* ERDC/CRREL TR-13-9. US Army Corps of Engineers, Environmental Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH. http://acwc.sdp.sirsi.net/client/search/asset/1030080

Clausen J.L., T. Georgian, K. Gardner, and T. Douglas. (2014a in review). Poor precision of conventional grab samples for characterization of metal contamination at small-arms ranges. Risk Analysis.

Clausen J.L., T. Georgian, K. Gardner, and T. Douglas. (2014b submitted). Conventional grab and incremental sample error for soils containing metallic residues. Soil and Sediment Contamination.

Clausen J.L., T. Georgian, K. Gardner, and T. Douglas. (2014c submitted). The necessity for sample processing of soils containing metallic residues when using the Incremental Sampling Methodology (ISM). Chemosphere.

Cline, M.G. (1944). Principles of soil sampling. Soil Science. **58**(4):275-288.

Cochran, W.G. (1977). Sampling Techniques. 3rd Ed. John Wiley and Sons. New York, NY.

Cohen A.C., Jr. *(*1959)*.* Simplified estimators for the normal distribution when samples are singly censored or truncated*.* Technometrics. **3***:* 217–237.

Craig, J. R., D. Edwards, J.D. Rimstidt, P.F. Scanlon, T.K. Collins, O. Schabenberger, and J.B. Birch. (2002). Lead distribution on a public shotgun range. Environmental Geology. **41**:873-882.

Crepin, J. and R.L. Johnson. (1993). Soil Sampling for Environmental Assessment. In *Soil Sampling and Methods of Analysis*. Ed. M.R. Carter. Lewis Publishers. Boca Raton, FL.

CSSS. (2007). Soil Sampling and Methods of Analysis. 2nd Ed. M.R. Carter and E.G. Gregorich Ed. Canadian Society of Soil Science. CRC Press. London.

Dorfman, R. (1943). The detection of defective members of a large population. Annals of Mathematical Statistics. **14**(4):436-440.

Drees L.R. and L.P. Wilding. (1973). Elemental variability within a sampling unit. Soil Science Society of America Proceedings. **37**:82-87.

DSB. (2003). *Report of the Defense Science Board Task Force on Unexploded Ordnance.* Office of the Under Secretary of Defense for Acquisition and Technology. Defense Science Board. Washington, DC.

Duggan, J. and A. Dhawan. (2007). Speciation and vertical distribution of lead and lead shot in soil at a recreational firing range. Soil and Sediment Contamination. **16**:351-269.

Duncan, A.J. (1962) Bulk sampling: Problems and lines of attack. Technometrics, **4**:319-344.

Elder, R.S., W.O. Thompson, and R.H. Myers. (1980) Properties of composite sampling procedures. Technometrics. **22**:179-186.

El Khoukhi, T., R. C. El Moursli, A. Chouak, Z. Moutua, N. Lferde, A. Senhau, A. Gaudry, S. Ayrault, and M. Chaki. (2005). Comparative use of tree leaves, needles, tree barks and lichens for air pollution biomonitoring. Physica Scripta. 118:211.

Felt, D. R., A. J. Bednar, and T. Georgian. (2008). The effects of grinding methods on metals concentrations in soil. Talanta. **77**:380-387.

Florida. (2013). Draft Incremental Sampling Methodology (ISM) Guidance. Florida Department of Environmental Protection, Bureau of Waste Cleanup. Tallahassee, FL.

GAO. (2003). *Environmental Compliance: Better DOD Guidance Needed to Ensure That The Most Important Activities Are Funded.* GAO-03-639. Government Accounting Office. Washington, DC.

Garrett, R.G. and R. Sinding-Larsen. (1984). Optimal composite sample size selection, applications in geochemistry and remote sensing. Journal of Geochemical Exploration. **21**:421-35.

Gerlach, R.W. , D.E. Cobb, G.A. Robb, and J.M. Nocerino (2002). Gy sampling theory in environmental studies. 1. Assessing soil splitting protocols. Journal of Chemometrics. **16**:321-328.

Gerlach, R.W. and J.M. Nocerino (2003). *Guidance for Obtaining Representative Laboratory Analytical Subsamples from Particulate Laboratory Samples.* U.S. Environmental Protection Agency. EPA 600/R-03/027. Washington, DC.

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

Gilbert, RO. and P.G. Doctor. (1985). Determining the number and size of soil aliquots for assessing particulate contaminant concentrations. Journal of Environmental Quality. **14**(2) 286-292.

Gomez, K.A. and A.A. Gomez. (1984). Statistical Procedures for Agricultural Research. 2nd Edition. John Wiley and Sons, Inc. Singapore.

Gore, S.D. and G.P. Patil. (1994). Identifying extremely large values using composite sample data. Environmental and Ecological Statistics. **1**:227-245

Grant, C.L., T.F. Jenkins and A.R. Mudambi. (1996*). Comparison Criteria for Environmental Chemical Analyses of Split Samples Sent to Different Laboratories: Corps of Engineers archived data.* CRREL Special Report 96-9. US Army Corps of Engineers, Environmental Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH.

Griffiths, J.C. (1971). Problems of sampling in geoscience. Transactions of the American Institute of Mining and Metallurgy. **80B**:B346-B355.

Gy, P.M. (1982). Sampling of Particulate Materials: Theory and Practice. Elsevier Scientific Publishing Company. New York, NY.

Gy, P.M. (1986). The analytical and economic importance of correctness in sampling. Analytica Chimica Acta. **190**:13-23.

Gy, P.M. (1992). Sampling of Heterogeneous and Dynamic Material Systems. Elsevier Scientific Publishing Company. New York, NY.

Gy, P.M. (1999). Sampling for Analytical Purposes. John Wiley & Sons. New York, NY.

Hadley, P.W., E. Crapps, and A.D. Hewitt. (2011). Time for a change of scene. Environmental Forensics. **12**:312-318.

Hadley, P.W. and S.D. Mueller. (2012). Evaluating “Hot Spots” of soil contamination (Redux). Soil and Sediment Contamination. **21**:335-350.

Hansen, M.H., W.N. Hurwitz, and W.G. Madow. (1953). Sample Survey Methods and Theory, Vol. 1. John Wiley and Sons. New York, NY.

Harrison, A.F. (1979). Variation of four phosphorus properties in woodland soils. Soil Biology and Biochemistry. **11**:393-403.

Hartman, H. L. (1992). *SME Mining Engineering Handbook*: Vol.2. Seeley W. Mudd Memorial Fund of AIME., Society for Mining, Metallurgy and Exploration. Littleton, CO

Hawaii. (2008). *Technical Guidance Manual.* Hawaii Department of Health, Office of Hazard Evaluation and Emergency Response. http://www.hawaiidoh.org/

Helsel, D.R. (2005) Nondetects and Data Analysis: Statistics for Censored Environmental Data. Wiley-Interscience. Hoboken, NJ.

Herzig, C.S. (1914). Mine Sampling and Valuing. Mining and Scientific Press. San Francisco, CA.

Hewitt, A. D., and M.E. Walsh. (2003). *On-site homogenization and sub sampling of surface samples for analysis of explosives.* ERDC/CRREL TR 03-14. US Army Corps of Engineers, Environmental Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH.

Hewitt, A.D., T.F. Jenkins, C.A. Ramsey, K.L. Bjella, T.A. Ranney, and N.M. Perron. (2005). *Estimating energetic residue loading on military artillery ranges: Large decision units.* ERDC/CRREL TR-05-7, U.S. Army Corps of Engineers, Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH. http://www.crrel.usace.army.mil/techpub/CRREL_Reports/reports/TR05-7.pdf

Hewitt, A.D., T.F. Jenkins, M.E. Walsh, M.R. Walsh, S.R. Bigl, and C.A. Ramsey. (2007). *Protocols for Collection of Surface Soil Samples at Military Training and Testing Ranges or the Characterization of Energetic Munition Constituents.* ERDC/CRREL TR-07-10. US Army Corps of Engineers, Environmental Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH.

Hewitt, A.D., T.F. Jenkins, M.E. Walsh, S.R. Bigl, and S. Brochu. (2009). *Validation of Sampling Protocol and the Promulgation of Method Modifications for the Characterization of Energetic Residues on Military Testing and Training Ranges.* ERDC/CRREL TR-09-6. US Army Corps of Engineers, Environmental Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH.

Hewitt, A.D., T.F. Jenkins, S.R. Bigl, J.L. Clausen, H. Craig, M.E. Walsh, R. Martel, and K. Nieman, S. Taylor, and M.R. Walsh. (2012). *EPA Federal Facilities Forum Issue Paper: Site Characterization for Munitions Constituents.* EPA-505-S-11-01. U.S. Environmental Protection Agency, Solid Waste and Emergency Response, Federal Facilities Forum Issue. Washington, DC.
http://www.epa.gov/fedfac/pdf/site_characterization_for_munitions_constituents.pdf

Holmgren, G.G.S. (1988). Division S-5 soil genesis, morphology, and classification: The point representation of soil. Soil Science Society of America Journal. **52**:712-716.

Ingamells, C.O. (1974). New approaches to geochemical analysis and sampling. Talanta. **21**:141-155.

Ingamells, C.O. and F.F. Pitard. (1986). Applied Geochemical Analysis. John Wilely and Sons. New York, NY.

Isaaks, E.H. and R.M. Srivastava. (1989). An Introduction to Applied Geostatistics. Oxford University Press. New York, NY.

ITRC. (2003). *Characterization and Remediation of Soils at Closed Small-arms Firing Range.* Interstate Technology and Regulatory Council, Small-arms Firing Range Team, Washington, DC.

ITRC. (2005). *Environmental Management at Operating Outdoor Small Arms Firing Ranges.* Interstate Technology and Regulatory Council, Small-arms Firing Range Team, Washington, DC.

ITRC. (2012). *Technical and Regulatory Guidance: Incremental Sampling Methodology*. ISM-1. February 2012, Interstate Technology and Regulatory Council, Incremental Sampling Methodology Team, Washington, DC. http://itrcweb.org/ism-1/

Jenkins, T.F., C.L. Grant, G.S. Brar, P.G. Thorne, T.A. Ranney, and P.W. Schumacher. (1996). *Assessment of Sampling Error Associated with Collection and Analysis of Soil Samples at Explosives-Contaminated Sites.* CRREL Special Report 96-15. US Army Corps of Engineers, Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH.

Jenkins, T.F., M.E. Walsh, P.G. Thorne, S. Thiboutot, G. Ampleman, T.A. Ranney, and C.L. Grant. (1997a). *Assessment of Sampling Error Associated with the Collection and Analysis of Soil Samples at a Firing Range Contaminated with HMX.* CRREL Special Report 97-22. U.S. Army Corps of Engineers, Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH.

Jenkins, T.F., C.L. Grant, G.S. Brar, P.G. Thorne, P.W. Schumacher, and T.A. Ranney. (1997b). Sampling error associated with collection and analysis of soil samples at TNT contaminated sites. Field Analytical Chemistry Technology, 1:151-163.

Jenkins, T.F., M.E. Walsh, P.G. Thorne, P.H. Miyares, T.A. Ranney, C.L. Grant, and J.R. Esparza. (1998). *Site Characterization for Explosives Contamination at a Military Firing Range Impact Area.* CRREL Special Report 98-9. US Army Corps of Engineers, Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH.

Jenkins, T.F., C.L. Grant, M.E. Walsh, P.G. Thorne, S. Thiboutot, G. Ampleman, and T.A. Ranney. (1999). Coping with spatial heterogeneity effects on sampling and analysis at an HMX - contaminated antitank firing range. Field Analytical Chemistry Technology. **3**:19-28.

Jenkins, T. F.; J. C. Pennington, T. A. Ranney, T. E. Berry, P. H. Miyares, M. E. Walsh, A. D. Hewitt, N. M. Perron, L. V. Parker, C. A. Hayes, and E. G. Wahlgren. (2001). *Characterization of Explosives Contamination at Military Firing Range.* ERDC TR-01-5. U.S. Army Corps of Engineers, Environmental Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH.

Jenkins, T.F., M.E. Walsh, P.H. Miyares, A.D. Hewitt, N.H. Collins, and T.A. Ranney. (2002). Evaluation of the use of snow-covered ranges to estimate the explosives residues that result from high order detonations of army munitions. Thermochima Acta. **384**:173-185.

Jenkins, T.F., T.A. Ranney, A.D. Hewitt, M.E. Walsh, and K.L. Bjella. (2004a). *Representative sampling for energetic compounds at an antitank firing range.* ERDC/CRREL TR-04-7. U.S. Army Corps of Engineers, Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH. http://www.crrel.usace.army.mil/library/technicalreports/TR04-7.pdf

Jenkins, T.F., A.D. Hewitt, T.A. Ranney, C.A. Ramsey, D.J. Lambert, K.L. Bjella, and N.M. Perron. (2004b). *Sampling strategies near a low-order detonation and a target at an artillery impact area.* ERDC/CRREL TR-04-14. U.S. Army Corps of Engineers, Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH. http://www.crrel.usace.army.mil/library/technicalreports/TR04-14.pdf

Jenkins, T.F., S. Thiboutot, G. Ampleman, A.D. Hewitt, M.E. Walsh, T.A. Ranney, C.A. Ramsey, C.L. Grant, C.M. Collins, S. Brochu, S.R. Bigl, and J.C. Pennington. (2005a). *Identity and Distribution of Residues of Energetic Compounds at Military Live-Fire Training Ranges.* ERDC-TR-05-10. U.S. Army Corps of Engineers, Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH.

Jenkins, T.F., A.D. Hewitt, M.E. Walsh, T.A. Ranney, C.A. Ramsey, C.L. Grant, and Kevin L. Bjella. (2005b). Representative sampling for energetic compounds at military training ranges. Environmental Forensics, **6**:45-55.

Jenkins, T.F., A.D. Hewitt, C.A. Ramsey, K.L. Bjella, S.R. Bigl, and D.J. Lambert. (2006). *Sampling Studies at an Air Force Live-Fire Bombing Range Impact Area.* ERDC/CRREL TR-06-2. U.S. Army Corps of Engineers, Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH.

Jenny, H. (1941). Factors of Soil Formation: A System of Quantitative Pedology. Dover Publications, Inc. New York.

Johanson, J.R. (1978). Particle segregation and what to do about it. Chemical Engineering, **May**:183-188.

Johnson, N.J. (1978). Modified t-tests and confidence intervals for asymmetrical populations. The American Statistician. **73**:536-544.

Journal, A.G. and C. Huijbregts. (1978). Mining Geostatistics. The University of California Academic Press. San Francisco, CA.

King, A.J. and G.D. Simpson. (1940). New developments in agricultural sampling. Journal of Farm Economics. **22**(1): 341-349.

Knechtenhofer, L., I. Xifra, A. Scheinost, H. Flühler, R. Kretzschmar. (2002). Fate of heavy metals in a strongly acidic shooting-range soil: Small-scale metal distribution and its relation to preferential water flow. Journal of Plant Nutrition and Soil Science. **166**, 84-92.

Leutwyler, K. (1993). Shaking conventional wisdom. Scientific American. **September**:24.

Land, C.E. (1971). Confidence intervals for linear functions of the normal mean and variance. Annals of Mathematical Statistics. **42**:1187-1205.

Land, C.E. (1975). Tables of confidence limits for linear functions of the normal mean and variance. In Selected Tables in Mathematical Statistics. Vol. III. American Mathematical Society. Providence, RI. pp. 385-419.

Lyon, T.L. (1932). Is the soil homogenous with respect to its fertilizer needs? Journal of the American Society of Agronomy. **28**:801-804.

Maclean, A. J. and R. Summerby. (1945). A study of the variability of certain chemical properties in soils. Scientific Agriculture. **25**:221-228.

Markert, B. (1990). Multi-element analysis in ecosystems – Analytical tools and biological questions. International Conference on Metals in Soil., Waters, Plants, and Animals. April 30-May3. Orlando, FL.

Mason, B.J. (1983). *Preparation of Soil Sampling Protocols: Techniques and Strategies*. U.S. Environmental Protection Agency. EPA-600/4-83-020. Las Vegas, NV.

Mason. B.J. (1992). *Preparation of Soil Sampling Protocols: Sampling Techniques and Strategies.* U.S. Environmental Protection Agency. EPA/600/SR-92/128. Las Vegas, NV.

Mathur, S.P. and R.B. Sanderson. (1978). Relationships between copper contents of some Histosols in an area of the southwestern Quebec in the summer and the fall. Canadian Journal of Soil Science. **58**(5):125-134.

Matzke B.D., L.L. Nuffer, J.E. Hathaway, L.H. Sego, B.A. Pulsipher, S. McKenna, J.E. Wilson, S.T. Dowson, N.L. Hassig, C.J. Murray, and B. Roberts. (2010). *Visual Sample Plan Version 6.0 User's Guide.* PNNL-19915, Pacific Northwest National Laboratory, Richland, Washington.

Mausbach, M.J. B.R. Brasher, R.D. Yeck, and W.D. Nettleton. (1980). Variability of measured properties in morphologically matched pedons. Soil Science Society of America Journal. **44**: 358-363.

Mercer, W.B. and A.D. Hall. (1911). The experimental error of field trials. Journal of Agricultural Science. **4**:107-132.

Miesch, A.T. (1967). *Methods of Computation for Estimating Geochemical Abundance.* U.S. Geological Survey, Professional Paper 574-B. Washington, DC.

Moore, G.K. (1995). *Inorganic Soil and Groundwater Chemistry Near Paducah Gaseous Diffusion Plant, Paducah, Kentucky.* ORNL/TM-12897. Oak Ridge National Laboratory. Oak Ridge, TN.

Moore, K. (2011). History of Soil-Testing in South Carolina. Clemson Agriculture Service Laboratory. http://www.clemson.edu/public/regulatory/ag_svc_lab/soil_testing/history.html Accessed on 4/9/13.

Moran, P.A.P. (1950). Notes on continuous stochastic phenomena. Biometrika. **37**(1):17–23.

Mullins, C.E. and B.J. Hutchison. (1982). The variability introduced by various sub-sampling techniques. Journal of Soil Science. **33**:547-561.

Nam, S.I., B.D. Roebuck and M.E. Walsh. (1994). Uptake and loss of white phosphorus in American Kestrels. Environmental Toxicology and Chemistry, **13**:637-641.

Neptune, D. (1990). Practical statistical applications to planning superfund site remediation. *EPA Statistician*. **5**(1):1-3.

Nielsen, D.R., J.W. Biggar, and K.T. Erh. (1973). Spatial variability of field-measured soil-water properties. Journal of Agricultural Science. **42**(7):215-259.

Ogden. (1998). *Interim Results Report for the Camp Edwards Impact Area Groundwater Quality Study Massachusetts Military Reservation Cape Cod, Massachusetts*. Ogden Environmental and Energy Services. Westford, MA.

Onate, B.T. (1953). Some statistical aspects of the use of composites in soil sampling. The Philippine Agriculturist. **2**:41-57.

Pennington, J.C., T F. Jenkins, A.D. Hewitt, J.A. Stark, D. Lambert, N.M. Perron, S. Taylor, G. Ampleman, S. Thiboutot, J. Lewis, A. Marois, C. Gauthier, P. Brousseau, R. Martel, R. Lefebvre, J. Ballard, S. Brochu, J. Clausen, J.E. Delaney, M.A. Hollander, H.Q. Dinh, I. Davis, T.A. Ranney, and C.A. Hayes. (2004). *Distribution and Fate of Energetics on DoD Test and Training Ranges: Interim Report 4.* ERDC TR-04-4. Strategic Environmental Research and Development Program. Arlington, VA

Petersen, L., C.K. Dahl, and K.H. Esbensen. (2004). Representative mass reduction in sampling – a critical survey of techniques and hardware. Chemometrics and Intelligent Laboratory Systems. **74**:95-114.

Peterson, R.G. and L.D. Calvin. (1982). Sampling. In Methods of Soil Analysis, Part I. Physical and Mineralogical Methods. Ed. A. Klute. 2nd Edition. American Society of Agronomy, Inc. and Soil Science Society of America, Inc. Madison, WI.

Pèwè, T.L, J.W. Bell, R.B. Forbes, and F.R. Weber. (1976). *Geologic map of the Fairbanks D-2 SE Quadrangle, Alaska.* Miscellaneous Investigations series Map I-942. US Geological Survey. Washington, D.C.

Pitard, F.F. (1993). Pierre Gy's Sampling Theory and Sampling Practice: Heterogeneity, Sampling Correctness, and Statistical Process Control. CRC Press, Boca Raton, Florida.

Racine, C.H., M.E. Walsh, C.M. Collins, S.T. Taylor, B.D. Roebuck, L. Reitsma, B. Steele. (1993). *Remedial Investigation Report for White Phosphorus Contamination in an Alaskan Salt Marsh.* CRREL Report 93-17. U.S. Army Corps of Engineers, Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH.

Racine, C.H., M.E. Walsh, B.D. Roebuck, C.M. Collins, D.J. Calkins, L. Reitsma, P. Buchli, and G. Goldfarb. (1992). White phosphorus poisoning of waterfowl in an Alaskan Salt Marsh. Journal of Wildlife Diseases. **28**:669-673.

Ramsey, C. (2009). Multi Increment Sampling. SERDP/ESTCP Symposium. December 2-7. Washington, DC.

Ramsey, C. (2006). Sampling for Defensible Environmental Decisions. Training Class presented July 25-28, Omaha, NE. EnviroStat Inc.

Rasemann, W. (2000). Industrial Waste Dumps, Sampling and Analysis. In Encyclopedia of Environmental Chemistry. John Wiley and Sons, Inc. New York, NY.

Reed, J.F. and J. A. Rigney. (1947). Soil sampling from fields of uniform and nonuniform appearance and soil type. Journal of the American Society of Agronomy. **39**:26-40.

Reimann C. and P. Filzmoser. (2000). Normal and lognormal data distribution in geochemistry: death of a myth. Consequences for the statistical treatment of geochemical and environmental data. Environmental Geology. **39**(9):1001-1014

Rickard, T.A. (1906). The Sampling and Estimation of Ore in a Mine. The Engineering and Mining Journal. New York, NY.

Robinson, G.W. and W.E. Lloyd (1915). On the probable error of sampling in soil surveys. The Journal of Agricultural Science. **7**:144-153.

Roebuck, B.D., S.I. Nam, D.L MacMillan, K. Baumgarner, and M.E. Walsh. (1998). Toxicology of white phosphorus (P_4) to Ducks, and risk to their predators: Effects of particle size. Environmental Toxicology and Chemistry. **17**(3):511-518.

Rooney, C.P., R.G. Mclaren, and R.J. Cresswell. (1999). Distribution and phytoavailability of lead in a soil contaminated with lead shot. Water, Air and Soil Pollution. **116**:534-548.

Sampford, M.R. (1962). An Introduction to Sampling Theory with Applications to Agriculture. Oliver and Boyd. London.

Scheinost, A., A. Rossberg, D. Vantelon, I. Xifra, R. Kretzschmar, A-K, Leuz, H. Funke, and C.A. Johnson. (2006). Quantitative antimony speciation in shooting-range soils by EXAFS spectroscopy. Geochimica et Cosmochimica Acta. **70**(13): 3299-3312.

Singh, A.K., A. Singh, M. Engelhardt. (1997). *The Lognormal Distribution in Environmental Applications.* EPA/600/R-97/006. U.S. Environmental Protection Agency, Office of Research and Development, Office of Solid Waste and Emergency Response. Washington, DC.

Singh, A., A.K. Singh, and R.J. Laci. (2002). *Estimation of the Exposure Point Concentration Term Using a Gamma Distribution.* EPA/600/R-02/084. U.S. Environmental Protection Agency, Office of Research and Development, Office of Solid Waste and Emergency Response. Washington, DC.

Singh, A. and A. K. Singh. (2003). *Estimation of the Exposure Point Concentration Term (95% UCL) Using Bias-Corrected Accelerated (BCA) Bootstrap and Several Other Methods for Normal, Lognormal, and Gamma Distributions.* Draft USEPA Internal Report. U.S. Environmental Protection Agency, Office of Research and Development, Office of Solid Waste and Emergency Response. Washington, DC.

Smith, A.M. and E.G. Prentice. (1929). Investigation on *Heterodera schachtii* in Lancashire and Cheshire. Annals of Applied Biology. **16**:324-339.

Snedecor, G.W. (1940). Statistical Methods. Iowa State College Press, Ames, Iowa.

Snedecor, G.W. and W.G. Cochran. (1967). Statistical Methods (6th ed.). Iowa State University Press. Ames, Iowa.

Soil Survey Staff. (1975). Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys. USDA-SCS Agricultural Handbook436. U.S. Government Printing Office. Washington, DC.

Sorvari, J., R. Antikainen, and O. Pyy. (2006). Environmental contamination at Finnish shooting ranges- the scope of the problem and management options. Science of the Total Environment. **366**(1):21-31.

Splitstone, D.E. (2001). Sample support and related scale issues in composite sampling. Environmental and Ecological Statistics. **8**:137-149.

Stephen, F.F. (1949). History of the uses of modern sampling procedures. Journal of the American Statistical Association. **43**(241):12-39.

Student. (1909-10). The distribution of the means of samples which are not drawn at random. Biometrika. **7**:210.

Studt, T. (1995). For material researchers, It's back to the sandbox. R&D Magazine. **July**:41-42.

Taylor, S., A. Hewitt, J. Lever, C. Hayes, L. Perovich, P. Thorne and C. Daghlian. (2004). TNT particle size distributions from detonated 155-mm howitzer rounds. Chemosphere. **55**:357–367.

Taylor, S.,.S. Bigl, C. Vuyovich, J. Roningen, A. Wagner, N. Perron, S. Daly, M.E. Walsh, J. Hug. (2011). *Explosives Dissolved from Unexploded Ordnance.* ADA562287. U.S. Army Corps of Engineers, Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH.

Thiboutot, S., G. Ampleman, A. Gagnon, A. Marois, T.F. Jenkins, M.E. Walsh, P.G. Thorne, and T.A. Ranney. (1998a). *Characterization of Antitank Firing Ranges at CDB Valcartier, WATC Wainwright and CFAD Dundurn.* Defence Research Establishment – Valcartier. Valcartier, Quebec, Canada.

Thiboutot, S., G. Ampleman, P. Dube, J. Hawari, B. Spencer, L. Paquet, T. F. Jenkins, and M. E. Walsh. (1998b). *Protocol for the Characterization of Explosive-Contaminated Sites.* DREV-R-9721. Defense Research Establishment Department of National Defense. Valcartier, Quebec, Canada.

Thiboutot, S., G. Ampleman, A. Gagnon, and A. Marois. (2000a). *Characterization of an Unexploded Ordinance Contaminated Range (Tracadie Range) for Potential Contamination by Energetic Materials.* DREV-TR-2000-102. Defence Research Establishment – Valcartier: Valcartier, Quebec, Canada,

Thiboutot, S., G. Ampleman, P. Dube, C. Dubois, R. Martel, R. Lefebvre, M. Mailloux, G. Sunahara, P. Y. Roubidoux, and J. Hawari. (2000b). *Characterization of DND Training Ranges Including Anti-Tank Firing Ranges and Ecotoxicological Assessment.* Defence Research Establishment – Valcartier. Valcartier, Quebec, Canada.

Thiboutot, S., G. Ampleman, and A.D. Hewitt. (2002). *Guide for Characterization of Sites Contaminated with Energetic Materials.* ERDC/CRREL Technical Report TR-02-1. U.S. Army Corps of Engineers, Environmental Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH.

Thiboutot, S., G. Ampleman, J. Lewis, D. Faucher, A. Marois, R. Martel, J.M. Ballard, S. Downe, T.F. Jenkins, A. Hewitt. (2003a). *Environmental Conditions of Surface Soils and Biomass Prevailing in the Training Area at CFB Gagetown, New Brunswick.* DREV-TR-2003-152. Defence Research Establishment – Valcartier. Valcartier, Quebec, Canada.

Thiboutot, S., G. Ampleman, S. Brochu, R. Martel, G. Sunahara, J. Hawari, S. Nicklin, A. Provatas, J. Pennington, T.F. Jenkins, and A. Hewitt. (2003b). *Protocol for Energetic Materials-Contaminated Sites Characterization.* NATO KTA 4-28 Final Report. The Technical Cooperation Program.

URS. (2010). *Remedial Investigation/Feasibility Study 1000" Rifle Range Fort Eustis, Virginia Military Munitions Response Program, Volume 1: Remedial Investigation*. URS Group Inc., Gaithersburg, MD.

USACE. (2009). *Implementation of Incremental Sampling (IS) of Soil for the Military Munitions Response Program, Environmental and Munitions Center of Expertise.* Interim Guidance Document (IGD) 09-02. Department of the Army, Huntsville Center, Corps of Engineers. Huntsville, AL.

USDoD. (2005). *Operational Range Assessments*. DoD Instruction 4715.14. Department of Defense. Washington, DC. http://www.dtic.mil/whs/directives/corres/pdf/471514p.pdf.

USDoD. (2007). *Environmental and Explosives Safety Management on Operational Ranges Within the United States*. DoD Directive 4715.11. Department of Defense. Washington, DC. http://www.dtic.mil/whs/directives/corres/pdf/471511p.pdf.

USEPA. (1984). *Characterization of Hazardous Waste Sites—A Methods Manual: Volume II. Available Sampling Methods, Second Edition.* EPA-600/4-84-076. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory. Las Vegas, NV.

USEPA. (1989). *Risk Assessment Guidance for Superfund, Volume I – Human Health Evaluation Manual (Part A).* Interim Final. U.S. Environmental Protection Agency, Office of Emergency and Remedial Response. EPA/540/1/89/002. Washington, DC. http://www.epa.gov/superfunf/programs/risk/ragsa

USEPA. (1991). *Description and Sampling of Contaminated Soils: A Field Pocket Guide.* EPA/625-12-91/002. U.S. Environmental Protection Agency. Cincinnati, OH.

USEPA. (1992*). A Supplemental Guidance to RAGS: Calculating the Concentration Term.* U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response. Publication 9285.7-081. Washington, DC. http://www.deq.state.ms.us/newweb/opchome.nsf/pages/HWDivisionFiles/$file/uclmean.pdf

USEPA. (1994a). Method 8330: Nitroaromatics, nitramines, nitrate esters by high performance liquid chromatography (HPLC). In *Test Methods for Evaluating Solid Waste, Physical/Chemical Methods, Office of Solid Waste and Emergency Response.* SW-846. U.S. Environmental Protection Agency. Washington, DC.

USEPA. (1994b). Method 200.7: Determination of Metals and Trace Elements in Water and Wastes by Inductively Coupled Plasma-Atomic Emission Spectrometry. In *Test Methods for Evaluating Solid Waste, Physical/Chemical Methods, Office of Solid Waste and Emergency Response.* SW-846. U.S. Environmental Protection Agency. Washington, DC.

USEPA. (1995). *Superfund Program Representative Sampling Guidance, Volume 1: Soil.* EPA 540/R-95/141. U.S. Environmental Protection Agency. Office of Solid Waste and Emergency Response. Washington, DC.

USEPA. (1996a). SW-846 Method 3050B: Acid Digestion of Sediments, Sludges, and Soils. In *Test Methods for Evaluating Solid Waste, Physical/Chemical Methods*. U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response. Washington, DC. http://www.epa.gov/osw/hazrad/testmethods/sw846/pdfs/3050b.pdf

USEPA. (1996b). SW-846 Method 3051A: Microwave Assisted Acid Digestion of Sediments, Sludges, Soils, and Oils. In *Test Methods for Evaluating Solid Waste, Physical/Chemical Methods*. U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response. Washington, DC.

USEPA. (1996c). SW-846 Method 3052: Microwave Assisted Acid Digestion of Siliceous and Organically Based Matrices: In *Test Methods for Evaluating Solid Waste, Physical/Chemical Methods*. U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response. Washington, DC.

USEPA. (1996d). Method 8330B: Nitroaromatics, nitramines, nitrate esters by high performance liquid chromatography (HPLC). In *Test Methods for Evaluating Solid Waste, Physical/Chemical Methods, Office of Solid Waste and Emergency Response.* SW-846. U.S. Environmental Protection Agency. Washington, DC.

USEPA. (1997). Massachusetts Military Reservation Training Range and Impact Area: EPA Region I Administrative Order. SDWA 1-97-1019. U.S. Environmental Protection Agency. http://www.epa.gov/region1/mmr/pdfs/448135.pdf

USEPA. (2000a). *USEPA Environmental Response Team, Standard Operating Procedures for Soil Sampling.* U.S. Environmental Protection Agency. http://www.epa.gov/region6/qa/qadevtools/mod5_sops/soil_sampling/ertsop2012-soil.pdf

USEPA. (2000b). *Guidance for Data Quality Assessment: Practical Methods for Data Analysis.* U.S. Environmental Protection Agency, Office of Environmental Information. EPA QA/G-9, QA00 Update. Washington, DC. http://www.epa.gov/r10earth/offices/oea/epaqag9b.pdf

USEPA. (2002a). *Guidance on the Calculation of Upper Confidence Limits for Exposure Point Concentrations at Superfund Sites.* U.S. Environmental Protection Agency, Office of Emergency and Remedial Response. OSWER 9285.6-10. Washington, DC. http://www.epa.gov/oswer/riskassessment/pdf/ucl.pdf

USEPA. (2002b). *Guidance of Choosing a Sampling Design for Environmental Data Collection for Use in Developing A Quality Assurance Project Plan.* U.S. Environmental Protection Agency. EPA/240/R-02/005. Washington, DC.

USEPA. (2005). *Best Management Practices for Lead at Outdoor Shooting Ranges.* U.S. Environmental Protection Agency. EPA-02-B-01-001. Division of Enforcement and Compliance Assistance, RCRA Compliance Branch. New York, NY.

USEPA. (2006a). Method 8330B: Nitroaromatics, nitramines, nitrate esters by high performance liquid chromatography (HPLC). In *Test Methods for Evaluating Solid Waste, Physical/Chemical Methods, Office of Solid Waste and Emergency Response.* SW-846. U.S. Environmental Protection Agency. Washington, DC.

USEPA. (2006b). Method 6020A: Inductively coupled plasma-mass spectrometry. In *Test Methods for Evaluating Solid Waste, Physical/Chemical Methods, Office of Solid Waste and Emergency Response.* SW-846. U.S. Environmental Protection Agency. Washington, DC.

USEPA. (2006c). Method 6010C: Inductively coupled plasma-atomic emission spectrometry. In *Test Methods for Evaluating Solid Waste, Physical/Chemical Methods, Office of Solid Waste and Emergency Response.* SW-846. U.S. Environmental Protection Agency. Washington, DC.

USEPA. (2013). *ProUCL Version 5.0.0 User Guide, Statistical Software for Environmental Applications for Data Sets with and without Nondetect Observations.* U.S. Environmental Protection Agency. EPA/600/R-07/41. Washington, DC.

USGS. (2002). *Analytical Methods for Chemical Analysis of Geologic and Other Materials.* J.E. Taggert Jr. Ed. U.S. Geological Survey, Open-File Report 02-0223. Denver, CO. http://pubs.usgs/gov/of/2002/ofr-02-0223/OFR-02-0223.pdf.

van Ee, J. and L. Blum. (1990). *A Rational for the Assessment of Errors in the Sampling of Soils.* U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory. USEPA/600/4/-90/013. Las Vegas, NV.

Vantelon, D., A. Lanzirotti, A. Scheinost, and R. Kretzschmar. (2005). Spatial distribution and speciation of lead around corroding bullets in a shooting range soil studied by micro-X-ray fluorescence and absorption spectroscopy. Environmental Science and Technology. **39**:4808-4815.

Visman, J. (1969). A general sampling theory. Material Research Standards. **9**(11):8-13, 51-56, 62, 64, 66.

Visman, J. (1971). A further discussion; a general sampling theory. Material Research Standards. 32-37.

Wallace, D. and B. Kratochvil (1985) Use of a mill and spinning riffle for sub-sampling laboratory samples of oil sand. Australian Journal of Research. **2**: 233-239.

Walsh M.E. and C.M. Collins. (1993). *Distribution of White Phosphorus in Residues from the Detonation of 81 mm Mortar WP Smoke Rounds.* CRREL Special Report 93-18. U.S. Army Corps of Engineers, Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH.

Walsh M.E. and D.J. Lambert. (2006). *Extraction Kinetics of Energetic Compounds from Training Range and Army Ammunition Plant Soils: Platform Shaker versus Sonic Bath Methods.* ERDC/CRREL TR-06-6. US Army Corps of Engineers, Environmental Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH.

Walsh, M.E., C.M. Collins, and C.H. Racine. (1996). Persistence of white phosphorus particles in salt marsh sediments. Environmental Toxicology and Chemistry. **15**:846-855.

Walsh, M.E., C.I. Collins, R.N. Bailey, and C.L. Grant. (1997). *Composite Sampling of Sediments Contaminated with White Phosphorus.* CRREL Special Report 97-30. U.S. Army Corps of Engineers, Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH.

Walsh, M.E., C.A. Ramsey, and T.F. Jenkins. (2002). The effect of particle size reduction on sub sampling variance for explosives residues in soil. Chemosphere. **49**:1265-1271.

Walsh, M.E., C.M. Collins, T.F. Jenkins, A.D. Hewitt, J. Stark, and K. Myers. (2003). Sampling for explosives-residues at Ft. Greely. Soil and Sediment Contamination. **12**:631-645.

Walsh, M.E., C.M. Collins, A.D. Hewitt, M.R. Walsh, T.F. Jenkins, J. Stark, A. Gelvin, T.S. Douglas, N. Perron, D. Lambert, R. Bailey and K. Meyers. (2004). *Range characterization studies at Donnelly Training Area, Alaska: 2001 and 2003.* ERDC/CRREL TR-04-3. U.S. Army Corps of Engineers, Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH. http://www.crrel.usace.army.mil/library/technicalreports/TR04-3.pdf

Walsh, M.E., C.A. Ramsey, C.M. Collins, A.D. Hewitt, M.R. Walsh, K. Bjella, D. Lambert, and N. Perron. (2005). *Collection methods and laboratory processing of samples from Donnelly Training Area Firing Points Alaska 2003*. ERDC/CRREL TR-05-6. US Army Corps of Engineers, Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory. Hanover, NH. http://www.crrel.usace.army.mil/techpub/CRREL_Reports/reports/TR05-6.pdf

Walsh, M.E., C.A. Ramsey, S. Taylor, A.D. Hewitt, K. Bjella, C.M. Collins. *(*2006). Sub sampling variance for 2,4-DNT in firing point soils. Soil and Sediment Contamination: an International Journal. **16**(5):459-472.

Walsh, M.R. (2009). *User's Manual for the CRREL Multi-Increment Sampling Tool.* ERDC/CRREL SR-09-1. Hanover, NH: US Army Engineer Research and Development Center. http://acwc.sdp.sirsi.net/client/default.

Waynick, and Sharp, L.T. (1919). Variability in soils and its significance to past and future soil investigations: II. Variations in nitrogen and carbon in field soils and their relation to the accuracy of field trials. California University Publication in Agricultural Science. **4**:120-139.

White, G.C. and T.E. Hakonson. (1979). Statistical Considerations and Survey of Plutonium Concentration Variability in Some Terrestrial Ecosystem Components. Journal of Environmental Quality. **8**(2):176- 182.

Wilding, W.P. and L.R. Drees. (1983). Spatial variability and pedology. In Volume: II Issue: Part A, pp. 83-116. Elsevier. New York.

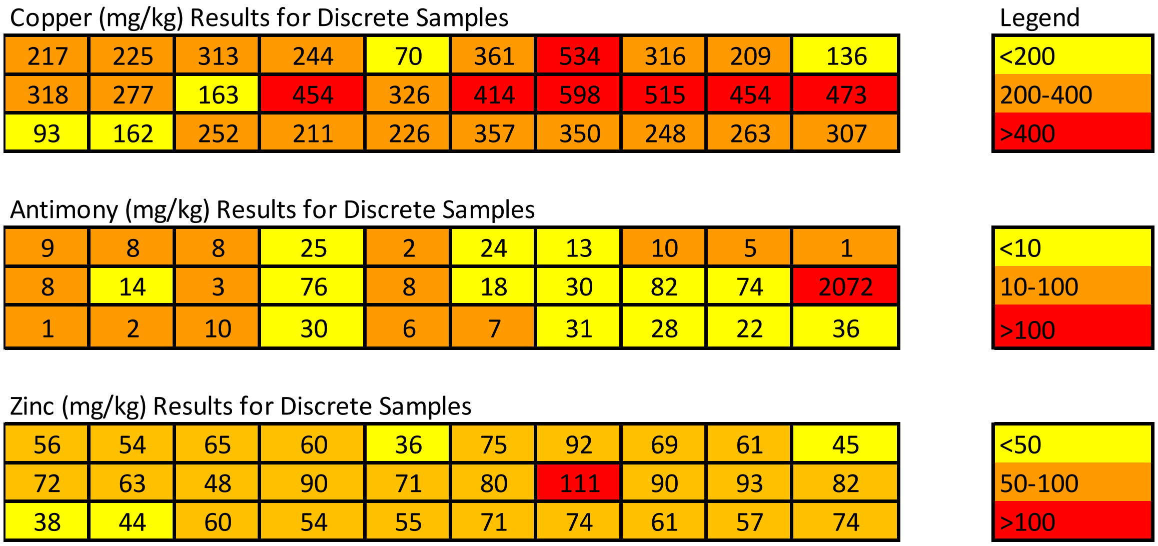
Wood, T.B. and F.J.M. Stratton. (1910). The interpretation of experimental results. Journal of Agricultural Science. **3**:415-440.

Wright, L.T. (1910). The element of chance in the sampling of ores. The Mining Magazine. **3**:353-358.

Yates, F. and I. Zacopany. (1935). Estimation of the efficiency of sampling with special reference for yield in cereal experiments. Journal of Sci. Agr. **25**:545-577.

APPENDIX

The following is supplemental information provided to support the various chapters in the dissertation, which is referenced in the text.



Copper (mg/kg) Results for Discrete Samples

217	225	313	244	70	361	534	316	209	136
318	277	163	454	326	414	598	515	454	473
93	162	252	211	226	357	350	248	263	307

Legend
<200
200-400
>400

Antimony (mg/kg) Results for Discrete Samples

9	8	8	25	2	24	13	10	5	1
8	14	3	76	8	18	30	82	74	2072
1	2	10	30	6	7	31	28	22	36

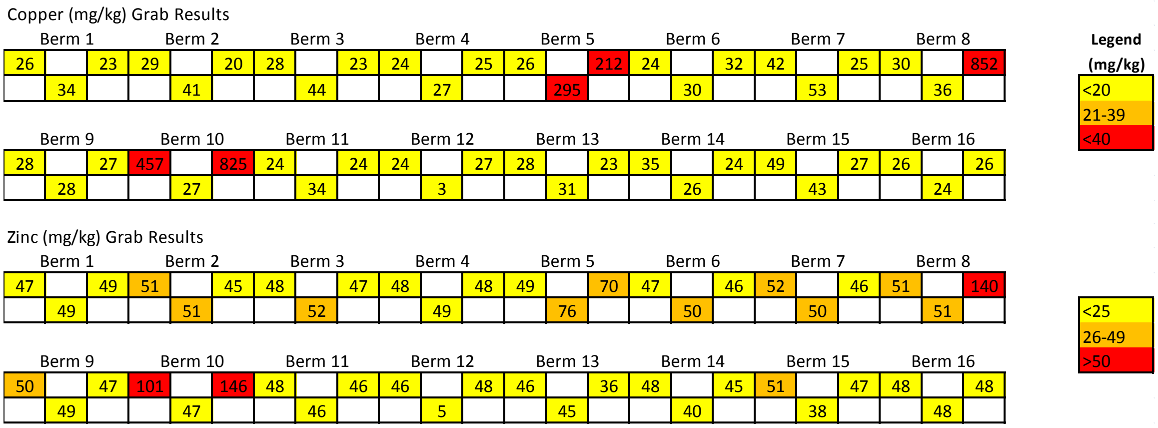
<10
10-100
>100

Zinc (mg/kg) Results for Discrete Samples

56	54	65	60	36	75	92	69	61	45
72	63	48	90	71	80	111	90	93	82
38	44	60	54	55	71	74	61	57	74

<50
50-100
>100

Figure SI1. Selected anthropogenic metal results for grab samples collected from the berm face of Range 4-3 at Camp Ethan Allen.



Copper (mg/kg) Grab Results

Berm 1			Berm 2			Berm 3			Berm 4			Berm 5			Berm 6			Berm 7			Berm 8		
26		23	29		20	28		23	24		25	26		212	24		32	42		25	30		852
	34			41			44			27			295			30			53			36	

Berm 9			Berm 10			Berm 11			Berm 12			Berm 13			Berm 14			Berm 15			Berm 16		
28		27	457		825	24		24	24		27	28		23	35		24	49		27	26		26
	28			27			34			3			31			26			43			24	

Legend (mg/kg)
<20
21-39
<40

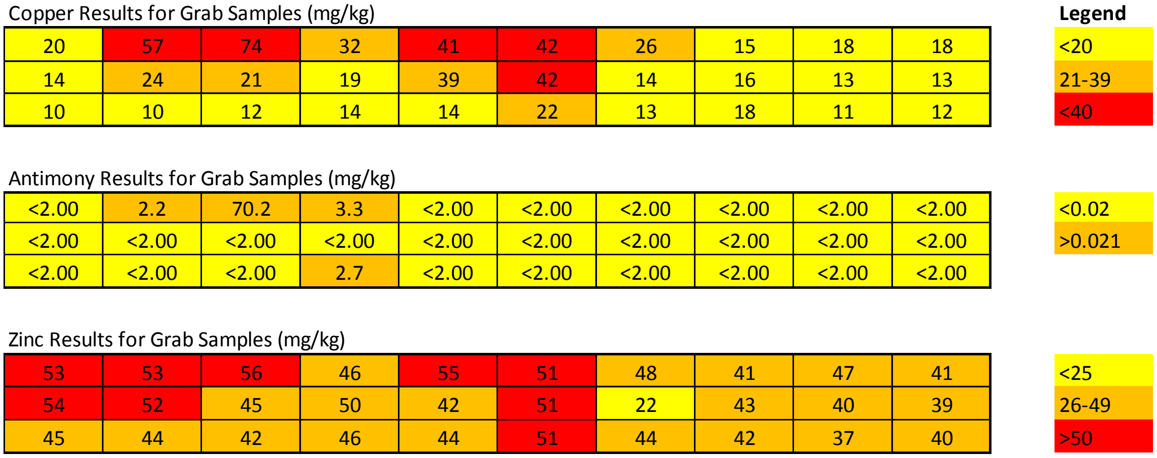
Zinc (mg/kg) Grab Results

Berm 1			Berm 2			Berm 3			Berm 4			Berm 5			Berm 6			Berm 7			Berm 8		
47		49	51		45	48		47	48		48	49		70	47		46	52		46	51		140
	49			51			52			49			76			50			50			51	

<25
26-49
>50

Berm 9			Berm 10			Berm 11			Berm 12			Berm 13			Berm 14			Berm 15			Berm 16		
50		47	101		146	48		46	46		48	46		36	48		45	51		47	48		48
	49			47			46			5			45			40			38			48	

Figure SI2. Selected anthropogenic metal results for grab samples collected from the berm face Range 16 at Fort Wainwright.



Copper Results for Grab Samples (mg/kg)

20	57	74	32	41	42	26	15	18	18
14	24	21	19	39	42	14	16	13	13
10	10	12	14	14	22	13	18	11	12

Legend
<20
21-39
<40

Antimony Results for Grab Samples (mg/kg)

<2.00	2.2	70.2	3.3	<2.00	<2.00	<2.00	<2.00	<2.00	<2.00
<2.00	<2.00	<2.00	<2.00	<2.00	<2.00	<2.00	<2.00	<2.00	<2.00
<2.00	<2.00	<2.00	2.7	<2.00	<2.00	<2.00	<2.00	<2.00	<2.00

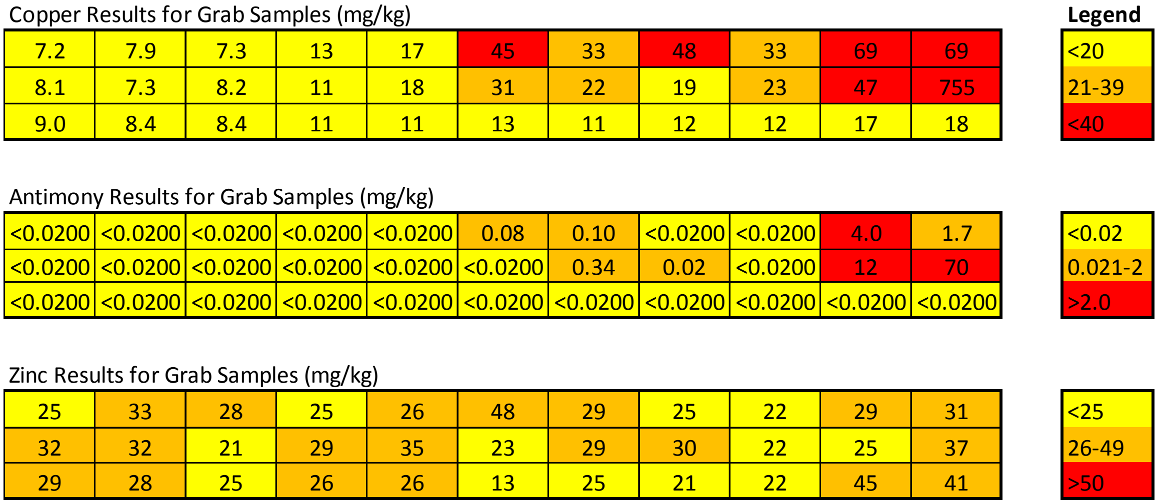
<0.02
>0.021

Zinc Results for Grab Samples (mg/kg)

53	53	56	46	55	51	48	41	47	41
54	52	45	50	42	51	22	43	40	39
45	44	42	46	44	51	44	42	37	40

<25
26-49
>50

Figure SI3. Selected anthropogenic metal results for grab samples collected from the western berm face at the Kimama training site.



Copper Results for Grab Samples (mg/kg)

7.2	7.9	7.3	13	17	45	33	48	33	69	69
8.1	7.3	8.2	11	18	31	22	19	23	47	755
9.0	8.4	8.4	11	11	13	11	12	12	17	18

Legend
<20
21-39
<40

Antimony Results for Grab Samples (mg/kg)

<0.0200	<0.0200	<0.0200	<0.0200	<0.0200	0.08	0.10	<0.0200	<0.0200	4.0	1.7
<0.0200	<0.0200	<0.0200	<0.0200	<0.0200	<0.0200	0.34	0.02	<0.0200	12	70
<0.0200	<0.0200	<0.0200	<0.0200	<0.0200	<0.0200	<0.0200	<0.0200	<0.0200	<0.0200	<0.0200

<0.02
0.021-2
>2.0

Zinc Results for Grab Samples (mg/kg)

25	33	28	25	26	48	29	25	22	29	31
32	32	21	29	35	23	29	30	22	25	37
29	28	25	26	26	13	25	21	22	45	41

<25
26-49
>50

Figure SI4. Selected anthropogenic metal results for grab samples collected from the berm face of the 1000-inch Range at Fort Eustis.

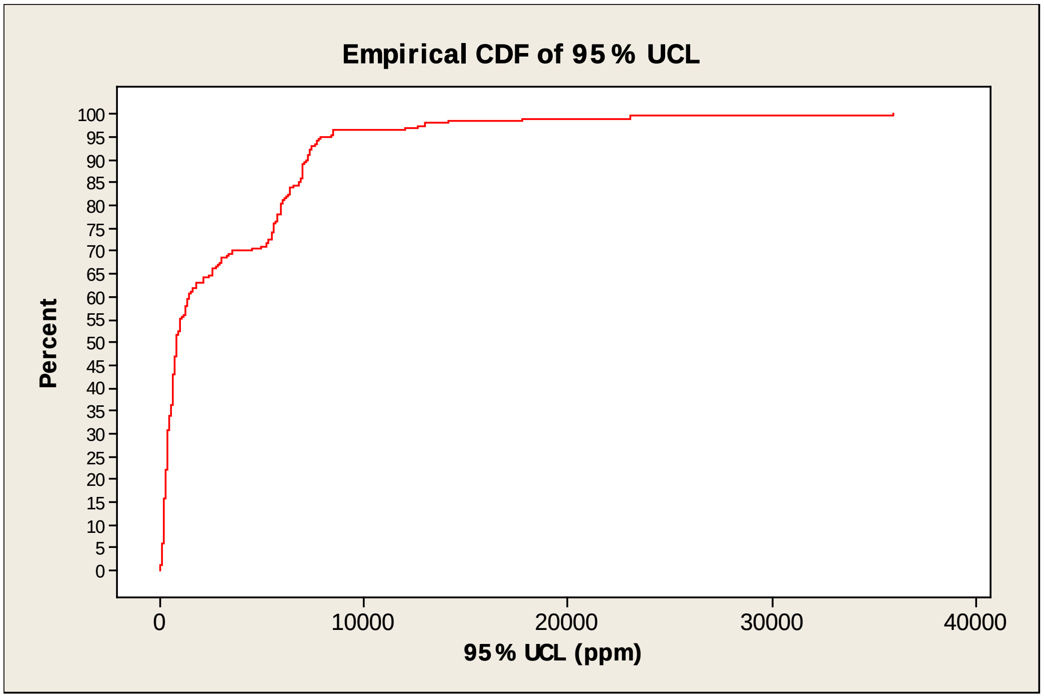


Figure SI5. A cumulative distribution function for the simulated 95% upper confidence limits for lead.

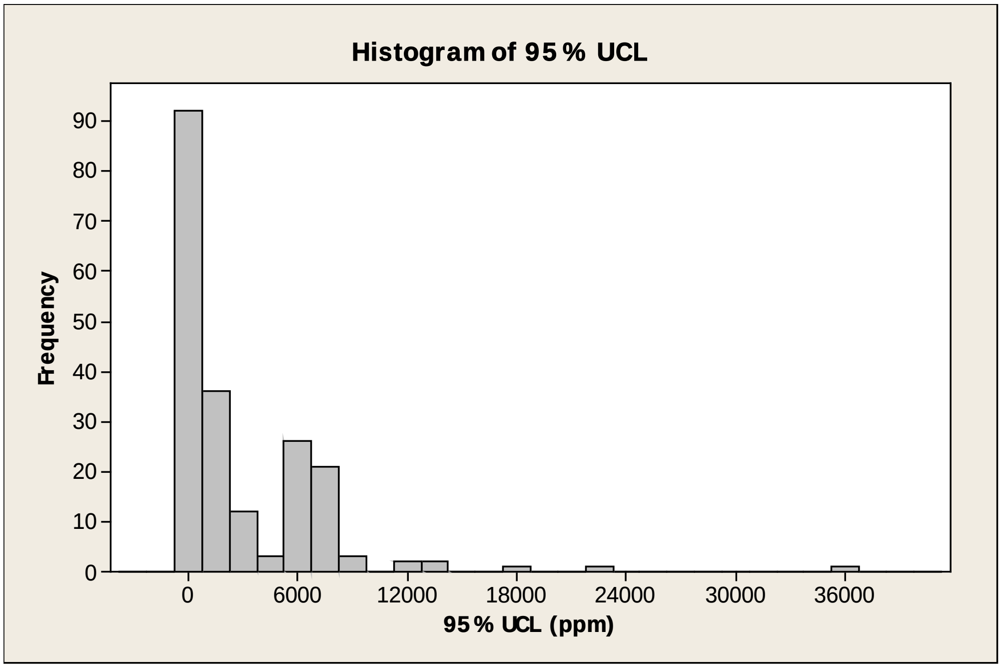


Figure SI6. Histogram for the simulated 95% upper confidence limits for lead.

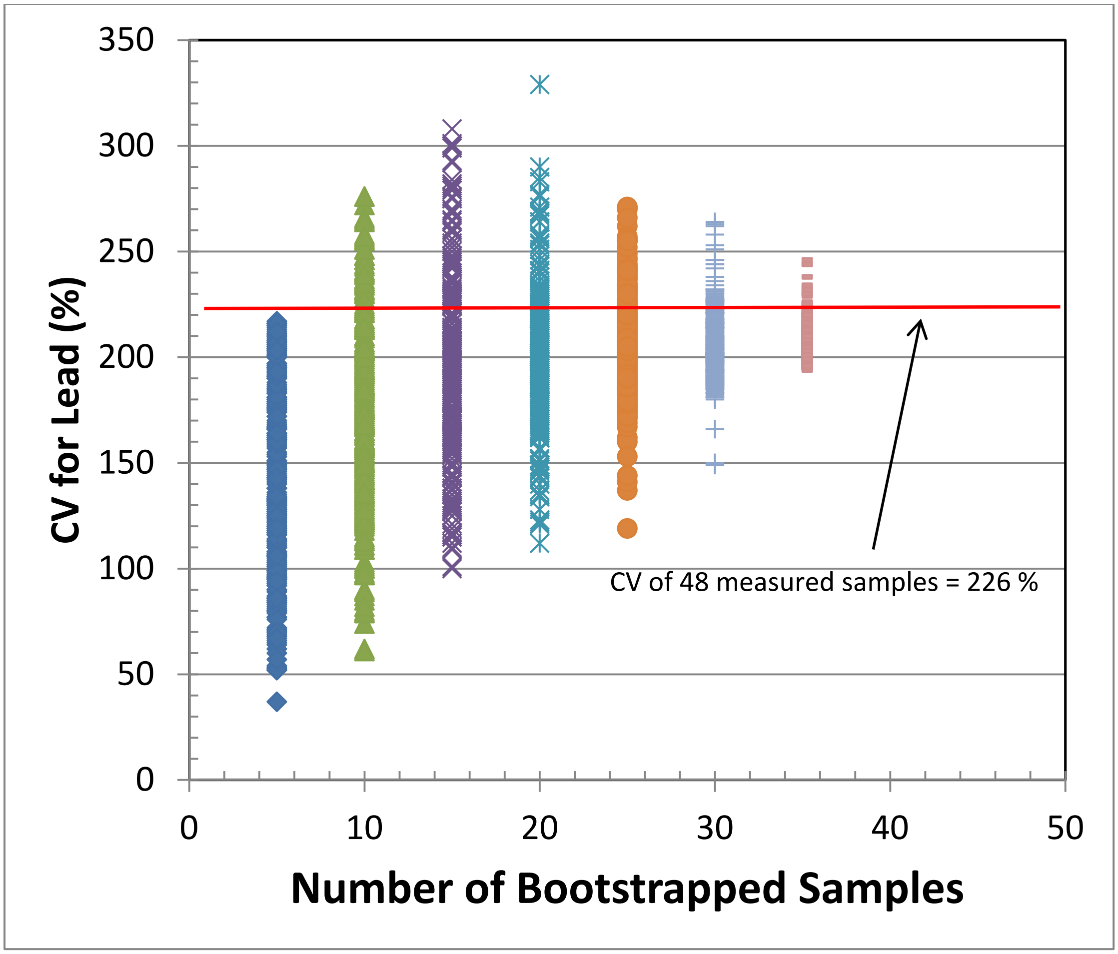


Figure SI7. Estimated covariance levels versus the number of bootstrapped samples for Range 16 at Fort Wainwright. Three hundred simulations were performed for each value of *m* (number of bootstrapped samples).

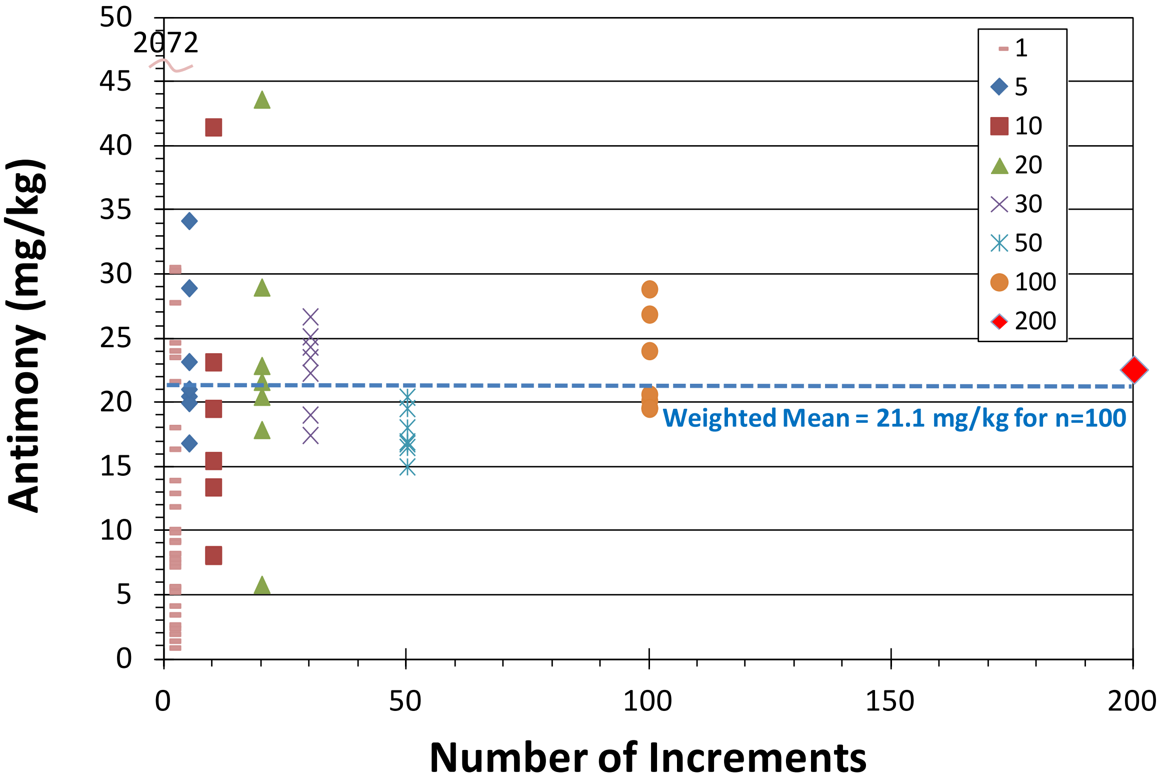


Figure SI8. Distribution of antimony results by number of increments. An increment of one is a grab result whereas increments of 5, 10, 20, 30, 50, 100, and 200 are incremental sampling method results. A weighted incremental sampling methodology mean was calculated for those samples with more than 30 increments

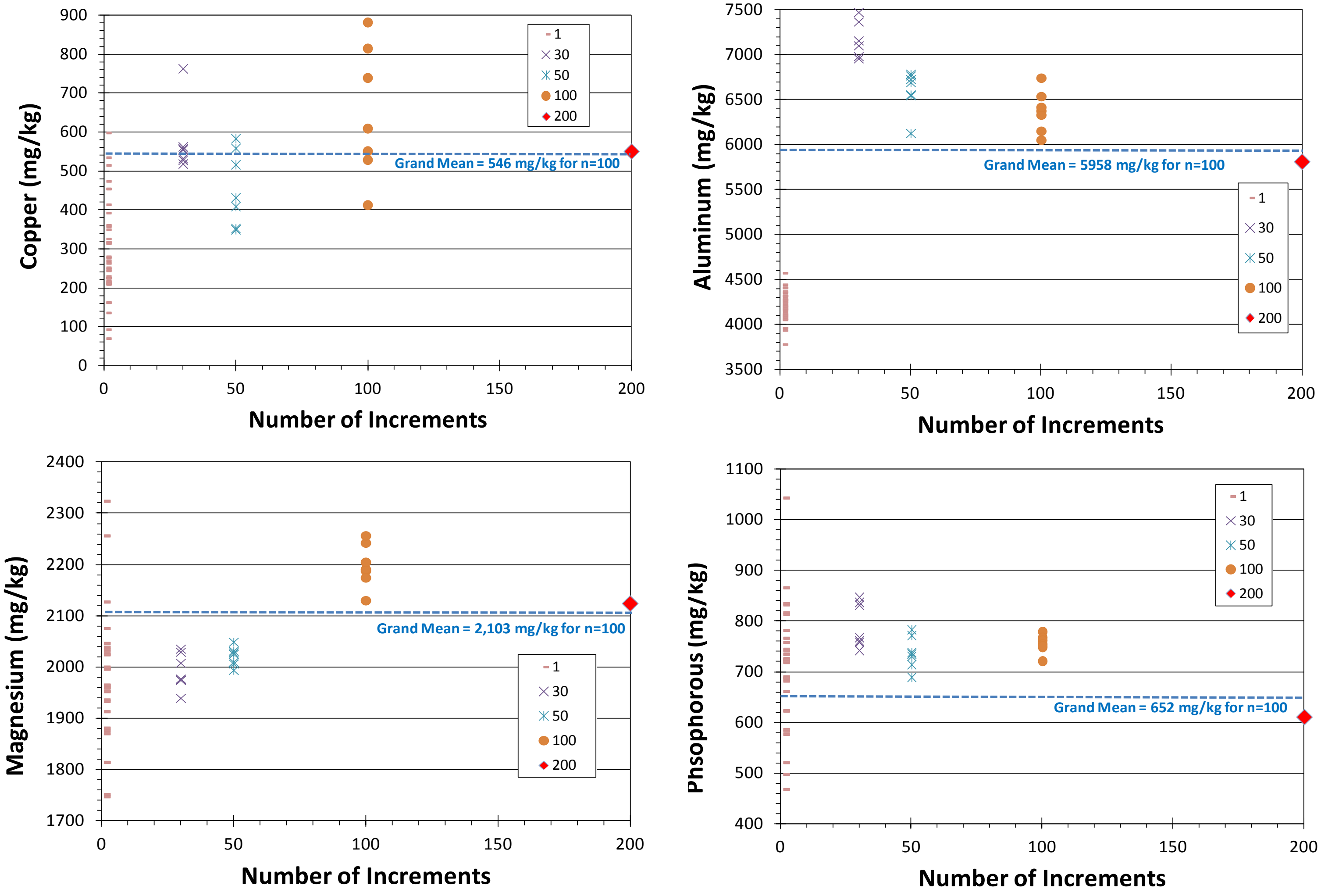


Figure SI9. Distribution of copper, aluminum, magnesium, and phosphorous by number of increments. An increment of one is a grab result, whereas increments of 30, 50, 100, and 200 are incremental sampling methodology results.

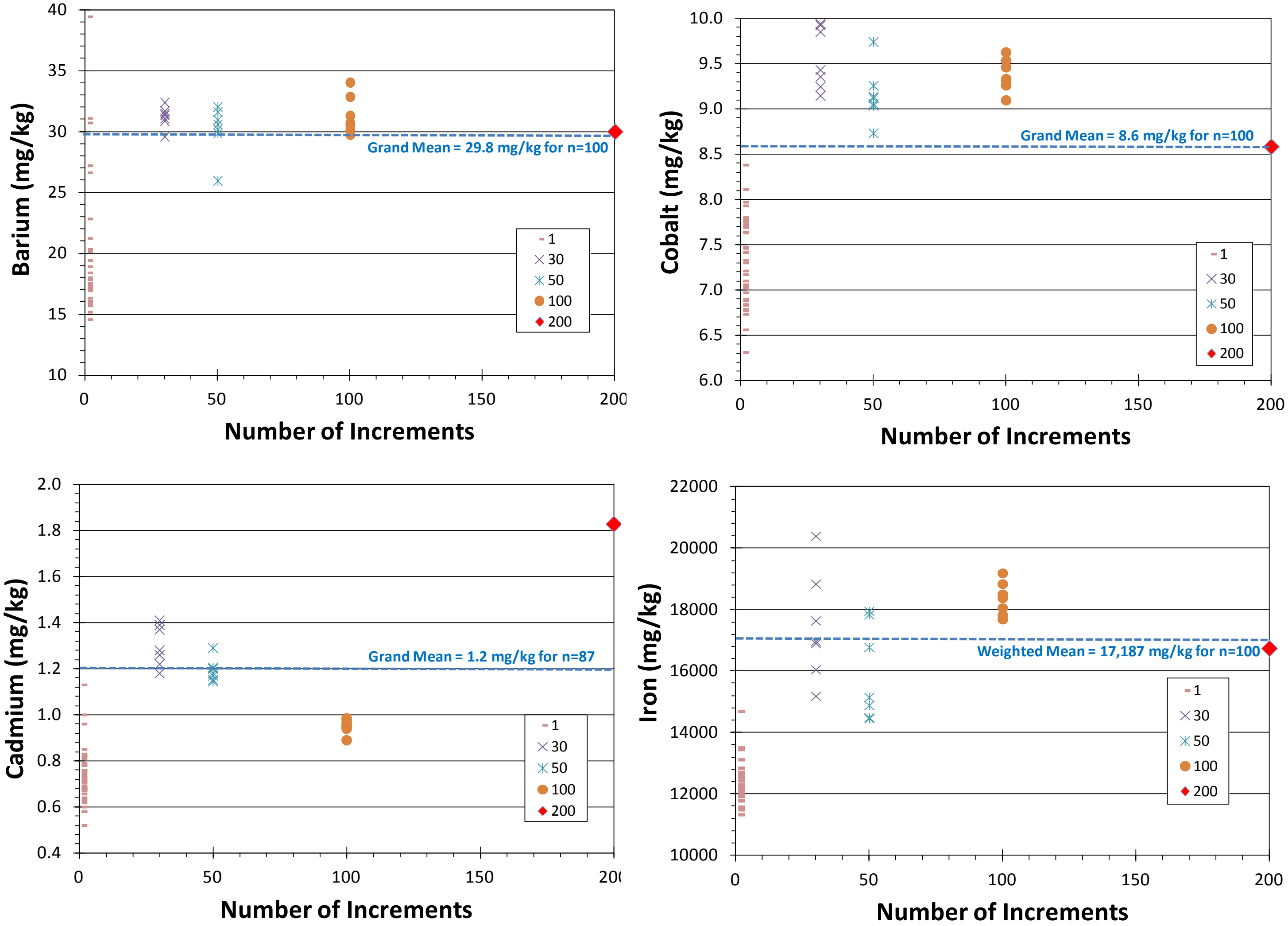


Figure SI10. Distribution of barium, cadmium, cobalt, and iron by number of increments. An increment of one is a grab result, whereas increments of 30, 50, 100, and 200 are incremental sampling methodology results.

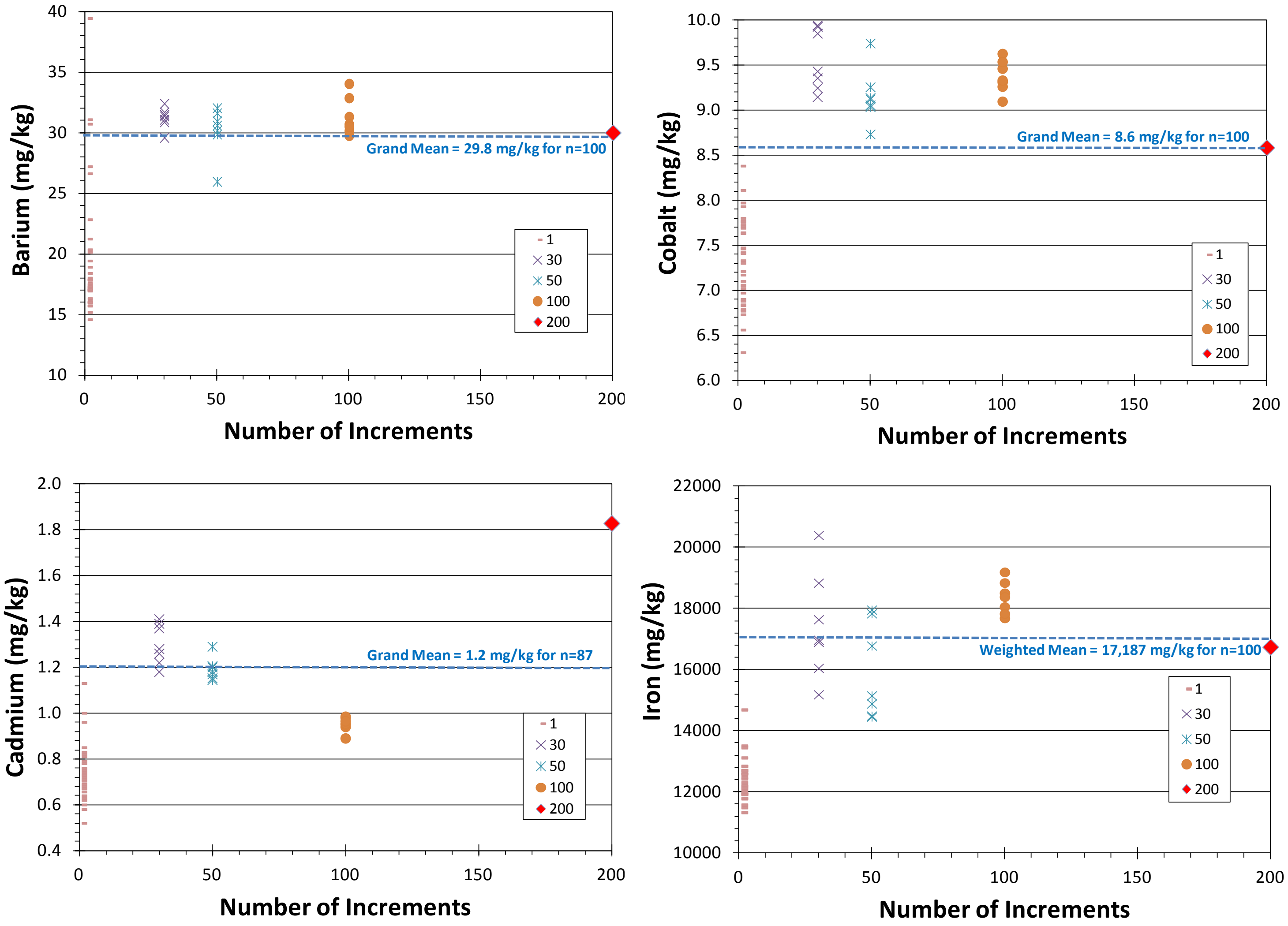


Figure S11. Distribution of manganese, nickel, strontium, and vanadium by number of increments. An increment of one is a grab result, whereas increments of 30, 50, 100, and 200 are incremental sampling methodology results.

Table SI12. Number of lead particles by aliquot mass and concentration for milled and unmilled 1 kg soil samples.

			Number of lead particles by aliquot mass (grams)							
Soil Type	Lead Concentration (mg/kg)	Number of Lead Particles in 1 kg Soil Sample	0.5	1	2	5	10	25	50	100
Unmilled	1	21	0	0	0	0	0	1	1	2
	10	210	0	0	0	1	2	5	11	21
	100	2105	1	2	4	11	21	53	105	210
	400	8419	4	8	17	42	84	210	421	842
	1000	21049	11	21	42	105	210	526	1052	2105
	10000	210485	105	210	421	1052	2105	5262	10524	21049
	100000	2.10E+06	1052	2105	4210	10524	21049	52621	105243	210485
Milled	1	3.99E+08	199571	399143	798286	1995714	3991428	9978570	19957141	39914281
	10	3.99E+09	2.00E+06	3.99E+06	7.98E+06	2.00E+07	3.99E+07	9.98E+07	2.00E+08	3.99E+08
	100	3.99E+10	2.00E+07	3.99E+07	7.98E+07	2.00E+08	3.99E+08	9.98E+08	2.00E+09	3.99E+09
	400	1.60E+11	7.98E+07	1.60E+08	3.19E+08	7.98E+08	1.60E+09	3.99E+09	7.98E+09	1.60E+10
	1000	3.99E+11	2.00E+08	3.99E+08	7.98E+08	2.00E+09	3.99E+09	9.98E+09	2.00E+10	3.99E+10
	10000	3.99E+12	2.00E+09	3.99E+09	7.98E+09	2.00E+10	3.99E+10	9.98E+10	2.00E+11	3.99E+11
	100000	3.99E+13	2.00E+10	3.99E+10	7.98E+10	2.00E+11	3.99E+11	9.98E+11	2.00E+12	3.99E+12

The sample calculations in Table SI12 assume an average particle size of 2-mm and 75 microns for the unmilled and milled soil samples, respectively. The total soil mass (M_{Tot}) assumed for the calculations was 1 kg. The radius of the soil particles was calculated assuming a spherical shape using the following equation;

3} $$r = {}^{d}/_{2}$$

where

r = particle radius (cm), and

d = particle diameter (cm).

The total lead (mg) mass in the 1 kg soil was calculated at variety of lead concentrations (C_{lead}) using units of mg/kg);

4} $M_{Pb} = C_{PB} M_{Tot}$

where

M_{Pb}= total lead mass (mg) in 1 kg of soil,

C_{PB} = concentration of lead (mg.kg), and

M_{Tot} = total soil mass, assumed equal to 1 kg.

The volume of the soil assuming a sphere was calculated using;

5} $V = {}^{4}\!/_{3}\, \pi r^3$

where

V = volume of soil (cm^3), and

r = particle radius (cm).

Using equation 3, the unmilled soil volume assuming a 2-mm particle size is 4.2E-06 cm^3 and the milled volume is 2.2E-13 cm^3. The mass of an individual lead particle was determined using;

6} $M_{Part-Pb} = V\rho_{Pb}$

where

$M_{Part\text{-}lead}$ = mass of lead particle (mg/particle),

V = volume of soil (cm^3), and

P_{lead} = density of lead (g/cm3).

The number of lead particles in 1-kg of soil is calculated as follows;

7} $n_{Pb-Part} = {}^{C_{Pb}}/_{M_{Part-Pb}}$

where

$n_{Pb-Part}$ = number of particles of lead in 1-kg of soil,

C_{PB} = concentration of lead (mg/kg), and

$M_{Part\text{-}lead}$ = mass of lead particle (mg/particle).

Table SI13. Probability of encountering one lead particle by aliquot mass and concentration for milled and unmilled 1 kg soil samples.

		Probability of Encountering 1 Particle by aliquot mass (grams)							
	Lead Concentration (mg/kg)	0.5	1	2	5	10	25	50	100
Unmilled	1	0	0	0	0	0	0.013155	0.052621	0.210485
	10	0	0	0	0.005262	0.021049	0.131553	0.526214	1
	100	0.000526214	0.002105	0.008419	0.052621	0.210485	1	1	1
	400	0.002104855	0.008419	0.033678	0.210485	0.841942	1	1	1
	1000	0.005262137	0.021049	0.084194	0.526214	1	1	1	1
	10000	0.052621367	0.210485	0.841942	1	1	1	1	1
	100000	0.526213667	1	1	1	1	1	1	1
Milled	1	1	1	1	1	1	1	1	1
	10	1	1	1	1	1	1	1	1
	100	1	1	1	1	1	1	1	1
	400	1	1	1	1	1	1	1	1
	1000	1	1	1	1	1	1	1	1
	10000	1	1	1	1	1	1	1	1
	100000	1	1	1	1	1	1	1	1

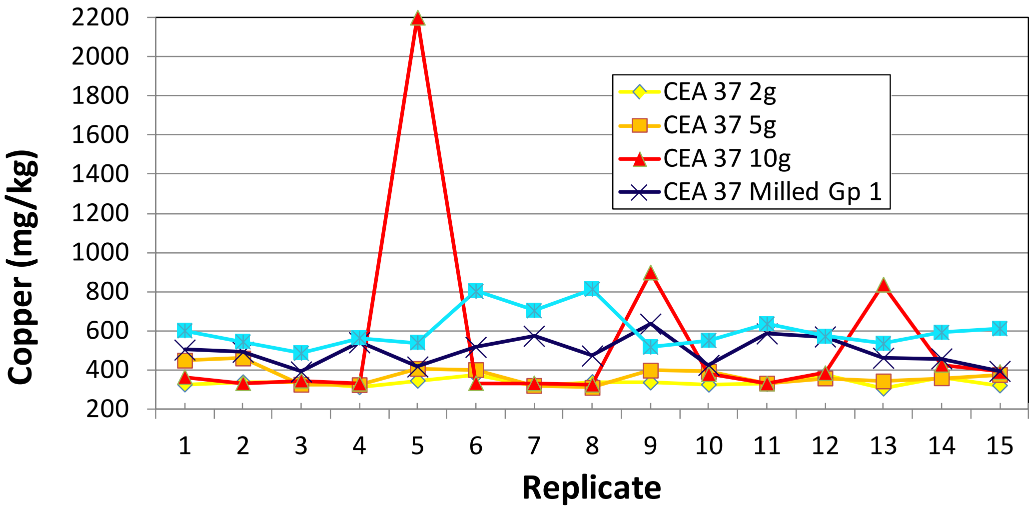


Figure SI14. Replicate sample comparison of milled and unmilled soil copper results by mass of digested material.

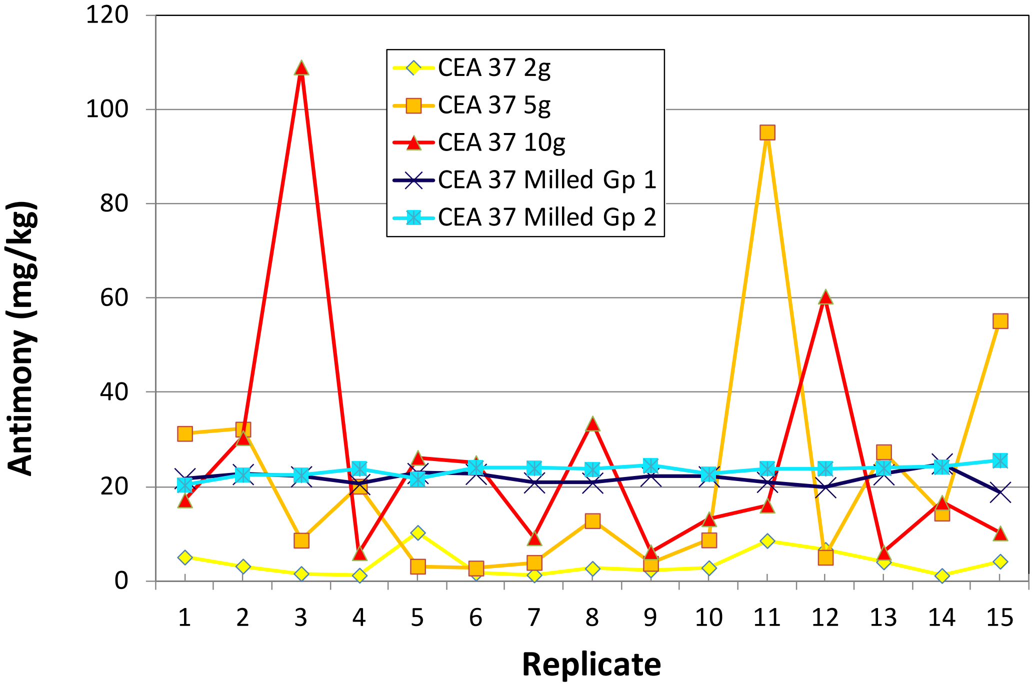


Figure SI15. Replicate sample comparison of milled and unmilled soil antimony results by mass of digested material.

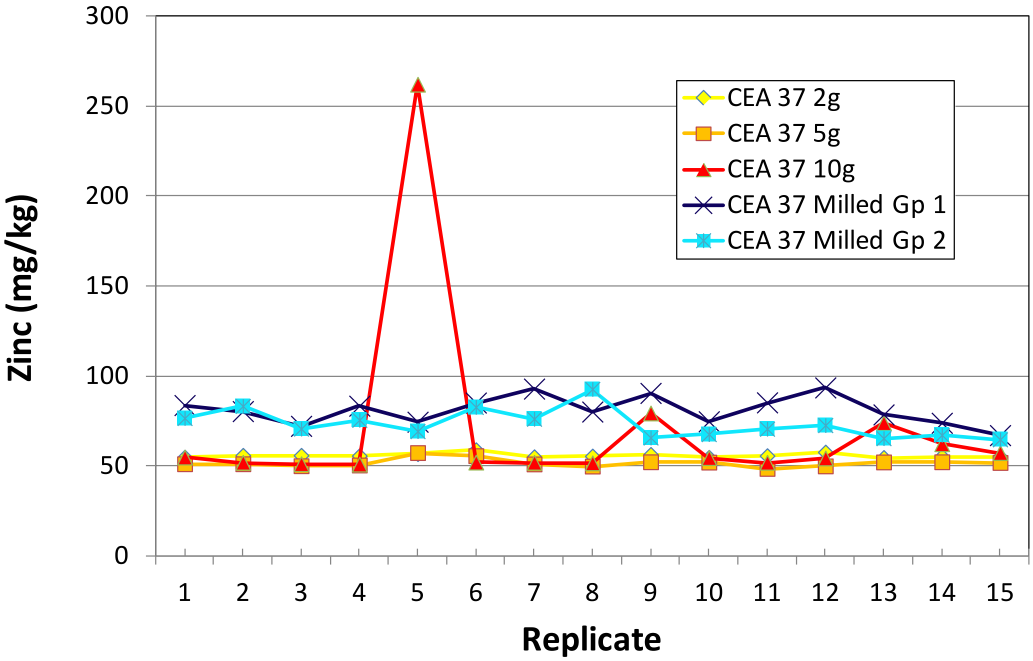


Figure SI16. Replicate sample comparison of milled and unmilled soil zinc results by mass of digested material.

Table SI17. Descriptive statistics of metal concentrations for the n =15 replicates for the milled and unmilled splits.

	Al	Ba	Cd	Co	Cr	Cu	Fe	Mg	Mn	Ni	P	Pb	Sb	Sr	V	Zn
	Unground #1 (mg/kg)															
Mean	3538	11.9	ND	4.37	4.47	1061	9173	NA	179	7.45	NA	2043	15.9	NA	6.19	83.7
Median	3560	11.9	ND	4.33	4.50	314	9210	NA	180	7.40	NA	1600	9.64	NA	6.20	48.7
Min	3230	10.9	ND	4.10	3.91	299	8320	NA	161	6.90	NA	1030	4.42	NA	5.49	46.2
Max	3690	12.7	ND	4.86	4.80	10900	9730	NA	194	9.15	NA	5190	60.9	NA	6.58	574
STD	125	0.560	ND	0.202	0.226	2726	327	NA	7.54	0.530	NA	1251	18.4	NA	0.272	136
RSD (%	4	5	ND	5	5	257	4	NA	4	7	NA	61	116	NA	4	162
	Unground #2 (mg/kg)															
Mean	4124	17.1	1.08	6.48	6.12	357	9307	2018	201	9.09	559	1600	14.2	12.6	NA	66.1
Median	4118	16.4	1.08	6.51	6.14	329	9325	2010	200	9.08	558	1432	10.9	12.7	NA	63.7
Min	3954	15.2	1.04	6.26	5.72	312	9120	1983	197	8.85	548	1004	6.86	11.8	NA	60.0
Max	4266	23.4	1.11	6.73	6.74	676	9455	2055	206	9.65	573	3340	44.3	13.7	NA	107
STD	94.7	2.48	0.017	0.139	0.282	89.5	98.5	24.7	2.56	0.191	8.14	626	9.84	0.546	NA	11.3
RSD (%	2	15	2	2	5	25	1	1	1	2	1	39	69	4	NA	17
	Mortar and Pestle (mg/kg)															
Mean	4173	17.3	ND	6.31	6.26	372	13818	2025	184	8.37	596	1359	10.5	13.5	10.3	71.5
Median	4111	17.4	ND	6.37	6.27	329	14010	2032	185	8.41	598	1156	7.83	13.5	10.4	67.0
Min	3753	15.4	ND	5.86	5.63	278	12665	1921	175	7.85	522	973	6.36	10.7	9.19	61.2
Max	4599	18.7	ND	6.73	6.53	886	14470	2139	194	8.75	644	2371	26.7	16.9	10.7	142
STD	227	0.832	ND	0.231	0.236	145	551	73.8	5.48	0.26	31.9	440	5.74	1.43	0.458	19.9
RSD (%	5	5	ND	4	4	39	4	4	3	3	5	32	55	11	4	28
	Puck Mill # 1 (mg/kg)															
Mean	5509	31.6	1.82	8.75	229	496	16676	2242	235	12.4	590	2760	21.8	20.6	15.2	81.0
Median	5505	31.5	1.82	8.75	230	492	16915	2255	235	12.4	589	2779	22.2	20.6	15.2	80.4
Min	5425	29.4	1.79	8.46	222	391	14775	2141	225	12.2	571	2509	18.9	20.1	14.7	67.0
Max	5615	34.8	1.86	8.94	236	636	17620	2290	240	12.6	610	2936	24.9	21.2	15.6	93.5
STD	58	1.31	0.024	0.11	4.52	74.3	700	41.1	4.00	0.098	10.3	120	1.45	0.323	0.230	7.79
RSD (%	1	4	1	1	2	15	4	2	2	1	2	4	7	2	2	10
	Puck Mill # 2 (mg/kg)															
Mean	6123	28.4	1.84	8.42	217	605	16828	2008	215	11.9	631	2673	23.4	23.4	15.2	73.3
Median	6185	28.6	1.83	8.38	217	572	16925	2009	215	11.9	632	2706	23.8	23.6	15.1	70.6
Min	5740	25.9	1.77	8.30	214	486	15530	1905	205	11.8	615	2440	20.3	21.9	14.7	64.7
Max	6800	30.5	1.95	8.66	221	813	17715	2100	223	12.7	655	2788	25.5	26.3	15.5	92.7
STD	276	1.38	0.044	0.108	1.94	98	543	54.6	5.30	0.228	13.1	104	1.28	1.12	0.228	7.94
RSD (%	5	5	2	1	1	16	3	3	2	2	2	4	5	5	2	11
	Puck Mill # 3 (mg/kg)															
Mean	2930	14.6	ND	4.25	108	279	7702	1520	142	6.46	335	3041	11.2	8.80	5.72	49.9
Median	2934	14.6	ND	4.23	107	271	7693	1503	142	6.39	331	2990	10.5	8.74	5.70	49.2
Min	2727	13.3	ND	3.91	102	242	7317	1448	135	6.10	319	2500	7.93	8.38	5.17	45.7
Max	3251	17.1	ND	4.60	116	356	8325	1668	155	6.98	365	3840	15.9	9.60	6.14	55.3
STD	133	0.895	ND	0.170	4.32	28.5	270	57.9	5.11	0.229	11.7	444	2.33	0.356	0.264	2.50
RSD	5	6	ND	4	4	10	4	4	4	4	3	15	21	4	5	5
	Puck and Ring Mill (mg/kg)															
Mean	5707	40.6	ND	6.92	353	346	14200	2144	221	13.2	384	2349	5.20	NA	14.3	50.5
Median	5725	40.7	ND	6.91	356	344	14100	2145	221	13.2	383	2345	5.22	NA	14.4	50.1
Min	5120	36.4	ND	6.30	317	308	13200	1930	199	12.0	345	2110	4.51	NA	13.1	45.0
Max	6150	43.6	ND	7.46	379	368	15300	2320	238	14.2	413	2550	5.93	NA	15.4	54.9
STD	317	2.28	ND	0.322	17.2	18.8	620	109	10.8	0.603	19.4	116	0.402	NA	0.655	3.00
RSD (%	6	6	ND	5	5	5	4	5	5	5	5	5	8	NA	5	6
	Ball Mill (mg/kg)															
Mean	4983	31.89	ND	6.00	9.37	319	17240	2212	226	9.38	427	1687	3.47	NA	10.81	51.6
Median	4990	31.80	ND	5.98	9.38	317	17200	2210	226	9.41	428	1690	3.46	NA	10.80	51.4
Min	4840	31.10	ND	5.93	9.19	309	17000	2160	223	9.22	409	1640	3.14	NA	10.60	49.9
Max	5060	32.70	ND	6.10	9.54	349	17500	2270	232	9.53	438	1710	4.02	NA	11.00	53.8
STD	55.0	0.375	ND	0.054	0.09	9.53	145	27.8	2.40	0.087	7.27	17.6	0.261	NA	0.144	1.15
RSD (%	1	1	ND	1	1	3	1	1	1	1	2	1	8	NA	1	2

n = 15 for all populations, NA - not analyzed, ND = not detected

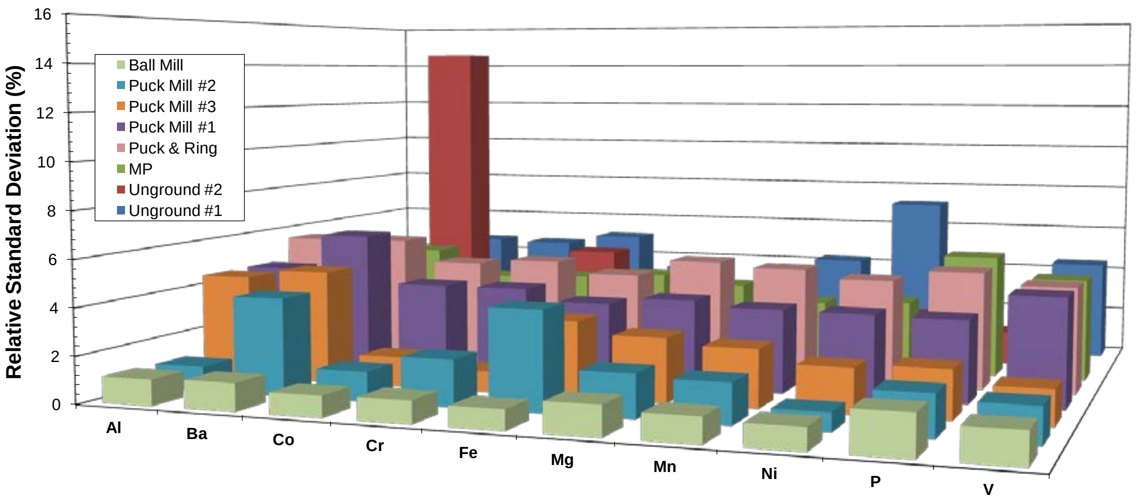


Figure SI18. Percent relative standard deviations by milling device for the non-anthropogenic metals.

Table SI19. Statistical summary of surface area measurements by Brunauer Emmett Teller (BET) by milling device and milling interval for the puck mill.

	Surface Area - Brunauer-Emmett-Teller (m^2/g)							
Process	Mortar and Pestle	Puck Mill					Puck and Ring Mill	Ball Mill
Time	300 sec	30 sec	60 sec	90 sec	120 sec	300 sec	300 sec	20 hr
n	2	6	3	3	2	3	3	2
Mean	0.56	0.28	1.14	1.41	1.65	1.36	2.32	1.75
Median	0.56	0.29	1.14	1.43	1.65	1.37	2.31	1.75
Minimum	0.55	0.001	1.12	1.38	1.64	1.34	2.30	1.70
Maximum	0.56	0.52	1.16	1.43	1.66	1.37	2.35	1.80
STD	0.01	0.22	0.02	0.03	0.02	0.018	0.02	0.07
RSD	1.5	76	1.6	2.0	1.0	1.4	1.0	3.8

RSD - percent relative standard deviation, STD - standard deviation

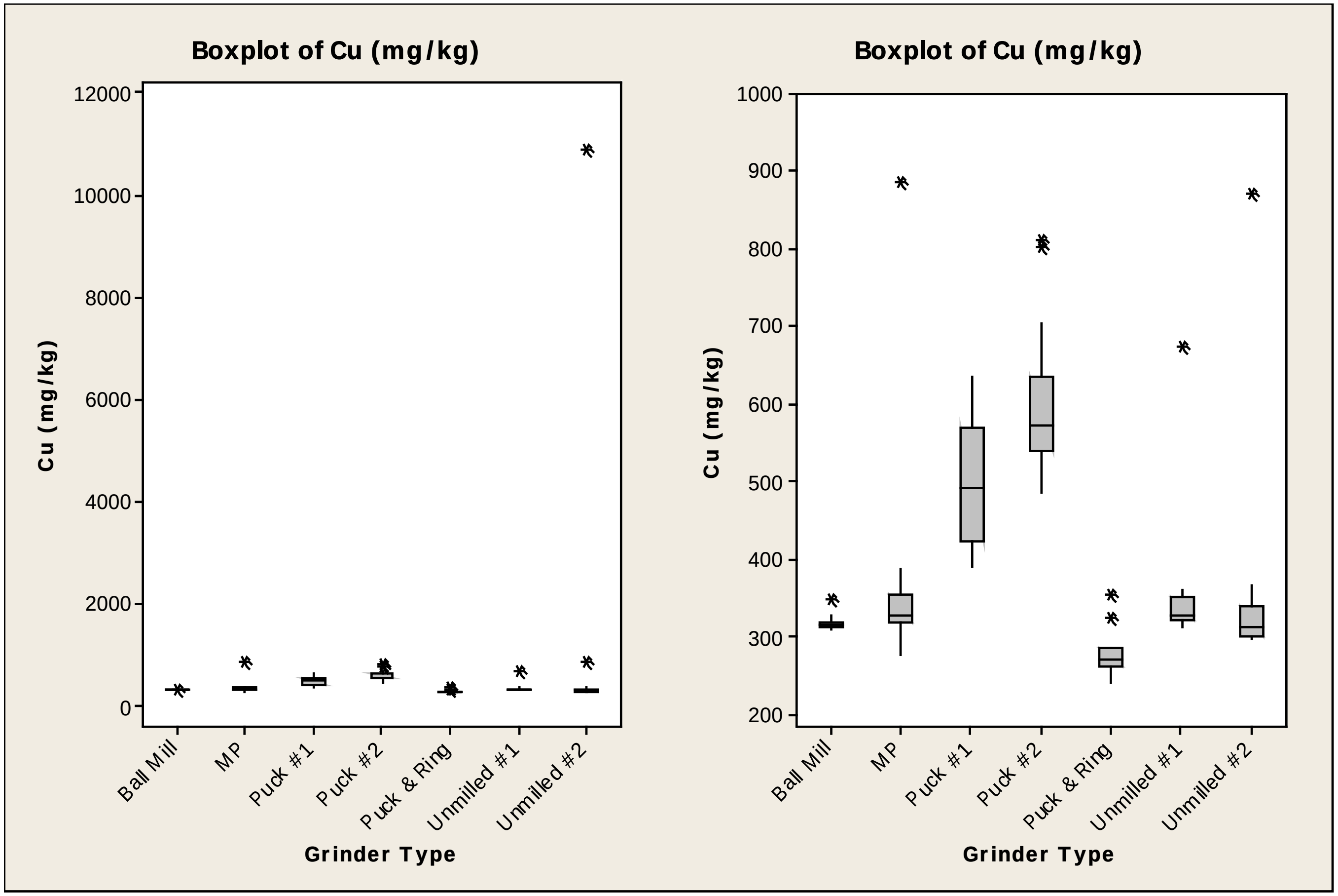


Figure SI20. Box plots for copper by milling device.

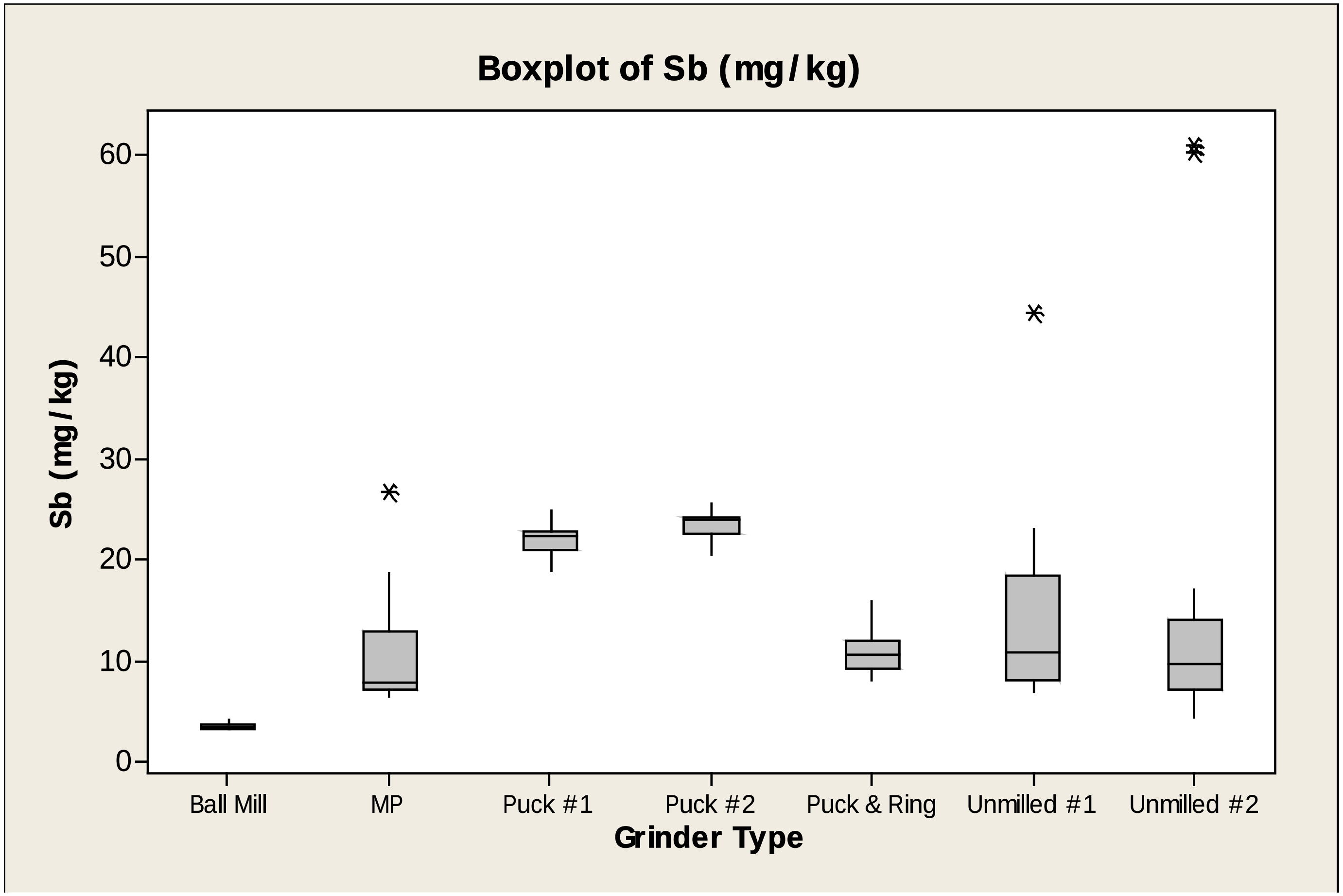


Figure SI21. Box plot for antimony by milling device.

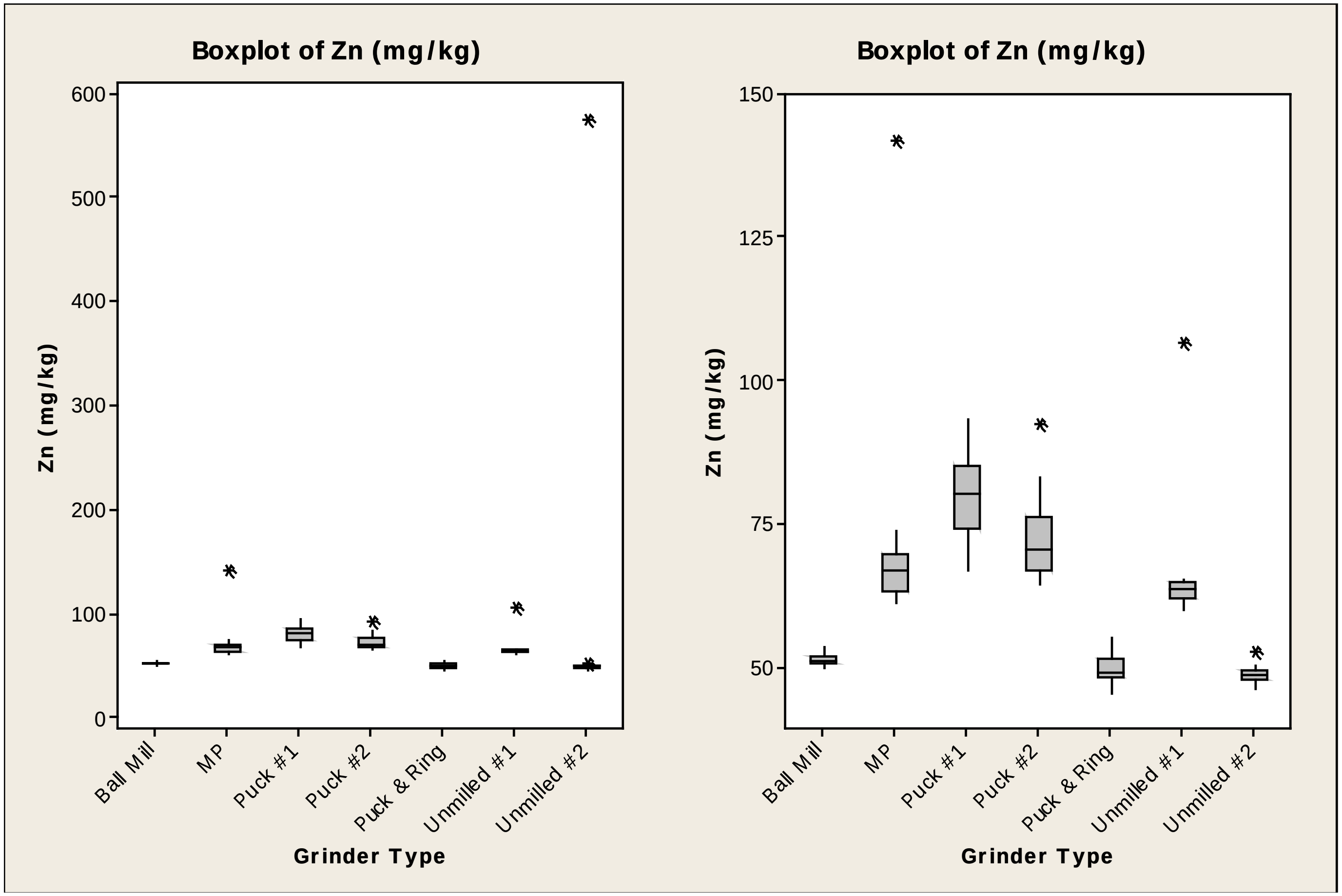


Figure SI22. Box plots for zinc by milling device.

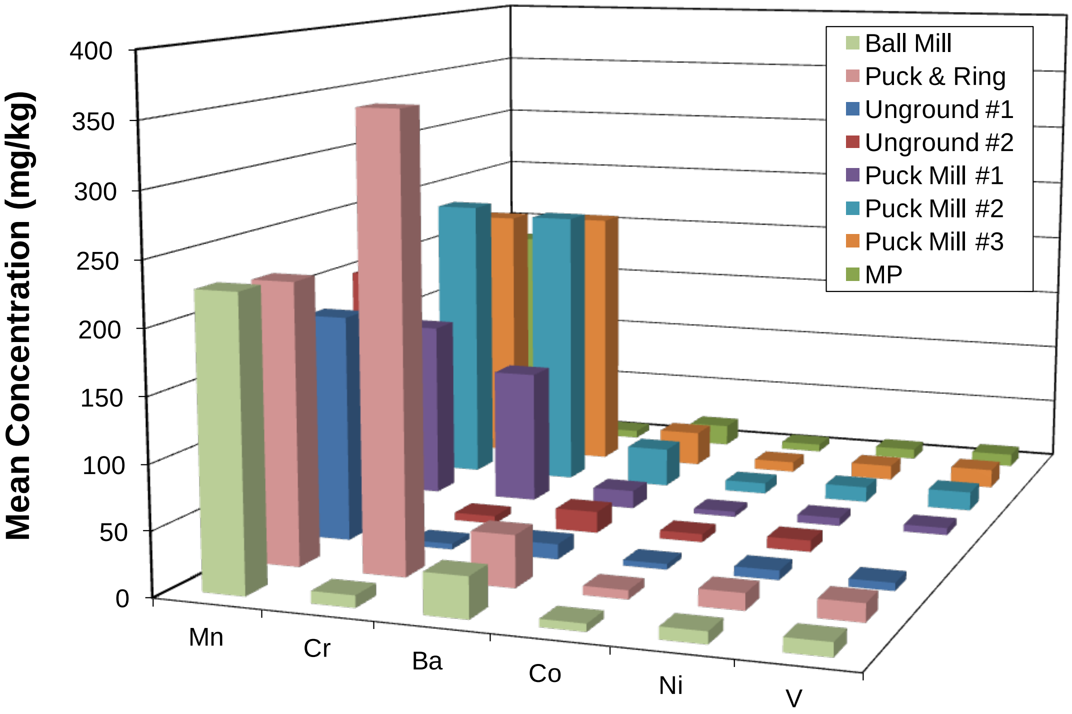


Figure SI23. Percent relative standard deviation by milling interval for non-anthropogenic metals.

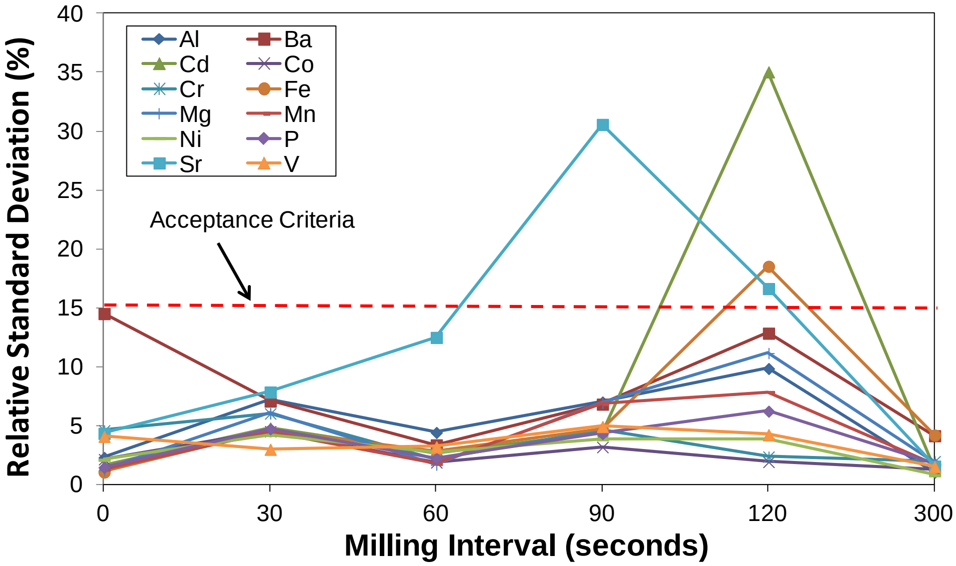


Figure SI24. Percent relative standard deviation by milling interval for the puck mill and non-anthropogenic metals.

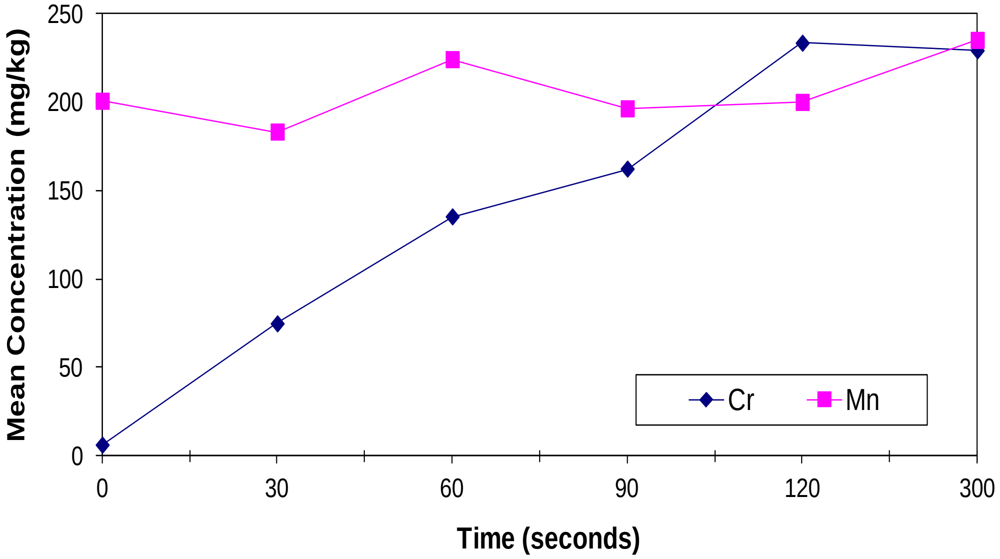


Figure SI25. Mean concentrations of chromium and manganese by milling interval for the puck mill.

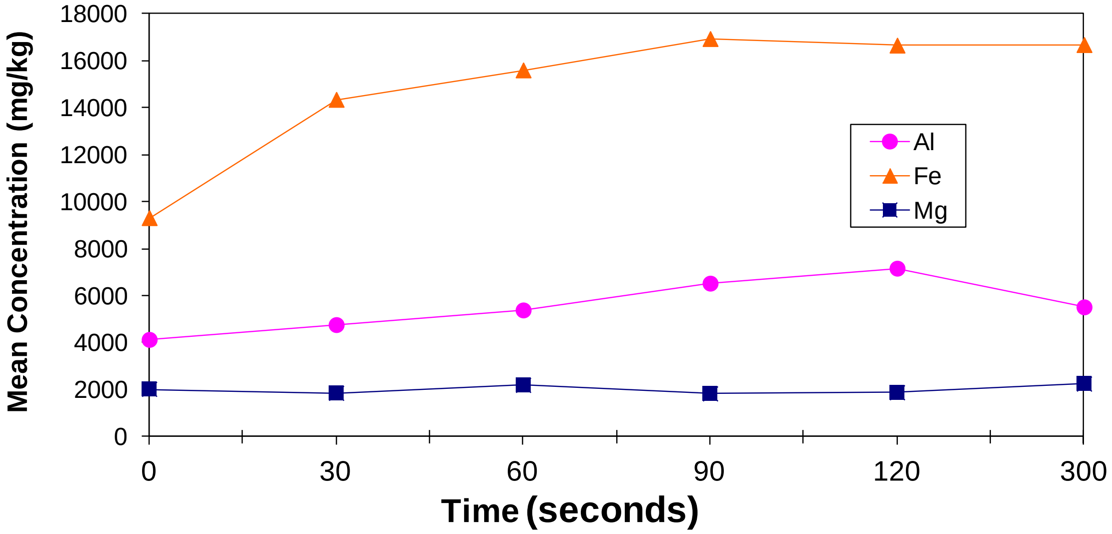


Figure SI26. Mean concentrations of aluminum, chromium, and manganese by milling interval for the puck mill.

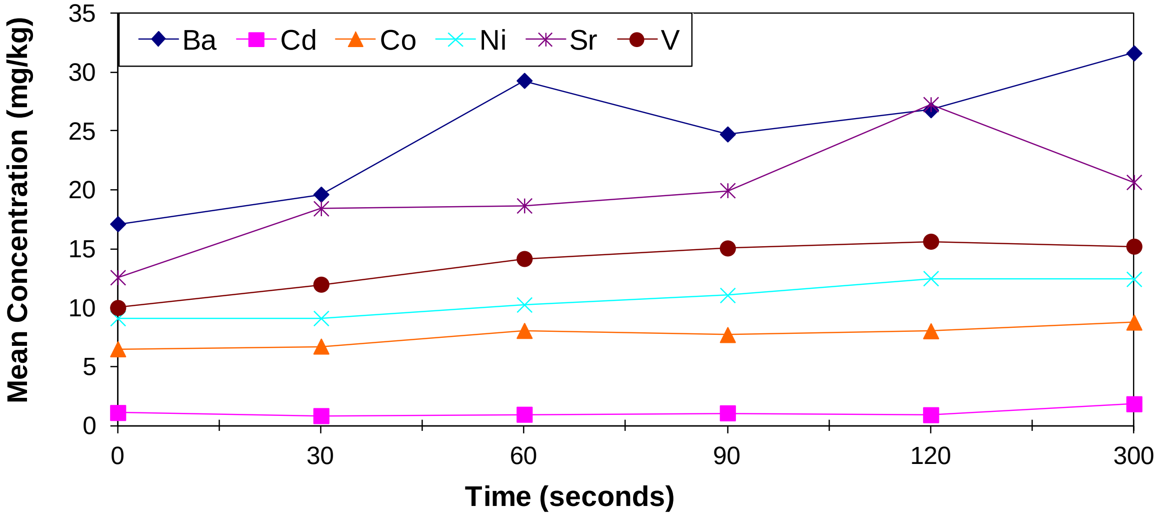


Figure SI27. Mean concentrations of barium, cadmium, cobalt, nickel strontium, and vanadium by milling interval for the puck mill.