DEMONSTRATION REPORT FOR VISUAL SAMPLE PLAN (VSP) VERIFICATION SAMPLING METHODS AT THE NAVY/DRI SITE

MUNITIONS RESPONSE PROJECT
STATISTICAL VERIFICATION AND REMEDIATION SAMPLING METHODS (200837)

August 2011

Pacific Northwest National Laboratory

Brent Pulsipher, John Hathaway,
Kevin Anderson and John Wilson

1. REPORT DATE *(DD-MM-YYYY)*
xx-09-2011

2. REPORT TYPE
Final Demonstration Report

3. DATES COVERED *(From - To)*
01/2010 - 09/2011

4. TITLE AND SUBTITLE
DEMONSTRATION REPORT FOR VISUAL SAMPLE PLAN (VSP) VERIFICATION SAMPLING METHODS AT THE NAVY/DRI SITE

5a. CONTRACT NUMBER
W74RDV10390801

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
D1441D

5d. PROJECT NUMBER
MM-200837

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Pulsipher BA, JE Hathaway, KK Anderson, and JE Wilson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Pacific Northwest National Laboratory
Richland, WA 99352

8. PERFORMING ORGANIZATION REPORT NUMBER
PNNL-20654

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Department of Defense Environmental Security Technology Certification Program

10. SPONSOR/MONITOR'S ACRONYM(S)
ESTCP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The U.S. Department of Defense (DoD) is currently removing, or will be removing, munitions and explosives of concern (MEC) from millions of acres of land and sea. The DoD has established and continues to improve the full process of transferring military land for public use. As a part of this continued improvement and to provide greater confidence that the MEC threat was removed in accordance with established quality assurance/quality control (QA/QC) standards, the DoD incorporated a verification procedure to be employed after all the remediation work is completed. Pacific Northwest National Laboratory (PNNL) has developed a statistical sampling and analysis software, Visual Sample Plan (VSP), which facilitates verification sampling approaches to validate MEC site remediation. This report summaries the objectives and results of a demonstration of these post-remediation verification methods in VSP during an actual remediation verification process.

15. SUBJECT TERMS
Post-Remediation Verification; Verification Transects; Sampling

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Brent Pulsipher
			UU	93	19b. TELEPHONE NUMBER *(Include area code)* 509-375-3989

Reset

Table of Contents

Figures

Tables

Acronyms

AOZ	accept on zero
AVS	attribute verification sampling
bgs	below ground surface
CERCLA	comprehensive environmental response, compensation, and liability act
DGM	Digital geophysical mapping
DoD	U.S. Department of Defense
DQO	data quality objectives
DRI	Denver Research Institute
DMM	discarded military munitions
DTA	DRI test area
EM	electromagnetic
EOD	explosive ordnance disposal
EPA	U.S. Environmental Protection Agency
ESTCP	Environmental Security Technology Certification Program
FLTA	Former Lowry Training Annex
GPO	geophysical prove-out
GPS	global positioning system
HE	high explosive
ID	identification
IVS	instrument verification strip
MC	munitions constituents
MDAS	material documented as safe
MDEH	material documented as an explosive hazard
MEC	munitions and explosives of concern
MR	munitions response
mV	millivolt
NAD	North American datum
PRV	post remediation verification
QA/QC	quality assurance and quality control
RTK	real-time kinematic
SERDP	Strategic Environmental Research and Development Program
TOI	targets of interest
TtEC	Tetra Tech EC, Inc.
UTL	upper tolerance limit
UTM	universal transverse mercator
UXO	unexploded ordnance
VSP	Visual Sample Plan
VTA	vehicle towed array

Acknowledgments

Pacific Northwest National Laboratory acknowledges the Environmental Security Technology Certification Program for their support of this demonstration. Specifically, we acknowledge Herb Nelson and Anne Andrews for their management, technical guidance, and leadership during site meetings. The Navy provided access to the site, their contractors, and pertinent remediation data. Karen Holmes (formerly of the Navy) and Justin Peach were quick to respond to any requests and were strong advocates of this demonstration. Many Tetra Tech EC staff supported this work under the supervision of Michael McGuire. Michael was thorough and responsive throughout this demonstration. Finally, we would like to thank the local participation from the relevant regulators. Ken Vogler and Jeff Swanson from the Colorado Department of Health have been involved since the site was selected for the demonstration and provided insightful review and discussion. David Rathke, U.S. Environmental Protection Agency, Region 8, similarly provided thoughtful review and suggestions.

Executive Summary

The U.S. Department of Defense (DoD) is currently removing, or will be removing, munitions and explosives of concern (MEC) from millions of acres of land and sea. The DoD has established and continues to improve the full process of transferring military land for public use. As a part of this continued improvement and to provide greater confidence that the MEC threat was removed in accordance with established quality assurance/quality control (QA/QC) standards, the DoD incorporated a verification procedure to be employed after all the remediation work is completed.

With support from the Strategic Environmental Research and Development Program (SERDP) and Environmental Security Technology Certification Program (ESTCP), Pacific Northwest National Laboratory (PNNL) has developed a statistical sampling and analysis software, Visual Sample Plan (VSP), which facilitates verification sampling approaches to validate MEC site remediation. This report summaries the objectives and results of a demonstration of these post-remediation verification (PRV) methods in VSP during an actual remediation verification process on the South Tract (~157 acres) of the Navy/Denver Research Institute (DRI) Site located just east of Denver, Colorado.

The primary objectives of this demonstration were to evaluate, illustrate, and determine acceptability of the VSP-PRV sampling methodology as it is applied to two of three possible application scenarios. The first scenario evaluated (Scenario A) was for a recently remediated site where the remediation effectiveness and QA/QC required verification. The second scenario evaluated (Scenario B) was for a site that has undergone remediation sometime in the past but historical records on remediation signal thresholds, QA/QC procedures followed, items discovered, or other related information is vague or unavailable. Because there had been previous, but poorly documented, remediation efforts on this site and the Navy was planning on a new round of remediation, this site allowed an evaluation of both Scenarios A and B. The third scenario, verification on sites where no previous munitions use is suspected, is the subject of a future demonstration report.

Specific performance objectives included the following (section where addressed in parentheses):

1. Clearly illustrate the applicability of verification sampling based on the site history and objectives (Sections 5 and 6)
2. Evaluate the feasibility and cost-effectiveness of a less than 100% survey or sampling verification approach versus a complete re-survey and digging of 100% of anomalies (Section 7.2)
3. Demonstrate the utility and regulator acceptance of the VSP-PRV sampling modules applied to QC and QA for remediated sites (Section 7.3)
4. Provide an example of how to account for signal variation to establish the verification remediation signal threshold upper tolerance limit that defines out of compliance anomalies (Section 6.2)
5. Evaluate the performance, feasibility, and costs of the VSP-PRV transect survey QA/QC approaches verses the VSP-PRV 100% re-survey approaches (Section 7.5)
6. Evaluate the performance of various transect sampling schemes that result from varying the required confidence, the required minimum percentage of transects that must be

proven to be free of any out of compliance targets, and the size of the remediated site (Section 6.3)

7. Evaluate the acceptance of and performance of various transect aggregation schemes (Section 6.3).

In support of the objectives of this demonstration, the Navy performed a 100% geophysical survey on the South Tract and remediated all anomalies above a certain remediation signal threshold. The results of this initial 100% survey were used to demonstrate the performance of the VSP-PRV modules for Scenario B applications. After remediation, we used the transect VSP-PRV methodology to derive a transect design. The geophysical surveys along these transects were then performed, and anomalies that exceeded the remediation signal threshold were dug and examined. The results of this final transect survey were used to demonstrate the performance of the VSP-PRV modules for Scenario A applications.

The key outcomes of this demonstration include the following.

- **Demonstrated Applicability for Recently Remediated Sites (Scenario A):** Based on this demonstration, it is apparent that these verification sampling methods are appropriate for sites that have recently been remediated (Scenario A) to provide confidence that the remediation process was performed appropriately and with adequate QC. The post-remediation verification sampling that was performed at DRI was successful at demonstrating, with 99% confidence, that at least 99.25% of all possible 9 X 200 ft (3 x 61 m) transects do not contain any unexplained out of compliance target (no failures), thereby providing strong evidence of an effective remediation process. Note that the 99%/99.25% parameters mentioned above defined a sample size higher than would typically be warranted for a PRV sampling exercise.

- **Demonstrated Applicability for Historical Site Remediation (Scenario B):** For Scenario B applications (historical remediation with limited records), the PRV methods were also shown to be appropriate if there was good planning and quality control applied during the remediation process. However, one must clearly state what would constitute a failure (out of compliance target) and recognize the implications of out of compliance target definitions, particularly how the probability of detecting true failures (however those failures are defined) is affected by the number of true failures on the site. For this Navy/DRI site demonstration under scenario B, we illustrated that if an out of compliance target was defined as UXO, discarded military munitions (DMM), or material documented as an explosive hazard (MDEH) then using the transect PRV approach likely would not have identified some of the very few failures (8) that existed on the site. However, if the out of compliance targets also included very high mV signal items that should have been detected during previous remediation efforts (54 out of compliance target examples shown), then the transect PRV approach would likely have concluded that the previous site remediation was a failure, thereby requiring additional remediation (probably 100% re-survey and digging, the same as was actually done on this site).

- **Developed Process for Applying VSP-PRV Methodology:** We outlined a detailed flowchart (Figure 19) that outlines the steps required for implementing either the transect or anomaly PRV approaches. This includes specifying Data Quality Objectives, using

Geophysical Prove Out results to help define the tolerable range of measured signals, and root cause analyses of failures.

- **Outlined a Methodology for Failure Definition that Accounts for Signal Variation**: We illustrated how the geophysical survey contractors typically derive a remediation signal threshold, above which defines a target of interest (TOI). Because the signal measurements can vary over multiple measurement readings, this variability must be considered when evaluating whether anomalies with mV readings above the remediation threshold qualify as failures (should have been detected and remediated previously). We derived the methodology to estimate and account for this signal variability by defining an upper tolerance limit (UTL). Any measured anomaly signals on the PRV survey that exceed this UTL would bring into question the effectiveness of the remediation process. For scenario A, the project team would define a failure as any measured anomaly signals that exceed this UTL, baring acceptable findings during a root cause analysis. If no measured anomaly signals exceed the UTL, they would also dig any anomaly that exceeded the original remediation signal thresholds and although not a remediation process failure, any found ordnance or ordnance-related item of explosive hazard would trigger a root cause analysis and a re-evaluation of underlying process criteria, particularly the acceptability of the TOI threshold.

- **Confirmed the Acceptability of VSP-PRV Confidence Calculations and Transect Aggregation Schemes:** Through a series of simulations on varying site configurations, we demonstrated that the VSP-PRV confidence calculations are correct. We varied the as-designed confidence and then estimated the achieved confidence based on the simulations and found that they matched. Similarly, we examined the effect of various levels of transect aggregation and found that such aggregation schemes do not negatively affect the confidence levels achieved. Random transect placement algorithms were also shown to be truly random.

- **Improved the VSP-PRV Module User Interfaces:** Several needed VSP usability improvements were identified and implemented during the course of this demonstration. A complete list is shown in Section 7.7.1.

- **Demonstrated the Cost Effectiveness of Verification Sampling as Opposed to 100% Resurvey of Site**: Decisions regarding the effectiveness of remediation can be made with high confidence through use of the VSP-PRV sampling/surveying approach. These PRV approaches are shown to significantly reduce the cost that may have been incurred if a 100% resurvey procedure were required. Costs associated with implementing these PRV approaches are documented and seem reasonable given the added confidence that can be achieved.

- **Determined that the transect PRV approach is generally preferred to the anomaly PRV sampling approach:** PRV applications on sites that have been previously remediated will often only identify a small number of anomalies. As documented in Section 7.5, this fact, together with the 100% resurvey requirements, makes the anomaly-PRV approach much more cost-prohibitive as compared to the transect-PRV application.

- **Maintained Regulator Involvement and Guidance Throughout the Demonstration**: Although final regulator concurrence is pending, Colorado State and U.S. Environmental Protection Agency regulators were involved in the development of the remediation verification process that was implemented on this DRI site demonstration.

- **Identified Important Follow-Up Work:** Several recommendations for future consideration were derived from this demonstration:

    - Explore applicability of methods for sites where it is believed no previous munitions were used (Scenario C)
    - Add VSP capability to calculate upper tolerance limits on remediation thresholds
    - Modify methods for areas where clutter anomaly density is very high
    - Continue VSP training with updated material based on this report.

Some additional key points regarding these verification methods and VSP tools include the following.

- The most important part of verification in a munitions response project happens during planning and production and is based on process planning and process quality control (QC). If a project did not employ good planning, QC, and documentation during the remediation phase, verification sampling is not recommended since there would be a very high likelihood of failure requiring another 100% survey/digging operation anyway.
- For verification sampling to be used, good planning, quality control, and documentation must be employed during the verification phase. Only then can verification sampling be considered as an option.
- When using the VSP tools, the project team should be fully aware of the meaning of the confidence terms and avoid the temptation to use VSP in a black box fashion.
- Regulators and project managers need to understand the proper application of VSP and be vigilant to ensure that the VSP verification sampling methods and tools are used appropriately.

1 INTRODUCTION

1.1 Background

The U.S. Department of Defense (DoD) is currently removing, or will be removing, munitions of explosive concern (MEC) from millions of acres of land and sea. The DoD has established and continues to improve the full process of transferring military land for public use. As a part of this continued improvement and to provide greater confidence that the MEC threat was removed in accordance with established quality assurance/quality control (QA/QC) standards, the DoD incorporated a verification procedure to be employed after all the remediation work is completed. This verification procedure used the statistical principles developed in MIL-STD-1916, which identified DoD preferred statistical sampling methods for acceptance of product (Department of Defense, 1996). MIL-STD-1916 was the military's implementation of attribute verification sampling as described in Shilling (1982).

During remediation, the portion of the site to be remediated is often "gridded," and the remediation team proceeds from grid to grid digging up any targets of interest and "cleaning" each grid in its entirety before moving to the next grid. The application of MIL-STD-1916 to MEC remediation sites has required some adaptation of the definition of the "product" (the items of the population of interest) as defined within the military standard. An early implementation defined the product to be the individual grids used during the remediation process. However, later approaches have explored using transects or the entire population of anomalies within the site as the product or item of interest and verifying that all anomalies within the selected transects, or that all the selected anomalies from a 100% resurvey, are not out of compliance targets. Regardless, some form of verification sampling is employed to verify that the remediation has been effective. An example is provided by Williams (2003) for a site near Denver, Colorado. He describes the use of process quality control and verification sampling as it is laid out in MIL-STD-1916.

Used within the proper context of verifying the characteristics of items within the population, verification sampling designs are useful in the process of transferring military land for public use. In fact, the statistical process of verification sampling continues to be recommended for many other applications outside of DoD applications. However, in the recent past verification sampling designs were inappropriately used to characterize the spatial extent of MEC within the area of concern. Designs to identify spatial extent of unexploded ordnance (UXO) require a more systematic sampling approach and often require a 100% survey and analysis to provide enough information to make statements/decisions about the spatial extent and remediation plans. The verification tools in the statistical sampling and analysis software, Visual Sample Plan (VSP), are tailored to the attribute verification process that is done following a complete remediation.

With support from the Strategic Environmental Research and Development Program (SERDP) and Environmental Security Technology Certification Program (ESTCP), Pacific Northwest National Laboratory (PNNL) developed VSP, which facilitates verification sampling approaches to validate MEC site remediation. The underlying statistical methods, based on the same methodology as those found in MIL-STD-1916, are being applied to several other domains such as standard item sampling and building remediation (Matzke, et al., 2010). Although other

ESTCP- and SERDP-sponsored VSP methods for transect design, target area identification and delineation, anomaly density estimation, and mapping have been proven through several site demonstrations conducted over the past 5 years, these VSP post-remediation verification (PRV) methods have not yet been demonstrated on actual sites with PNNL's participation. This demonstration report outlines the steps that were taken to demonstrate the PRV methods in VSP during an actual remediation verification process on a real site located just east of Denver, Colorado. The Navy/Denver Research Institute (DRI) Site is within the Former Lowry training annex, and PNNL collaborated with the Navy for this demonstration.

PNNL developed VSP through funding from the Environmental Protection Agency (EPA), U.S. Department of Energy, U.S. Department of Homeland Security, and the DoD. This software provides the user with an easy-to-use visual interface to identify the appropriate number and location of samples to meet required data quality objectives (DQO) in an interactive software environment. With approximately 5,000 users, this free software continues to be used within the industry and the regulatory community as well as by other government agencies.

1.2 Objectives of the Demonstration

The primary objective of the described demonstration is to demonstrate the validity of the VSP-PRV sampling methodology through onsite demonstrations and with simulations while meeting the QA/QC needs of the Navy on this site. Similar verification sampling approaches have been used within the DoD (Sky Research. 2009) but a clear process has not been agreed upon within the DoD. It should be noted that the VSP-PRV approaches do not ensure that no MEC remains. The VSP-PRV approaches result in statistical statements of a specified confidence that at least Y% of all possible transects (or anomalies) do not contain out of compliance targets. An out of compliance target is defined as an item that should have been detected given the chosen remediation signal threshold while accounting for the inherent variation in signal response.

The VSP-PRV methods may be applicable under three different scenarios as presented below. Scenario A would apply if a remediated site where good planning, QA/QC, and documentation were employed and the QA/QC needed to be verified. This should include using the planning concepts from the Unified Federal Policy for Quality Assurance Project Plans (UFP-QAPP), the concepts described in ITRC Technical/Regulatory Guideline Quality Considerations for Munitions Response Projects (UXO-5), the use of physics-based detection capabilities and blind seeding, and the use of similar concepts found in other references that emphasize process control.

Scenario B would apply if a site has undergone previous remediation but that remediation was deemed unsubstantiated (due to a lack of planning and QA/QC documentation) and some additional verification that the previous remediation was acceptable is needed. Finally, Scenario C may apply if sites are presumed to be munitions-free but that assumption needs to be verified. This DRI site demonstration has allowed for an evaluation of the VSP-PRV methods for scenarios A and B. A summary of these scenarios is shown in Table 1.

Table 1. Varied scenarios where the PRV tools in VSP can be applied.

Scenario	Objective of VSP-PRV Application	Definition of TOI (Should be Dug)	Out of Compliance Target Definition & Failure Criteria
A: Recent Remediation; well documented, good QC.	Confirm Adequate Remediation	Anomaly above the pre-specified, pre-remediated signal threshold	Any unexplained anomaly that is significantly above pre-specified remediation signal threshold.
B: Previous Remediation; incomplete records	Confirm Adequate Remediation	Anomaly above the agreed upon signal threshold	Ordnance-related item of explosive hazard or other items with similar features that were clearly missed during the previous remediation.
C: Presumed No Munitions Used	Confirm No Evidence of Munitions Use	Anomaly above the agreed upon signal threshold	Any munitions related item.

Specifically, original objectives of this demonstration that were outlined in the demonstration plan included the following. Nearly all objectives were met but due to unforeseen Navy budget limitations, the objectives for #2 and #5 below were only partially met.

1. Clearly identify the applications of verification sampling based on the site history and objectives. Site history and future use will determine which application and decision criteria are appropriate. We have identified three application scenarios (outlined in Section 3.0), each with different decision/failure criteria that could warrant the use of the VSP-PRV module. This demonstration allowed for an evaluation and demonstration of two of the scenarios.

2. Evaluate the feasibility and cost-effectiveness of a less than 100% survey or sampling verification approach verses a complete re-survey and digging of 100% of anomalies.

3. Demonstrate the utility and regulator acceptance of the VSP-PRV sampling modules applied to QC and QA for remediated sites.

    a. Obtain a baseline case study of the application of the VSP-UXO verification sampling modules that can be used as a basis for discussions with regulators and site managers and for comparisons with future applications using enhanced methods.
    b. Demonstrate that the methodology and software tools are appropriate for post-remediation verification objectives, provide regulators with good assurance of remediation effectiveness, and solicit feedback from regulators on acceptability of approach.

4. Provide an example of how to account for signal variation in the chosen verification signal limit, which defines anomalies, and illustrate how to identify appropriate definitions for out of compliance targets.

5. Evaluate the performance, feasibility, and costs of the VSP-PRV transect survey QA/QC approach verses the VSP-PRV 100% re-survey approach. Provide cost guidelines for each approach and identify when each should be used.

6. Evaluate the performance of various transect sampling schemes that result from varying the required confidence, the required minimum percentage of transects that must be proven to be free of any out of compliance targets, and the size of the remediated site.
    a. Evaluate resulting confidence using VSP sampling routines on irregular shaped sample areas.
    b. Provide guidance about site dimensions (acres > X) where statistical statements based on verification sampling can be used instead of census (100%) sampling to provide cost savings.

7. Evaluate the acceptance and performance of various transect aggregation schemes.
    a. Identify transect aggregation sample selection schemes that perform similar or better than completely random selection.

Some of these performance objectives were accomplished using the actual data from the Navy/DRI Site demonstration, whereas others required a simulation study.

The process of PRV sampling has been documented, verified, and applied using tools within VSP. The interface and application of PRV sampling in VSP has been improved as necessary to accommodate the unique applications to remediated sites. These improvements in the software have been incorporated into the UXO VSP course and the standard VSP courses.

1.3 Regulatory Drivers

All munitions response (MR) projects are required to have a uniform federal policy quality assurance project plan (UFP-QAPP) to establish minimum specifications for data quality activities for all phases and data uses to be aligned with the comprehensive environmental response, compensation, and liability act (CERCLA) (Intergovernmental Data Quality Task Force, 2005). Because QA checks are required to verify that remediation efforts have been effective, most projects employ some form of verification sampling. The specific approach for the type of verification sampling to be used should be documented in the MR project work plan.

Current best practices that emphasize process control must be used. Verification sampling may be employed on a case by case basis to increase stakeholder confidence, but the most important thing is to always have rigorous process quality control (see Section 3.6 in UXO-5).

A statistical PRV sampling approach was introduced as a way for project managers to demonstrate to regulators and stakeholders that an acceptable level of quality has been achieved. One of the original implementations identified that 10% of the remediated area should be subject to verification sampling. However, there was no strong basis for the 10% figure, so MIL-STD-1916 was accepted to incorporate a statistical basis for the amount of survey work required (ITRC Unexploded Ordnance Team, 2008). Although MIL-STD-1916 focuses on verification sampling, it also emphasizes the increasing importance of process control over that of end-of-the pipe verification sampling.

MIL-STD-1916 was developed for verification sampling of manufactured products that are generally subject to the same manufacturing process, have a clear product definition, and lot size. The adaption of MIL-STD-1916 to remediated MR sites has made a standard application across sites difficult due to the problems associated with defining product size (transect dimensions) and the resulting lot size. This demonstration addresses many of the difficulties with the application of MIL-STD-1916 and attempts to establish a standard application across sites using the VSP-PRV modules. With the VSP-PRV modules, the site managers and regulators specify the degree of confidence required and the acceptance criteria for verification, and then VSP will help derive an appropriate post-remediation survey/sampling scheme that achieves that desired confidence.

2 TECHNOLOGY

A statistically based approach for conducting a confirmation survey is to use “accept-on-zero” (AOZ) attribute verification sampling (AVS) (Shilling, 1982,1978; Squeglia, 1994), which has been named PRV sampling within the UXO modules in VSP. This approach involves conducting an inspection or sampling of a random selection of n “units” from among the total number of units (N) for a site. For example, the site may be divided into N non-overlapping 200 x 10 ft (60.96 x 3.05 m) (land areas (transects), n of which are randomly selected and inspected for out of compliance targets. The number n is statistically determined by specifying N and the confidence required that no more than Y% of the 200 x 10 ft (60.96 x 3.05 m) units contain out of compliance targets. The details of this statistical model are described in Section 2.1.

We have developed a sampling module in VSP that allows a user to implement AOZ AVS on sites where the clean-up work is believed to be completed and needs to be statistically verified. This demonstration provided an opportunity to show the appropriate use of this module and to make improvements associated with the unique application of transect PRV sampling to remediated sites. Section 2.2 outlines the current tools for transect PRV sampling in VSP, and Section 2.3 identifies the current developments that were necessary to address some of the limitations of PRV sampling for use on DoD sites.

2.1 Attribute Verification Sampling Model Description

Attribute verification sampling involves sampling discrete units from a finite population of units such that the Hypergeometric Distribution applies (Schilling 1982, pp. 40-42). In the typical statistical literature on this approach, the population, usually called a “lot,” consists of *N* units, *n* of which are selected without replacement (i.e., a given unit can only be selected once) using simple random sampling. These *n* units are then inspected to determine how many are “defective.” If the number of defective units exceeds an acceptance number, U, then the lot is “rejected,” which may trigger increased inspection and an investigation to determine why defectives are present and how to prevent them in the future. It is assumed that once an item is dug, the inspection procedure (not the detection procedure) is infallible; i.e., a dug non-defective unit is never mistaken as a defective unit and vice-versa (i.e., all detected TOI items can be dug, inspected, and properly classified).

Let *D* denote the number of the *N* units that are defective, where *D* is unknown in practice but expected to be small. When AOZ AVS is used, the null hypothesis, H_o, and the alternative hypothesis, H_a, are

$$\begin{aligned} H_o &: D = 0 \\ H_a &: D \geq D_a \end{aligned} \tag{1}$$

where D_a = the specified maximum number of defective units among the *N* units in the population (lot) that could be tolerated without rejecting the lot.

Note that H_o and H_a can also be stated as

$$\begin{aligned} H_o &: P = 0 \\ H_a &: P \geq P_a \end{aligned} \tag{2}$$

where $P =$ percent of the *N* units that are defective, and

$$P_a = 100(D_a / N)\% \tag{3}$$

is the specified maximum percent of *defective* units that can be tolerated.

When AOZ AVS is used, *n* of the *N* units are inspected and H_o is rejected in favor of H_a if one or more of the *n* units are defective. That is, the acceptance number is U = 0. Two types of decision errors can be made: falsely rejecting H_o and falsely accepting H_o. The probabilities of making these two decision errors are denoted by α and β, respectively, where

α = probability of falsely *rejecting* H_o
β = probability of falsely *accepting* H_0 or, conversely, of falsely rejecting H_a.

It is shown in Bowen and Bennett (1987, pp. 886-887) that when in reality D = 0, i.e., when none of the *N* units in the lot are defective, then $\alpha = 0$ and β is given by the following equation:

$$\beta = \frac{\binom{D_a}{0}\binom{N-D_a}{n}}{\binom{N}{n}} = \frac{(N-D_a)!n!(N-n)!}{n!(N-D_a-n)!N!}. \tag{4}$$

Jaech (1973) showed that β can be approximated as

$$\beta \approx \left(1 - \frac{2n}{2N - D_a + 1}\right)^{D_a} \tag{5}$$

The corresponding approximation for *n* is given by (Bowen and Bennett, 1987, page 887) as

$$n \approx 0.5\left(1 - \beta^{1/D_a}\right)\left(2N - D_a + 1\right) \tag{6}$$

In summary, *n* of the *N* units are selected using simple random sampling, dug, and inspected. If none of the *n* units are defective, then we reject H_a with 100(1 - β)% confidence and can state that we are 100(1 - β)% confident that no more than $100(D_a / N)$ % of the *N* units are defective, or equivalently we can state that we are 100(1-β)% confident that at least 100(1-(D_a/*N*))% of the *N* units are not defective. AOZ AVS can be applied to many stages of any process control, but our implementation primarily focuses on the PRV stage and is named such for emphasis.

2.2 Current Implementation in VSP

The attribute sampling methodology has been implemented within the VSP-PRV modules. The computation of *n* (number of units that must be sampled) for PRV sampling can be quickly accomplished using the "Post Remediation Verification Sampling (UXO)" sampling goal option in VSP (Matzke et al., 2007).

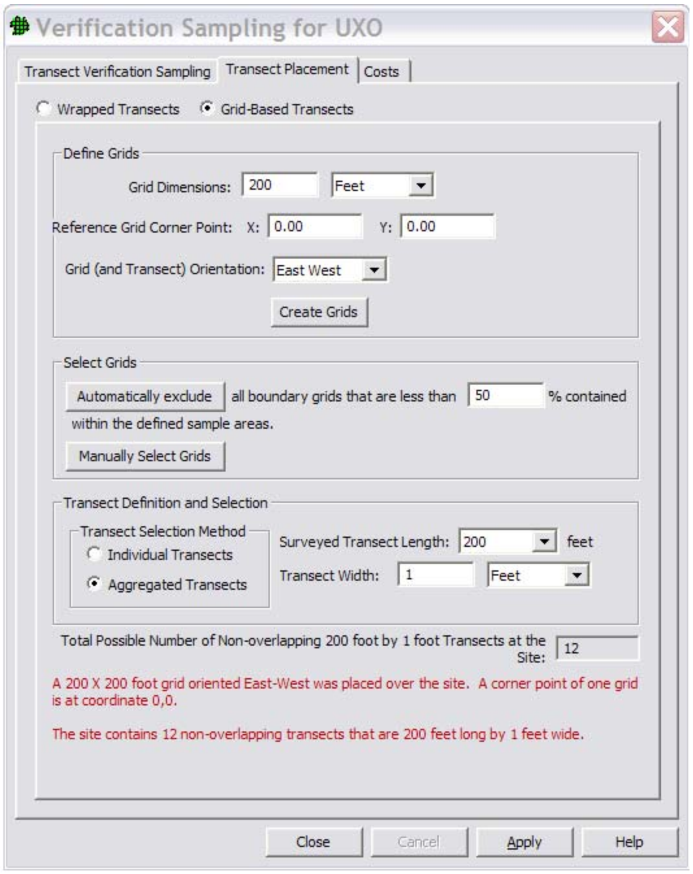
What constitutes the sampling unit in the context of TOI remediation verification could take on various definitions. The sampling unit could be any anomaly that is identified through a 100% survey of the remediated site (anomaly PRV approach). Otherwise, the sampling unit could be any transect of a given width/length of all possible transects of that width/length onsite (transect PRV approach). Finally, the sampling unit might be defined as a single grid area where the entire site is divided or "gridded-up" (grid PRV approach; the UXO Estimator tool can use this approach where the grid size is 1 square acre [.405 hectare]). The VSP-PRV modules provide options for either the anomaly PRV approach or the transect PRV approach. Although VSP can be manipulated to support a grid PRV approach, we do not recommend that approach due to the significant issues that arise relating to inadequate spatial coverage. Additional information on what constitutes a failure or success and the resulting confidence statements that can be made if no failures are identified is presented in Section 3.

2.2.1 Transect PRV Approach

The VSP user interface for the transect PRV approach is shown in Figure 1 and Figure 2, and an example of a finalized transect PRV design is displayed in Figure 3. Users can identify the desired statistical statements they would like to make using the “High Confidence Few Transects Contain UXO” tab shown in Figure 2. This tab shows the user the number of total transects on their site (lot) and then identifies how many transects must be incorporated into the PRV study to meet the design objectives.

During FY2009, we updated the transect PRV sampling to allow the user more flexibility in identifying transect dimensions and sampling routines. Figure 1 shows the new user interface tab that has been included in the PRV sampling dialog. This tab requires the user to identify grid features (dimension and orientation) and the transect dimensions, including transect aggregation.

We anticipate that the grid features will be based on the remediation grids used during the previous stages of the MR (often 200 x 200 ft [60.96 x 60.96 m] grids). However, the PRV transect dimensions for a site can be based on the remediation grids or defined independently. The “Define Grids” region on the “Transect Placement” tab will let the user identify this grid dimensions, which then defines the length of the transects. After grid dimensions and layout have been defined, the boundary grids can be selected or deselected for the final design. The last section on this tab requires the user to input their transect selection method and transect width. The applicability and validity of the transect aggregation option was shown in Hathaway et al. (2009) but is evaluated further during this demonstration.



○ Wrapped Transects ● Grid-Based Transects

Grid Dimensions: 200 Feet

Reference Grid Corner Point: X: 0.00 Y: 0.00

Grid (and Transect) Orientation: East West

all boundary grids that are less than 50 % contained within the defined sample areas.

○ Individual Transects
● Aggregated Transects

Surveyed Transect Length: 200 feet

Transect Width: 1 Feet

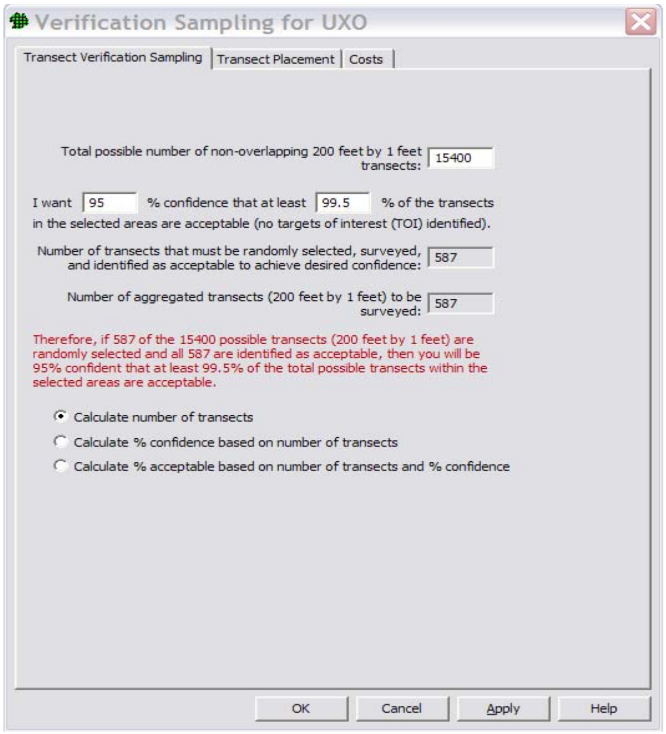
Total Possible Number of Non-overlapping 200 foot by 1 foot Transects at the Site: 12

A 200 X 200 foot grid oriented East-West was placed over the site. A corner point of one grid is at coordinate 0,0.

The site contains 12 non-overlapping transects that are 200 feet long by 1 feet wide.

Figure 1. "Transect Placement" tab within the "Post Remediation Verification Sampling (UXO)" dialog. This tab identifies the transect unit size based on the remediation grids and other user inputs.

After the population of transects on the site has been defined and the transect selection method has been identified, the "Transect Verification Sampling" tab allows the user to identify specific DQO and the resulting survey design requirements. In the example shown in Figure 2, the confidence was 95% and the percent of transects required to be acceptable was 99.5. This resulted in 587 of the 15,400 200 ft X 1 ft (60.96 x 0.305 m) transects being selected and surveyed. VSP allows for the 587 transects to be aggregated into groups; for example, we could randomly select 59 locations and survey 10 aggregated transects at each location (2,000 ft [609.6 m] survey length) if desired. One realization of this selection is shown in Figure 3. Transect aggregation provides a more cost effective, practical approach to transect surveys. Studies on the effect of transect aggregation on detection probabilities and confidence statements were performed previously (Hathaway et. al. 2009) and some aspects of transect aggregation are further evaluated within this demonstration (see performance objective 6.2 below). A process similar to this transect aggregation idea was implemented on other parts of the Former Lowry Bombing Range (Sky Research, 2009).



Total possible number of non-overlapping 200 feet by 1 feet transects: 15400

I want 95 % confidence that at least 99.5 % of the transects in the selected areas are acceptable (no targets of interest (TOI) identified).

Number of transects that must be randomly selected, surveyed, and identified as acceptable to achieve desired confidence: 587

Number of aggregated transects (200 feet by 1 feet) to be surveyed: 587

Therefore, if 587 of the 15400 possible transects (200 feet by 1 feet) are randomly selected and all 587 are identified as acceptable, then you will be 95% confident that at least 99.5% of the total possible transects within the selected areas are acceptable.

● Calculate number of transects

○ Calculate % confidence based on number of transects

○ Calculate % acceptable based on number of transects and % confidence

Figure 2. "Transect Verification Sampling" tab that identifies the desired statistical statements and resulting number of transects that must be sampled.

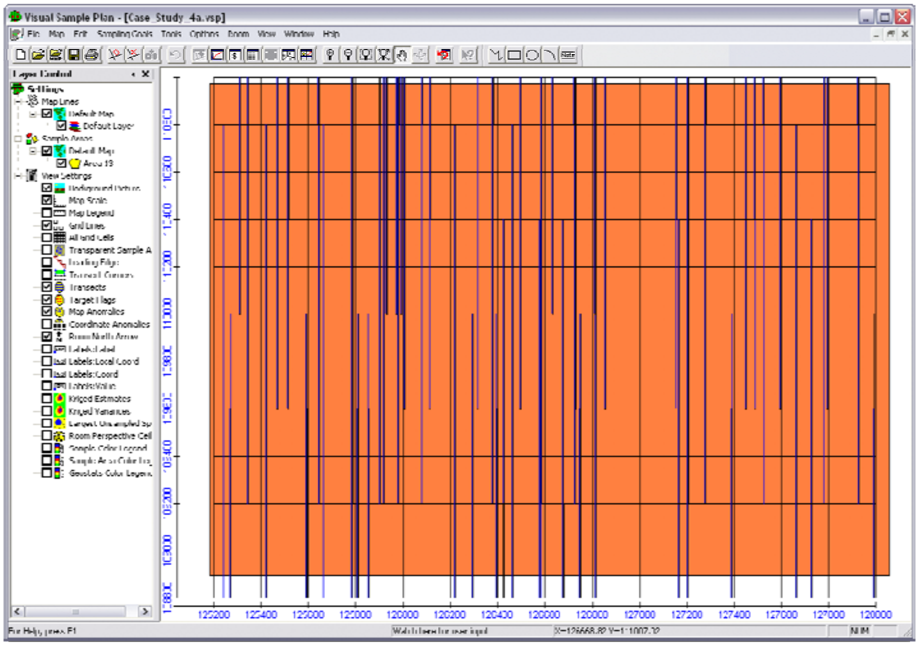
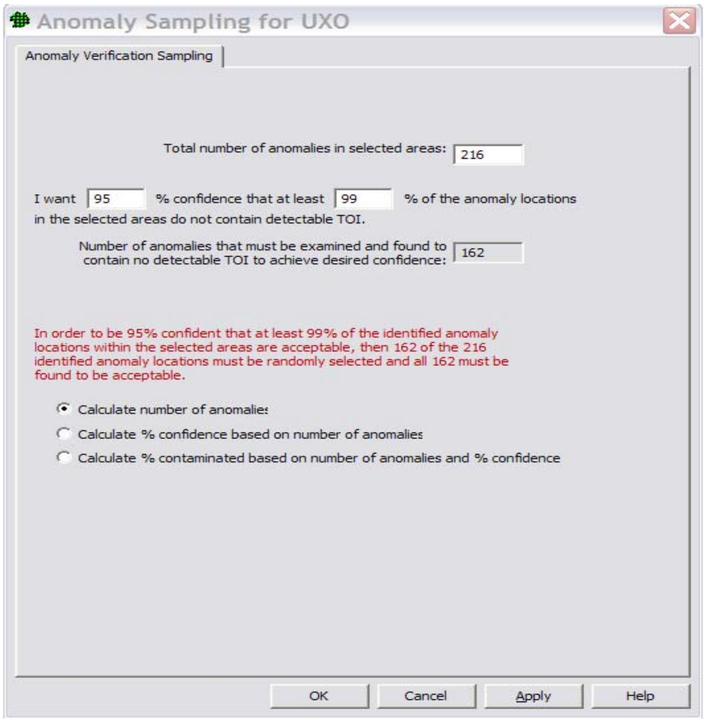


Figure 3. Example site with 59 aggregated survey transects placed on a site with a 656.17 x 656.17 ft (200 x 200 m) remediation grid. The 656.17 x 6.56 ft (200 x 2 m) transects are aggregated in groups of 10.

2.2.2 Anomaly PRV Approach

The anomaly PRV approach in VSP allows the user to take the identified anomalies from a 100% resurvey and select a subset of those anomalies that should be dug and verified to not be targets of interest. If there are anomalies that are deemed “significantly” above the remediation signal threshold, all of those anomalies should be dug and a root cause analysis performed to determine the cause of failure. If no anomalies are deemed “significantly” above the remediation signal thresholds then a subset of the anomalies that are above the remediation signal thresholds should be selected and dug to determine if any ordnance is identified. Figure 4 shows the VSP dialog and an example of one scenario where the total number of anomalies found on the resurvey that were above the remediation signal thresholds (but not “significantly above) was 216. In this case, if we want to state that we are 95% confident that at least 99% of the anomalies are not targets of interest, then 162 of the 216 anomalies need to be randomly selected, dug, and evaluated. As the number of identified anomalies increases, the number of required digs also increases but at a much lower rate (e.g., with 2000 total anomalies, only 277 need to be dug to meet the same 95%/99% objective).



Total number of anomalies in selected areas: 216

I want 95 % confidence that at least 99 % of the anomaly locations in the selected areas do not contain detectable TOI.

Number of anomalies that must be examined and found to contain no detectable TOI to achieve desired confidence: 162

In order to be 95% confident that at least 99% of the identified anomaly locations within the selected areas are acceptable, then 162 of the 216 identified anomaly locations must be randomly selected and all 162 must be found to be acceptable.

● Calculate number of anomalies
○ Calculate % confidence based on number of anomalies
○ Calculate % contaminated based on number of anomalies and % confidence

Figure 4. Anomaly PRV module in VSP with a total number of detected anomalies of 216.

2.3 Advantages and Limitations of the PRV Model

Advantages and limitations exist for both the anomaly PRV approach and the transect PRV approach. The anomaly PRV approach is simple in its random selection of the anomalies that must be dug. However, it does require a 100% re-survey of the site. The advantages and limitations of the anomaly and transect PRV tools as compared to each other mainly relate to the relative number of anomalies that are expected after the remediation. If this number is low, then transect PRV will be the most cost-effective method but an expectation of many anomalies as a result of the re-survey can make anomaly PRV the more cost-efficient approach.

The transect PRV approach also has some limitations based on the application of AOZ AVS. The main limitation is the difficulty of identifying an appropriate transect dimension and resulting "lot size." This demonstration supported an evaluation of this issue. Changing the transect dimension does have some effect on the percent of the site that needs to be resurveyed. Some have suggested that they can "game" the system by selecting a very small transect length and therefore survey only a very small proportion of the site based on the number of transects that this methodology requires. To control for this, we have proposed that the transect dimension should be based on the grid dimensions used when remediating the site. This provides a fairly standardized approach that will limit manipulation of the methodology. Another assumption that is required for the transect PRV approach is that the size of all possible transects are equivalent. Slight departures from this assumption would have little effect but many sites are irregularly shaped such that transects that lie along the edge of the site may not be full-sized. The VSP transect placement options correct for this by completing each partial transect by wrapping onto the adjacent transect lane, as shown in Section 2.2.1. The single-most advantage for any of these PRV approaches is that they require a limited amount of sampling as compared to other attribute sampling routines (Hathaway et al., 2009). MIL-STD-1916 is the most relevant alternative to using the VSP module, and the advantages/disadvantages of the PRV model previously described apply to it as well. However, MIL-STD-1916 is a technical document identifying a sampling method for manufacturing verification sampling within DoD and does not provide a clear method for sampling from spatial sites using transects. MIL-STD-1916 also uses tables with qualitative descriptions for sample size calculations whereas VSP allows the user to pick from quantitative values for their sample sizes associated with their DQO.

3 PERFORMANCE OBJECTIVES

The primary objective of the described demonstration is to demonstrate the validity of the VSP-PRV sampling methodology through onsite demonstrations and with simulations. Similar verification sampling approaches have been used within the DoD (Sky Research, 2009), but a clear process has not been agreed upon within the DoD. Specific objectives include:

1. Clearly identify the applications of verification sampling based on the site history and objectives. Site history and future use will determine which application and decision criteria are appropriate. We have identified three application scenarios, each with different decision/failure criteria that could warrant the use of the VSP-PRV module. Each scenario is defined below and summarized in Table 1.

   Scenario A. A site remediation was completed using recent survey technologies and mapping capabilities with a specific objective to dig all anomalies above a pre-specified remediation signal threshold. In this case, if the VSP-PRV module is used to support QC/QA of the completed remediation, any anomaly identified during re-survey that was significantly above the pre-specified remediation signal threshold used during the original survey, accounting for variation in the signal, would constitute a compliance failure (regardless of whether or not it is UXO).

Scenario B. According to historical records, a site remediation was completed previously but no detailed records are available to ascertain the remediation signal threshold used. Under this scenario, using the VSP-PRV module, a complete or partial transect resurvey would be conducted. A remediation signal threshold is then established that could be based on remediation objectives from other similar areas or by using a remediation signal threshold established for potential remediation activities on this site. Anomalies identified during the resurvey would be dug and compared to a site-specific definition of out of compliance targets. That site-specific out of compliance target may be defined as any unexploded ordnance and, perhaps, any item that the team feels should have been detected and dug during the initial remediation. Any items that were identified as an out of compliance target would constitute a compliance failure.

Scenario C. No site remediation has occurred because the interested parties accept a presumption of no munitions use. Under this scenario, a complete or partial transect resurvey would be conducted using the VSP-PRV module and anomalies identified based on an agreed upon remediation signal threshold would be dug and evaluated. The primary difference between this scenario and scenario “B” is in the definition of an out of compliance target. With the presumption of no munitions use, an out of compliance target would generally be defined as any munitions related item that is identified as part of the survey. Once again, any items that were identified as an out of compliance target would constitute a failure.

Before this demonstration, the southern section of the Navy/DRI Site had undergone previous investigations/remediations but there was varied information about survey equipment, remediation signal thresholds, and survey regions used. There were four different phases of work completed previous to the current remediation. Each phase surveyed different subsets of the south tract and had varied remediation signal thresholds and remediation objectives. In addition the Colorado Department of Public Health and Environment believed that the survey results from the third phase were unreliable (Malcom Pirnie, Inc. ,2008). As outlined in the demonstration plan, the Navy conducted a 100% survey of this southern area for remediation purposes. We use this 100% survey to demonstrate the VSP-PRV modules under Scenario B. In addition, the Navy conducted a QC/QA of their 100% survey and resulting remediation of the southern portion of the site. PNNL provided the verification transect design that allowed for the demonstration of the VSP tools under Scenario A as well.

2. Evaluate the feasibility and cost effectiveness of a less-than 100% survey or sampling verification approach verses a complete re-survey and digging of 100% of anomalies.

3. Demonstrate the utility and regulator acceptance of the VSP-PRV sampling modules applied to QC and QA for remediated sites.
    a. Obtain a baseline case study of the application of the VSP-UXO verification sampling modules that can be used as a basis for discussions with regulators and site managers and for comparisons with future applications using enhanced methods.

    b. Demonstrate that the methodology and software tools are appropriate for post-remediation verification objectives and provide regulators with good assurance of remediation effectiveness.

4. Provide an example of how to account for signal variation in the chosen verification remediation signal threshold that defines anomalies and illustrate how to identify appropriate definitions for out of compliance targets.

5. Evaluate the performance, feasibility, and costs of the VSP-PRV transect survey QA/QC approaches versus the VSP-PRV 100% re-survey approaches. Provide cost guidelines for each approach and identify when each should be used.

6. Evaluate the performance of various transect sampling schemes that result from varying the required confidence, the required minimum percentage of transects that must be proven to be free of any out of compliance targets, and the size of the remediated site.
    a. Evaluate resulting confidence using VSP sampling routines on irregular shaped sample areas.
    b. Provide guidance about site dimensions (acres > X) where statistical statements based on verification sampling can be used instead of census (100%) sampling to provide cost savings.

7. Evaluate the acceptance of and performance of various transect aggregation schemes.
    a. Identify transect aggregation sample selection schemes that perform similar or better than completely random selection.

Some of these performance objectives are accomplished using the Navy/DRI Site demonstration whereas others required a simulation study demonstration. The performance objectives for each of these demonstrations are shown in Table 2.

Table 2. Performance objectives for this demonstration.

Performance Objective	Metric	Data Required	Success Criteria
Navy/DRI Site VSP-PRV Demonstration Performance Objectives			
1. Scenario-based TOI and failure criteria developed	Application of two identified scenarios to the gathered data from the Navy/DRI Site	• Scenario-specific TOI and failure definitions developed. • All identified anomalies to be dug	A clear process for the application of VSP-PRV modules to differing site histories and objectives. Agreed-upon TOI and failure definitions

2. VSP-PRV methods versus 100% re-survey and digging all anomalies	Cost and time comparison between VSP-PRV or complete resurvey/dig	• Survey costs/time • Digging costs/time • Number of anomalies on this site (initial and verification phases) • Range of site sizes and anomaly densities to consider	Conditions identified for when VSP-PRV approach is recommended over 100% re-survey and digging approach
3. Demonstrate utility and regulator acceptance	Applicability and ease of use of VSP-PRV modules and feedback from regulators; VSP-PRV methods easy to use and meet design objectives.	• Full regulator participation and feedback	Positive feedback from regulators with stated intention of use at other sites. Final report documenting applicability of VSP-PRV methods.
4. Illustrate the process of selecting failure criteria accounting for signal threshold variation.	Agreed-upon signal variation adjustment to failure criteria	• Anomaly identification • Anomaly signals • Remediation signal threshold and variation	Agreement on mV threshold upper limit to use with verification sampling that accounts for signal variation
5. Performance, feasibility, and cost comparisons between transect surveys and 100% re-survey/sampling	Cost comparison between the two VSP-PRV sampling approaches	• Range of X%/Y% values to investigate • Survey costs • Dig costs • Analysis costs • Anomalies on this site and range of anomalies to evaluate	Determination of conditions under which one VSP-PRV method is preferred over the other
Performance Objective	**Metric**	**Data Required**	**Success Criteria**
PNNL Statistical Simulation Performance Objectives			
6.1 Demonstrate that desired confidences are achieved using VSP-PRV methods	Desired confidence equals achieved confidences for simulations where site size and anomaly numbers/patterns are varied	• Results from initial 100% survey/digs. • Simulated site data, varying site dimensions, anomaly numbers/patterns, %UXO (see details in Section 3.4)	Desired or statistically designed confidence equals the evaluated confidence from simulations

6.2 Aggregation schemes with similar performance to simple random selection	Resulting aggregations schemes (number of transects surveyed in a row) and associated statistical confidence	• Live site data on which to base simulations • Simulated remediated sites	Achieved confidence from aggregations schemes (number of transects surveyed in a row) same as confidence without aggregation based on calculations from simulation study

3.1 Objectives 1 and 4: Scenario-based TOI and failure criteria developed, accounting for signal variation

The three possible scenarios were presented above, and the definition of the TOI and what would constitute a failure may vary, depending on the scenario. Again, it is important to note that the VSP-PRV methods do not make any explicit statement about one's confidence that no MEC remain, although if none is found and no out of compliance targets are found, then implicitly one's confidence that no MEC remain is increased. The VSP-PRV approaches may allow one to state that they are X% confident that at least Y% of all possible transects (or anomalies) do not contain out of compliance targets, where an out of compliance targets is defined as any item that should have been detected given the chosen remediation signal threshold while accounting for the inherent variation of the signal. Thus, it is important that all parties come to some agreement on what constitutes a TOI and what is considered a failure (would not allow the X%/Y% confidence statement and would cause some further action). We have worked with the Navy and regulators to get agreement on the appropriate scenarios and decision criteria.

3.1.1 Metric, Data Requirements, and Success Criteria

We used this demonstration on the Navy/DRI Site to illustrate the application of scenarios A and B. The southern section of the Navy/DRI Site has undergone previous remediation but there is little detailed information about survey equipment signal thresholds used. A successful outcome is an agreement between the Navy and regulators on the appropriate scenarios and decision criteria for each scenario.

For this demonstration, the Navy contractors first performed a 100% geophysical survey of the southern section of the DRI site. This constituted the "initial survey" and allows us to evaluate the VSP-PRV methods as they relate to Scenario B. All anomalies above their pre-defined remediation signal threshold were dug and identified. The southern area then underwent a geophysical "verification survey," which consisted of a selected set of VSP-PRV-generated transects, and all detected anomalies above the same pre-defined remediation signal threshold were dug. This allows us to demonstrate the VSP-PRV sampling for Scenario A and to evaluate how to best account for the signal variation (standard deviation). We establish the process to identify the signal failure criteria for anomalies found during the transect verification survey.

3.2 Objective 2: VSP-PRV methods verses 100% re-survey and digging all anomalies

QA/QC of the remediation process could be accomplished a number of ways including performing a 100% re-survey and digging all anomalies above some threshold. The statistical VSP-PRV methods were developed to reduce the time, effort, and costs while providing sufficient confidence in remediation effectiveness. However, under some conditions, a 100% re-survey/dig may not significantly increase the cost and time so the VSP-PRV methods may not be warranted. The objective is to perform an evaluation of the costs of each approach on this site and determine whether any general guidelines can be extrapolated to other sites/conditions regarding when the VSP-PRV methods would be recommended over a 100% re-survey/dig.

3.2.1 Metric, Data Requirements, and Success Criteria

The metric is a cost and time comparison between the VSP-PRV methods and a 100% re-survey/dig approach. Data required include the survey cost and time (per linear foot), survey interpretation/anomaly analysis cost/time (average per anomaly), cost/time for digging, and number of anomalies per survey. We use this information from this site to explore how the site size, anomaly density, and costs all affect whether the VSP-PRV methods would be recommended over a 100% re-survey/dig for verification of remediation effectiveness and quality.

3.3 Objective 3: Demonstrate the utility of the PRV sampling modules and gain regulator acceptance

3.3.1 Metric, Data Requirements, and Success Criteria

This performance objective is dependent on the other objectives listed in this section being successful. However, we also see this objective as a measure on the usability of the VSP interface by regulators, stakeholders, and other VSP users. We document the final PRV module design and identify any feedback we received.

The current demonstration has included the EPA Region 8 representatives as well as Colorado Department of Public Health and the Environment regulators. We have met with these regulators and the Navy remediation manager, and they have been very open to use of the VSP-PRV modules for this site. We also anticipate that regulators on the ESTCP review board will provide feedback and insight into the utility of PRV sampling with VSP. The stated acceptance of these tools by those regulators involved and their stated plans for use on other sites will help identify if this objective is a success.

3.4 Objectives 5 and 6.1: Performance, feasibility, and cost comparisons between transect surveys and 100% re-surveys and demonstrate that desired confidences are achieved using VSP-PRV methods

This demonstration primarily focused on the transect PRV sampling, but we were able to evaluate the performance of the anomaly PRV sampling module within VSP. This performance objective identifies the advantages of each and cost considerations that would direct which one to use.

3.4.1 Metric, Data Requirements, and Success Criteria

During the demonstration accurate survey, dig, and analyst cost information were maintained. These costs are used to identify the associated costs for anomaly and transect PRV designs based on a desired confidence. The original plan called for a 100% re-survey that would have allowed us to have a ground-truth data set that we could use to evaluate the performance of the VSP-PRV methods. Because we did not get a 100% re-survey, we addressed this item using a combination of the 100% remediation survey and the limited PRV transect survey (~15%). A comparison between the two was examined to determine whether the performance of either is better. We also demonstrated the cost benefits of the anomaly and transect PRV approaches relative to each other.

3.5 Objective 6.2: Demonstrate that aggregation schemes have similar performance to simple random selection

An initial study of the aggregation routines was done in Hathaway et al., (2009). We have augmented this study based on the updated sampling routine in VSP. Using the 100% initial survey data, we varied the transect aggregation schemes in much the same way as described above and evaluated the achieved verses the as designed confidence

3.5.1 Metric, Data Requirements, and Success Criteria

Using the ground-truth data on this site, we vary the aggregation scheme starting with a basis of a 200 ft x 10 ft (61 x 3 m) transect. This basis transect length is consistent with the actual size of the remediation grid cell of 200 x 200 ft (60.96 x 60.96 m). We vary the aggregation from none to 10 transects (200 up to 2,000 ft [60.96 up to 609.6 m] survey transects) and evaluated the performance of each aggregation level. The North Tract, South Tract, and central portion of the DRI site define different site boundaries and sizes for our evaluation. For each, we use the VSP sampling algorithm to run Monte Carlo simulations to identify which theoretically derived statistical confidence statements are maintained. The main deliverable from this objective is a table of the different site conditions with the associated amount of transect aggregation that can be used. Those aggregation schemes that successfully meet or exceed the designed confidence are recommended.

4 SITE DESCRIPTION

Most of the information within this section was liberally copied from the Preliminary Assessment/Site Inspection (Malcom Pirnie, Inc., 2008). The Navy/DRI Site is 20 miles (32.2 km) southeast of Denver, Colorado and is in the southeast part of the 3,833 Former Lowry Training Annex (FLTA), which is a Formerly Used Defense Site (Figure 5). The 379-acre (153.4 hectares) Navy/DRI is separated into three munitions response sites shown in Figure 6:

- North Tract,~118 acres (47.8 hectares)
- DRI Test Area (DTA),~104 acres (42.1 hectares)

- South Tract, ~157 acres (63.5 hectares)

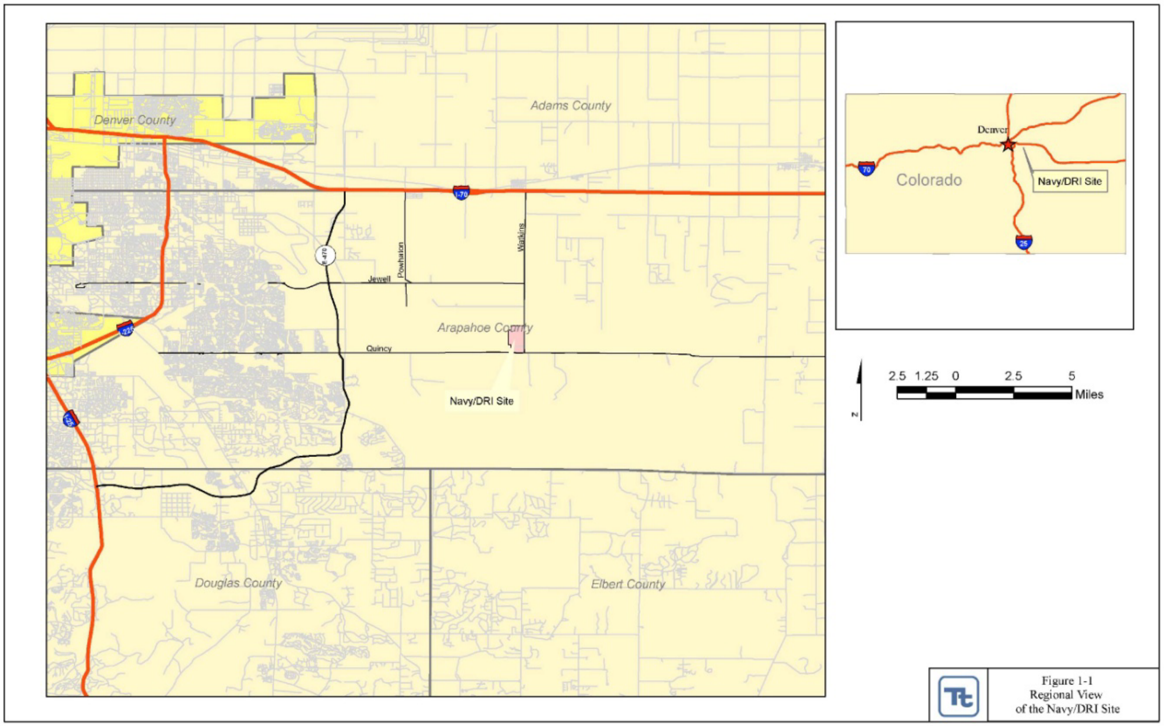


Figure 5. Map of the location of the Navy/DRI Site.

4.1 Site Selection

At the beginning of FY2009, we started looking for an appropriate site on which the PRV tools in VSP could be demonstrated. During our short course provided at the annual SERDP/ESTCP Symposium, we asked those in attendance to talk with us if they thought they had a site that could be used for a VSP demonstration of the PRV modules. A few participants talked with us after the meetings. Karan Holmes, the Navy project lead for the Navy/DRI Site, offered her site as a potential location for our demonstration. This site was selected because it is anticipated to provide a successful application of the verification sampling routines in VSP. The South Tract is generally level and open, which provided the consistent use of a vehicle towed array (VTA) system throughout the area. In addition, the regulators in Colorado have some experience with verification sampling on other parts of FLTA, and the location is convenient for travel and work. The Navy is allowing us to provide input and direction into the verification sampling done after remediation and is allowing for a 100% re-survey to be done on the South Tract of the site.

We participated in an onsite meeting during April 2009, which included the Navy, EPA, and the Colorado Department of Public Health and Environment. During this meeting, we gave a short presentation on the PRV tools in VSP and the goals we had from our involvement. All in attendance were positive about our participation and saw the use of VSP as a valuable piece of the complete remediation process.

4.2 Site History and Munitions Contamination

The Navy/DRI Site consists of three separate areas: the DTA, the South Tract, and the North Tract. The property is located within the southeastern corner of the 3,833-acre (1551.2 hectares) FLTA. The Navy owned and leased this property until 1991, when the entire FLTA was transferred to the Colorado State Land Board. The DRI, an affiliate of the University of Denver, conducted explosives and weapons testing and research at the DTA from 1959 to 1994. The DRI vacated the Navy/DRI Site in December 1995.

Explosives and munitions research and testing were focused within the central areas of the site within the DTA. The DTA consisted of numerous areas designed for testing different munitions and explosives effects. The test areas were located in and along the ravines within DTA. Other test areas were created outside of the ravines by the creation of 20 - 30 ft (6.1 – 9.1 m) manmade soil berms.

In addition, portions of the North Tract, South Tract, and DTA were used as a training facility during World War II as a part of a larger camp training system. The site currently contains no structures and is seasonally used for cattle grazing. Future plans for development of the site and the adjacent property within the FLTA is ongoing, to include development of mixed-use commercial and residential property. For a more detailed description of the entire FLTA and the Navy/DRI Site, see Malcom Pirnie, Inc. (2008). The South Tract is described in more detail below as the demonstration of the VSP-PRV tools were performed in this area.

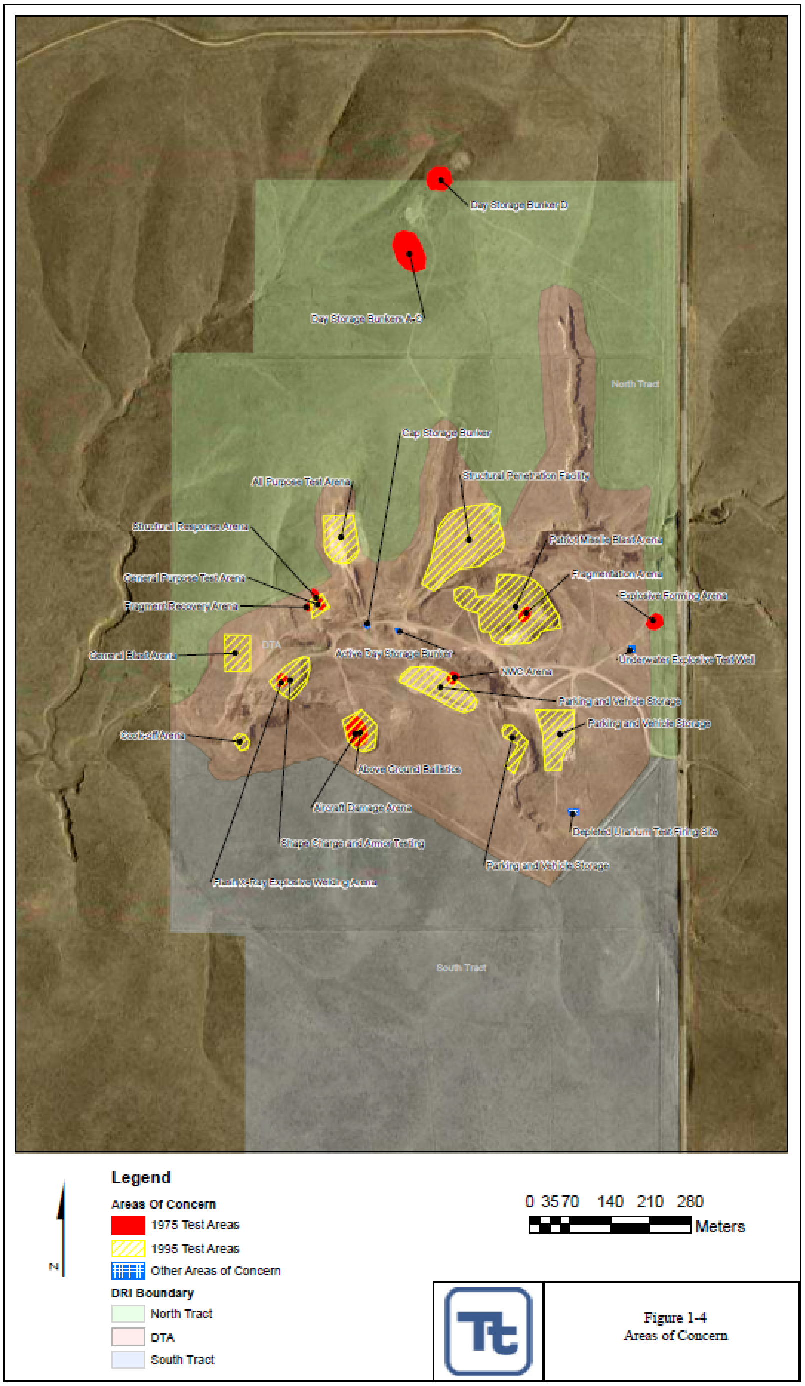


Figure 6. Map of the three areas within the Navy/DRI Site.

4.2.1 South Tract History and Munitions Contamination

The South Tract (Figure 6) represents the portion of the Navy/DRI Site extending south of the previously defined extent of the DTA. The South Tract consists of approximately 157 acres (63.5 hectares) and is bounded on the east by Watkins Road, the south by Quincy Avenue, and the west by the FLTA Eastern Extension Area. The area includes sections of the Navy/DRI Site referred to in past investigations as the Southwest Corner, South Gallery, Southwest Corner, and East Bottom areas, which are summarized in Malcom Pirnie, Inc. (2008). Table 3 summarizes most of the conceptual site information provided in Malcom Pirnie, Inc. (2008) and can be used for a more complete understading of the South Tract.

Explosives testing activities at the Navy/DRI Site were conducted within the DTA. However, fragmentation from explosives and weapons testing activities may have been deposited across the South Tract, particularly in areas directly south of the southwest corner of the DTA. While the majority of the explosives used at the DTA were bulk explosive charges, which typically result in high-order detonations, other munitions items such as MK-118 submunitions, CROW warheads, and 20 and 30 mm projectiles may have been deposited on the South Tract. Munitions associated with Camp 2 (See Range/Site History in Table 3) may also be present at the South Tract and may include small arms ammunition (.22 cal, .30 cal, and .50 cal), dynamite, and hand grenades. The exact locations of the component ranges comprising Camp 2 cannot be verified, and there is limited potential for extensive residual small arms ammunition and/or munitions to remain at the South Tract. Many aircraft-associated projectiles were recovered from the South Tract during the Phase 1 MEC investigations. It is likely that munitions originating from the Rocket and Gunnery Range south of the Navy/DRI Site are the source of these projectiles. Munitions associated with the training activities at the Rocket and Gunnery Range, which was used for gunnery, bombing, and rocketry, may include small arms ammunition (.50 cal), 20 mm projectiles, rockets (2.25-, 2.75-inch), and several varieties of bombs (practice, high-explosive, fragmentation, incendiary, and photoflash) ranging from 3 pounds to 100 pounds. Many of these munitions items were handled at the Air Force explosive ordnance disposal (EOD) Range for destruction; discoveries of many of these items in the FLTA Extension Area may be attributable the use of either the Air Force EOD Range or the Rocket and Gunnery Range.

MEC consisting of a photoflash cartridge, 30 mm high explosive (HE) projectile, HE-filled burster tube, and a MK-118 Rockeye submunition have been identified and removed from the South Tract. Because many targets meeting the 10-mV selection criteria were not investigated during Phase 3 of the MEC investigation within the South Tract, the potential exists for MEC to be present there.

Table 3. Conceptual Site Model Information Profiles for the South Tract of the Navy/DRI Site.

Profile Type	Information Needs	Preliminary Assessment Findings
Range/Site Profile	Installation	Navy/DRI Site
	Installation Location	The Navy/DRI Site is located in Arapahoe County, Colorado, approximately 8 miles (12.9 km) east of Aurora.
	Range/Site Name	South Tract
	Range/Site Location	The site is located in the southern portion of the Navy/DRI Site.

Profile Type	Information Needs	Preliminary Assessment Findings
	Range/Site History	Camp 2, a former World War II-era training area, was established within the bounds of the Navy/DRI Site and the FLTA in 1943. The area was considered part of the Buckley Field Bombing and Gunnery Range (currently referred to as the FLBGR). Following World War II, the site remained unused until 1959. The Navy/DRI Site was owned and leased by the Navy starting in 1959. It occupied the east side of the ~3,800-acre (1537.8 hectares) FLTA. DRI first conducted explosives research on the DTA (within the Navy/DRI Site boundaries) in 1959. The DRI continued to conduct explosive research at the site under contract to the Navy, private companies, local law enforcement agencies, and other government agencies through 1995. Operations at the DTA ceased in 1995 as a result of the land swap agreement between the DoD and the State of Colorado.
	Range/Site Area and Layout	The South Tract is approximately 157 acres (63.5 hectares) and is located in the southern portion of the Navy/DRI Site.
	Range/Site Structures	An industrial complex is located approximately 0.25 miles (0.4 km) east of the South Tract across Watkins Road, and several residences are located approximately 1.3 miles north of the site.
Munitions/ Release Profile	Range/Site Boundaries	N: DTA and North Tract S: Quincy Avenue W: FLTA Eastern Extension Area; former Air Force EOD Range E: Watkins Road; industrial complex
	Munitions Types	Since this was an explosives research facility, primarily shape charges were tested at the site, but the following munitions were also tested: •MK-118 Rockeye submunitions •Patriot missile warheads •Squibs

Profile Type	Information Needs	Preliminary Assessment Findings
		Other munitions associated with the Former Camp 2 and adjacent rocket and EOD ranges may include: •Small arms ammunition (up to .50 cal) •Hand grenades •20 mm projectiles •Bombs •Rockets
	Maximum Probability Penetration Depth	Based on the activities conducted in the vicinity of the South Tract, it is anticipated that munitions would not penetrate the surface. Fragmentation of metallic test targets and munitions from the DTA would likely be dispersed across the surface of the site. Small arms ammunition associated with the former Camp 2 training activities and the Rocket and Gunnery Range would likely have minimal penetration (about 6 in [15.24 cm] bgs). Rockets and bombs fired in association with the Rocket and Gunnery Range would have greater penetration depths (e.g., several feet), depending on the height of release and angle of entry into the surface. The maximum frost penetration depth in this portion of Colorado is reported to be 36 inches (91.4 cm). Given the likely depth of penetration of the munitions types at the site, resurfacing of MEC and munitions debris may occur.
	MEC Density	MEC has not been encountered or confirmed within the specific boundaries of the South Tract. Prior to this demonstration, four MEC items have been recovered across the entire Navy/DRI Site.
	Munitions Debris	Metallic fragmentation has been observed across the site during past investigations, though much has been removed. Other munitions debris consisting of 20 and 30 mm projectiles, 0.50 cal ammunition, grenades, and other munitions items have been recovered.

Profile Type	Information Needs	Preliminary Assessment Findings
	Migration Routes/Release Mechanisms	Erosion and bioaccumulation are potential natural release mechanisms and migration routes. Human interaction that act as potential migration routes or release mechanisms include: construction, excavation, plowing or tilling, and surface soil redistribution. Resurfacing of munitions due to frost heave may potentially occur, given the anticipated shallow depths of munitions and the maximum frost penetration for the region (up to 36 inches [91.4 cm]).
Physical Profile	Climate	The climate is generally mild and semi-arid with large variations in daily temperature. Temperatures range from summer averages in the 90s (30s Celsius) to below freezing in the winter. The average annual precipitation is approximately 17 inches/year (43.18 cm). Thunderstorms are fairly frequent in spring and summer, though winter and early spring see the most precipitation. Blowing dust can develop during an abnormally dry season.
	Topography	Two hilltops are located on the central and southern portions of the South Tract. The “gently rolling” hills on the northern portion of the South Tract are vaguely oriented to the east and west. The hills on the northern portion of the site slope to the north and the northwest while the hills on the central and southern portions of the site slope to the north, west, and south, peaking at the hilltops. The highest point at the South Tract is approximately 5,940 feet (1,810 m) above mean sea level while the lowest point is approximately 5,800 feet (1,768 m) above mean sea level.
	Geology	The site is located in the Colorado Piedmont section of the Great Plains Province within the Denver Basin. The Denver Basin consists of Quaternary alluvium and Upper Cretaceous formations. The Cretaceous formations include (from youngest to oldest): the Dawson Arkose, Denver Formation, Arapahoe Formation, Laramie

Profile Type	Information Needs	Preliminary Assessment Findings
		Formation, and Fox Hills Sandstone.
	Soil	The soils present on the South Tract range in composition from clay loams to loamy coarse sand. The northern portion of the site consists of Renohill-Litle-Thedalund complex, the southern portion of the site consists of Nunn-Bresser-Ascalon complex, the western portion of the site consists of Bresser-Truckton sandy loams, and the eastern portion of the site consists of Renohill loam. The soils in the northern portion of the site consist mostly of loamy clay, which is considered well-drained with a moderately low to a moderately high water transmissivity. The soils in the southern portion of the site consist of sandy loam and sandy alluvium, which are considered well-drained with moderately low to high water transmissivities.
	Hydrogeology	The South Tract is located within the Denver Basin aquifer system. The Denver Basin holds five major confined heterogeneous aquifers, including (from youngest to oldest): the Dawson aquifer, Denver aquifer, Arapahoe aquifer, and the Laramie-Fox Hills aquifer.
	Hydrology	Various drainage canals are located on the North Tract that drain to the south into the unnamed stream located on the DTA. This unnamed stream is the primary surface water feature near the site. Surface water within this tributary flows to the west. Various smaller tributaries merge with this primary drainage feature from the north and south. The unnamed stream flows west and discharges to Coal Creek, which cuts across the southwestern corner of the FLTA. Coal Creek eventually discharges to the South Platte River. No permanent surface water bodies are located on the South Tract. The nearest permanent surface water body is the Aurora Reservoir, located approximately three miles southwest of the South Tract.
	Vegetation	Vegetation on the South Tract is composed of grazed foothill and Piedmont grassland

Profile Type	Information Needs	Preliminary Assessment Findings
		functioning as short grass prairie. Trees on the South Tract are scarce, as they are mainly located in the ravine areas where surface water collects during rain events.
Land Use and Exposure Profile	Current Land Use	The South Tract is currently owned by the Colorado State Land Board. It has not been developed and is not currently in use; however, portions of the FLTA immediately west of the site are being used by ranchers as grazing land for their cattle.
	Current Human Receptors	Potential receptors include state employees, trespassers, and contractors. Contactors may come into contact with MEC and MC when performing intrusive activities.
	Current Activities (frequency, nature of activity)	Cattle currently graze at the South Tract.
	Potential Future Land Use	There are currently plans to develop the entire FLTA, to include the Navy/DRI Site and the North Tract. The type of development that will occur at the site has not yet been determined due to the ordnance clearance work that is yet to be completed; however, it is likely that the development will be similar to that of the rest of the FLTA, consisting mainly of mixed-use residential and commercial areas.
	Potential Future Human Receptors	Future receptors include residents that will live in the newly developed residential areas and their visitors.
	Potential Future Land Use-Related Activities:	Potential activities that may take place at the South Tract include construction, outdoor recreation, and other activities associated with day-to-day living in residential areas.
	Zoning/Land Use Restrictions	There are no known zoning or land use restrictions. These may be evaluated and/or required prior to development of the area, based on the completion of munitions investigations and clearances at the Navy/DRI Site.
	Demographics /Zoning	The South Tract is located in Arapahoe County. Based on 2006 data from the U.S. Census Bureau, Arapahoe County has a population of 537,197 with an overall population density of 608 people per square mile (people/mi^2). However, the population density in the vicinity of the DTA is

Profile Type	Information Needs	Preliminary Assessment Findings
		approximately 12 people/mi^2. Population density is expected to greatly increase following development of the site.
	Beneficial Resources	The South Tract consists of Piedmont grasslands that provide a suitable ecosystem for various mammals, birds, reptiles, and amphibians that inhabit the area. The FLBGR, located south of the Navy/DRI Site, is a planned natural resource conservation area for natural habitat.
Ecological Profile	Habitat Type	The natural habitat is described as semi-arid grasslands. Trees are mostly found in ravine areas where surface water collects during rain events.
	Degree of Disturbance	The current degree of disturbance at the site is low, as the site has not been developed and is not currently in use. However, once development of the site begins, there will be a high degree of disturbance.

5 TEST DESIGN

The performance objectives described in Section 3 that apply to the test design and data analysis plan can be grouped into the demonstration performed on the Navy/DRI Site (Section 5.1) or the simulated sites demonstration (Section 5.8).

5.1 Conceptual Experimental Design

We worked in conjuction with the Navy during this demonstration. As such, we were dependent upon their procedures and budget constraints for surveys and data acquisition (Department of the Navy, 2009). The Navy is remediating the entire Navy/DRI Site. During this demonstration, PNNL only participated with the Navy on the verification sampling that occurred on the South Tract. The Navy performed a 100% geophysical survey on the South Tract and remediated all anomalies above the defined remediation signal threshold. This threshold was based on site conditions to identify all "metalic targets at or above the performance objective of a 20 mm projectile at 6 inches below ground surface." (Tetra Tech EC, Inc. 2010a). The Navy also planned on conducting a 100% post-remediation re-survey of the entire South Tract, followed by digging any anomalies above the remediation threshold. We planned on using this 100% re-survey data as a baseline to illustrate the likelihood of sampling out of compliance targets that exist in the baseline data set using various VSP-PRV transect designs. We also planned to use the costs associated with the survey and digging work to compare anomaly and transect PRV sampling. However, due to budget shortfalls and an unanticipated abundance of anomalies identified in the Central Tract of the Navy/DRI Site, the Navy was unable to perform the 100% re-survey in the South Tract. Instead, we provided them with a transect design, funded by ESTCP, and they conducted geophysical surveys along these transects. This modified re-survey approach still allowed us to meet our objectives as described in Section 6. The initial 100% survey and the post-remediation transect survey events are described below.

5.2 Geophysical Survey Instruments Used and Systems Specifications

Two geophysical survey events were performed during this demonstration. The initial survey was a 100% survey and the second was the transect verification survey. The same survey equipment, personnel, and procedures were used for each survey. All surveys and data processing were performed by Tetra Tech EC, Inc. (TtEC) under contract with the Navy. The geophysical survey was performed using time domain electromagnetic techniques with five Geonics EM-61-MK2 units configured on their VTA. shown in Figure 7. Areas not suitable for execution of the geophysical survey (i.e., areas not accessible for geophysical survey utilizing the VTA configurations or where the power lines and fences created significant electromagnetic [EM] noise) were avoided. The specific system specifications and parameters used are described in the TtEC work plan (Tetra Tech EC, 2010).

This workplan was conducted using good process quality control throughout the production portion of the project using the concepts described in UXO-5 and using the UFP-QAPP as a basis for the design.

Figure 7. The VTA used during both surveys in the South Tract by Tetra Tech.

5.3 Site Preparation

Surface clearance was performed to identify and remove munitions (and related hazardous items) or metallic debris on the ground surface that may have posed a safety hazard to site personnel, interfere with the operation of the geophysical equipment, or otherwise affect the quality of the geophysical survey. The instrument-aided surface clearance was designed to clear 100 percent of the site of surficial ferrous and nonferrous items. The site was divided into 200-ft by 200-ft (60.96-m by 60.96-m) grids and the surface clearance was performed by grid. Using an approved handheld metal detector, lanes of approximately 5 ft (1.52 m) were swept and all surface contacts and metal debris was removed.

Vegetation removal was performed across the site to improve surface visibility and access for the radiological survey, geophysical mapping, and intrusive investigation. TtEC UXO personnel used a brush hog or commercial mower to remove vegetation. Mower heads were maintained at a minimum distance of 4 in (10.16 cm) from the ground surface to eliminate contact with potentially present MEC items. Vegetation was cut to a maximum of 6 in (15.24 cm) above grade and left onsite as ground cover.

5.4 Calibration Activities

A geophysical prove-out (GPO) test bed was constructed during the initial survey for equipment and procedure validation and to ensure the geophysical data acquisition personnel had the requisite experience to perform the work. Digital geophysical mapping (DGM) data were collected over the GPO test grid before data were collected in the South Tract to ensure the equipment and procedures met the project QC performance metrics. TtEC collected data over the GPO multiple times in the east-west direction and once in the north-south direction. These GPO data were used to ascertain the variations in instrument response for the seed items. In addition to the GPO testing, DGM data were collected over the instrument verification strip (IVS) at the beginning and end of each day to ensure the instruments were operating correctly. Specific details about the configuration and successful application of the GPO and IVS are documented in the TtEC work plan (Tetra Tech EC, 2010).

5.5 Geophysical Surveys Performed

The initial 100% geophysical survey of the South Tract was completed in an east-west direction. The identified anomalies that met the the signal theshold and rules (outlined in Section 5.6) were remediated. The results of the intrusive investigation of these anomalies is shown in Figure 8 and summarized in Table 4.

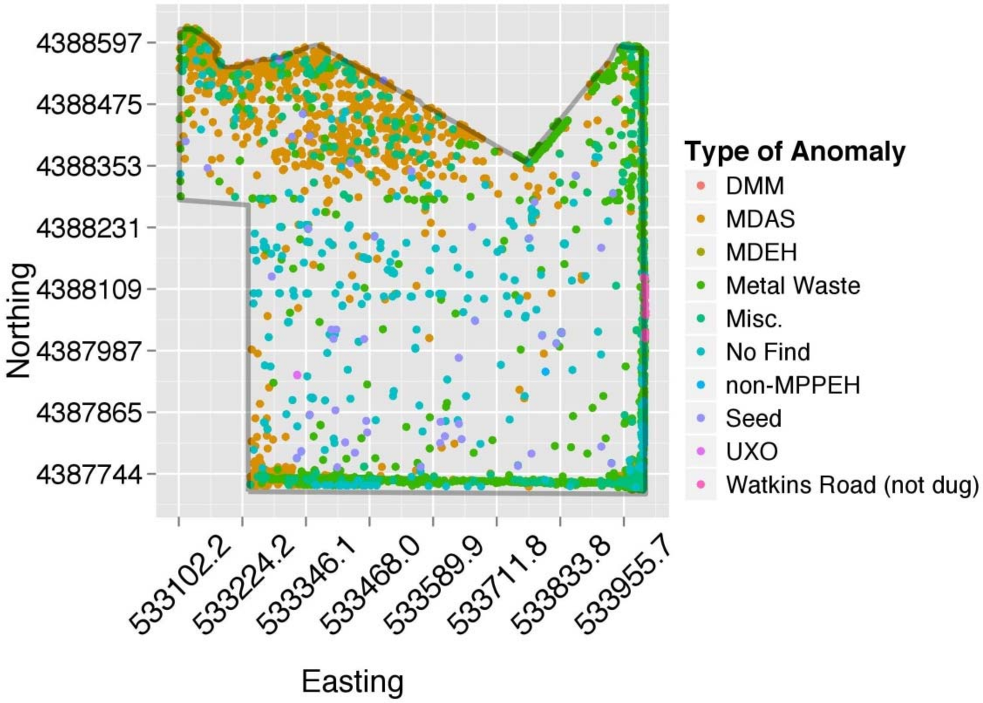


Figure 8. Anomaly locations and type based on the intrusive investigation of the South Tract of the Navy/DRI Site.

Table 4. Summary of the anomaly types resulting from the remediation of the Navy/DRI South Tract.

Discarded Military Munitions (DMM)	Material Documented as Safe (MDAS)	Material Documented as Explosive Hazard (MDEH)	Metal Waste	Misc.	No Find	non-MPPEH	Seed	UXO	Watkins Road (not dug)
5	1121	1	1348	316	258	1	46	1 item/ 2 anomalies	16

For the second (verification transects) geophysical survey, PNNL used VSP to derive a transect survey plan. If no out of compliance targets are found, then we can state that we are 99% confident that at least 99.25% of all possible 9.84 x 200 ft (3 x 60.96 m) transects do not contain out of compliance targets. We used lengthwise aggregated random sampling and collected transects in groups of five to form survey transects that were 1,000 ft (305 m) long. If the 1000-ft-long survey transect did not fit within the southern tract grid boundary, the transect was wrapped around at the edge. The final proposed transect design resulted in 21.2 line-miles (34.14 line-km) for the entire transect survey (Figure 9). Note that if a transect on the as-designed survey fell outside the defined northern boundary of the South Tract, TtEC shifted the transects south to the boundary line. The surveyed transects are shown in Figure 10.

Figure 9. The South Tract of the Navy/DRI Site and the proposed transects to be surveyed. 21.2 line-miles (34.12 line-km) were proposed with 9.64% coverage.

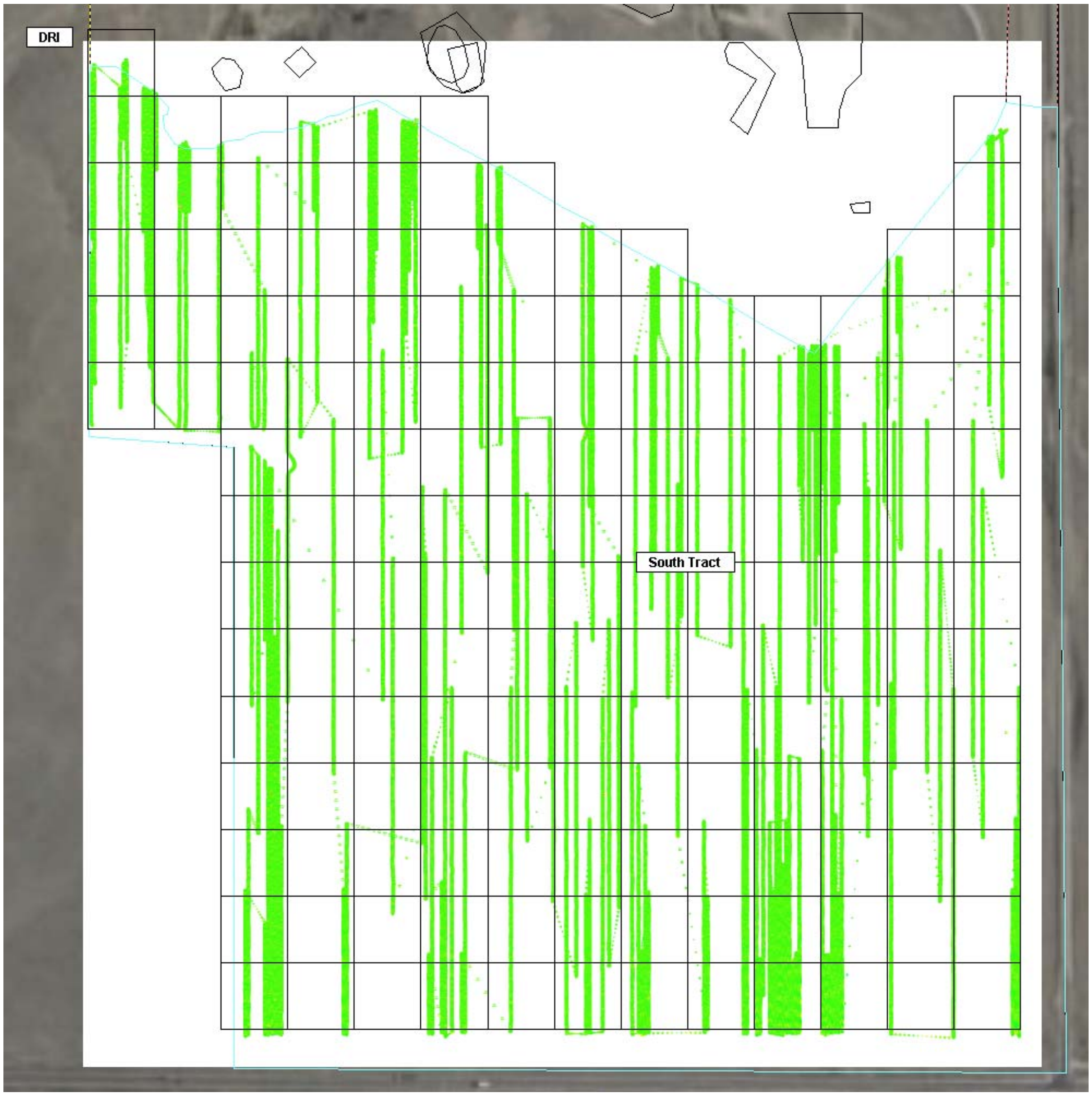


Figure 10. Geophysical data from the surveyed transects within the South Tract of the Navy/DRI Site.

5.6 Target Remediation signal thresholds and Digging Procedures

Geophysical data processing and TOI identification was accomplished to meet a performance objective of a 20 mm projectile buried 6 in. (15.24 cm) below ground surface (bgs). The anomaly dig selection criteria that were used for both surveys are:

- Dig anomaly with greater than 5 mV response on EM61-MK2 Channel 2 (366 microsecond [µs] time gate) and visible on at least two adjacent acquisition lines (some flexibility was given to analysts to include some that were greater than 4 mV to account for shape, noise, etc., if uncertain)

or

- Dig anomaly with greater than 8 mV response on EM61-MK2 Channel 1 (216 µs time gate) and visible on at least two adjacent acquisition lines (some flexibility was given to analysts to include some that were greater than7 mV to account for shape, noise, etc., if uncertain) .

Anomalies that met these criteria on the initial 100% survey were dug, identified, and remediated and are shown in Figure 8. A total of 30 anomalies were dug from the verification transects shown in Figure 10. Of the 30 dug anamolies, 17 met the thresholds bulleted above. The additional 13 anomalies were generally anomalies just below the threshold or only appeared within one acquisition line. Normally in practice, these additional 13 anomalies would not be dug but because this was a demonstration, we chose to select additional anomalies slightly below the remediation signal thresholds for examination to provide greater assurance that the methodology has no unexpected flaws. The available budget only allowed for the additional 13 digs selected by PNNL.

As planned, we worked with the Navy, its contractors, regulators, and ESTCP to understand the appropriate variation associated with the TtEC remediation signal threshold. Our process for estimating this variation and providing uncertainty bounds on the threshold is described in Section 6.

For both the 100% survey and the verification survey, a two-person team using the real-time kinematic (RTK) global positioning system (GPS) performed anomaly reacquisition. The procedure for reacquiring the location of the anomalies was to;

1. obtain the universal transverse mercator (UTM) coordinates of the anomalies in question from the geophysically interpreted dig sheets,
2. load the target anomalies onto the RTK GPS positioning system in the correct format, and
3. place a non-metallic pin flag marked with the unique anomaly identification (ID) using indelible ink at the target location.

The positioning system was checked for proper coordinate location by reacquiring and comparing (in the field) a minimum of one known grid corner before reacquiring any anomaly locations. This procedure allowed for early identification of potential errors in the reacquisition process.

All digging was performed with appropriate levels of safety and instrument testing. After the intrusive teams had received their briefings and conducted their daily vehicle inspections and equipment checks in the instrument test strip, they mobilized to the worksite. The intrusive team verified that no unauthorized personnel were present within the exclusion zones. The intrusive team utilized Vallon handheld instruments to initially excavate the flagged target location. Once the initial investigation removed the TOI from the anomaly location, the target location was rechecked with the Vallon instrument. As each anomaly was excavated, the team leader recorded target information (e.g., description, depth, and size) found at each anomaly flag (and photographed the items dug during the VSP transect survey). Each hole was subsequently rechecked with the Vallon to ensure no additional metal was present.

5.7 Data Validation

Geophysical measurements and position data were stored on digital media during data acquisition. Each day after acquisition, the data were transferred to the processing center for processing and evaluation. A TtEC geophysicist performed the geophysical and position data processing and QC checks. Processing of the data was performed with a combination of industry-standard Geosoft Oasis montaj mapping software (version 7.2.1) and TtEC-developed software specifically produced to integrate and assess digital geophysical data acquired with the VTA. The processing included such steps as merging of EM and position data, instrument bias removal, and instrument latency corrections. These processed data were output to Geosoft Oasis montaj mapping software for further processing (e.g., leveling), QC analysis, and gridding, and to create color-coded images of sensor intensity for interpretation. Data were recorded in North American datum 1983 (NAD 83) UTM Zone 13 (meters) coordinate system. All data processing parameters were stored in digital files (*.chk) and in the Oasis montaj log file (*.log). All DGM data met the QC performance criteria required in the TtEC Work Plan.

The Navy also considered using their standard verification sampling procedure with an independent contractor then providing us with these results. Again, because of budget constraints this was not accomplished. Had these results been available, we intended to use these data in conjuction with the 100% survey to make comparisons between the VSP-PRV transect and anomaly sampling modules versus the Navy's current practice.

5.8 VSP-PRV Simulation Demonstration

For these simulations we did not incorporate the actual survey work done by TtEC but used the boundaries of the Navy/DRI Site and the transect dimension (3 x 60.96 m (200 ft) as a basis for the simulation demonstrations within VSP. The Navy/DRI Site was separated into the South Tract, Denver Research Institute Test Area (Central), and North Tract (Figure 11). We used different combinations of these three regions as a basis for varying the site boundaries and sizes. These combinations are the

- South Tract (blue),
- Central Area (yellow),
- North Tract (red),
- South Tract and Central Area combined, and
- all three regions combined.

The map in Figure 11 also shows the assumed locations, randomly selected, for the out of compliance targets that are in each area. The number of out of compliance targets within each scenario was developed to maintain 20 out of compliance targets per 100 acres (40.5 hectares). The counts for each scenario, shown in Table 5 and Figure 11, resulted in ~1% of the 9.84 x 200 ft (3 x 60.96 m) transects containing out of compliance targets. With these baseline scenarios, the Monte Carlo simulations were run with 99% and 95% confidence to answer the following:

- Do the VSP selection algorithms use an accurate random selection routine when using transects?
- Are the statistical confidence statements maintained when transects are length aggregated up to 10 times the original transect dimension (3 x 60.96 m [200 ft])?
- How do varied site boundaries and sizes affect the statistical confidence statements and theoretical assumptions in VSP when verification sampling is used?

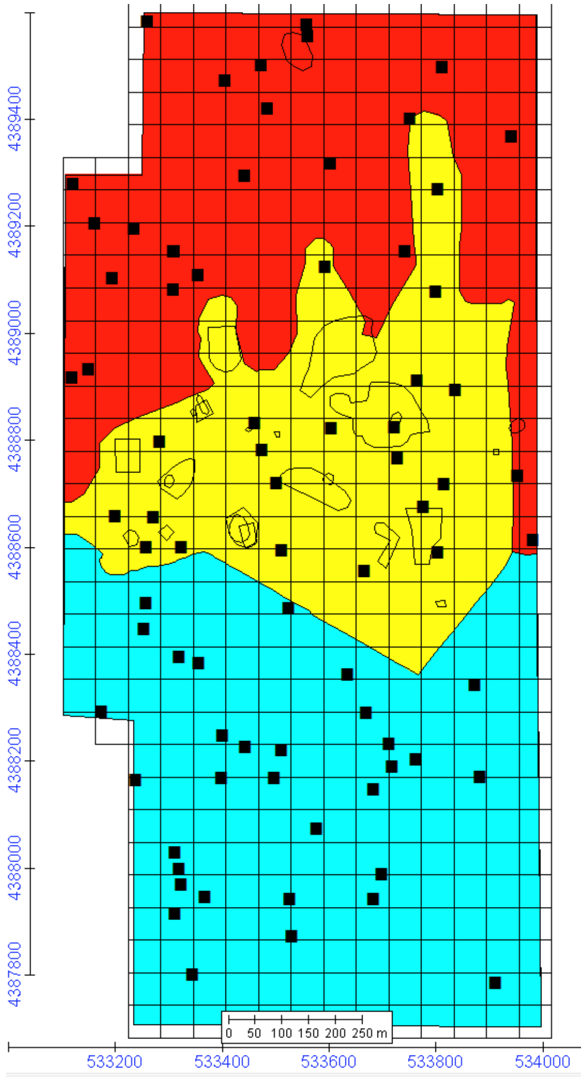


Figure 11. Baseline site assumptions for the simulation scenarios. The defined out of compliance targets locations are shown for all three areas (South, Central, and North). Combinations of these three parts of the Navy/DRI Site were used to define site boundaries for the simulations.

Table 5. Summary of the Acres and out of compliance targets for each of the scenarios used for the simulations.

Scenario	**Full**	**Central**	**South+Central**	**North**	**South**
Out of compliance targets	77	21	53	24	32
Acres	385	104	263	122	159

6 DEMONSTRATION RESULTS AND ANALYSES

This demonstration project allowed an investigation of the performance of the VSP verification sampling modules for two of the three scenarios outlined in Table 1 and repeated in Table 6. Scenario B could be applicable for the initial 100% geophysical survey because there were previous site investigations and removals performed between 1994 and 2002 (Tetra Tech EC, 2010). Scenario A is primarily applicable for the transect resurvey that was performed after this final remediation was completed to verify that the remediation process was successful. This demonstration also allowed for the development of a statistical methodology to define "significantly" above the geophysical remediation signal threshold, as stated in Table 6, using an upper tolerance limit on the remediation signal threshold for the verification transect survey. This section outlines the results of each scenario evaluation and the necessary out of compliance target definition required for each. It should be noted that partial grids to the east of 533955.7 Easting UTM and to the south of 4387744 Northing UTM were excluded from our Scenario A analysis because the south portion of the site was traversed by a power line and the east portion followed a road. These were largely excluded as a cost reduction measure for our demonstration.

Table 6. Summary of the different scenarios (A and B), methods applied, and out of compliance target definitions

Scenario	**VSP Methods Applied**	**Definition of Out of Compliance Target**
A: Recent Remediation	Transect Verification Sampling	Any item that is "significantly" above the geophysical signal threshold used for remediation
B: Previous Remediation and/or incomplete records for area	Anomaly and Transect Verification Sampling	Ordnance-related item of explosive hazard or other items with similar features that were clearly missed during the previous remediation (See Section 6.1)

6.1 Scenario B: Verification of Historical Remediation Efforts at DRI

Under a Scenario B application of the VSP verification modules, the objective would be to verify that historical remediation efforts were sufficient such that there were no remaining out of compliance targets, where out of compliance target is defined as any ordnance related item of explosive hazard or other items with similar features that were clearly missed during the previous remediation (i.e. a large metallic item found below the surface that clearly predates the previous remediation). There are two possible approaches within VSP for meeting this objective: anomaly sampling (anomaly PRV) and transect sampling (transect PRV). For the anomaly sampling approach, one can do 100% geophysical survey of the area and then select some of the anomalies (VSP determines how many) to be dug and evaluated. For the transect sampling approach, VSP will identify transects where geophysical surveys must be performed and all identified anomalies dug and evaluated.

The Navy contracted to do a complete remediation for the entire Navy/DRI Site and completed the remediation for the South Tract. The results from the remediation are shown in Figure 8 and Figure 12. Figure 12 depicts the identified anomalies and their associated channel 2 mV readings from the survey. As documented in the work plan for the current remediation, previous projects did not have as clear a remediation basis; however, investigations and removals were performed between 1994 and 2002. For demonstration purposes, we apply the post remediation verification tools implemented in VSP using the Navy/DRI remediation data as ground truth (100% survey and all 3,098 anomalies investigated). With this ground truth, we document the implementation and comparison of both transect verification sampling and anomaly verification sampling as would be used to verify a historical remediation effort.

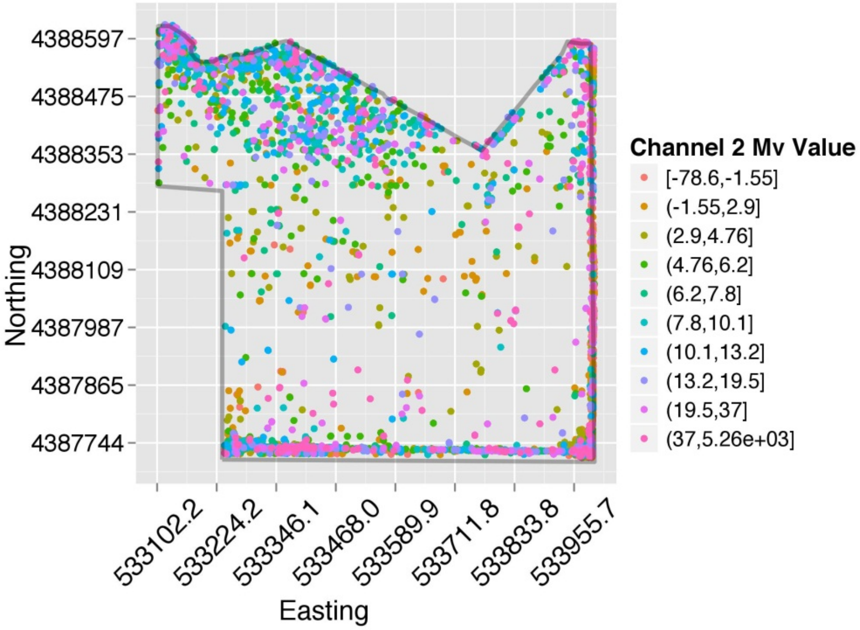


Figure 12. Anomalies that were identified during the 100% geophysical survey. All anomalies shown were investigated and remediated as a part of the Navy/DRI remediation.

6.1.1 Decision Criteria and Failure Definition

Under a Scenario B (Historical Remediation Efforts), an out of compliance target can be defined as any ordnance-related item of explosive hazard or similarly featured anomaly (based on site specific conditions) believed to predate the previous remediation. Because the Navy chose to perform a 100% survey and digging for the South Tract, we can use those results as a basis for evaluating the performance of various VSP-PRV methods. For the anomaly PVR approach, we determine the number of anomalies that must be dug and found to be in compliance in order to state that we are X% confident that at least Y% of all anomalies on site are in compliance. For the transect PVR approach, we determine the number of transects that must be surveyed with all identified anomalies within the transects dug and evaluated and found to be in compliance in order to state that we are X% confident that at least Y% of all transects do not contain out of compliance targets. Thus, the decision rule we wish to employ is:

- If any dug anomaly in the <100% sample/survey is determined to be an out of compliance target, then we cannot state with X% confidence that at least Y% of the anomalies or transects contain no out of compliance targets, thereby failing the acceptance criteria.
- Otherwise, if all dug anomalies do not contain out of compliance targets, then we can conclude that we are X% confident that at least Y% of the anomalies or transects do not contain out of compliance targets, thereby passing the acceptance criteria.

In this demonstration, for both the transect and anomaly verification designs we determine the number of anomalies that must be dug or transects that must be surveyed and evaluated necessary to meet a specified confidence that at least 99% of the samples are "clean" (no out of compliance targets present). We evaluate results for 95% and 99% confidence and show the performance of the implemented designs based on the 100% survey ground truth results shown in Figure 8.

6.1.2 Geophysical Survey and Dig Results from 100% Survey

The geophysical survey and dig results are shown in Figure 8 and Table 4. A more complete listing of the remediation results for the South Tract is available and will be included in the Navy report when the remediation work is completed at the Navy/DRI Site. Using an out of compliance target definition of any ordnance-related item of explosive hazard, Table 4 shows that there are 8 out of compliance targets that were within the South Tract boundaries (if seeded items are used, there are a total of 54 out of compliance targets). The specific items identified as having an explosive hazard (or seeded) and used as out of compliance targets are shown in Figure 13 (Note that a few of the non-Seed out of compliance targets are close enough to be indistinguishable on this figure).

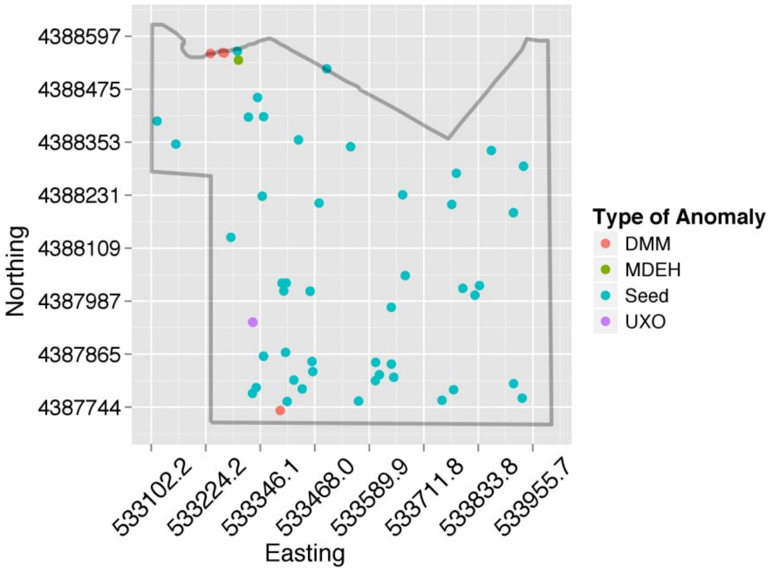


Figure 13. South Tract boundary and the remediated items that were marked as an explosive hazard and used as the out of compliance targets in the demonstration.

6.1.3 Performance of Anomaly Verification Sampling

Using the 100% survey results as a ground-truth data set, there are 3,098 identified anomalies, 54 of which are out of compliance targets. In practice, if we were using the anomaly verification module in VSP as shown in Figure 4, we would specify the confidence and percent acceptable requirements to determine the number of anomalies that would need to be dug and evaluated. For this demonstration, we chose to evaluate the performance of the VSP anomaly verification sampling module when the percent acceptable is 99% (inversely 1% defective) and the confidence is either 95% or 99%. The choice of confidence level would generally be one of the two options selected, but the percent acceptable is a site specific DQO and could vary from what was selected for this demonstration. Ideally one would want to state confidence in a 100% acceptable criteria, but that would require digging virtually all of the anomalies. Thus, one should be aware that the number of required digs will increase significantly as this percent acceptable parameter gets closer to 100%. The selected 99% value was chosen as a reasonable compromise between cost and risk but again, the percent clean choice is a site specific DQO determined by the stakeholders. Table 7 provides a summary of the two sample sizes and the necessary inputs to attain them. While the ground truth is known for the actual percent clean on site, we implemented a design as if the ground truth was not known. The resulting sample sizes of 285 and 426 were repeatedly sampled (10,000 times) from the 100% survey results, and we determined how many times the test failed the decision criteria where at least one out of compliance target was identified out of the 285 or 426 sampled anomalies. These probabilities of detecting at least one out of compliance target in various sample sizes from a population of 3,098 anomalies with 54 out of compliance targets (98.26% in compliance TOI) or alternatively 8 out of compliance targets (99.74% in compliance TOI) are determined theoretically and compared against the simulated results. The simulated results agreed with the theoretical results.

Figure 14 and Table 8 summarize the results from the anomaly verification evaluation study. Table 8 shows the final confidence performance for our sample sizes of 285 and 426. The confidence values shown identify the probability that at least one out of compliance target is included in the collected number of samples. As would be expected, the probability that one out of compliance target is included in the sample (confidence performance) falls below the designed confidence when the site-specific percentage in compliance anomalies is larger than assumed. Similarly, the confidence performance is above the designed confidence when the site-specific percentage in compliance anomalies is smaller than the assumed percentage of 99%. Figure 14 depicts the confidence performance for varied sample sizes under three different percent in compliance conditions. The blue line provides confidence values for the case when 99.74% of the anomalies are in compliance. The red line and black line are based on 98.26% and 99% in compliance assumptions, respectively.

Figure 15 summarizes the simulation results under the four different evaluation scenarios. Each histogram depicts the distribution of out of compliance targets found in a sample (x-axis) based on the specified sample size and number of out of compliance targets onsite (see label above each histogram) over 10,000 runs. The bin to the left of the zero identifies the empirically based estimate of percent of time that the sample did not contain out of compliance targets. All bins represent the proportion of 10,000 runs that contained the specified number of out of compliance targets for each sample size that was applied.

The results shown in Figure 14 and Figure 15 illustrate that anomaly compliance sampling performs correctly in VSP when the assumed parameters line up with site conditions. Additionally, we show how the user's confidence is affected when the assumed percent of in compliance anomalies is different from the design inputs during an actual evaluation. For the scenario that did not include the seeded items such that the percent of out of compliance targets is much less than 1%, the chances of an out of compliance target being in the designed sample drops off dramatically from the desired 95% or 99%.

Table 7. Summary of the DQO parameters and resulting sample size for anomaly verification sampling.

Confidence	Percent Acceptable	Assumed number of defectives	N (Population Size)	Number of anomalies (n/sample size)
95%	99%	31	3,098	285
99%	99%	31	3,098	426

Table 8. Summary of the site-specific confidence based on the 8 and 54 out of compliance targets scenarios for anomaly compliance with 3,098 total anomalies and an as designed acceptable criteria of 99% clean or no more than 31 anomalies are out of compliance targets.

		Achieved Confidence for Each Design Criteria	
		95% Confidence / 99% In Compliance-Targets	**99% Confidence / 99% In Compliance Targets**
Number of samples required based on designed criteria		285	426
Confidence using designed sample size when number of TOI is as specified	**8 TOI (99.74% in compliance)**	53.83%	69.51%
	31 TOI (99% in compliance)	95%	99%
	54 TOI (98.26% in compliance)	99.50%	99.97%

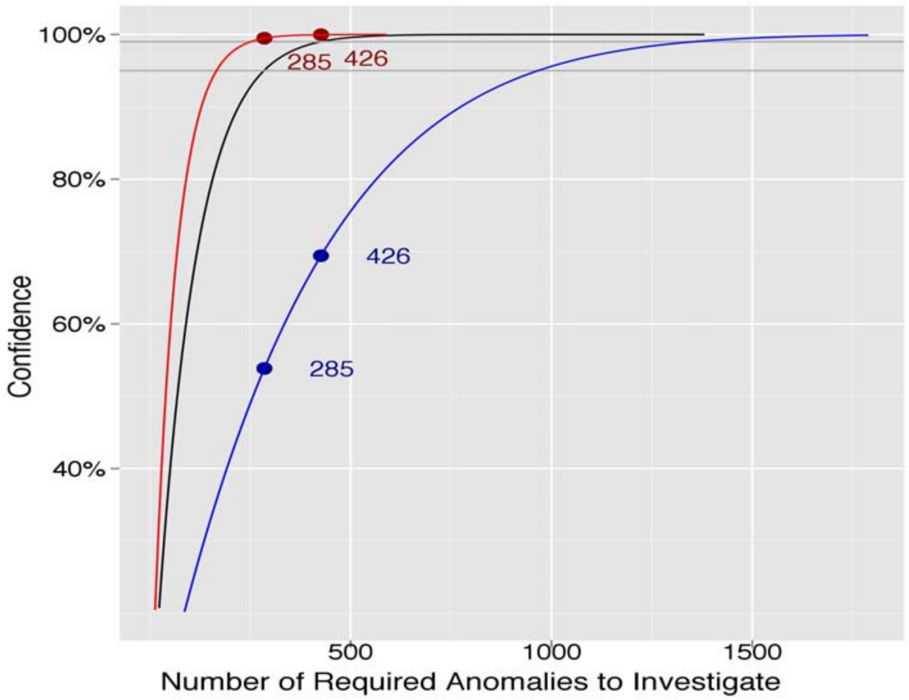


Figure 14. Confidence curves that identify the probability (y-axis) that at least one out of compliance target would be identified in the defined number of samples (x-axis) if sampling were repeated many times from a defined population of 3,098 total anomalies. Black curve shows the as-designed confidence assuming percent in compliance anomalies of 99% (31 potential out of compliance targets), the red curve represents the confidence if the actual percent in compliance targets is 98.26% (54 out of compliance targets), and the blue curve shows the confidence if the actual percent in compliance is 99.74% (8 out of compliance targets).

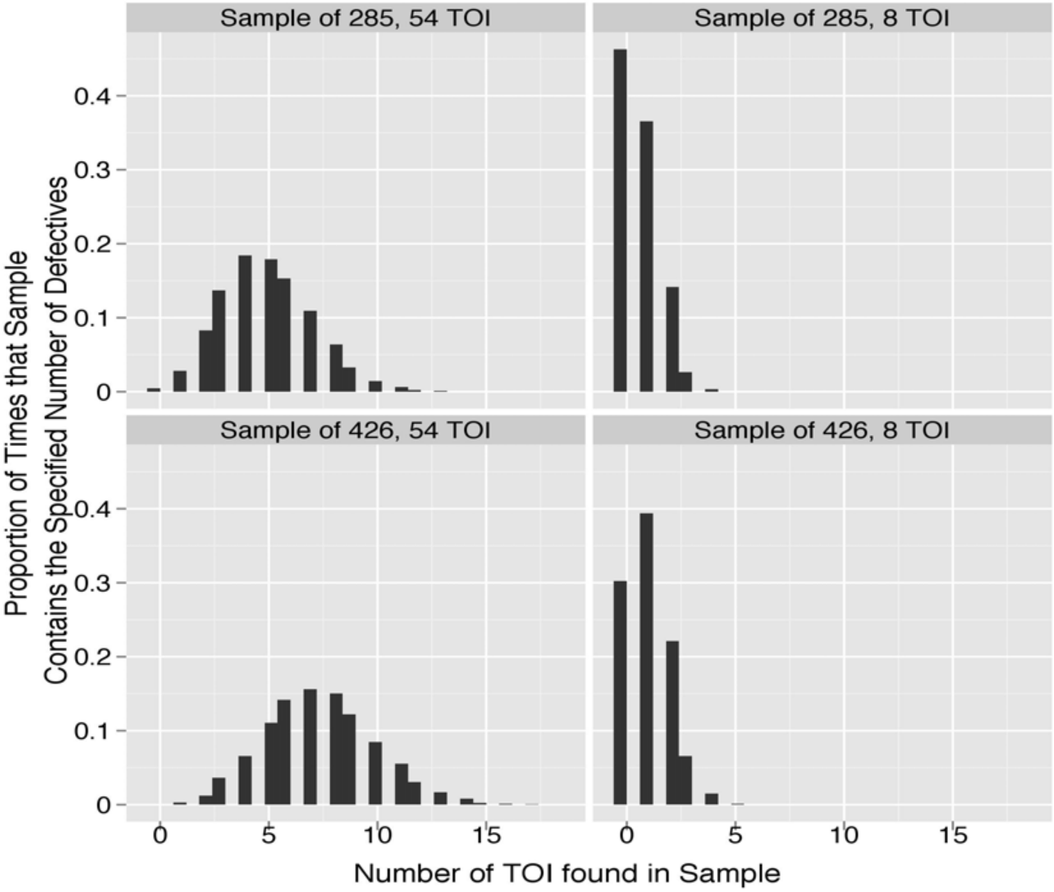


Figure 15. Histograms that depict the distribution of out of compliance targets found (x-axis) in the specified sample size over 10,000 runs. The bin to the left of the zero would identify the empirically based estimate of percent of time that the sample did not contain out of compliance targets.

The lack of performance in the confidence in Figure 14 and Table 8 when there are fewer unacceptable items than assumed is to be expected. Digging a few anomalies will not ensure detection of a very few out of compliance targets. However, if enough anomalies are dug with no out of compliance targets found, then one can ensure that there are no more than a small percentage of the remaining undug anomalies that might be out of compliance targets. Verification sampling is focused on providing a defensible method for reducing the number of items that need to be reevaluated from a 100% resurvey. The percent acceptable (or number of defective) as well as the confidence assumed is used to quantify the allowable uncertainty that results from doing a less than 100% resurvey. In practice, one never knows the true number of unacceptable items unless 100% of the anomalies are dug. This methodology is focused on a more affordable approach involving sampling that controls the risk of incorrect decisions to acceptable levels. The stakeholders select the required confidence and the minimum desired percent in compliance targets (or alternatively the maximum possible undetected number of out of compliance targets) and VSP can then be used to determine the number of anomalies that should be dug.

As illustrated in the analyses within this section, when selecting the minimum desired percent in compliance targets, the maximum possible undetected number of out of compliance targets could be quite large. For example 99% desired percent in compliance sounds very high but with digging 285 of the 3098 anomalies, we could have up to 31 of the 3098 that are out of compliance and we could still have a small chance (5%) that the remediation is deemed acceptable. It should however be remembered that we would still have a fairly good chance (78%) of finding at least one out of compliance target even if there were only a maximum of 15 of the 3098 that were out of compliance (we can be 78% confident that at least 99.5% of the anomaly locations do not contain detectable out of compliance targets). It is important to understand how these percentages translate into the number of possible out of compliance, undetected targets when using this anomaly verification sampling procedure.

6.1.4 Performance of Transect Verification Sampling

The transect verification sampling application to the 100% remediation data is similar to the anomaly verification illustration shown in Section 6.1.3. Our DQO and resulting number of transects required are listed in Table 9. The confidence and percent acceptable values chosen are identical to those in our evaluation of anomaly compliance sampling. However, the population is now 9.84 x 200 ft (3 x 60.96 m) transects, and the population size (number of possible transects of that length/width) is 3,170. Because the population sizes are similar, the resulting sample sizes and potential number of out of compliance targets present between the two verification methods (anomaly and transect) are almost identical as well. Figure 16 shows examples of a random selection of length-aggregated transects required to meet 95% confidence (top) and 99% confidence (bottom) that 99% of the transects are free of out of compliance targets. The amount and location of out of compliance targets are also shown in Figure 16 as an example of how the out of compliance targets could be located within a sample of transects. The out of compliance targets locations excluding seeds are shown on the right, and the left side includes the seeds as out of compliance targets.

Table 10 and Figure 17 summarize the results from the transect verification evaluation study. Table 10 shows the final confidence performance for our sample sizes of 285 and 427. The confidence values shown reflect the probability that at least one out of compliance target is included in the collected number of transects. As would be expected, the probability that one of the sampled transects includes at least one out of compliance target (confidence performance) falls below the designed confidence when the site-specific percentage in compliance targets is larger than assumed. Similarly, the confidence performance exceeds the designed confidence when the site-specific percentage in compliance TOI is smaller than the assumed percentage of 99%.

A key difference with transect sampling lies in how the out of compliance targets fall within the transects. In the example with eight out of compliance targets, only four different transects contain out of compliance targets. For the Navy/DRI Site, two anomalies were created by one TOI item. One other transect contained three cartridge-activated devices (CAD) within a short distance of each other. All the seeded items were spread apart such that all fell in unique transects. Figure 17 depicts the confidence performance for varied sample sizes under three different percentage of transects clean conditions. The blue line provides confidence values that

99.88% of the anomalies are not out of compliance targets. The red line and black line are based on 98.49% and 99% clean confidence designs, respectively.

Table 9. Summary of the DQO parameters and resulting sample size for transect verification sampling.

Confidence	Percent Accept-able	Potential Number of out of compliance targets from 1% of population	N (Population Size)	Number of 3x60.96m transects (n/sample size)	Amount of aggregation	Number of random aggregated sampling locations
95%	99%	32	3,170	285	5	57
99%	99%	32	3,170	427	5	86

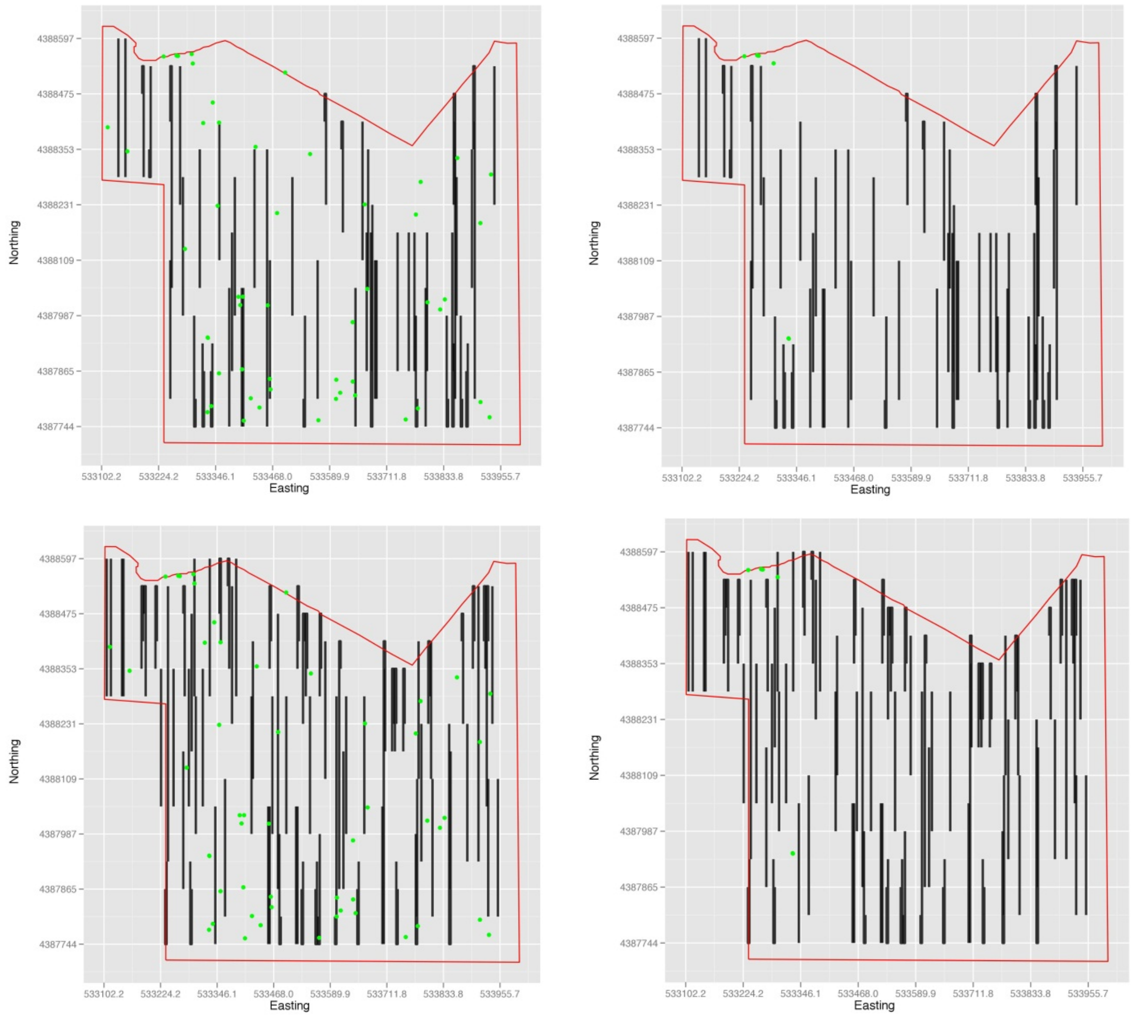


Figure 16. Example of the statistically based transect verification designs for the DQO of 99% clean. The two upper plots show the sample size of 285 (57 length-aggregated transects), and the two lower plots exemplify the sample size of 427 (86 length-aggregated transects). The green dots show the actual locations of out of compliance targets including seeds (left side) and excluding the seeds (right side).

Table 10. Summary of the site-specific confidence based on the 4 and 48 out of compliance targets scenarios for transect verification sampling with 3,170 total possible transects and an as-designed acceptable criteria of 99% in compliance targets or that no more than 32 transects contain out of compliance targets.

		Achieved Confidence for Each Design Criteria	
		95% Confidence / 99% In compliance	99 Confidence / 99% In compliance
Number of samples required based on designed criteria		285	427
Confid-ence using designed sample size when number of out of compli-ance targets is as specified	4 defective transects (99.88% in compliance)	31.40%	43.95%
	32 defective transects (99% in compliance)	95%	99%
	48 defective transects (98.49% in compliance)	98.95%	99.90%

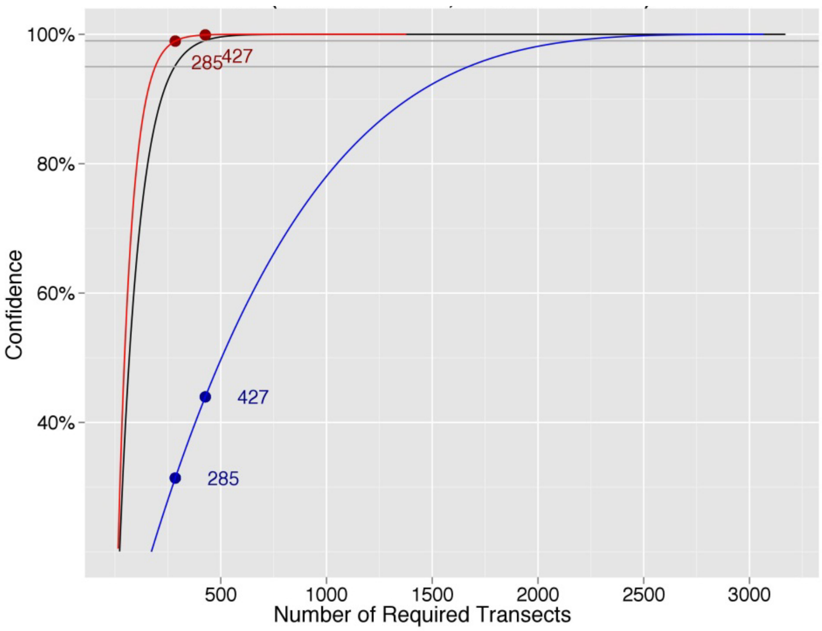


Figure 17. Confidence curves that identify the probability (y-axis) that at least one out of compliance target would be identified in the sampled number of transects (x-axis) if sampling were repeated many times from the specified population. Black curve represents assumed percent clean of 99% (32 potential transects that contain out of compliance targets) and the red curve is the confidence curve when the number of transects containing out of compliance targets is 48 (includes the seeds). The blue curve represents the actual confidence curve when only 4 transects contain out of compliance targets within the survey area (excluding seeds).

Using this Navy/DRI remediation data as an example, there were very few ordnance-related items of explosive concern (eight anomalies). There is a high probability that a verification design similar to those designed in this example, which defined out of compliance targets as solely items of explosive concern, would conclude that no more remediation work was needed in the South Tract. To illustrate this, in Figure 18 sample sizes of 285 and 427 were obtained to meet designed confidence (95% and 99% respectively) and the 99% in compliance TOI assumption (31 out of compliance targets). The probability that at least one out of compliance transect is in the sample when the actual number of out of compliance transects changes is shown for a sample size of 285 (red line) and 427 (green line). As the number of out of compliance transects changes from the assumed value, the confidence can change quite dramatically. If a verification design is used to validate historical remediations, the confidence statements will hold. However, the interpretation of the assumed percent in compliance transects should be made clear.

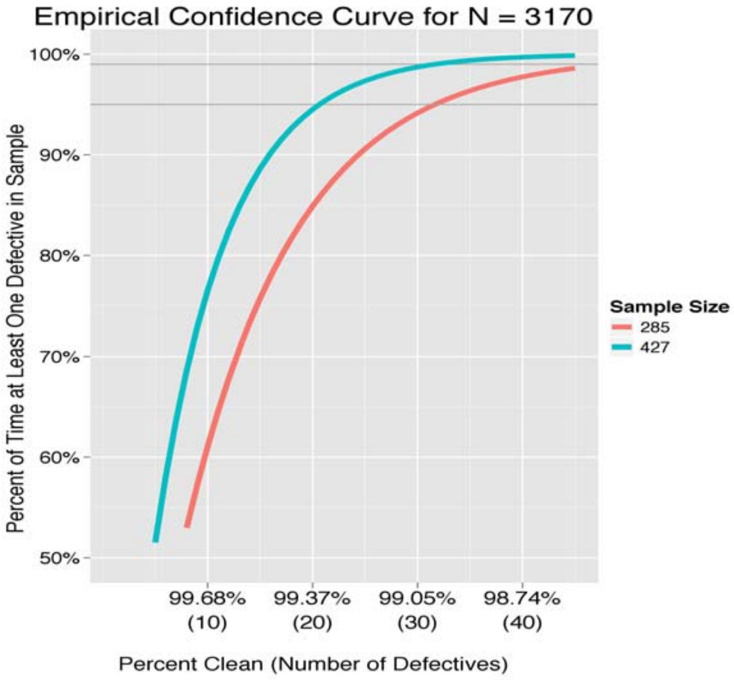


Figure 18. Confidence performance for two sample sizes when the percent clean (number of defectives) varies.

6.2 Scenario A: Verification of Current Remediation Effort for the South Tract of Navy/DRI

A flow chart of the process of applying post remediation verification sampling is shown in Figure 19. This chart shows the processes that can be used to verify an appropriate remediation process. Either anomaly or transect verification sampling can be used. As originally stated in our demonstration plan, after remediation we anticipated collecting a complete 100% resurvey of the South Tract to provide a similar suite of comparisons (shown in Section 6.1) and to demonstrate both paths (shown in Figure 19). TtEC and the Navy encountered significant cost and time delays during the remediation of the central portion of the Navy/DRI Site. These delays restricted the time and funding available for resurvey work in the South Tract. As such, the demonstration was modified to survey only a limited number of transects based on the design parameters described in Section 6.2.1. As can be seen in Figure 19, under Scenario A applications, an additional step of determining an appropriate upper tolerance levels for the remediation signal thresholds is warranted (described in Section 6.2.2). The final decision statement in the flow chart, “Were all dug anomalies acceptable?”, is a little vague. More specifically, “acceptable” would be defined as all dug items being identified as anomalies that have an appropriate explanation as to how the mV signal could have been below the remediation threshold during the first survey and above it during the verification survey.

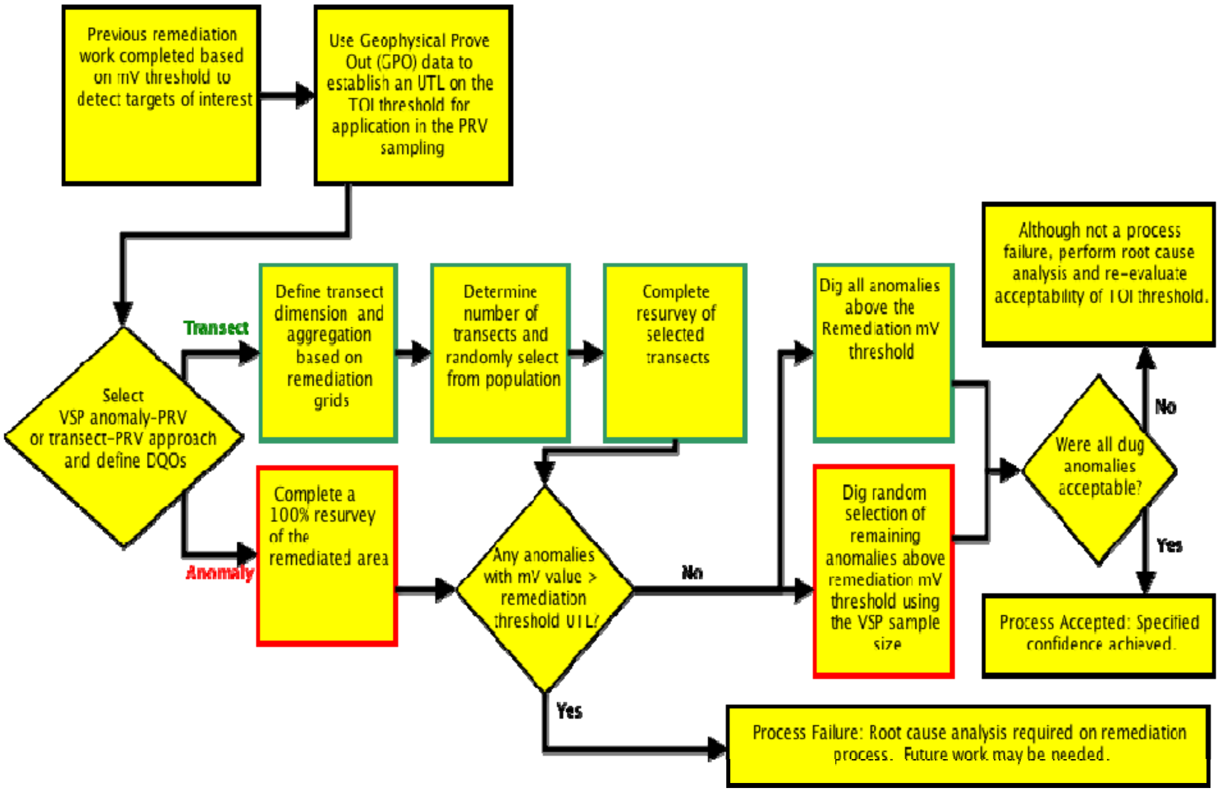


Figure 19. Flow chart of the application of PRV sampling.

6.2.1 Transect Survey Design and Parameter Settings

Figure 9 and Figure 10 in Section 5.5 show the proposed transect design and the actual surveyed data, respectively. This design was set up to obtain as many transects as possible for this demonstration given the budget limitations. As such, the number of transects requested, the desired confidence level, and the % acceptable transects may be much higher than what might typically be needed for most applications. Data quality objectives are unique to each project, and care should be taken to make sure that they are appropriately based on the specific site conditions and needs.

To reiterate, the transect design was set up to allow one to state with 99% confidence that at least 99.25% of all possible transects are in compliance if none of the randomly selected 557 9.84 x 200 ft (3 x 60.96 m) transects contain out of compliance targets. We used lengthwise-aggregated random sampling and collected transects in groups of five to form survey transects that were 1000 ft (305 m) long. If the 1000-ft-long survey transect did not fit within the southern tract grid boundary, the transect was wrapped around at the edge. The final proposed transect design resulted in 34.14 line-km for the entire transect survey (see Figure 9). Note that if a transect on the as-designed survey fell outside the defined northern boundary of the South Tract, TtEC shifted the transects south to the boundary line. The surveyed transects are shown in Figure 10.

6.2.2 Establishing the Remediation Signal Threshold Variation Estimates and the Upper Tolerance Limit that Defines Out of Compliance Thresholds

For remediation projects that use geophysical survey data (i.e., towed array systems), the TOI definition is generally defined in terms of a mV threshold related to the lowest readings from the smallest munitions of interest on the site and the specific site conditions. This remediation signal mV threshold is often established using geophysical data from multiple passes over a GPO area located onsite. Once this remediation signal threshold is established, it is applied to determine what constitutes an anomaly and should therefore be dug. During the post-remediation verification sampling, this same remediation signal mV threshold should be used to identify anomalies. Because the mV signals can vary over multiple measurement readings, this variability must be considered when evaluating whether anomalies with mV readings above the remediation signal threshold qualify as out of compliance targets. We account for this signal variability by defining a UTL on items with mV values expected at the TOI remediation signal threshold (see Section 6.2.3 below).

To establish the remediation signal threshold, the Navy and TtEC teams agreed on geophysical data processing and TOI identification to meet a performance objective of a 20 mm projectile buried 6 in (15.24 cm) bgs. The anomaly dig selection criteria that was used for the remediation were:

- Greater than 4 to 5 mV response on EM61-MK2 Channel 2 (366 microsecond [μs] time gate) or
- Greater than 7 to 8 mV response on EM61-MK2 Channel 1 (216 μs time gate) and
- Anomaly visible on at least two adjacent acquisition lines.

Using the GPO area, multiple readings from each of the buried items was obtained. The mV values for each pass over the 20mm items are shown in Figure 20 and Figure 21. The plot of the mV values by each seeded item in Figure 20 provides an understanding of the variations in mV readings over multiple passes of the same item. Figure 21 shows the same data plotted over time to help reveal whether there are any trends associated with when the survey was performed. There does not appear to be any strong effect associated with when the survey was performed. This figure also highlights that there is little if any difference between the east-west survey and the north-south surveys performed over the GPO.

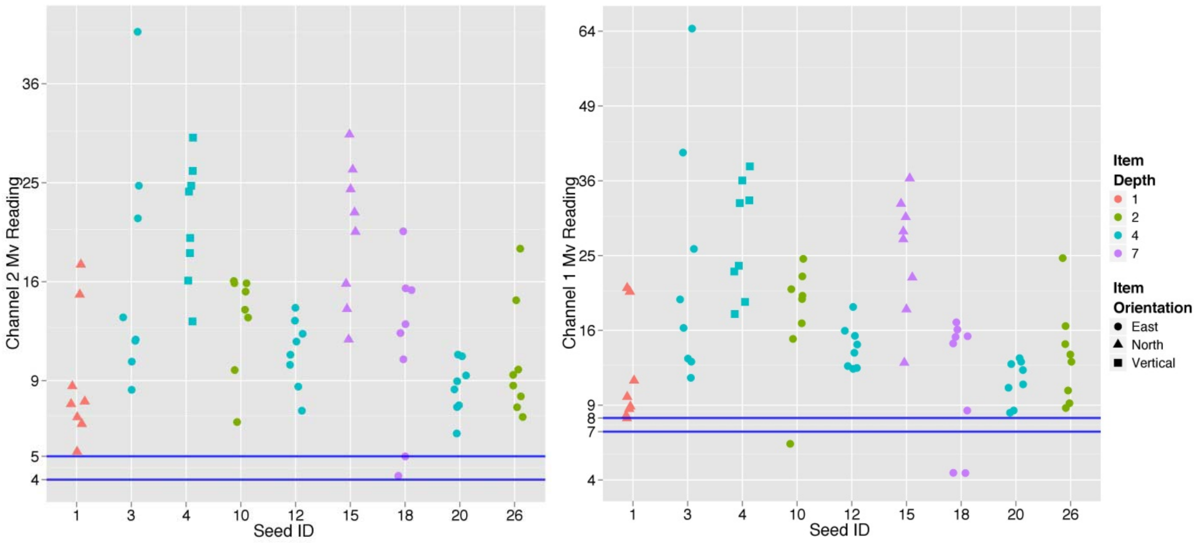


Figure 20. GPO survey data for the 20mm items. These data were used to define the mV value for the remediation signal threshold (blue lines) by TtEC and the Navy for Channel 2 (left) and Channel 1 (right). Item depths are in inches.

The data from the GPO surveys are an important element in establishing the site-specific survey variation with respect to the same item. The Navy and TtEC established the remediation signal thresholds to ensure remediation of all 20 mm items at a specified depth by visually selecting a low mV value that tended to include nearly all of the multiple GPO survey readings for the 20 mm items. This evaluation was the basis for the EM61-MK2 Channel 2 threshold of 4 to 5 mV and the EM61-MK2 Channel 1 threshold of 7 to 8 mV.

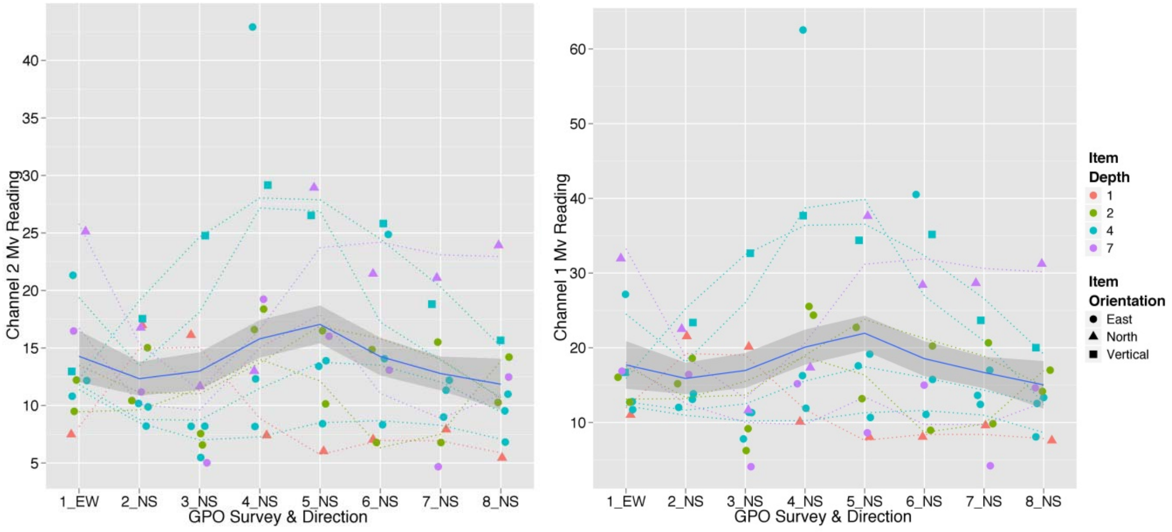


Figure 21. GPO survey data for the 20mm items. Plotted by survey to evaluate the effect on mV values as it relates to the time and type of survey. Channel 2 (left) and Channel 1 (right) are shown with a smoothed fit over time. Item depths are in inches.

6.2.2.1 Remediation Signal Threshold Variation Estimates Based on GPO Analyses

For the transect verification sampling a second survey is performed. As Figure 20 shows, the same item will result in a range of mV values from one survey to the next. Because of this inherent variation, we expect this second verification survey to reveal some anomalies above the remediation signal thresholds that were measured as below the remediation signal thresholds during the remediation survey. Just because these verification survey anomalies are above the remediation signal thresholds, we do not want to suggest that we failed the remediation process when in fact the exceedences could be due to inherent signal variation.

We establish a rigorous process to calculate an upper tolerance limit on the remediation signal threshold that accounts for this variation, above which may constitute a failure in the remediation process (the anomaly should have been detected, investigated, and removed during remediation). This process is largely dependent on the GPO surveys and the remedation contractor's effort to minimize signal variation in defining the remediation signal threshold. The remediation contractor should have an incentive to minimize the signal variation in order to maximize the remediation signal threshold so the number of required remediation and verification digs are minimized, thereby conserving their resources. Additional data collected during the verification sampling survey can also be used to estimate variation in the process, as will be shown in Section 6.2.2.2.

The mean $(\overline{x})$ and standard deviation (s_x) of the mV readings of each of the 20 mm items were calculated from the 8 GPO surveys and plotted in Figure 22 (left side). The right side of Figure 22 plots the average mV reading against the standard deviation of the logged mV values $(s_{\log(x)})$. For small values of $s_{\log(x)}$ this value is nearly equivalent to the relative standard deviation (RSD) estimates $(s_x/\overline{x})$. The estimates for channel 1 and 2 are shown in each plot, and the solid blue line shows the pooled standard deviation estimate from the eight items for both cases. As depicted by the apparent upward trend in the standard deviation (s_x) plot (left side), the standard deviation increases as the average increases. Although there are some extreme values, the upward trend does not appear in the $s_{\log(x)}$ plot. From these results, we assume the mV readings have a multiplicative error structure and the uncertainty and resulting upper tolerance limit should be established using the logged mV readings from the GPO. The upper tolerance limit in log space will be transformed back into mV units for the final determination. The pooled standard deviation estimates of the logged mV readings are 0.382 and 0.359 for channel 1 and 2, respectively.

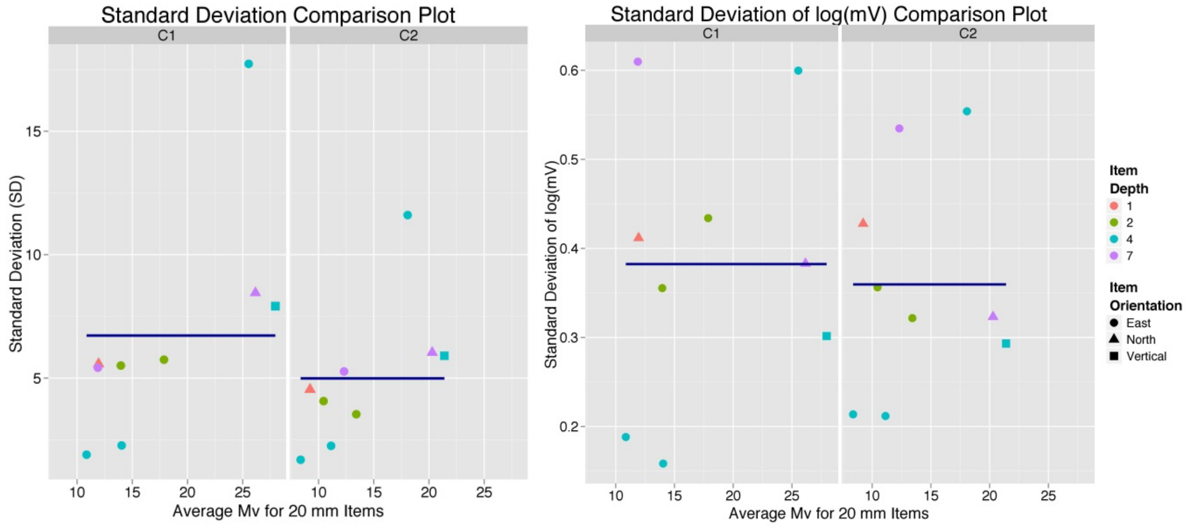


Figure 22. Standard deviation of mV readings (left) and log(mV) (right) are plotted against the average mV value (x-axis) for each 20 mm item.

6.2.2.2 Signal Threshold Variation Estimates Based on Anomaly Pre/Post Matching

As an additional check on our estimates of signal variation, we used the identified anomalies from the verification transect survey matched against their readings from the remediation survey to provide a comparison estimate of the variation. We selected only those anomalies whose average mV readings from the verification surveys and remediation survey were above 6 mV and 4 mV for channel 1 and 2, respectively. The results, shown in Figure 23, follow the same pattern as the standard deviation of the logged mV values shown in Figure 22 for channels 1 and 2. As shown in Figure 23, the pooled $s_{\log(x)}$ estimate from the GPO data (section 6.2.2.1, 0.382 and 0.359 for channel 1 and 2, blue lines) is not significantly different from the Pre/Post $s_{\log(x)}$ estimate (0.326 and 0.491 for channel 1 and 2, red dashed lines).

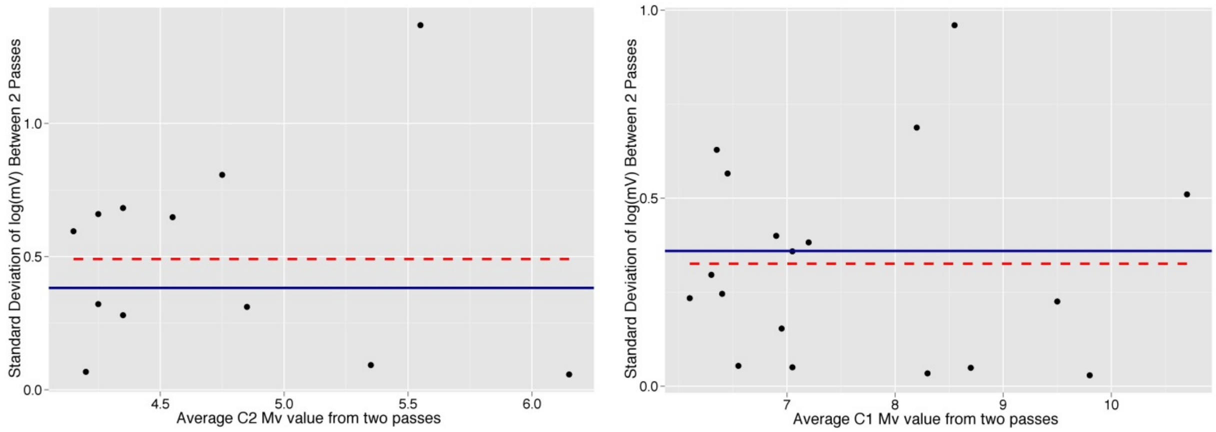


Figure 23. Each point on the plots represents the estimation of variation between surveys of the same low mV anomaly. The pooled $s_{\log(x)}$ estimate from the GPO data (0.382 and 0.359 for channel 1 and 2, blue lines) and the Pre/Post $s_{\log(x)}$ estimate (0.326 and 0.491 for channel 1 and 2, red dashed lines) are shown.

6.2.3 Verification Transect Survey Results

The Navy and TtEC identified the anomalies from the verification survey that would qualify as items to be dug based on the 100% remediation signal mV thresholds. Those 17 anomalies are marked “dig” in the “Geologist Notes” column of Table 11. Although most of these exceed the 5mV (channel 2) or 8mV (channel 1) thresholds, a few were included based on the analyst’s interpretation of the signal but none below 4mV(channel 2) or 7mV (channel 1) were included. As described previously, because this was a demonstration, for additional verification purposes PNNL requested that they dig an additional 13 locations where the remediation signal threshold was exceeded on one or more of the channels but the anomaly was only visible on one line. Information about each of the 30 anomalies is provided in Table 11.

Figure 24 shows the channel 1 and 2 readings for all potential anomalies identified during the transect verification survey. The red points are the 17 anomalies identified by TtEC that met the remediation signal threshold criteria (would have been selected for digging and remediation during the remediation phase allowing for analyst interpretation) and would be required to be investigated as a part of the verification work. The green points are the additional digs requested by PNNL solely for study purposes. The black lines are the remediation signal thresholds and the grey lines depict the lowest values that the analyst could consider for inclusion as described in Section 5.6.

Table 11. Summary information of all 30 items dug during the verification transect survey. Images are included in the appendix.

Grid	Easting (meters)	Northing (meters)	C2 mV	Geologist Notes	Description	Diameter/Length/ Depth (inch)	Weight (oz)	Photo #
I011	533631.8	4388284.5	9.7	dig/edge	(1) 100 lb. bomb fragmentation square	0.5/ 0.5/ 2	0.1	I-11-VSP001
					(1) Wire	0.1/ 2/ 2	0.01	I-11-VSP001
J011	533691	4388286.4	7.2	dig	(4) Wire	0.2/ 7/ 2	0.1	J-11-VSP002
F014	533401.8	4388460.9	6.5	dig	(1) CROW fragmentation	0.25/ 4/ 2	0.1	F-14-VSP003
A015	533132.8	4388555.7	6.4	dig	(1) CROW fragmentation	0.25/ 3.75/ 1	0.1	A-15-VSP004
					(1) Fragmentation	0.25/ 1.5/ 1	0.15	A-15-VSP004
C013	533259.1	4388421.5	6.1	dig	(1) CROW fragmentation	0.25/ 4.5/ 1	0.1	C-13-VSP005
N006	533894.3	4387983.6	6	dig	(1) Wire	0.1/ 4/ 1	0.1	N-06-VSP006
E013	533389.8	4388403.5	5.8	dig	(1) .50 caliber bullet tip	0.4/ 2/ 4	0.2	E-13-VSP007
K006	533736.2	4387960.7	5.4	dig	Photoflash end cap	1.5/ 1/ 3	0.2	K-06-VSP008
E013	533362.7	4388397.3	5	dig	(1) .50 caliber bullet tip	0.4/ 2/ 2	0.2	E-13-VSP009
H008	533529.6	4388137.3	4.7	dig/edge	No Find	N/A/ N/A/ N/A	N/A	NA
G014	533463.4	4388499.1	4.7	dig/edge	(1) Fragmentation	1/ 2/ 2	0.1	G-14-VSP011
L012	533776	4388356.3	4.6	dig	(1) Wire	0.1/ 14/ 2	0.1	L-12-VSP012
H013	533554.5	4388423.3	4.6	dig	(1) CROW fragmentation	0.25/ 4/ 4	0.1	H-13-VSP013
F009	533443	4388179.2	4.5	dig/edge	(1) Wire	0.1/ 1/ 1	0.01	F-09-VSP014
B015	533153.7	4388563.7	4.4	dig	(1) CROW fragmentation	0.25/ 2/ 2	0.1	B-15-VSP015
					(1) Fragmentation	1/ 1.5/ 2	0.15	B-15-VSP015
					(1) 100 lb. bomb fragmentation square	0.5/ 0.5/ 2	0.1	B-15-VSP015
D013	533295.9	4388400.2	4.3	dig	(1) CROW fragmentation	0.25/ 4/ 2	0.2	D-13-VSP016
D013	533296.9	4388411.9	4.3	dig	(1) 20mm projectile fragmentation	0.75/ 2.5/ 2	0.2	D-13-VP017
M013	533834.8	4388428.3	7.7	no dig/ single line	(1) Wire	0.2/ 2/ 1	0.1	M-13-VSP021

B014	533155.7	4388474	6.3	no dig/ single line	(1) CROW fragmentation	0.25/ 3.5/ 2	0.1	B-14-VSP022
					(1) Fragmentation	0.5/ 1/ 2	0.15	B-14-VSP022
B013	533104.8	4388435.7	6.1	no dig/ single line	(1) CROW fragmentation	0.25/ 2/ 1	0.1	B-13-VSP023
A016	533138.9	4388600.6	5.9	no dig/ single line	(1) CROW fragmentation	0.25/ 5/ 4	0.1	A-16-VSP024
					(1) Small arms casing 7.62mm	0.5/ 3/ 3	0.05	A-16-VSP024
B015	533185.8	4388518.2	5.7	no dig/ single line	(1) CROW fragmentation	0.25/ 3/ 2	0.1	B-15-VSP025
					(2) 100 lb. bomb fragmentation square	0.5/ 0.5/ 2	0.1	B-15-VSP025
B013	533105.8	4388446.8	5.6	no dig/ single line	(4) Wire	0.2/ 6/ 2	0.1	B-13-VSP026
O005	533953.8	4387897.1	5.2	no dig/ single line	(1) Wire	0.1/ 6/ 1	0.1	O-05-VSP-027
E013	533364.2	4388419.5	5.1	no dig/ single line	(1) Fragmentation	0.75/ 1/ 2	0.1	E-13-VSP028
					(1) 100 lb. bomb fragmentation square	0.5/ 1/ 2	0.15	E-13-VSP028
E013	533108.8	4388451	5.1	no dig/ single line	(1) 100 lb. bomb fragmentation square	0.5/ 0.5/ 1	0.1	E-13-VSP029
					(1) Wire	0.1/ 48/ 1	0.1	E-13-VSP029
A016	533138.9	4388569.6	4.7	no dig/ single line	(1) CROW fragmentation	0.25/ 3/ 2	0.1	A-15-VSP030
E015	533359.7	4388509.8	4.3	no dig/ single line	(1) Aluminum scrap metal	2/ 2/ 2	0.1	E-15-VSP035
M013	533847.6	4388442.9	3.5	no dig	(10) Barbed wire	0.25/ 24/ 4	0.25	M-13-VSP049
H010	533556	4388216	3.1	no dig	(1) Wire	0.1/ 20/ 1	0.1	H-10-VSP076

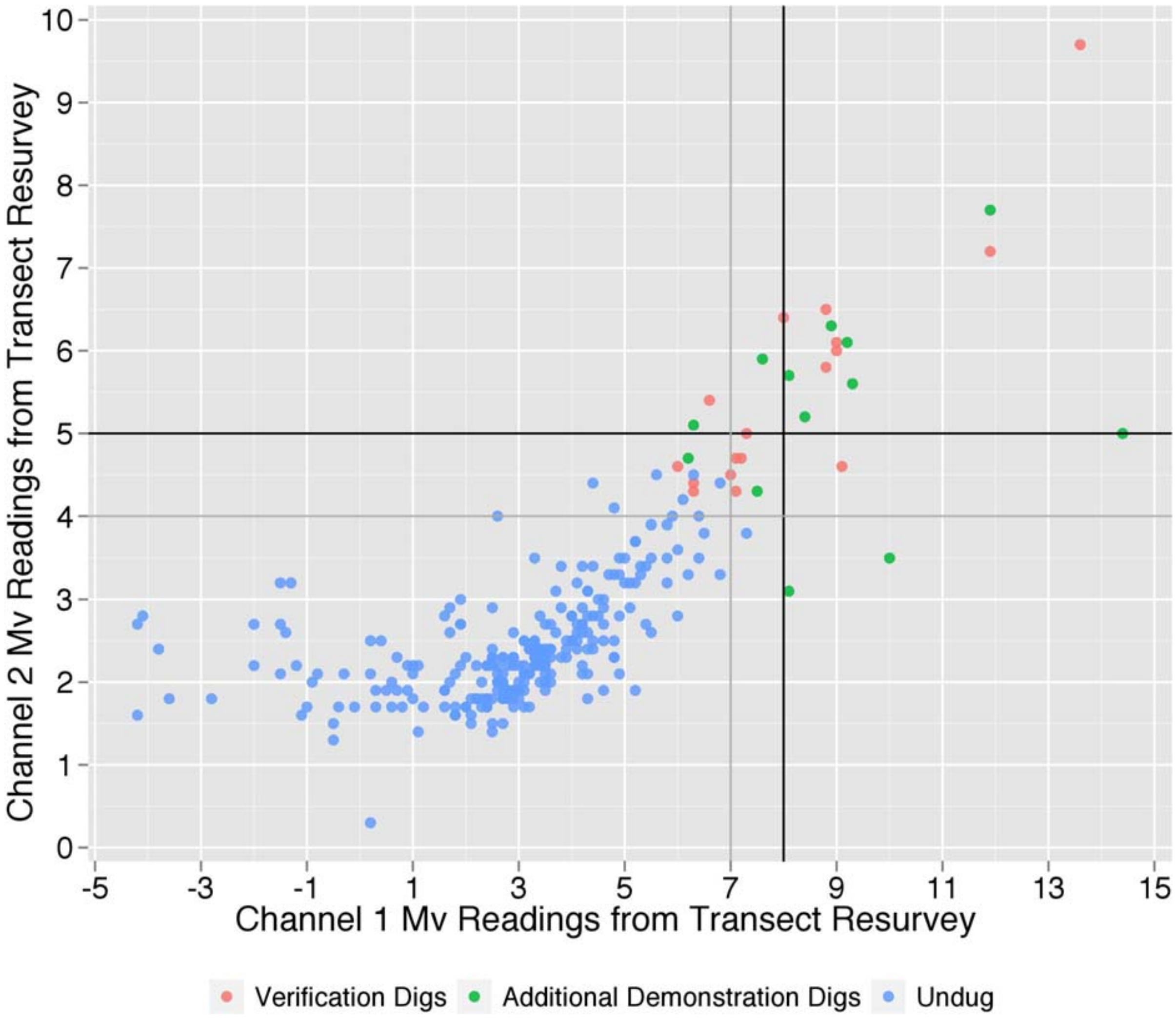


Figure 24. Plot of the channel 1 and 2 mV readings for the candidate anomalies from the transect verification survey.

6.2.4 Remediation Signal Threshold Upper Tolerance Limits for Process Verification

Our objective in performing post-remediation verification surveys and digs was to verify that the entire remediation process was effectively executed. As discussed in sections 6.2.2 and 6.2.2.1, when determining whether any particular anomaly constitutes a remediation process failure, we need to account for the inherent signal variations.

Section 6.2.2.1 provides site-specific estimates of the anomaly signal variation. Estimates were used together with the remediation signal thresholds for each channel (5 and 8 mV) to calculate 95%/99% upper tolerance limits for each of the channel mV signal thresholds. We establish the final channel-specific UTL such that we would expect that 95% of the time at least 99% of all measured mV values would be below the limit if the true mV reading were at the remediation signal thresholds. For example, if an anomaly had a true mV reading for channel 2 of 7 mV, then 99% of the time a mV reading from a survey would be below the upper bound with 95% confidence. This process factors in the signal variation and protects against declaring a process failure when a high mV reading could just be a reflection of inherent signal variations. The final upper bound equation used to calculate the maximum mV value that could result from an item with an expected mV reading at the remediation signal threshold is

$$TOI\ upper\,bound = \exp\left\{ \log(mV\ threshold) + Z_{percentile} s_{\log(mV)} \sqrt{\frac{\nu}{\chi^2_{(1-confidence;\nu)}}} \right\}. \tag{7}$$

Where the following elements of Equation (7) are defined as,

$Z_{percentile}$	is the value from the standard normal distribution at which the specified percentile (i.e., 99%) of values would be smaller
$s_{\log(mV)}$	is the standard deviation of the logged mV values,
ν	is the degrees of freedom used to estimate $s_{\log(mV)}$, and
$\chi^2_{(1-confidence;\nu)}$	is the value from the Chi-squared distribution resulting from 1-confidence percentile with ν degrees of freedom.

Appendix C provides the details of the process used to derive Equation (7) and relies on methodology shown in Hahn and Meeker (1991). The final upper bounds based on 95% confidence for channel 1 and 2 are shown in Equations (8) and (9), respectively.

$$\begin{aligned} \text{Channel}\,1_{95} &= 16.87 = \exp\left\{ \log(8) + Z_{.95} \times 0.382 \sqrt{\frac{63}{\chi^2(1-.95;63)}} \right\}, \\ \text{Channel}\,1_{99} &= 22.98 = \exp\left\{ \log(8) + Z_{.99} \times 0.382 \sqrt{\frac{63}{\chi^2(1-.95;63)}} \right\} \end{aligned} \tag{8}$$

$$\begin{aligned} \text{Channel}\,2_{95} &= 10.08 = \exp\left\{ \log(5) + Z_{.95} \times 0.36 \sqrt{\frac{63}{\chi^2(1-.95;63)}} \right\}, \\ \text{Channel}\,2_{99} &= 13.49 = \exp\left\{ \log(5) + Z_{.99} \times 0.36 \sqrt{\frac{63}{\chi^2(1-.95;63)}} \right\} \end{aligned} \tag{9}$$

Figure 25 shows the mV readings for each identified anomaly in the post remediation verification survey. As can be seen from Figure 25, no anomalies found during the post remediation verification survey were above the UTL. The anomaly with the largest channel 2 mV reading (9.7 mV) is the first anomaly listed in Table 11, which the geologist noted was found on the edge of a transect with an additional piece of wire. This anomaly is pictured in Figure 26, and the images of all 30 digs can be found in the appendix.

In practice, all anomalies that have higher mV readings than the selected UTL criteria for either channel should be dug and, barring explanatory findings during the root cause analysis, would constitute a failure. In addition, all anomalies that exceed the remediation criteria should be dug and examined. If any are found to be items of concern (MEC), although the remediation process is not considered a failure, a root cause analysis should be conducted and a re-evaluation of the basis for the remediation signal threshold should be considered.

This procedure was used on the South Tract of the Navy/DRI verification survey. All of the identified anomalies (red points in Figure 24) were below the UTLs on the remediation signal threshold, and the anomalies that were dug were not ordnance. Thus, we conclude that this DRI south tract remediation was a success. Based on these results and the number of transects surveyed, we conclude that we are 99% confident that at least 99.25% of all possible 200 x 9.84 foot transects do not contain out of compliance targets.

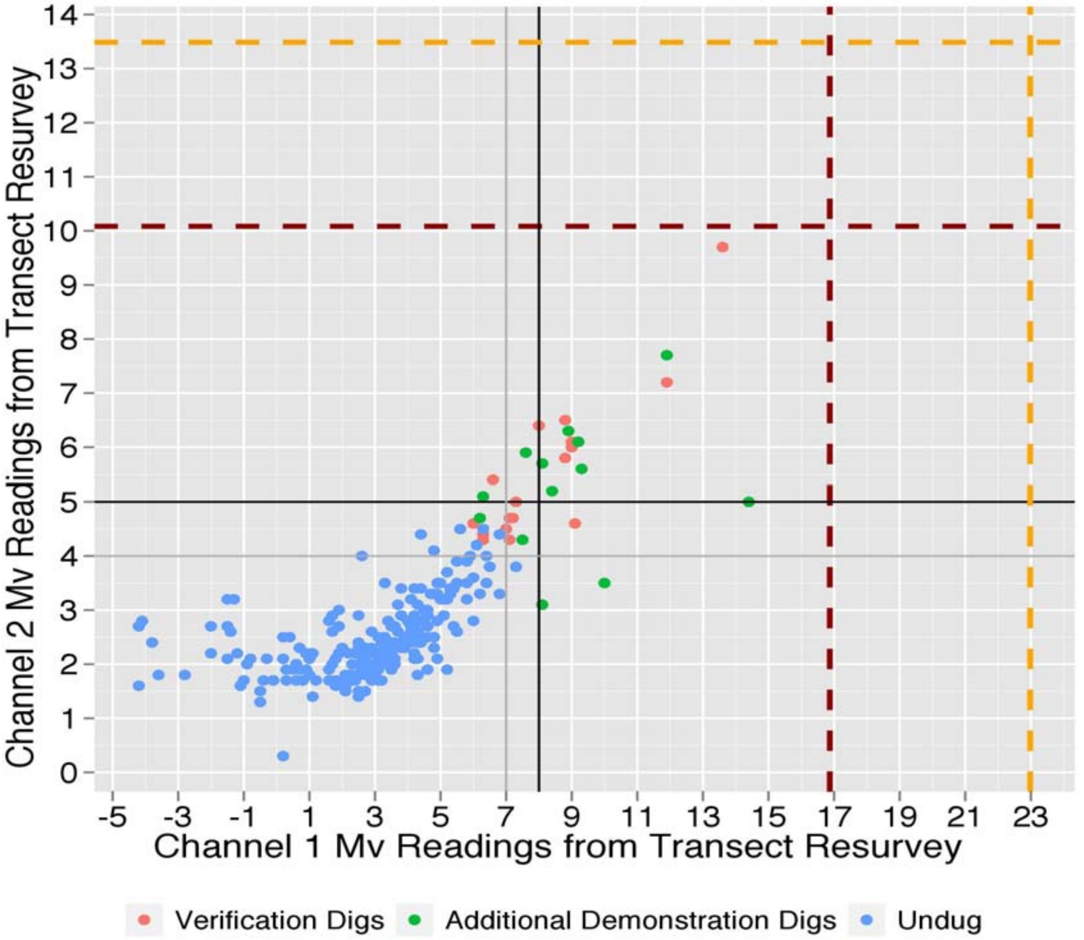


Figure 25. Plot of the channel 1 and 2 mV readings for the candidate anomalies from the transect verification survey. The dashed lines depict the upper bound estimates for 5 and 8 mV (black lines) with 95% confidence that 95 (maroon) and 99 (orange) percent of items would be below them.

Figure 26. Image of the items found from the single location with a mV reading above the verification remediation signal threshold (channel 2 reading of 9.7).

6.3 Transect Sampling Simulation Evaluation

Verification sampling is well documented in the literature as was summarized in Section 2.1. If model assumptions are met, the theoretical framework is sound and reliable. Our implementation of verification sampling has some slight differences from traditional applications. Two of the more important differences are the artificial establishment of the sampling unit and the use of aggregated random sampling. Hathaway et al. (2009) evaluated these differences and documented that when verification sampling is appropriately used accounting for these differences the defined statistical properties are maintained. For this demonstration, we evaluated the actual implementation in VSP and tested the reliability on varied site properties (i.e., size and shape of site).

For these simulations, we used the boundaries of the Navy/DRI Site and the transect dimension (3 x 60.96 m [200 ft]) as a basis for the simulation demonstrations within VSP. The Navy/DRI Site was separated into the South Tract, Denver Research Institute Test Area (Central), and North Tract. We used different combinations of these three regions as a basis for varying the site boundaries and sizes. The map in Figure 11 shows the assumed locations, which were randomly placed, for the out of compliance targets that could be within a surveyed transect for each area. The number of out of compliance targets within each scenario was developed to maintain 20 out of compliance targets per 100 acres (.20 per acre [.2 per 0.4 hectares]). The counts for each scenario, shown in Table 5 and Figure 11, resulted in ~1% of the 3 x 60.96 m (200 ft) transects containing out of compliance targets. With these baseline scenarios, the Monte Carlo simulations were run with 99% and 95% confidence to answer the following:

- Do the VSP transect selection algorithms use an accurate random selection routine for selecting from all possible transects?

- Are the statistical confidence statements maintained when transects are length aggregated up to 10 times the original transect dimension?
- How do varied site boundaries and sizes affect the statistical confidence statements and theoretical assumptions in VSP when verification sampling is used?

6.3.1 Accurateness of the Random Selection Routine in VSP

Figure 27 shows the selection results from 10,000 sample selection runs in VSP, and similar results could be shown for the other 5 example sites. In this example, a sampling design was built to select 285 of the 9.84 x 200 ft (3 x 60.96 m) transects length aggregated for sampling in sets of 5 transects (1,000 ft [304.8 m] survey length). For the South Tract boundary, shown in Figure 27, there were 3,170 possible transects from which a sample could be taken. Thus, each transect had an 8.99% (285/3170) chance of being selected in a given sample. Both images in Figure 27 show that this selection probability was maintained. Any variation from 8.99% was a result of the uncertainty associated with using simulations (10,000) as an evaluation medium. Also, no apparent spatial preferential bias is observed in Figure 25. Thus, based on these results and similar results from the other simulated sites, the VSP random selection routine are performing correctly.

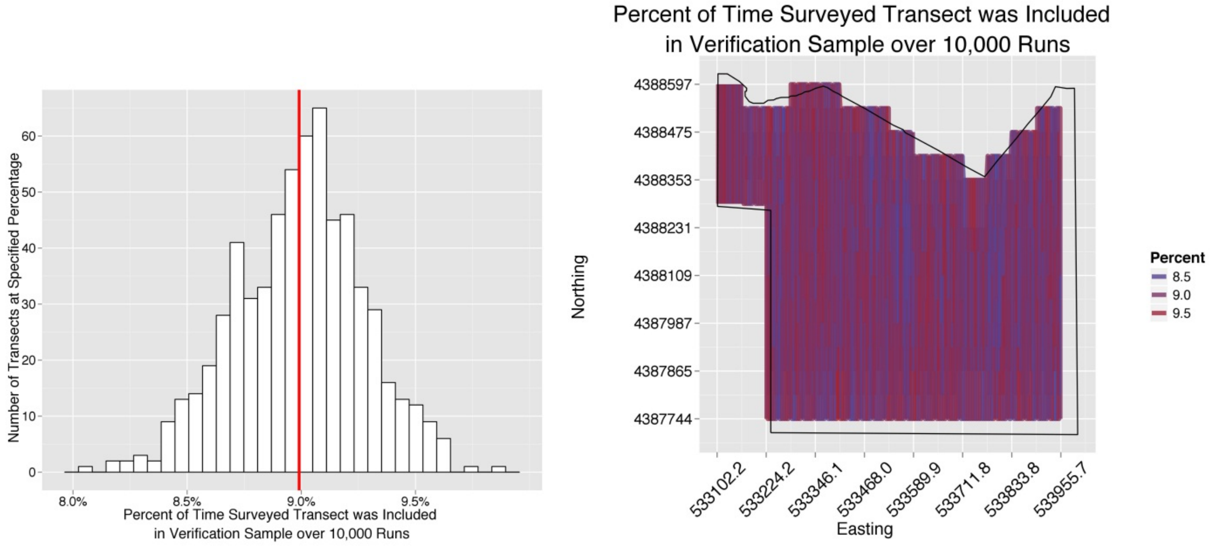


Figure 27. Summary of the selection probabilities for each transect from the South Tract of the Navy/DRI Site.

6.3.2 Performance of Statistical Confidence Statements with Length Aggregation on Varied Site Sizes

While much of the performance and issues associated with aggregated random sampling were documented in Hathaway et al. (2009), our work in this report differs in that we use VSP to do the simulation comparisons and vary the size and shape of the site within which sampling will occur. We limited our evaluation to a range of site conditions but only a few transect size variations. As VSP was used to run the simulations, length aggregated random sampling and simple random sampling were the only two types of sampling evaluated. As stated at the beginning of Section 6.3, we used the transect dimension of 3 x 60.96 m (200 ft) for this evaluation. Based on the work in Hathaway et al. (2009), similar results would occur if the base transect dimension were changed.

Figure 28 provides an example of the different aggregations scenarios that were evaluated. The South Tract is the only one shown, but the aggregation schemes were evaluated for each map shown in Figure 11. All of the transect survey designs shown below for the South Tract covered ~8.12% of the site (17.37 line km). A summary of the coverage for each site is shown in Table 12 for the 95% and 99% confidence designs.

Table 12. Summary of the transect verification designs that were simulated.

Site	**Size (Acres)**	**Total Number of Transects (N)**	**95% Confidence, 99% Clean**		**99% Confidence, 99% Clean**	
			Sample Size (n)	*Coverage (%)*	*Sample Size (n)*	*Coverage (%)*
South	158.59	3170	285	8.12	427	12.25
South-Central	262.64	5319	290	4.99	439	7.55
Central	104.15	2231	279	12.15	415	18.01
North	122.03	2782	283	10.55	423	15.74
Full	384.70	7880	293	3.44	446	5.44

Figure 29 shows the difference between the designed (or desired) confidence and the empirical confidence achieved based on 10,000 simulations. The results demonstrate that the aggregations routines maintain designed confidence across the different types of site boundaries and sizes. These results also highlight that length aggregated random sampling generally maintains the designed confidences as well. The designs that were built around 99% confidence (top of Figure 29) all stayed generally close to the designed confidence and did not show any trends associated with increased aggregation. Those designs established to obtain 95% confidence (bottom of Figure 29) also performed well. The length aggregation of 10 for the south site did fall farther away from the designed criteria and outside of the reasonable range expected (dashed lines) from the set of 10,000 simulations. Even with this most extreme case, the difference decreased the confidence by only 1.4%. The length aggregation results matched the results from Hathaway et al. (2009) even with the inclusion of varied site boundaries and sizes. Finally, VSP routines were validated as well.

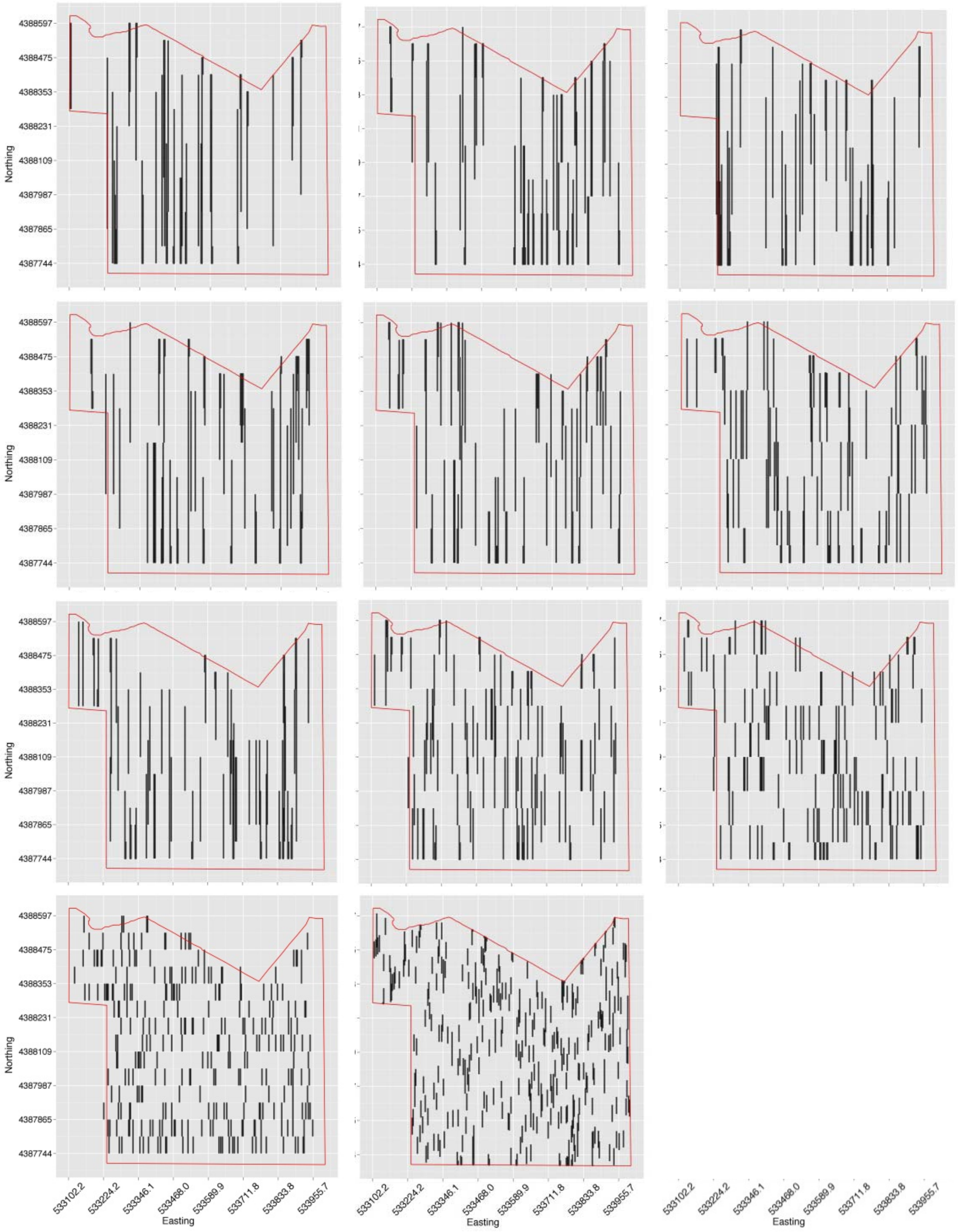


Figure 28. The 11 different random sampling routines evaluated during the simulations. Pictures shown are one realization of a sample of transects based on 95% confidence. All designs are based on a design transect of 3 x 60.96 m (200 ft). The bottom 2 maps are simple random sampling using the remediation grid locations (left) and ignoring the remediation grid locations (right). The remaining maps are based on aggregated random sampling ranging from 2 to 10 transects. Aggregation increases from 2 in the lower right to 10 transects in the upper left.

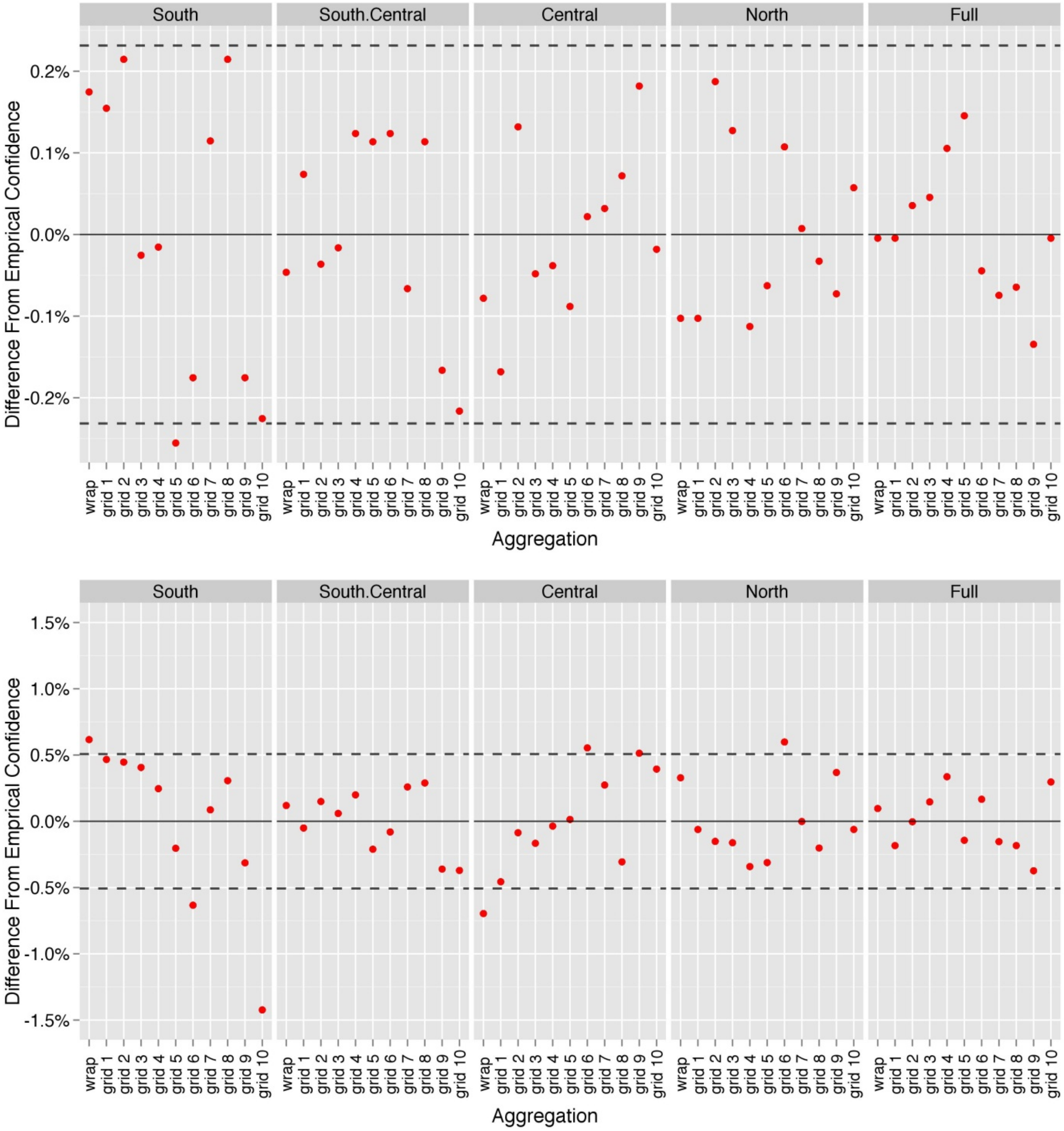


Figure 29. Final simulation results of the empirical confidence differences from 10,000 runs in VSP for both 99% confidence (top) and 95% confidence (bottom). The dashed lines mark the reasonable range within which the differences should typically fall.

7 Performance Assessment

The primary objective of this demonstration was to evaluate and illustrate the validity of the VSP-PRV sampling methodology on an actual site and on simulated sites. As outlined in Section 3, the following seven performance objectives were stated:

1. Clearly identify the applications of verification sampling based on the site history and objectives (see Section 7.1).
2. Evaluate the feasibility and cost-effectiveness of a less-than 100% survey or sampling verification approach verses a complete re-survey and digging of 100% of anomalies (see Section 7.2)
3. Demonstrate the utility and regulator acceptance of the VSP-PRV sampling modules applied to quality control and quality assurance for remediated sites (see Section 7.3).
4. Provide an example of how to account for signal variation in the chosen verification remediation signal threshold that defines anomalies and illustrate how to identify appropriate definitions for out of compliance targets (see Section 7.4).
5. Evaluate the performance, feasibility, and costs of the VSP-PRV transect survey QA/QC approaches versus the VSP-PRV 100% re-survey approaches (see Section 7.5).
6. Evaluate the performance of various transect sampling schemes that result from varying the required confidence, the required minimum percentage of transects that must be proven to be free of any out of compliance targets, and the size of the remediated site (see Section 7.6).
7. Evaluate the acceptance of and performance of various transect aggregation schemes (see Section 7.6).

In this section we evaluate how well these performance objectives were met and discuss each outcome.

7.1 Clearly Identify Verification Sampling Applications Based on Site History

This remediation work provided the ability to demonstrate the use of PRV sampling for an area with historically documented remediation work (Scenario B) and for an area that was recently remediated (Scenario A). Table 1 lists the three scenarios that are applicable to the PRV methods in VSP. Sections 6.1 and 6.2 document our implementation of PRV sampling for scenario A and B and establish a clear process for their use. In Section 6.1, we highlight the implications of sampling design performance when the actual number of out of compliance targets on site differs from the designed criteria. In Section 6.2 we document the process for an actual implementation of PRV sampling within an area with recent remediation. This section also includes a flowchart (Figure 19) that documents the process to be used in implementing PRV sampling.

This performance objective was met for both scenario A and B applications. Scenario C applications were not demonstrated and are expected to the subject of a future site demonstration. Based on this demonstration, it is apparent that these verification sampling methods are appropriate for sites that have recently been remediated (Scenario A) to provide confidence that the remediation process was performed appropriately and with adequate quality control. The post-remediation verification sampling that was performed at DRI was successful at demonstrating with 99% confidence that at least 99.25% of all possible transects do not contain any unexplained out of compliance targets (no failures), thereby providing strong evidence of an effective remediation process. Note that the 99%/99.25% parameters mentioned above are much higher than would typically be warranted for a PRV sampling exercise.

For Scenario B applications (historical remediation with limited records), the PRV methods would also be appropriate. However, one must clearly state what would constitute a failure and recognize the implications of TOI definitions, particularly how the probability of detecting failures (UXO or anomalies exceeding the UTLs) is affected by the actual number of failures on the site. In other words, one must recognize that when surveying/sampling less than 100% of the site, it is likely that failures will not be detected in the transect survey if only a very few failures exist on the site. For this DRI site demonstration, if the PRV had been applied rather than the 100% survey and reacquisition, there would have been a chance that none of the failures (excluding seed items) that were found with the 100% survey would have been identified.

Under Scenario B, there has been a previous remediation but it is unclear how good it was or whether they followed rigorous QA/QC protocols. Because the remediation thresholds and UTLs are set based on GPO results from the verification survey phase, they aren't affected by possible poor quality of the remediation. However, if the remediation was poorly controlled, then there would be a much greater likelihood of finding "failures" during the verification sampling phase. If they happened to have had a very well controlled remediation process but just lacked documentation/proof of that, then the verification sampling phase would help prove that the remediation was adequate because the number of "failures" would be negligible or non-existent.

We would argue that if they didn't have good QA/QC, then their chances of "passing" the verification sampling without failures is very low so they probably shouldn't waste their time on verification sampling but instead should just perform another remediation following accepted QA/QC protocols. If PRV surveys are considered under Scenario B assumptions, we strongly recommend that the % acceptable requirement should be increased well beyond the 99% level to better protect against those few failures that could go undetected.

7.2 Feasibility and Cost-effectiveness of Complete Resurvey and Anomaly Investigation Verses PRV Approaches

Although detailed cost information for transect PRV methods implementation is provided in Section 8, this section evaluates the trade-offs of using PRV methods for different population sizes and compares the percent of the population that must be surveyed and relative costs to meet the PRV design with the specified confidence and percent clean. We use the anomaly counts from the South Tract surveys as a specific example for the results in Table 13. This table highlights some of the ranges in sample sizes for different PRV designs when the total number of possible transects (population size) is 3,098. Additionally, total costs and surveyed amounts are based on the work in the South Tract of the Navy/DRI Site for transect-PRV sampling. Specifically, we assumed the transect dimension was 9.84 x 200 ft (3 x 60.96 m) and the costs were $400 per line km (excluding mobilization and other misc. costs). The general percentage of population surveyed would apply to anomaly-PRV sampling as well if there were 3098 possible anomalies. Figure 30 demonstrates the percent of the population that must be sampled for a range of population sizes from 1 to 20,000. For either anomaly or transect verification sampling projects, the percent of the population that must be sampled becomes an important cost driver. Very small populations will necessitate an almost 100% sample, which negates the need for statistical sampling. Even moderately sized populations can require a large proportion to be sampled. When resampling is required to verify a remediation process, these results demonstrate the cost effectiveness of statistical sampling.

Table 13. Summary of the percentage of population sampled for different PRV designs with a population size of 3,098 possible 9.84 x 200 ft (3 x 60.96 m) transects.

Confidence	Percent In Compliance Transects	Percent of Population Sampled/Surveyed	Sample Size	Total Line Kilometers *	Estimated Cost*
99%	95%	2.88%	89	5.43	$2,170
99%	99%	13.75%	426	26	$10,388
99%	99.9%	77.40%	2,397	146.1	$58,448
95%	95%	1.87%	58	3.54	$1,414
95%	99%	9.20%	285	17.4	$6,949
95%	99.9%	62%	1,920	117	$46,817
90%	95%	1.45%	45	2.74	$1,097
90%	99%	8.13%	221	13.5	$5,389
90%	99.9%	52.45%	1,625	99.1	$39,624

*Assumed use of transect-PRV sampling with a 9.84 x 200 ft (3 x 60.96 m) transect and survey costs of $644 per line mile ($400 per line kilometer).

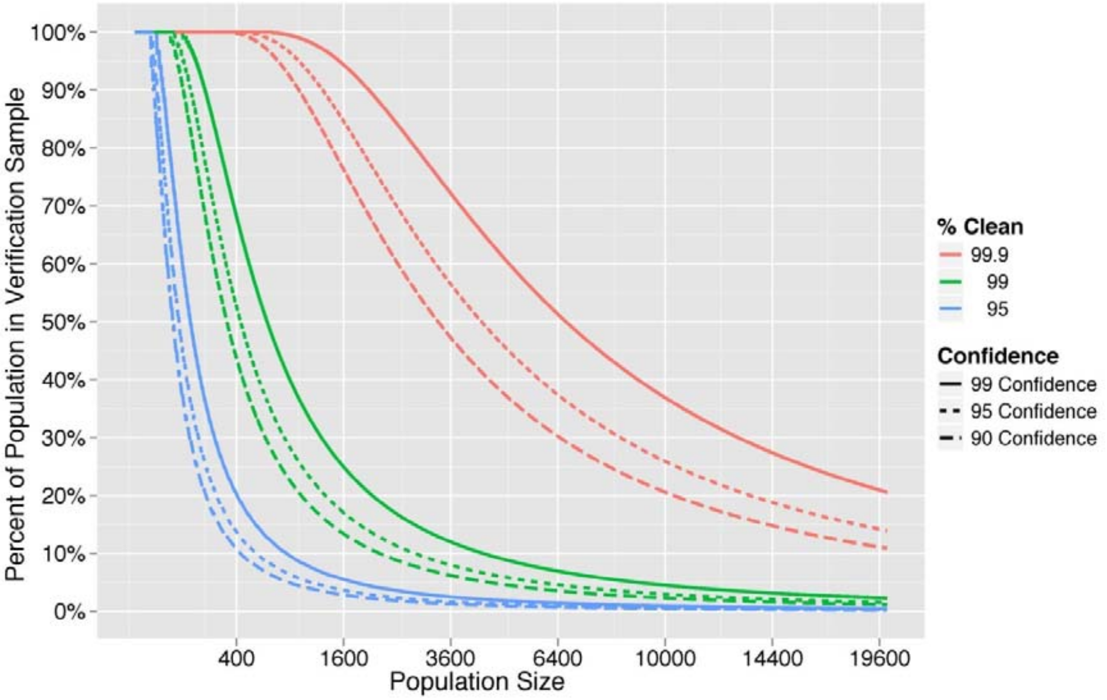


Figure 30. Plot of the relationship between population size and the percent of sample required to meet

7.3 Regulator Involvement and Understanding

During the entire process of this demonstration, EPA (David Rathke) and Colorado regulators (Ken Vogler and Jeff Swanson) have participated in the site meetings and methodology implementation, including a kickoff meeting, a mid-project meeting, and several phone and email interactions. Those that have participated have provided feedback on when these methods are appropriate. We also will be holding a final brief-out on this project with regulators, site managers, contractors and ESTCP program managers. This baseline case study of the application of the VSP-UXO verification sampling modules will be used in our current courses to provide a more complete basis for discussions with regulators, site managers, and contractors. We have demonstrated that the methodology and software tools are appropriate for post-remediation verification objectives and have provided regulators with an understanding of the remediation effectiveness.

7.4 UTL Estimation for PRV Applications

Section 6.2.2 and Appendix C provide a thorough development of the process to use on future implementations of PRV sampling for determining the upper bound. This process accounts for the signal uncertainty associated with anomaly mV readings. Equation (7) is the final equation proposed for use in selecting the upper bound estimate. Equations (8) and (9) document the specific calculation process for application on the South Tract of the Navy/DRI Site and were used to define the UTL on the mV signals, above which generally constitutes a failure (i.e. a root cause analysis may identify an allowable reason for this anomaly).

This process relies on results from the site-specific GPO, although these estimates of signal uncertainty may prove approximate for future site applications. It would be prudent to implement a simple tool within the VSP software to assist the user of the PRV methods with the UTL calculations.

7.5 Contrast and Compare the Anomaly and Transect Verification Sampling Tools

The mathematical model for both methods is built on the same hyper-geometric model on which most verification sampling relies. While both methods have the same TOI definition process for specific anomalies found during the resurvey work, the primary difference lies in how a sampling unit is defined for use with the hyper-geometric model. The sampling unit is highlighted in our naming of each PRV method (anomaly-PRV and transect-PRV). This results in statistical confidence statements that are slightly different as they are in terms of the specified sampling unit.

If the difference in the sampling unit is not of concern, then the primary motivation to select one of these two methods would be the cost of implementation. We use the data from the Scenario A (recent remediation; limited verification transects data) and B (historical remediation with limited documentation; 100% resurvey data) examples documented in Section 6 to provide a comparison. These data, provided in Table 14, show the benefits of transect-PRV sampling. Specific monetary costs are not shown, but the specified survey amounts and digs are adequate to document the cost savings. For scenario B, the same number of anomalies must be dug, but the anomaly verification survey requires a 100% resurvey. Scenario A would only require 10 digs with 9.2% of the site needing to be resurveyed, while the anomaly-PRV method would require 100 anomalies to be investigated with a 100% resurvey.

There will generally never be a scenario where anomaly-PRV sampling would be cost-effective for scenario A. However, as the number of anomalies expected per transect increases in a scenario B project, there will be a point where anomaly-PRV sampling could be a cost-effective option.

Table 14. An example of the dig/survey amounts for transect and anomaly-PRV sampling using the data from the South Tract of the Navy/DRI Site.

				Sample Size for 95% Conf./ 99% Clean design [Number of anomalies to investigate]	
Scenario	**Total number of anomalies on site**	**Total number of possible 200 x 9.84 ft transects**	**Approximate # of transects surveyed for every one anomaly found**	**Transect-PRV (9.2% Resurvey)**	**Anomaly-PRV (100% Resurvey)**
B	3,098	3,170	~1	285 [~285]	285 [~285]
A	106*	3,170	~30	285 [~10]	100 [~100]

* This estimate is based on the number of digs (17) identified with a survey that had 16% of the transects on site surveyed.

7.6 Simulation Evaluations of Statistical Designs in VSP

Section 6.3 detailed the results from the simulation evaluation performed in this demonstration using VSP. The conclusions from the simulation studies met the stated success criteria listed in Table 2. These success criteria are:

- Desired or statistically designed confidence equals the evaluated confidence from simulations.
- Achieved confidence from aggregations schemes (number of transects surveyed in a row) same as confidence without aggregation based on calculations from simulation study.

Figure 29 shows the confidence performance for PRV sampling and validates that desired confidences are maintained when aggregated random sampling is used for varied site boundaries (five different test cases) and aggregation amounts (up to 10). These results corroborate and augment the work in Hathaway et al. (2009). In addition, all simulations were done within VSP to validate the software's implementation of PRV sampling. Figure 27 verifies that the sampling routines in VSP meet the equal probability of selection criteria required for random sampling.

7.7 Visual Sample Plan (VSP) Modifications and Recommendations

As this demonstration proceeded, we identified several needed modifications to the VSP-PRV software modules. These modifications, listed in Section 7.7.1, were necessary for both ease of use and accurate application of the methodology. In addition, several other modifications were identified that would better support user needs, and these are listed in Section 7.7.2.

7.7.1 VSP Modifications During the Demonstration

The process for defining grid dimensions within the transect PRV dialog was implemented during this demonstration. We implemented these tools to address the dilemma of selecting an appropriate transect dimension. This process proved useful in defining the default transect dimension for use during the verification process. This implementation works well for sites that have had a recent remediation that used grids during the process, like the Navy/DRI South Tract. As we used VSP in this demonstration, we also found that numerous usability improvements were needed. Some of these are listed below:

- Added ability to select between grid-based and wrapped transects
- Allow the user to change the transect orientation for the “Wrapped Transects” option.
- Can turn off and turn on individual transects, even if the transects are outside of the sample boundary
- Set the default units on the Costs tab to km and acres.
- Removed “Surveyed” from “Surveyed Transect Length” on the “Wrapped Transects” radio button options.
- Under the “Wrapped Transects” radio button options, changed the default value for transect length to be blank and transferred over the grid dimension if the user input one.
- Show the total number of transects for the “Wrapped Transects” option.
- Set default grid dimension to be blank.
- When the window resizes, the images resize as well.

7.7.2 VSP Proposed Modifications from Demonstration

There was an implementation issue in how VSP handled transects in partial grids within the remediation area. Currently, the transects will cover the entire grid and pass outside the boundary of the remediated area when partial grids are included in the remediation area. This implementation caused problems within the South Tract as a few of the selected transects went into the central area that had not been remediated. For this demonstration, TtEC manually moved the transects south to line up with the north side of the South Tract boundary. We propose to update VSP to support the process for dealing with partial grids within the site.

The process we developed for establishing the upper bound on the remediation signal threshold, provided in Section 6.2.2 and Appendix C, would be a nice complimentary option to be used with the verification sampling dialogs currently in VSP. We propose adding this functionality this next year.

It is possible that scenario C and scenario B implementations will result in a very large number of anomalies per transect. If the area of investigation is also large, a 100% survey may be cost-prohibitive. These cost prohibitions make the implementation of either PRV approach unfeasible. Faced with this dilemma, many contractors want a statistically based method that uses transect sampling and but does not necessitate that all anomalies are dug in each transect. We propose evaluating if such methods can be developed that are within a statistical framework beneficial to both DoD and the stakeholders.

8 COST ASSESSMENT

We document the costs to implement PRV methods as a final step of this demonstration. It should be noted that the VSP software that facilitates PRV method implementation is freely available at no cost and can be downloaded from http://vsp.pnnl.gov .

8.1 COST MODEL

Table 15 lists the costs associated with the application of a PRV project within a remediated site. These costs do not include equipment purchase, mobilization, maintenance, and demobilization as those would have generally been included in the original remediation costs. We should highlight that PRV approaches on previously remediated sites may have higher costs per unit if the transects and anomalies are sparsely spaced over the entire remediated area, thereby requiring a lot of driving between transect or anomaly locations. However, the limited amount of site to be surveyed and digs required should more than offset the potential increases in cost per unit. For example, on this DRI site demonstration, there were only 17 anomalies identified to be dug with about 15% of the site being resurveyed.

Table 15. Potential costs of using VSP

Cost Element	*Data*	*Cost*
Use of VSP	Obtaining the software and learning how to use it 1. VSP Training hours 2. VSP Training costs 3. VSP Software	Costs limited to learning PRV dialogs and use. 1. 4-8 hours 2. Free 3. Free
Data Quality Objectives Design	Using VSP to develop the implemented surveyed transects for DQO.	• From one to two stakeholder meetings. • VSP use and transfer of design to survey team is minimal (< 2 hours)
Geophysical Survey Costs	Costs include all events of a daily routine (i.e., QC, morning meetings, etc). Both the equipment and maintenance are not included in the costs. 1. Cost per linear kilometer for conducting geophysical surveys.	Costs will be different from a 100% survey as the transects are spread randomly over a large area. 1. $400 per km $644 per line mile

	2. km per hour	2. 0.83 hours per km (acquisition speed was between 3 to 4 km an hour)
Interpretation/Analysis of Geophysical Survey Data	Average cost in time of evaluating the geophysical surveys per anomaly.	1. $6 per anomaly 2. 3 minutes
Digging and Anomaly Identification Costs	Cost per anomaly: Costs include QC, equipment preparation, morning meetings along with digging and identifying anomalies. 1. Cost per anomaly 2. Time per anomaly	1. $300 per anomaly 2. 30 minutes per anomaly

8.1.1 Use of VSP and DQO Design

The VSP software has many different methods and tools. To gain a basic understanding of the range of tools and how to use each could require up to two weeks of training. However, a user interested in only the PRV tools within VSP could reasonably learn how to implement the methods in 4-8 hours. The VSP website (vsp.pnl.gov) provides guidance documents and users manuals. There are instructional materials built into VSP using the expert mentor functionality. We also provide training courses free of charge to the students. The VSP software is free and can be downloaded from our website as well.

The specific DQO implementation costs would add one or two additional meetings to a project. However, it is likely that these meetings would be combined with other meetings associated with the remediation. The actual time required to use VSP to build the PRV design and transfer the design to the survey team is minimal (< 2 hours).

8.1.2 Geophysical Survey, Interpretation, and Digging Costs

The specific costs are listed in Table 15. These per unit costs are meant to include the costs associated with the PRV implementation. The time and monetary costs include such things as morning kickoff meetings, daily QC, daily GPO, oversight, etc. Project management, project support, reports, work plans, GPO use, and GIS are examples of items that are not included in the costs. As noted in Table 15, the actual transect acquisition speeds would be closer to 3 to 4 km/hr. Additionally, the time to specifically dig each anomaly would be shorter than reported in the table as the reported time includes the time required to move between and locate the sparsely located anomalies.

8.2 Cost Drivers

An ideal verification method would use a third party to do the evaluation. However, bringing in a third party to perform the PRV work would require additional costs of another contract, mobilization, and demobilization. If sufficient controls are in place, the team that performed the original remediation work could be tasked to do the PRV study. In this demonstration, the remediation contractor did perform the PRV work as well.

8.3 Cost Benefits

As with any verification QA process, additional costs are accrued in performing the additional surveys and anomaly digging. However, the primary benefit is additional assurance and greater confidence that the remediation process was effective and properly controlled. It is also beneficial to conduct this post-remediation verification sampling before completely demobilizing the remediation efforts. If inconsistencies are found in the previous remediation work, significant savings can occur by addressing the problem while all tools are currently mobilized and before the land is approved for development. It is always cost-effective to identify deficiencies early to help avoid litigation and distrust if those deficiencies are identified later. One might also argue that a verified remediation process could add value to the property and provide greater confidence in future remediation projects.

Through this demonstration, PRV sampling has been shown to be a cost-effective technique for evaluating the effectiveness of past remediation processes. Through this demonstration, the remediation performed on the DRI South Tract was proven to be effective and adequately controlled.

9 REFERENCES

Bowen, W., and Bennett, C. (1987). *Statistical Methods for Nuclear Material Management.* Washingtion: Nuclear Regulatory Commission.

Department of Defense. (1996). *U.S. Department of Defense Test Methods Standard, DoD Preferred Methods for Acceptance of Product, MIL-STD 1916.* Washington: U.S. Department of Defense, U.S. Printing Office.

Department of the Navy. (2009). *Explosives Safety Submission: Non-Time Critical Removal Action, Navy/Denver Research Institute.* Silverdale, WA: Department of the Navy, Naval Facilities Engineering Command Northwest.

Gilbert, R., Wilson, J., O'Brien, R., Carlson, D., Pulsipher, B., and Bates, D. (2003). *Version 2.0 Visual Sample Plan (VSP): UXO Module Code Description and Verification.* Richland: Pacific Northwest National Laboratory.

Hahn, G. J., and Meeker, W. Q. (1991). *Statistical Intervals: A Guide for Practitioners*, New York: Jonh Wiley and Sons, Inc.

Hathaway, J., Gilbert, R., and Wilson, J. P. (2009). Evaluation of spatially clustered ordnance when using compliance sampling surveys after clean-up at military training sites. *Stochastic Environmental Research and Risk Assessment , 23* (2), 253-261.

Intergovernmental Data Quality Task Force. (2005). *Uniform Federal Policy for Quality Assurance Project Plans: Part 2B, Quality Assurance/Quality Control Compendium: Minimum QA/QC Activities.* DoD. Washington, D.C.: DoD.

ITRC Unexploded Ordnance Team. (2008). *Quality Consideration for Munitions Response Projects.* Washington, D.C.: Interstate Technology and Regulatory Council.

Jaech, J. (1973). *Statistical Methods in Nuclear Material Control.* Oakridge: Technical Information Center, Office of Information Services, U.S. Atomic Energy Commission.

Malcom Pirnie, Inc. (2008). *Final Preliminary Assessment / Site Inspection Report: Navy /Denver Research Institute (DRI) Site Former Lowry Training Annex, Colorado.* Sliverdale, WA: Naval Facilities Engineering Command Northwest.

Matzke B., Wilson J., Nuffer L., Dowson S., Hathaway J., Hassig N., Sego L., Murray C., Pulsipher B., Roberts B, McKenna S. (2010). *Visual Sample Plan Version 6.0 User's Guide*. PNNL-19915, Pacific Northwest National Laboratory, Richland, Washington.

MIL-STD-1916. (1996). *DOD Test Method Standard: DOD prefereed Methods for acceptance of product.*

Shilling, E. (1978). A Lot Sensitive Sampling Plan for Compliance Testing and Acceptance Inspection. *Journal of Quality Technology , 10* (2), 47-51.

Shilling, E. (1982). *Acceptance Sampling in Quality Control.* New York: Marcel Dekker.

Sky Research. (2009). *Final Former Lowry Training Annex - East Extension Area Post Removal Verification Report.* Omaha: U.S. Army Corps of Engineers Omaha District.

Squeglia, N. (1994). *Zero Acceptance Number Sampling Plans.* Milwakee: American Society for Quality Press.

Tetra Tech EC, Inc. (2010a). Explosives Safety Submission, *Non-Time Critical Removal Action Navy/Denver Research Institute, Former Lowry Training Annex, Aurora, Colorado*. Washington: Naval Facilities Engineering Command Northwest.

Tetra Tech EC, Inc. (2010). *Non-Time Critical Removal Action Work Plan*. *Navy/Denver Research Institute, Former Lowry Training Annex, Aurora, Colorado*. Washington: Naval Facilities Engineering Command Northwest.

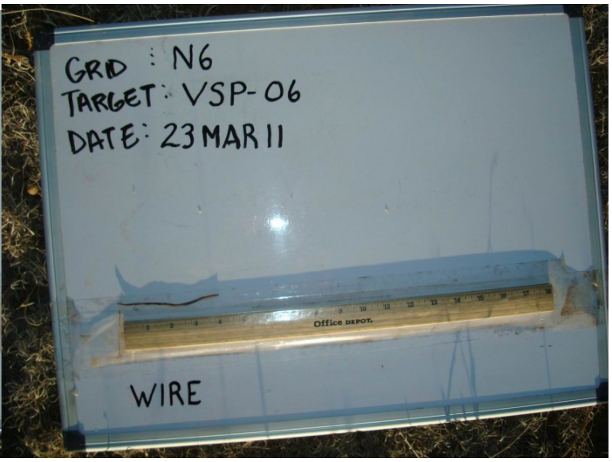
Williams, G. (2003). *Echo/Foxtrot Sector Completion Report, Beacon Point Ordnance Clearance.* Lakewood: Fostwer Wheeler Environmental Corporation.

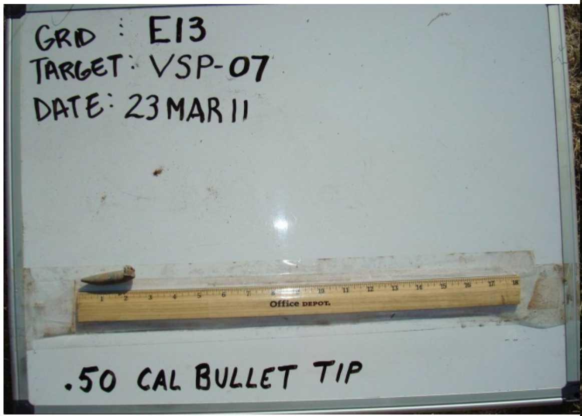
Appendix A: Points of Contact

POINT OF CONTACT Name	ORGANIZATION Name Address	Phone	Role in Project
Brent Pulsipher	PNNL	(509) 375-3989	Principal Investigator
John Hathaway	PNNL	(509) 372-4970	Demonstration Lead
John Wilson	PNNL	(970) 270-2998	VSP Lead Programmer
Justin Peach	NAVFAC	(360) 396-0082	Navy/DRI Project Lead
Michael McGuire	Tetra Tech EC	(303) 988-2202	Contractor Project Lead
Ken Vogler	Colorado Department of Health	(303) 692-3383	Regulator
Jeff Swanson	Colorado Department of Health	(303) 692-3416	Regulator
David Rathke	EPA Region 8	(303) 312-6016	Regulator

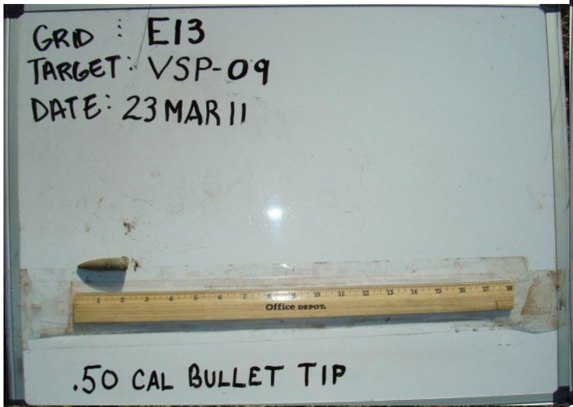
Appendix B: Images of Investigated Anomalies from Verification Survey

The 30 images below document the items dug as a part of the transect-PRV survey work. The photo ID column in Table 11 matches the Target ID listed in each picture. Table 11 also provides additional information about each item.

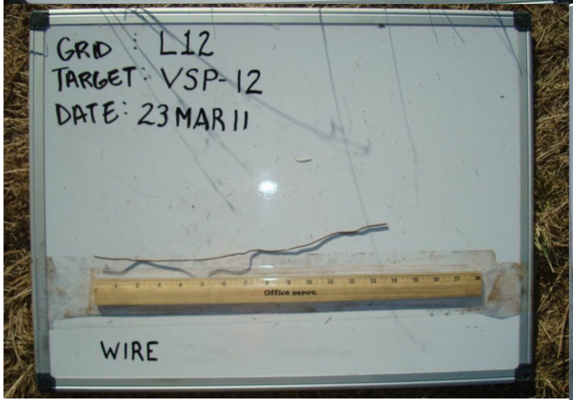



GRID : E13
TARGET: VSP-07
DATE: 23 MAR 11
Office DEPOT
.50 CAL BULLET TIP

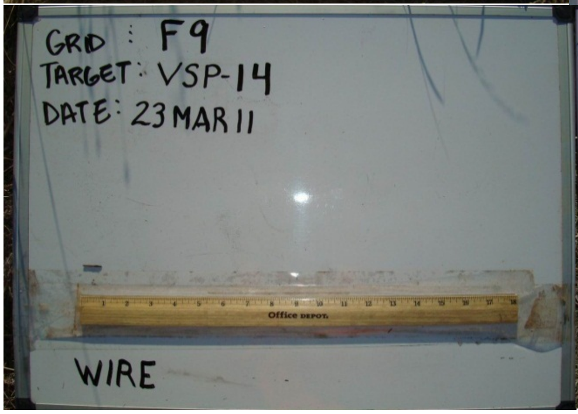

GRID : K6
TARGET: VSP-08
DATE: 23 MAR 11
Office DEPOT
PHOTO FLASH END CAP


GRID : E13
TARGET: VSP-09
DATE: 23 MAR 11
Office DEPOT
.50 CAL BULLET TIP


GRID : G14
TARGET: VSP-11
DATE: 23 MAR 11
Office DEPOT
FRAG


GRID : L12
TARGET: VSP-12
DATE: 23 MAR 11
Office DEPOT
WIRE


GRID : H13
TARGET: VSP-13
DATE: 23 MAR 11
Office DEPOT
CROW


GRID : F9
TARGET: VSP-14
DATE: 23 MAR 11
Office DEPOT
WIRE


GRID : B15
TARGET: VSP-15
DATE: 23 MAR 11
Office DEPOT
CROW/FRAG/FRAG SQUARE


GRID : D13
TARGET: VSP-16
DATE: 23 MAR 11
Office DEPOT.
CROW

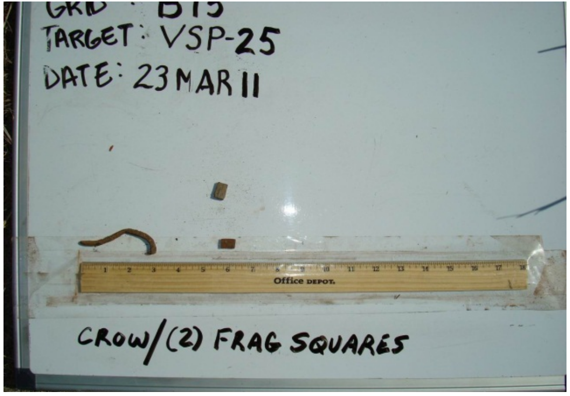

GRID : D13
TARGET: VSP-17
DATE: 23 MAR 11
Office DEPOT.
20MM FRAG

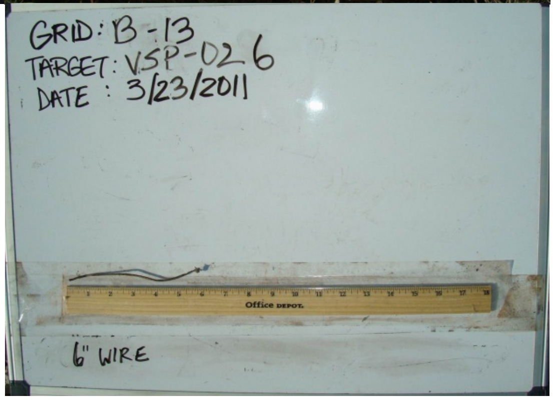

GRID : M13
TARGET: VSP-21
DATE: 23 MAR 11
Office DEPOT.
WIRE


GRID : B14
TARGET: VSP-22
DATE: 23 MAR 11
Office DEPOT.
CROW/FRAG

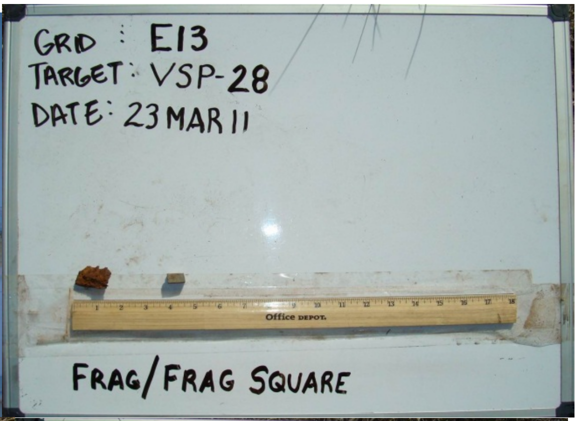

GRID: B-13
TARGET: VSP-023
DATE : 3/23/2011
Office DEPOT.
CROW

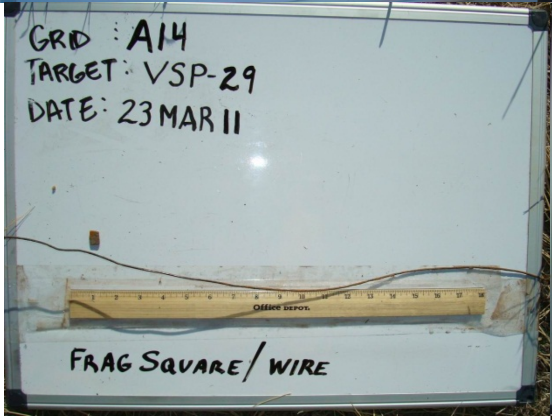

GRID : A16
TARGET: VSP-24
DATE: 23 MAR 11
Office DEPOT.
CROW/7.62 CASING

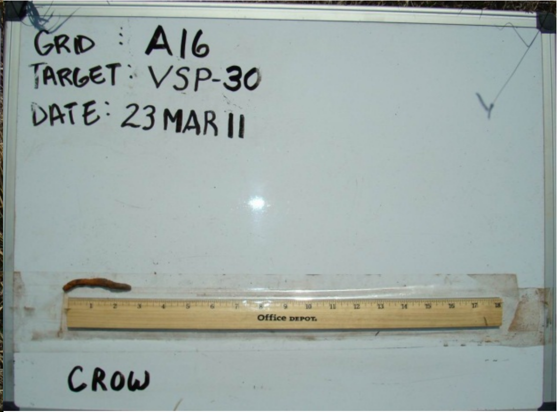

GRID : B15
TARGET: VSP-25
DATE: 23 MAR 11
Office DEPOT.
CROW/(2) FRAG SQUARES


GRID: B-13
TARGET: VSP-026
DATE : 3/23/2011
Office DEPOT.
6" WIRE


GRID : O5
TARGET : VSP-27
DATE : 23 MAR 11
WIRE


GRID : E13
TARGET : VSP-28
DATE : 23 MAR 11
FRAG/FRAG SQUARE


GRID : A14
TARGET : VSP-29
DATE : 23 MAR 11
FRAG SQUARE/WIRE


GRID : A16
TARGET : VSP-30
DATE : 23 MAR 11
CROW


GRID : E15
TARGET : VSP-35
DATE : 23 MAR 11
ALUMINUM SCRAP


GRID : M13
TARGET : VSP-49
DATE : 23 MAR 11
(10) BARB WIRE


GRID : H10
TARGET : VSP-76
DATE : 23 MAR 11
WIRE

Appendix C: Derivation of the mV Signal Upper Tolerance Limit

Lognormal model. The lognormal distribution is useful for characterizing the distribution of measurements with multiplicative error structure. The defining feature of the lognormal distribution is that the (natural) logarithmic transform has the normal (Gaussian) distribution. We say that a random variable or measurement X has the lognormal distribution with parameters μ and σ^2 if $Y = \log(X)$ has the normal distribution with mean μ and variance σ^2. The mean and variance of X are $\exp\{\mu + \sigma^2/2\}$ and $\exp\{2\mu + \sigma^2\}(\exp\{\sigma^2\} - 1)$, respectively. The median of X is $\exp\{\mu\}$.

There are two ways of expressing the relative standard deviation (RSD) of lognormally distributed measurements. The first way is to divide the standard deviation of X by its mean:

$$RSD_1 = \sqrt{\exp\{\sigma^2\} - 1}\,. \tag{1}$$

The second way is to simply use the standard deviation of the logarithm of X:

$$RSD_2 = \sigma\,. \tag{2}$$

These two expressions are nearly equivalent for small values of σ, that is, small relative standard deviations. Because it is more straightforward to estimate (from Y values) and simplifies the theoretical development in what follows, the second expression will be used hereafter.

Estimating the relative standard deviation. Suppose experimental data are available in which m items with varying true values are each measured n times. Denote these measurements as X_{ij}, the j-th measurement of the i-th item. Let $Y_{ij} = \log(X_{ij})$. Then the estimates of the individual item RSDs are

$$\hat{\sigma}_i = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_i)^2}\,, \tag{3}$$

where $\bar{Y}_i = (1/n)\sum_{j=1}^{n} Y_{ij}$. The pooled estimate of the RSD is

$$\hat{\sigma} = \sqrt{\frac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_i)^2}\,. \tag{4}$$

Probabilistic bounds on measurements. Suppose a specific item with true median value of $\exp\{\mu_0\}$ is to be measured. That is, with 50/50 odds, a measurement X of the item will be less than/greater than $\exp\{\mu_0\}$. The logarithmic transform Y will have mean μ_0 and variance σ^2. For an upper bound, with probability $100\beta\%$, Y will be less than $\mu_0 + Z_\beta \sigma$, where Z_β is the $100\beta\%$ percentile of the standard normal distribution (for example, $Z_{0.99} = 2.3263$).

If σ is unknown, then it must be estimated and the associated uncertainty must be incorporated into the probabilistic upper bound. The pooled estimate in Equation (4) is known to have a $100(1-\alpha)\%$ confidence upper bound of

$$\hat{\sigma}\sqrt{\frac{\nu}{\chi^2(\alpha;\nu)}}\,, \tag{5}$$

where $\chi^2(\alpha;\nu)$ is the $100\alpha\%$ percentile of the chi-square distribution with $\nu = m(n-1)$ degrees of freedom (see Hahn and Meeker 1991). Therefore, with 95% confidence, 99% of measurements of the specific item will have Y less than

$$\mu_0 + Z_{0.99}\hat{\sigma}\sqrt{\frac{\nu}{\chi^2(0.05;\nu)}}\,. \tag{6}$$

The factor multiplying $\hat{\sigma}$ in Equation (5) is tabled below for 95% confidence ($\alpha = 0.05$) and various values of the degrees of freedom. Back in the original measurement space, with 95% confidence, 99% of measurements of the specific item will have X less than

$$\exp\left\{\mu_0 + Z_{0.99}\hat{\sigma}\sqrt{\frac{\nu}{\chi^2(0.05;\nu)}}\right\}. \tag{7}$$

Table 1: 95% Confidence Scaling Factors

Degrees of Freedom	Scaling Factor
5	2.089257
10	1.593072
15	1.437306
20	1.357638
25	1.308049
30	1.273682
35	1.248190
40	1.228375
50	1.199274
60	1.178676
70	1.163158
90	1.141039
100	1.132789
300	1.072363
600	1.050036
1000	1.038295
2000	1.026757

Reference

Hahn, G. J., and Meeker, W. Q. (1991). *Statistical Intervals: A Guide for Practitioners*, New York: Jonh Wiley and Sons, Inc.