PNNL-13991

Version 2.0 Visual Sample Plan (VSP): Models and Code Verification

R. O. Gilbert
J. E. Wilson
R. F. O'Brien
D. K. Carlson
D. J. Bates
B. A. Pulsipher
C. A. McKinstry

August 2002

Prepared for the
U.S. Environmental Protection Agency
under a Related Services Agreement
with the
U.S. Department of Energy
under Contract DE-AC06-76RL01830

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC06-76RL01830

PNNL-13991

Version 2.0 Visual Sample Plan (VSP): Models and Code Verification

R. O. Gilbert
J. E. Wilson
R. F. O'Brien
D. K. Carlson
D. J. Bates
B. A. Pulsipher
C.A. McKinstry

August 2002

Prepared for the
U.S. Environmental Protection Agency
under a Related Services Agreement
with the
U.S. Department of Energy
under Contract DE-AC06-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Summary

Visual Sample Plan (VSP) is an easy-to-use visual and graphic software tool being developed by the Pacific Northwest National Laboratory (PNNL) to select the right number and location of environmental samples so that the results of statistical tests performed to provide input to environmental decisions have the required confidence and performance. It is a significant help in implementing the Data Quality Objectives (DQO) planning process that was developed by the U. S. Environmental Protection Agency (EPA).

Gilbert et al. (2001) documented the quality assurance (QA) procedures that were conducted to assure that Version 0.91 of VSP was operating correctly. Subsequently, Version 0.91 was renamed Version 1.0 and placed on the internet at http://dqo.pnl.gov/vsp . Since that time VSP has been enlarged and improved and is now available as Version 2.0. The current document is an expansion of Gilbert et al (2001) to include the QA procedures and testing that were conducted to assure the validity and accuracy of the new features added to Version 1.0 to obtain Version 2.0.

Features added to Version 1.0 include the following:

- Methods for determining the number of samples needed for three additional sampling designs:
  - Stratified Sampling (StS) to estimate a weighted mean or proportion over several subareas (strata) of the study area
  - Ranked Set Sampling (RSS) to estimate the mean of a population based on data from field locations that are selected using simple random sampling in combination with either expert opinion or quantitative screening measurements in the field
  - Adaptive Cluster Sampling (ACS), a phased sampling design, to obtain data to 1) estimate the mean of a population that contains one or more local areas of elevated contamination, and a 2) determine the boundary of the local areas of elevated contamination that are detected during the first phase of sampling
- Sequential sampling, wherein sampling and testing the sample mean against a threshold value is conducted repeatedly over time, each time combining new data with the previous data, until a decision can be made with specified confidence. Two statistical tests are implemented: (1) the Sequential Probability Ratio Test (SPRT), which is used when the standard deviation of the data is known, and (2) Barnard's sequential t-test, which is used when the standard deviation is not known.
- Transect (swath) sampling in which a field detector (e.g., radiation detector or a geophysical sensor of unexploded ordnance) makes measurements from the ground or from aircraft along transects laid out in a parallel, square, or rectangular grid design across the study area to search for objects or areas of contamination. Features include:
  - Calculation of the probability of traversing and detecting a target area of specified size, shape, and density of objects using the transect design developed in VSP
  - Probability that a target area exists even though none was found using the transect design
- Computing the number of transects that should be surveyed to achieve a specified confidence that the fraction of detective transects (e.g., those that contain unexploded ordnance) is less than a specified value
- A VSP-generated summary report of the sampling design developed by the VSP user. This report is available in Version 2.0 for most of the sampling designs in VSP.

- New help features including 1) a “Welcome to Visual Sample Plan” screen that provides quick tips on how to get started using VSP and 2) a “VSP Advisor” that gives guidance on using VSP and what it does
- Ability to add a “historical” flag to previously obtained samples that are placed on the study area map by the VSP user
- Improvements in operating with the MAP view that displays the VSP user’s site map and study areas

For each sampling design in Version 2.0 of VSP this report contains 1) the formulas and methods used to obtain the recommended number of samples, 2) the assumptions, technical basis, and scientific source (e.g., peer-reviewed papers and books) of the design, and 3) the QA activities conducted (e.g., independent hand and computer computations) to verify that the models and methods used are correctly programmed and implemented in VSP.

Acknowledgments

The authors are pleased to acknowledge the assistance of the following staff of the Research Triangle Institute in developing Version 2.0 of VSP: Lorraine Gallego for conducting quality assurance activities to verify that Version 2.0 is correctly computing the number of samples for most of the newly added designs, and Kara Morgan for her development of the "VSP Advisor" and for her comments and suggestions for improving the final product. In addition, we wish to thank John Warren of the Quality Staff, U.S. EPA, Office of Environmental Information, for his insight in how to make VSP more user friendly, and Tony Jover and Larry Zaragoza of the U.S. EPA, Office of Solid Waste and Emergency Response, for their continued support and interest in a high quality product. We also wish to thank David Bottrell, U.S. Department of Energy, for his continued support of VSP developments. A special thanks is also extended to the following individuals in the Statistical and Quantitative Sciences Group at PNNL: Stacey A. Hartley for assistance in developing the design report outputs of VSP; Lucille A. Walker for her project financial accounting support; and Mary H. Cliff for her assistance in preparing the final report.

Contents

Figure

Tables

Abbreviations and Acronyms

ACS	Adaptive Cluster Sampling
DOE	U.S. Department of Energy
DoD	U.S. Department of Defense
DQO	Data Quality Objectives
DPGD	Decision Performance Goal Diagram
DXF	Data Exchange Format
EPA	U.S. Environmental Protection Agency
GIS	Geographic Information System
GPS	Global Positioning System
MARSSIM	Multi-Agency Radiation Survey and Site Investigation Manual
MQO	Measurement Quality Objectives
OEI	Office of Environmental Information, U.S. Environmental Protection Agency
PNNL	Pacific Northwest National Laboratory
QA	Quality Assurance
QAPP	Quality Assurance Project Plan
RSS	Ranked Set Sampling
RTI	Research Triangle Institute
SAS	Statistical Analysis Systems (software code)
SRS	Simple Random Sampling
SPRT	Sequential Probability Ratio Test
StS	Stratified Sampling
TS	Transect Sampling
UXO	Unexploded Ordnance

VSP	Visual Sample Plan
WRS	Wilcoxon Rank Sum Test

1.0 Introduction

Visual Sample Plan (VSP) is a software tool under development at the Pacific Northwest National Laboratory (PNNL) with support from the U.S. Department of Energy (DOE), the U.S. Environmental Protection Agency (EPA), and the U.S. Department of Defense (DoD). VSP is used to select the right number and location of environmental samples in order to achieve various sampling objectives with required performance. Version 2.0 of VSP provides formulas and statistical algorithms to compute the number of samples needed for specific statistical tests, estimations, an, evaluations appropriate for the following environmental sampling goals:

- comparing an average or proportion to a fixed threshold value
- comparing an average or proportion to a reference area average or proportion
- constructing a confidence interval on a mean
- estimating a mean or a proportion within specified cost and variance constraints
- locating a hot spot
- finding a UXO target area (not yet peer-reviewed)

An overview of Version 2.0 is provided in Section 2.0. Section 3.0 presents (in Section 3.1) the equations, computation methods, and underlying assumptions used to compute the required number of samples (sample size) and other quantitative outputs in Version 2.0, as well as (in Section 3.2) the activities, computations, and other quality checks conducted by PNNL and the Research Triangle Institute (RTI) to:

- verify the technical and scientific basis of the sample-size equations and other computational algorithms used,
- verify that those equations and computational methods are correctly programmed and implemented, and
- evaluate the correctness of the non-statistical elements of the user interface and input/output procedures.

The verification activities and documentation of the non-statistical components of VSP are provided in Section 4.0. Section 5.0, which provides the reference list, is followed by the Appendices.

2.0 Visual Sample Plan Software Overview

Visual Sample Plan (VSP) is an easy-to-use, visual and graphic software tool. It is a significant help in implementing the sixth and seventh steps of the data quality objectives (DQO) planning process ("Specify Tolerable Limits on Decision Errors" and "Optimize the Design for Obtaining Data," respectively). These steps of the DQO process are needed to determine the number and field location of samples required for the sampling objective of interest, such as comparing a mean to a fixed threshold value, comparing a mean to a reference area mean, estimating a mean with specified accuracy and confidence within cost constraints, finding hot spots of a specified size and shape, and finding target areas that may contain unexploded ordnance (UXO). All of these objectives are included in Version 2.0 of VSP.

VSP is designed primarily for project managers and other environmental professionals who may not have extensive statistical training, although individuals with statistical expertise will also find the program useful. VSP is applicable to any two-dimensional geographical population to be sampled, including surface soil, a defined layer of subsurface soil, building surfaces, water bodies, or other similar applications.

When VSP is opened, the *Welcome to Visual Sample Plan* and *the VSP Advisor* screens appear that provide help on how to get started using VSP. These screens provide answers to questions like "What will VSP do?", "How do I get started using VSP?", and "How does VSP fit into the DQO process?". Then the user imports or draws a map of the facility that contains the geographical areas to be studied. After the VSP user selects the particular study areas of interest on the site map, the sampling goal of interest is selected from the *sampling goals* pull-down menu. The next step is to enter design specifications (DQOs) into the design dialogue box of the selected sampling goal. VSP uses these specifications (performance requirements) to determine the recommended number and location of samples (or spacing of sampling transects) in the field that are needed to achieve the desired performance. The output of VSP takes several forms, some of which are shown in Figure 2.1. This figure displays:

- the required number and location of sampling locations,
- the Decision Performance Goal Diagram (DPGD), which shows the VSP user's acceptable probabilities of making decision errors using a statistical test,
- a report that summaries the sampling design constructed by VSP based on user DQOs, and
- a list of the geographical sampling locations determined by VSP.

All of these outputs can be saved, copied, and pasted in project reports such as the Quality Assurance Project Plan (QAPP) and the Sampling and Analysis Plan, as well as in presentation slides and technical papers. The sampling locations can be exported as a Data Exchange Format (DXF) file for use in a geographical information system (GIS) or global positioning system (GPS) software package.

The sample locations at the study site are determined by the specific sampling goal and sampling design selected by the VSP user. Version 2.0 of VSP permits the user to select the following designs: simple random sampling (SRS), systematic (grid) sampling, stratified sampling (StS), ranked set sampling (RSS), adaptive cluster sampling (ACS), transect sampling (TS), and judgment (authoritative) sampling.

VSP contains two types of random number generators which are used to determine randomly selected sampling locations at the study site: 1) a pseudo-random number generator, for which each potential location has an equal and independent chance of being selected, and 2) a quasi-random number

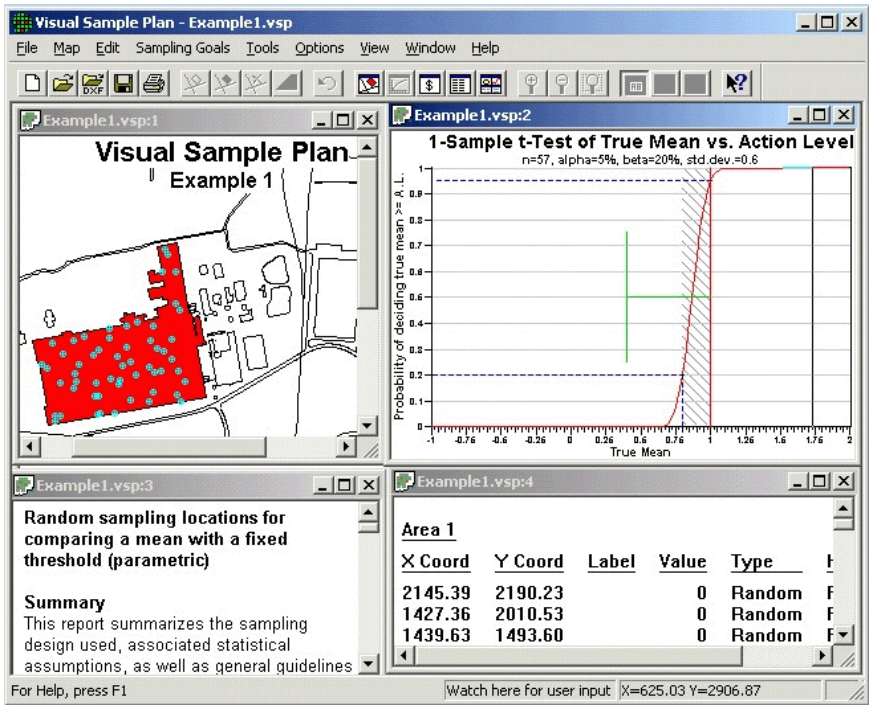


Figure 2.1. VSP Quad Window Display

generator (Press et al. 1992), for which locations are chosen to be somewhat more evenly spaced than what is typically obtained using the pseudo-random number generator. Version 2.0 also enables the user to add sample points to a current design either manually (subjectively) or using the Adaptive Fill algorithm. The Adaptive Fill algorithm adds data points in such a way as to avoid existing sampling locations.

Version 2.0 of VSP also contains a Measurements Quality Objectives (MQO) module that enables the user to assess the cost and benefits of two alternative analytical measurement protocols. The MQO allows the user to evaluate the tradeoffs between taking more samples or performing replicate analyses on each sample. The MQO module will compute the number and cost of field samples required for a specified analytical method when the user specifies the number ($r = 1, 2, \ldots m$) of replicate analyses being considered, the variability in the data due to only the analytical measurement process, and the variability due to sampling and all other sources of variability that enter prior to the analytical measurement process.

3.0 Verification and Documentation of Statistical Methods and Computations

The activities conducted in evaluating the statistical methods and computations contained within Version 2.0 of VSP are documented in this section.

Two criteria were used to assess the scientific basis of the sample-size equations in VSP:

- Were the sample-size equations and other statistical algorithms used published in peer-reviewed scientific publications?
- Did the EPA or other government or regulatory agency publications and guidance documents recommend the sample-size equations and algorithms for environmental applications?

The sample-size equations and algorithms in Version 2.0 of VSP are valid if the assumptions used in deriving them are valid for the study area. These assumptions are documented in Section 3.1. The user of VSP must determine if the assumptions are reasonable for the site of interest and seek advice from a statistical expert if assumptions are seriously violated.

To verify that the sample-size equations were correctly programmed and provide correct results, the number of samples computed using the equations in VSP were compared with hand calculations as well as those obtained by an independent exercise of the equations using an independent computer code. An S-PLUS [(R)] computer code was constructed for this purpose, as discussed in Section 3.2.1. Also, SAS (Statistical Analysis System) codes were written to verify VSP computations for sequential testing (Sections 3.2.4 and 3.2.5). Other such comparisons are discussed in Section 3.2.2. The accuracy of the VSP calculations of the number of samples needed to detect hot spots was also checked, as discussed in Section 3.2.3. Finally, the correctness of the four output "views" in VSP (see Figure 2.1) were checked for accuracy, as discussed in Section 3.2.6.

3.1 Technical Basis of Sample-Size Equations and Algorithms

Each of the subsections in this section correspond to one of the sampling goals listed in the "Sampling Goals" pull-down menu in Version 2.0. Within each subsection PNNL's and/or RTI's quality checking of the VSP sample-size equations and other algorithms is provided. Some of the common mathematical notation used in these equations and algorithms is defined in Table 3.1. Other mathematical notation is provided throughout the report as needed.

For the Measurement Quality Objectives (MQO) module, VSP uses a two-component additive variance model. In this model the total random variance of the measurements is equal to the sum of the sampling variance and the analytical variance. Any random variations that arise from activities outside the analytical laboratory are part of the sampling variance, σ^2_{sample}. Any random variations that occur within the analytical laboratory are part of the analytical variance , $\sigma^2_{analytical}$.

The sampling variance could include variations due to spatial differences, sample handling, field

[(R)] S-PLUS is a registered trademark of the Insightful Corporation, Seattle, Washington.

sampling methods and instruments, or field activities or other similar activities occurring outside the analytical laboratory. The analytical variance includes variations among r different portions of the field sample that undergo separate analytical treatment in the laboratory. Thus, the analytical variance may include variations due to sample preparation, dilution, fusion, sample splitting, calibration, measurement, or instrument variations or similar variations that are associated with the analytical laboratory. In the VSP variance model, it is assumed that the entire analytical process in the laboratory, from sample aliquoting and preparation through instrument measurements, is repeated for each of the r portions of the field sample. The VSP MQO module applies to any two-component variance model as long as the definitions of the r different portions of the field sample and the estimated analytical variance ($s^2_{analytical}$) are consistent with the VSP variance model. The VSP MQO module can be used when the estimated analytical variance ($s^2_{analytical}$) supplied by the VSP user includes only those parts of the total variance that are added during the analytical process applied to each of the r portions (analytical replicates) of a field sample.

Table 3.1. Some Common Notation for Sample-Size Equations Used in Version 2.0 of VSP

Notation	Description
n	the minimum recommended number of samples that should be collected from a site, as computed using one of the sample-size equations or algorithms in VSP
r	the number of measurements (analytical replicates) that will be obtained for each field sample
"	the probability that the VSP user is willing to tolerate that a Type I decision error will be made, i.e., that the data collected and used in the appropriate statistical test will falsely reject the null hypothesis. For example, if the null hypothesis is "the mean concentration at the site *exceeds* the action limit," then " is the probability the VSP user can tolerate that the statistical test computed using the n data will incorrectly indicate that the mean concentration does not exceed the action limit, in short, calling a dirty site clean.
$	the probability the VSP user is willing to tolerate that a Type II decision error will be made, i.e., that the data collected and used in the appropriate statistical test will falsely accept the null hypothesis. For example, if the null hypothesis is "the mean concentration at the site exceeds the action limit," then $ is the probability the VSP user can tolerate that the statistical test computed using the n data will falsely indicate the mean concentration does exceed the action limit, in short, calling a clean site dirty.

Table 3.1. Some Common Notation for Sample-Size Equations (cont.)

Notation	Description
$)$	the width of the "gray region" in the Decision Performance Goal Diagram (DPGD) used in the DQO process (EPA 2000a) and in VSP. For example, if the sampling objective is to compare the true mean of the site to the true mean of a background area, then $)$ is the difference between the true site mean and the true background mean that the VSP user specifies is important to detect with (high) probability 1 - β. Similarly, if the objective is to compare the mean of the site to a fixed action limit, then Δ is the difference between the true mean and the action limit that is important to detect with (high) probability 1 - β.
σ^2_{total}	the true "total variance" of the population of all possible measurements made on all possible samples collected from the study site. The model of the true total variance used in Version 2.0 of VSP is $\sigma^2_{total} = \sigma^2_{sample} + \sigma^2_{analytical}$, where $\sigma^2_{analytical}$ is the true variance component due to the analytical measurement process in the laboratory and σ^2_{sample} is the true variance component due to all other sources of variation, including variations in true concentrations at different study site locations and the variance added due to selecting, collecting, and transporting samples to the laboratory.
s^2_{total}	the computed *estimate* of the true total variance, σ^2_{total}. If r = 1 for all n field samples, then the quantity s^2_{total} is computed as $S^2_{total} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$, where x_i is the measurement obtained for the single aliquot from the ith field sample and $\bar{x}$ is the arithmetic mean of the n measurements, x_i.
s^2_{sample}	an estimate of the total variance of the data x_i that would be obtained if $\sigma^2_{analytical} = 0$
$s^2_{analytical}$	an estimate of the total variance of the data x_i that would be obtained if the only variability in the data were due to the analytical process, i.e., if $\sigma^2_{sample} = 0$
$t_{1-\alpha,df}$	the value of the Student's t-distribution with df degrees of freedom. By definition, the proportion of the distribution to the left of the value $t_{1-\alpha,df}$ is 1-". A table of the values of $t_{1-\alpha,df}$ is found in most statistics books, e.g., Gilbert (1987, Table A2).

Table 3.1. Some Common Notation for Sample-Size Equations (cont.)

Notation	Description
$Z_{1-\alpha}$	the value from the standard normal distribution for which the proportion of the distribution to the left of $Z_{1-\alpha}$ is 1-". A table of the values of $Z_{1-\alpha}$ is found in most statistics books, e.g., Gilbert (1987, Table A1). If the selected probability of a false rejection, ", is made smaller, then $Z_{1-\alpha}$ will be larger, leading to larger sample sizes. If the null hypothesis is that concentrations at the site *exceed* the action limit, i.e., that the site is "dirty," then $Z_{1-\alpha}$ can be thought of as an index number whose magnitude quantifies the strength of our desire to avoid deciding a dirty site is clean.
$Z_{1-\beta}$	the value of the standard normal distribution for which the proportion of the distribution to the left of $Z_{1-\beta}$ is 1-$. If the selected probability of a false acceptance, $, is made smaller, then $Z_{1-\beta}$ will be larger, leading to larger sample sizes. If the null hypothesis is that concentrations at the site *exceed* the action limit, i.e., that the site is "dirty," then $Z_{1-\beta}$ can be thought of as an index number whose magnitude quantifies the strength of our desire to avoid deciding a clean site is dirty.
$\Phi(z)$	the cumulative standard normal distribution function, i.e., $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}x^2} dx$. A table of $\Phi(z)$ values is provided in most statistics books, e.g., Gilbert (1987, Table A1).

3.1.1 Compare Average to a Fixed Threshold

Suppose a VSP user wants to determine the number of samples, n, that should be collected to use in a statistical test of whether the true average concentration for a specified geographical area exceeds a specified risk-based or regulatory-based threshold (upper limit). Version 2.0 of VSP provides several different equations and algorithms for computing n. The method used is selected by the VSP user on the basis of the shape of the underlying distribution of the data to be collected and the particular statistical test that will be used.

If the data are expected to be normally distributed, then Version 2.0 provides methods for determining the number of samples for the one-sample t test, the Sequential Probability Ratio Test (SPRT), and Barnard's Sequential Test. If the data are not expected to be normally distributed, then sample-size methods for the Wilcoxon Signed Ranks Test and the MARSSIM Sign test are provided.

The sample-size equation or algorithm for each test is provided below, along with the assumptions that underlie the use of that test. An assumption common to all tests is that the measurements at different locations are not correlated.

3.1.1.1 Data Are Normally Distributed

One-Sample t-Test

The one-sample t-test can be used to test if the true mean of the population exceeds a fixed upper limit. The equation used in Version 2.0 of VSP to compute the minimum recommended number of samples, n, needed for the test when the VSP user specifies that only r = 1 analytical replicate from each sample will be made is

$$n = \frac{s_{total}^2 \left(Z_{1-\alpha} + Z_{1-\beta}\right)^2}{\Delta^2} + 0.5 Z_{1-\alpha}^2 \tag{3.1}$$

If the MQO module in Version 2.0 is used, in which case r = 1, 2, ... m analytical replicates of each field sample can be made, then the equation that is used to compute n is

$$n = \frac{\left(s_{sample}^2 + \frac{s_{analytical}^2}{r}\right)\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2}{\Delta^2} + 0.5 Z_{1-\alpha}^2 \tag{3.2}$$

The notation used in Equations (3.1) and (3.2) is defined in Table 3.1. These equations are computed using the values of α , β , Δ , r, s_{total} , s_{sample} , and $s_{analytical}$ that are specified by the VSP user.

The assumptions that underlie the derivation of Equation (3.1) are that the data are normally distributed and representative of the study site, they are not spatially or temporally correlated, and that $\bar{x} - k s_{total}$ is normally distributed with mean $\mu - k\sigma$ and variance $(\frac{\sigma^2}{n})(1 + \frac{k^2}{2})$, where k is a given constant (Guenther 1981). The derivation of Equation (3.1) is found in Wallis (1947), Guenther (1977), EPA (2000a, Appendix A), and EPA (1992, pp. F-8, F-9, and F-10). Equation (3.1) is used in the statistics book by Bowen and Bennett (1988, pp. 155, 156), EPA (2000a), and EPA (2000b, pp. 3-7).

Guenther (1981) indicates that although Equation (3.1) is an approximation to the true minimum sample size required for the one-sample t-test, Equation (3.1) usually yields the exact solution for n. The exact solution is obtained using an iterative approach using tables of the non-central t distribution found in Owen (1965).

Equation (3.2) also should provide a very accurate approximation of n for a specified value of r because it is a straightforward extension of Equation (3.1) to the case of r > 1. For this case, it is easily shown that the total variance is estimated by computing

$$s_{total}^2 = s_{sample}^2 + \frac{s_{analytical}^2}{r}$$

Inserting this equation for s^2_{total} into Equation (3.1) yields Equation (3.2).

Sequential Probability Ratio Test (SPRT)

Sequential sampling over time is provided in Version 2.0 of VSP to test if the true mean exceeds a specified threshold value. The sampling is sequential in the sense that several sequential (in time) trips to the study site may be necessary to collect a sufficient number of samples in order for the statistical test to have the power specified by the VSP user to decide between the null or alternative hypothesis. Sequential tests are useful if samples can be obtained and analyzed sequentially in near-real time, as may be the case when Dynamic Work Plans or Expedited Site Characterization processes are feasible. These tests may be particularly useful when in-field sampling and analysis methods are employed.

In this section, the Sequential Probability Ratio Test (SPRT) is discussed (Wald 1947; Wetherill 1966). This test requires that the standard deviation of the measurements be known with great accuracy before the sampling study is conducted. In the next section of this report, Barnard's sequential test is discussed (Barnard, 1952; Wetherill, 1966). Barnard's test can be used in place of the SPRT when the standard deviation is not known with great accuracy.

The assumptions that underlie the SPRT test are that the data are normally distributed and representative of the study site, the data are not spatially or temporally correlated, and the standard deviation of the data to be collected is known with great accuracy.

The SPRT test in VSP works as follows if a map with at least one selected study area is provided to VSP:

1. The VSP user inputs the DQO parameters (α , β , null hypothesis, width of the gray region, known standard deviation (σ), and the threshold value) in the design dialog box. Also, the number of samples to collect on each trip to the field is specified. When the "Apply" button in VSP is pressed, then VSP places the required number of sampling locations for the first field trip on the map. VSP also provides a listing of the geographical coordinates of the required samples.

2. The samples are collected and analyzed for the chemicals of concern.

3. The VSP user reopens the SPRT design dialog box, presses the "Input Values" button, and inputs the measurements obtained for the samples collected. The VSP user then closes the data input dialog box and VSP computes the mean and determines whether more samples are needed before a decision can be made by the SPRT whether to accept the null hypothesis or the alternative hypothesis. The sample mean is plotted on a decision graph in the VSP "Graph View" for ease of interpretation.

4. If more samples are needed, the Apply button is pressed and VSP places the additional sampling locations on the map (avoiding existing sampling locations) and provides the geographical coordinates of the new samples.

Steps 2 through 4 are repeated until there is enough data so that the SPRT can either accept the null hypothesis or the alternative hypothesis with the performance specified by the VSP user. It is useful to watch the VSP Graph View during the sequential sampling and testing process to see how close the SPRT is to making a decision.

If no sample areas on the site map have been selected by the VSP user, or no map is being used, then the VSP user can enter as few or as many data values as desired. Also, in this case, it is not necessary to close the design dialog box in order to enter more data values.

Suppose the null hypothesis selected by the VSP user is that the site is dirty. Then the SPRT determines that there is enough evidence to accept the null hypothesis if the mean of the sample values is greater than UL_d. Also, the SPRT determines that there is enough evidence to accept the alternative hypothesis if the mean of the sample values is less than LL_d, where

$$UL_d = AL - \frac{\Delta}{2} + \frac{\sigma^2 A}{\Delta n}$$

$$LL_d = AL - \frac{\Delta}{2} - \frac{\sigma^2 B}{\Delta n}$$

and

$$A = \ln\frac{1-\beta}{\alpha}$$

$$B = \ln\frac{1-\alpha}{\beta}$$

$\ln$ is the natural logarithm

α is the maximum acceptable Type I decision error rate (probability of rejecting the null hypothesis when the null hypothesis is true)

β is the maximum acceptable Type II decision error rate (probability of accepting the null hypothesis when the null hypothesis is false)

Δ is the width of the gray region in the Decision Performance Goal Diagram (DPGD)

σ^2 is the variance of the data, assumed to be known with great accuracy

n is the number of samples collected thus far and used in the SPRT

AL is the action limit (threshold value)

If the null hypothesis is that the site is clean, then the SPRT determines that there is enough evidence to accept the null hypothesis if the mean of the sample values is less than LL_c and that there is enough evidence to accept the alternative hypothesis if the mean of the sample values is greater than UL_c, where:

$$UL_c = AL + \frac{\Delta}{2} + \frac{\sigma^2 A}{\Delta n}$$

$$LL_c = AL + \frac{\Delta}{2} - \frac{\sigma^2 B}{\Delta n}$$

The notation for these equations is defined above.

VSP projects the number of additional samples needed to make a decision by using the following algorithm:

1. Increase the value of n by 1.
2. Recalculate the upper and lower bounds using the new value of n.
3. If the sample mean falls outside the new boundaries, then the increase in n is given by VSP as the number of additional samples needed.
4. Repeat Steps 1 through 3 up to 100 times

Barnard's Sequential t-Test

Barnard's sequential t-test can be used in place of the Sequential Probability Ratio Test (SPRT) discussed in the previous section when the standard deviation is not known with great accuracy. The assumptions that underlie Barnard's sequential t-test are that the data are sampled sequentially from a normal distribution and that the data are representative of the study site and are not spatially or temporally correlated.

The VSP user goes through the same steps to use the sequential t-test using the VSP dialogue box as was described above for the SPRT except that the user must initially supply the measurements for 10 samples collected at random from the study site of interest. VSP uses these data to compute a sample standard deviation for the first iteration of the sequential t-test.

Null Hypothesis: Site is Dirty

If the null hypothesis selected by the VSP user is that the site is dirty, then the sequential t-test determines that there is enough evidence to accept the null hypothesis if

$$\ln(L_n) \geq \ln\left(\frac{1-\alpha}{\beta}\right)$$

where

L_n is the likelihood ratio test statistic, which is computed using the method described in Appendix A of this report

ln is the natural logarithm

α is the maximum acceptable Type I decision error rate (probability of rejecting the null hypothesis when the null hypothesis is true), and

β is the maximum acceptable Type II decision error rate (probability of accepting the null hypothesis when the null hypothesis is false).

The sequential t-test determines that there is enough evidence to reject the null hypothesis and accept the

alternative hypothesis if

$$\ln(L_n) \le \ln\left(\frac{\alpha}{1-\beta}\right)$$

Finally, the sequential t-test determines that the information from the n samples is not sufficient to make a decision if

$$\ln\left(\frac{\alpha}{1-\beta}\right) < \ln(L_n) < \ln\left(\frac{1-\alpha}{\beta}\right)$$

in which case additional samples are collected and the sequential test repeated using the full data set of all measurements collected to date.

Null Hypothesis: Site is Clean

If the null hypothesis selected by the VSP user is that the site is clean, then the sequential t-test determines that there is enough evidence to accept the null hypothesis if

$$\ln(L_n) \le \ln\left(\frac{\beta}{1-\alpha}\right) \ ,$$

that there is enough evidence to reject the null hypothesis and accept the alternative hypothesis if

$$\ln(L_n) \ge \ln\left(\frac{1-\beta}{\alpha}\right) ,$$

and that additional samples are needed if

$$\ln\left(\frac{\beta}{1-\alpha}\right) < \ln(L_n) < \ln\left(\frac{1-\beta}{\alpha}\right) .$$

Please see Appendix A for additional discussion of Barnard's sequential t-test.

3.1.1.2. Data Not Required to be Normally Distributed

This section presents the methods in Version 2.0 of VSP that test if the true average exceeds the threshold value for when the data are not necessarily normally distributed.

Wilcoxon Signed Ranks Test

The Wilcoxon Signed Ranks test can be used to test whether the true median or mean of the study-site population exceeds a fixed upper limit. The assumptions needed for this test are that the data are representative of the study site, are not spatially or temporally correlated, and have a symmetric (but not necessarily normal) distribution. Note that the test applies to either the mean or median when the assumption of symmetry is true, because those two parameters have identical values when the population of measurements has a symmetric distribution.

The equation used in Version 2.0 to compute a lower bound on the number of samples, n, needed for the test when the VSP user specifies that only r = 1 analytical replicates from each field sample will be obtained is

$$n = 1.16\left[\frac{s_{total}^2\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2}{\Delta^2} + 0.5Z_{1-\alpha}^2\right], \tag{3.3}$$

which is used in *Guidance for Data Quality Assessment* (EPA 2000b, page 3-12).

If the MQO module for this test is used, in which case r = 1, 2, or 3 can be used, then the equation for n that is used in VSP is

$$n = 1.16\left[\frac{\left(s_{sample}^2 + \frac{s_{analytical}^2}{r}\right)\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2}{\Delta^2} + 0.5Z_{1-\alpha}^2\right] \tag{3.4}$$

Equation (3.3) is identical to Equation (3.1) and Equation (3.4) is identical to Equation (3.2) except for the 1.16 multiplier. The constant 1.16 is used because it is known (Conover 1999, p.363) that the Wilcoxon Signed Ranks test will require no more than 1.16 times as many samples as the t-test to achieve the α and β decision error rate test performance specifications provided by the VSP user when the data (obtained using random sampling) are normally distributed.

Noether (1987, Section 2.2) developed an alternative to Equation (3.3) for computing n needed for the Wilcoxon Signed Ranks test. His method does not require that the data have a symmetric distribution. The assumptions that underlie his method are that the data are representative of the underlying population, the data are not correlated, and the computed Wilcoxon Signed Ranks test statistic (the quantity computed using the data to make the test) is approximately normally distributed. Noether (1987) indicates that the value of n computed using his method should achieve the performance requirements for the test (as specified by α and β) unless the n computed using his method is "quite small."

MARSSIM Sign Test

The sign test can be used to test whether the true median concentration of the population being sampled exceeds a fixed upper limit value. The assumptions that underlie this test are that the data are representative of the study site and they are not spatially or temporally correlated. The data need not be symmetric nor normally distributed.

The formula used to compute the approximate number of samples, n, needed for the sign test when only r = 1 analytical replicates per field sample will be obtained is given in the Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) (EPA 1997, page 5-33):

$$n = 1.20\left[\frac{\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2}{4\left(SignP - 0.50\right)^2}\right] \tag{3.5}$$

where

$$SignP = \Phi\left(\frac{\Delta}{s_{total}}\right) \tag{3.6}$$

Equation (3.6) is derived in Gogolak, Powers, and Huffert (1997, pp. 9-3 and 9-7).

VSP denotes the sign test as the MARSSIM sign test to signify that Equation (3.5) is taken from the MARSSIM document (EPA 1997). The function Φ (defined in Table 3.1) denotes the cumulative distribution function of the standard normal distribution (a normal distribution with mean zero and standard deviation 1). Hence, by definition, $\Phi\left(\frac{\Delta}{s_{total}}\right)$ is the fraction of the bell-shaped standard normal distribution that is less than or equal to the value of $\frac{\Delta}{s_{total}}$. Note that $\frac{\Delta}{s_{total}} > 0$ because Δ, the width of the gray region in the DPGD, must be greater than zero. VSP computes Sign P = $\Phi\left(\frac{\Delta}{s_{total}}\right)$ using the value of $\frac{\Delta}{s_{total}}$ specified by the VSP user. EPA (1997, Table 5.4, p. 5-32) provides values of Sign P for selected values of $\frac{\Delta}{s_{total}}$ between 0.1 and 3.0.

If the VSP user makes use of the MQO module, in which case r = 1, 2, ... m analytical replicates per field sample can be used, then

$$SignP = \Phi\left[\frac{\Delta}{\left(s^2_{sample} + \frac{s^2_{analytical}}{r}\right)^{1/2}}\right] \tag{3.7}$$

Equation (3.5) is based on the formula for n proposed by Noether (1987, Section 2.1) for the sign test. His equation for n is identical to Equation (3.5) except that Noether used the constant 1.00 in place of 1.20. The assumptions that underlie Noether's equation are that the data are representative of the underlying population, the data are not correlated, and the computed sign test statistic (the quantity computed using the data to make the test) is approximately normally distributed.

Noether (1987) indicates that the value of n computed using his method should achieve the performance requirements for the sign test (as specified by α , β and Δ) unless the computed n is "quite small." The MARSSIM report (EPA 1997) used 1.20 instead of 1.00 in Equation (3.5) to provide subjective added confidence that the larger value of n computed would result in achieving the performance requirements specified for the test by the VSP user.

Equation (3.5) is used in VSP primarily because it is used in EPA (1997), which is a multi-agency consensus document that was developed collaboratively by four federal agencies having authority and control over radioactive materials: the U.S. Departments of Defense and Energy, the U.S. Nuclear Regulatory Commission, and the EPA.

3.1.2 Compare Average to a Reference Average

Suppose a VSP user wants to determine the number of samples, n, that should be collected at a study site and also the number of samples, m, that should be collected in a reference site (e.g., a background area) to decide if the true average site concentration exceeds the true average reference area concentration. For this situation, VSP provides several different equations for computing the recommended minimum number of samples, n and m needed, with the restriction that n = m. The equation used by VSP depends on which statistical test is selected by the VSP user: the "two-sample" t-test, the Wilcoxon Rank Sum test, or the MARSSIM Wilcoxon Rank Sum (WRS) test. Each of these tests and the methods VSP used to compute the number of samples is now provided.

3.1.2.1 Data Are Normally Distributed

Two-Sample t-test

The two-sample t-test is a test to evaluate if the true site mean exceeds the true reference-area mean. The assumptions underlying the test are that the data are normally distributed, uncorrelated, representative of the underlying site and reference populations, and the variance (which can be unknown) of the site data equals the variance of the reference area data.

The equation used in VSP to compute the minimum recommended number of samples n and m for the study site and the reference area, respectively, when r = 1 analytical replicates per field sample will be made is

$$n = m = \frac{2s_{total}^2 \left(Z_{1-\alpha} + Z_{1-\beta}\right)^2}{\Delta^2} + 0.25Z_{1-\alpha}^2 \qquad (3.8)$$

If the MQO module in VSP is used, in which case r = 1, 2, or 3 analytical replicates at the site and reference-area is specified by the VSP user, then the equation used in VSP is

$$n = m = \frac{2\left(s^2_{sample} + \frac{s^2_{analytical}}{r} \right)\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2}{\Delta^2} + 0.25\, Z^2_{1-\alpha} \tag{3.9}$$

The key assumption that underlies the derivation of Equation (3.8), in addition to the assumptions given above, is that

$$\bar{x}_{site} - \bar{x}_{reference} - 2k\left(\frac{s^2_{site} + s^2_{reference}}{2} \right)^{1/2} \tag{3.10}$$

is normally distributed with mean $\mu_{site} - \mu_{reference} - 2k\sigma$ and variance $(\frac{\sigma^2}{n})(2 + k^2)$, where k is an arbitrary constant (Guenther 1981).

The derivation of Equation (3.8) is found in a more general context in Guenther (1977), where the problem Guenther considers includes the two-sample t-test. Equation (3.8) is used in the statistics book by Bowen and Bennett (1988, p. 168) and in EPA (2000b, p 3-24)

Guenther (1981) indicates that Equation (3.8) is an approximation to the true minimum sample sizes n and m required for the two-sample t-test. However, he states that Equation (3.8) usually yields the exact solution for n and m. The exact solution is obtained using an iterative approach using tables of the non-central t distribution found in Owen (1965). Equation (3.9) should provide a very accurate approximation of n and m for a specified r value because it is a straightforward extension of Equation (3.8) to the case of $r \geq 1$.

Methods other than Equations (3.8) and (3.9) also have been proposed for approximating n and m for the two-sample t-test. Desu and Raghavarao (1990, pp. 31-33) give an iterative procedure as well as a two-stage procedure that use the t distribution.

3.1.2.2 Data Not Required to be Normally Distributed

Wilcoxon Rank Sum Test

The Wilcoxon rank sum (WRS) test can be used to test whether the true median concentration at the study site is larger than the true median at the reference area. The assumptions that underlie the WRS test are that the site and reference data are representative of the site and reference areas, respectively, the data are not correlated, and the two population distributions are identical in shape and have the same (unknown) variance. The two distributions need not be symmetric or normally distributed.

The equation used in Version 2.0 VSP to compute the number of samples, n and m, for the site and reference areas is

$$n = m = 1.16\left[\frac{2s^2_{total}\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2}{\Delta^2} + 0.25 Z^2_{1-\alpha} \right] \tag{3.11}$$

Equation (3.11) is also used in EPA (2000b, p. 3-34). The constant 1.16 in Equation (3.11) is used because it is known (Conover 1999, p. 284) that the WRS test will require no more than 1.16 times as

many samples as the two-sample t-test to achieve the α and β test performance specifications provided by the VSP user if the two population distributions (for the site and reference area) are normally distributed and are identical except for their true median values.

If the MQO module for the WRS test is used and r = 1, 2, ... m analytical replicates per field sample is specified by the VSP user, then the equation for n and m used in VSP is

$$n = m = 1.16\left[\frac{2\left(s^2_{sample} + \frac{s^2_{analytical}}{r}\right)\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2}{\Delta^2} + 0.25Z^2_{1-\alpha}\right] \qquad (3.12)$$

Equation (3.11) is identical to Equation (3.8) and Equation (3.12) is identical to Equation (3.9) except for the 1.16 multiplier.

MARSSIM Wilcoxon Rank Sum (WRS) Test

Noether (1987, Section 3.0) developed an alternative equation for approximating the number of samples required for the WRS test. Noether's equation is

$$n + m = \frac{\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2}{3\left(P_r - 0.5\right)^2} \qquad (3.13)$$

where n + m is the sum of the minimum number of study area and reference area samples needed for the WRS test, assuming n = m. Noether (1987) assumed when deriving Equation (3.13) that the WRS test statistic was approximately normally distributed, although the data themselves need not be normally distributed. Equation (3.13) is a special case of a more general equation given by Noether (1987, Section 3.0) and used by EPA (1992) that allows the number of samples from the site and reference area to be unequal. Equation (3.13) is used in EPA (1997, p. 5-28, Equation 5-1) and in EPA (1992, p. 6.3). If the VSP user selects the MARSSIM WRS test option in VSP, then Equation (3.13) will be computed.

The parameter P_r in Equation (3.13) is the probability that a measurement of a sample collected at a random location at the study site is greater than a measurement of a sample collected at a random location in the reference area. The allowable values of P_r that can be specified range from 0.5 to 1.0. The difference, P_r - 0.5 , is a measure of the difference between the site and reference distributions, Δ, that the VSP user wishes to detect with high probability $1 - \beta$.

P_r is computed as

$$P_r = \Phi\left(\frac{\Delta}{\sqrt{2}\sigma_{total}}\right) \tag{3.14}$$

when only r = 1 analytical replicate per field sample is measured. If the MQO module is used, then P_r is computed as

$$P_r = \Phi\left(\frac{\Delta}{\sqrt{2}\sqrt{s^2_{sample} + \frac{s^2_{analytical}}{r}}}\right) \tag{3.15}$$

Equations (3.14) and (3.15) assume the data are normally distributed, which is necessary because the function Φ is the cumulative normal distribution function (see Table 3.1.). Hence, although the WRS test does not require that the data be normally distributed, the normality assumption is used to determine the number of samples needed. In addition, the validity of Equation (3.13) requires that the WRS test statistic is itself approximately normally distributed.

3.1.3 Confidence Interval for the Mean

One can assess whether the true mean exceeds a threshold value by computing a one-sided or a two-sided confidence interval for the mean and then looking to see if that interval overlaps the threshold value. For this case, VSP computes the number of samples required to obtain an interval that provides the required confidence specified by the VSP user.

The equation used in VSP to compute the minimum recommended value of n for a one-sided confidence interval, when only r = 1 analytical replicates per field sample is measured is as follows:

$$n = \left(\frac{t_{1-\alpha,df}}{d}\right)^2 s^2_{total} \tag{3.16}$$

where d denotes the VSP user's largest acceptable difference between the estimated mean, $\overline{x}$, and the true mean μ . In other words, the VSP user specifies that n should be large enough such that the absolute value of $\overline{x}$ - μ will be no larger than the specified value of d with 100 (1- α) percent confidence. The quantity d can also be interpreted as the width of the one-sided confidence interval on μ that the VSP user specifies should not be exceeded. If a two-sided confidence interval is used, then d is the *half*-width of that interval.

The notation $t_{1-\alpha,df}$ and s^2_{total} is defined in Table 3.1. As n appears on both sides of Equation (3.17) (on the right side of the equation n is involved in the degrees of freedom, df), the equation is solved iteratively in VSP to determine the smallest n that satisfies the equality sign. The iteration scheme used is

given in Gilbert (1987, p. 32). Equation (3.16) is a standard formula found in many statistics books, e.g., Hahn and Meeker (1991, p. 136), Iman and Conover (1983, p. 188), and Gilbert (1987, p. 32).

If the MQO module is used, in which case r = 1, 2, or 3 analytical replicates per field sample are measured, then n is computed using

$$n = \left(\frac{t_{1-\alpha,df}}{d} \right)^2 \left(s^2_{sample} + \frac{s^2_{analytical}}{r} \right) \qquad (3.17)$$

Equation (3.17) is a straightforward extension of Equation (3.16).

If the VSP user wants to use a two-sided confidence interval to test whether the true mean exceeds the action limit, then $t_{1-\alpha,df}$ in Equation (3.16) and Equation (3.17) is replaced by $t_{1-\alpha/2,df}$. Furthermore, d is now the half-width of the confidence interval that must be achieved.

Hahn and Meeker (1991, pp. 137-141) suggest two alternative methods for computing n. These methods have not been evaluated for use in VSP.

3.1.4 Estimate the Mean

Suppose the sampling goal is to estimate the mean of a variable of interest in a defined geographical area. The three designs discussed in this section, Stratified Sampling (StS), Ranked Set Sampling (RSS) and Adaptive Cluster Sampling (ACS), are sampling strategies that can provide better estimates of the mean than is possible using simple random sampling. These three methods are described in detail in EPA (2001). These methods, as well as their underlying assumptions and the process for determining the number of samples required are briefly described in the sections below.

3.1.4.1 Stratified Sampling (StS)

The primary purpose of the Stratified Sampling (StS) sampling design is to estimate the mean for a study area that has been divided into different subareas (strata) on the basis of pre-existing information about the past use of the site. The strata must not overlap and they should be more internally homogeneous than the entire study area (all strata combined). In this situation, the use of StS to determine the locations where samples will be collected in the strata should provide a more precise estimate of the mean of the total study area than can be achieved if simple random sampling is used to determine sampling locations over the study area. EPA (2001) and Gilbert (1987) discuss StS in the context of environmental sampling.

The assumptions that underlie these methods are (1) that the sampling locations are determined using either simple random sampling or a systematic grid whose starting position in the field is determined at random, and (2) the measurements obtained are statistically independent (not correlated). There is no requirement that the measurements be normally distributed in order to use them to estimate the mean. However, if a confidence interval for the estimated mean will be computed, then the number of samples should be large enough that the mean is normally distributed.

Version 2.0 of VSP uses inputs provided by the VSP user to compute the total number of samples needed (for all strata combined) as well as the optimal allocation of those samples to the strata. The VSP

user has a choice of 3 methods for determining the total number of samples and 2 choices for determining the optimum allocation of those samples to the strata. These methods are now described.

Equations to Compute the Total Number of Samples, n, in Stratified Sampling

Method 1: Minimize the Variance of the Sample Mean for Fixed Cost of Sampling and Analysis

VSP computes the total number of samples, n, over all strata in such a way that the precision of the estimated population mean is maximized for a given pre-specified fixed total cost, $C\text{-}C_0$, of collecting and measuring samples. The formula used to calculate n is (Gilbert 1987, page 51)

$$n = \frac{\left(C - c_o\right)\sum_{h=1}^{L} \frac{W_h s_h}{\sqrt{c_h}}}{\sum_{h=1}^{L} W_h s_h \sqrt{c_h}}$$

where

L is the number of strata, h = 1, 2, ... , L,

s_h is the estimated total standard deviation of the measured values in stratum h,

$W_h = N_h / N$ is the weight associated with stratum h,

N_h is the total number of possible sampling locations (units) in stratum h,

$N = \sum_{h=1}^{L} N_h$ is the total number of possible units in all strata combined,

$C = c_o + \sum_{h=1}^{L} c_h n_h$ is the total sampling budget,

c_o is the fixed overhead cost,

c_h is the cost of collecting and measuring a sample in stratum h, and

n_h is the number of samples collected in stratum h.

The VSP user inputs values for L, C, N_h, s_h, c_o, and c_h in the VSP dialogue box for StS. VSP computes the values of the n_h using one of the two methods described below.

Method 2: Minimize Cost for a Specified Required Variance of the Sample Mean

The total number of samples is computed to achieve the pre-specified precision of the estimated population mean without any specified upper limit on stratum costs. Note again, that n is the total number of samples for all strata combined.

The formula used to calculate n is (Gilbert 1987, page 51)

$$n = \frac{\left(\sum_{h=1}^{L} W_h s_h \sqrt{c_h}\right)\sum_{h=1}^{L} W_h s_h / \sqrt{c_h}}{V + \frac{1}{N}\sum_{h=1}^{L} W_h s_h^2}$$

where V is the pre-specified variance (precision). The other notation has been defined above.

Method 3: Predetermined Number of Total Samples, n

The VSP user supplies the total number, n. There is no statistical basis for claiming this value of n will result in an estimated mean that is adequate for any particular design goal. This method is provided in VSP for a situation where n has been predetermined by some non-statistical process and there is no option but to select n samples.

Equations to Determine the Allocation of the n Samples to the Strata in Stratified Sampling

The VSP user selects either Method 1 or Method 2 described below to determine the number of samples that should be collected in the h^{th} stratum. These methods assume that the following cost equation is applicable:

$$C = c_o + \sum_{h=1}^{L} c_h n_h$$

Method 1: Optimal Allocation of n Samples to the Strata such that the Variance of the Sample Mean is Minimized for a Fixed Cost C

Version 2.0 of VSP uses the following formula to compute the number of samples that should be collected in the h^{th} stratum in order that the sample mean based on n measurements has the minimum variance for a fixed total cost C (from Gilbert 1987, Equation 5.9, page 50):

$$n_h = n\frac{N_h s_h / \sqrt{c_h}}{\sum_{h=1}^{L} N_h s_h / \sqrt{c_h}}$$

The notation in this equation was defined above.

Method 2: Optimum Allocation when the Stratum Costs, c_h, are Equal for all Strata (Neyman Allocation)

When the VSP user selects this allocation method, the number of samples that should be collected from the h^{th} stratum, n_h, is computed as follows (Gilbert 1987, equation 5.10, page 50):

$$n_h = n \frac{N_h s_h}{\sum_{k=1}^{L} N_k s_k}$$

The notation used in this equation was defined above.

3.1.4.2 Ranked Set Sampling (RSS)

Ranked Set Sampling (RSS) was originally developed by McIntyre (1952) for estimating the mean. RSS combines simple random sampling with either professional knowledge and judgment or field screening measurements to select places to collect samples. RSS may provide for a better estimate of the mean at equal or less cost. EPA (2001) describes and illustrates RSS applications and methods.

A simple ecological example will illustrate the ranked set sampling method. Suppose we need to estimate the average age of trees on a property and that we want to measure 30 trees to estimate the average. The RSS design provides a way to select which 30 trees to measure using our judgment. Suppose we decide to visually (judgmentally) assess the size of a tree to get a rough idea of its age. Begin by randomly selecting, say, three trees and judge by eye which tree is the smallest. Mark the smallest tree to be measured and ignore the other two. Next, randomly select another set of three trees to rank. Mark the medium sized tree and ignore the other two. Next, randomly select another three trees. Mark the largest tree and ignore the other two. Repeat this procedure a total of 10 cycles for a total of 90 trees. 30 of the trees will have been marked and 60 ignored. Of the 30 marked trees, 10 are from a stratum of generally smaller trees, 10 are from a stratum of generally middle-sized trees and 10 are from a stratum of generally larger trees. Determine the age of each of the 30 marked trees by coring or some other appropriate measurement technique and use that measurement to estimate the average age of the trees on the lot.

In this illustration there are r = 10 cycles and m = 3 trees chosen per cycle. In practice, the number of sample locations chosen per cycle (the set size, m) and the number of cycles, r, is determined using a systematic planning process, which is described in this section. Version 2.0 of VSP implements the systematic process needed to determine the number of cycles, and hence, the number of locations to be ranked and the number of locations to be measured. VSP can also place ranking and sampling locations on the map of the study site.

The assumptions that underlie RSS are (1) the sampling locations ultimately chosen to be sampled are among those initially selected using simple random sampling, (2) either judgment or screening measurements are used to rank potential sampling locations, (3) the ranking method used does a very good job of ranking, and (4) the measurements are not correlated. The data need not be normally distributed for the purpose of estimating the mean, the application used in Version 2.0 of VSP. If the goal is to compute confidence limits on the mean, then the number of measurements used to compute the sample mean should be large enough that the sample mean itself is normally distributed.

Determine the Number of Locations to Sample in RSS

There are 2 major types of RSS designs: balanced and unbalanced. Balanced designs are used when the analytical measurements of interest are expected to be symmetrically distributed about the mean. Unbalanced designs are used when the distribution of measurements is skewed to higher values.

Determining the number of ranked set samples to collect is a two-part process: Part 1 and Part 2.

Part 1: Determine if RSS is More Cost-Effective than Simple Random Sampling Alone for Estimating the Mean

The cost ratio (cost of an analytical measurement divided by cost of ranking a field location) is compared to a cost ratio threshold to determine if RSS is more cost effective than SRS. If the cost ratio for the given study exceeds the cost threshold, then RSS is cost-effective relative to simple random sampling (SRS) and should be used in place of SRS. If the cost ratio does not exceed the threshold, then RSS is not cost-effective and will cost more than SRS to achieve the same precision in the estimated mean. However, if the additional cost is not prohibitive, RSS may still be desired because it will usually provide a more precise estimate of the mean than SRS.

If professional judgment is used to do the ranking, the cost ratio threshold is taken from the following table (based on Figure 3 in Mode et al. 1999).

Lab Data Distribution	Degree of Ranking Error	Set Size = 2	Set Size = 3	Set Size = 4	Set Size = 5
Symmetric	None	4	3	3	3
Symmetric	Substantial	7	6	6	6
Asymmetric	None	6	5	5	4
Asymmetric	Substantial	10	9	9	9

If field screening measurements are used to do the ranking of field locations, the cost ratio threshold is obtained from the following table (based on Figure 4 of Mode et al. 1999):

Analysis / Screening Correlation	Set Size = 2	Set Size = 3	Set Size = 4	Set Size = 5	Set Size = 6	Set Size = 7	Set Size = 8
1.0	5	4	3	3	2	2	2
0.9	6	6	5	5	5	5	5
0.8	7	8	8	8	8	9	9
0.7	12	12	12	13	14	15	16

Part 2: Determine the Number of Ranked Set Sampling Samples Needed to Estimate the Mean

Version 2.0 of VSP uses a five-step process to determine the number of ranked set samples.

Step 1: Determine the Number of Samples if Simple Random Sampling Alone is Used Instead of RSS

Balanced RSS Design

For balanced RSS designs, the number of samples needed for simple random sampling, n_o, is calculated such that the estimated mean will be within a pre-specified margin of error from the true mean at a specified confidence level. Either a one-sided or a two-sided confidence interval equation can be used. The assumptions used in Version 2.0 for this design is that the distribution of the data is normal and the measurements are not correlated. The normality assumption is used to determine the relative precision, as discussed below in Step 3.

If the VSP user selects a one-sided confidence interval and a balanced RSS design, then the equation used in Version 2.0 to calculate the number of samples needed for simple random sampling, n_o, is

$$n_o = s_{total}^2 \left(\frac{t_{1-\alpha,df}}{d} \right)^2$$

where

s_{total}^2 is the total standard deviation of the measurements from the laboratory that were collected from the study area

d is the maximum desired width of the one-sided confidence interval for the mean

$t_{1-\alpha,df}$ is the value of the Student's t distribution with n-1 degrees of freedom (df) such that the proportion of that distribution less than $t_{1-\alpha,df}$ is $1-\alpha$

For a two-sided confidence interval and a balanced RSS design, the equation used is identical to the one immediately above except that (1) d is the maximum desired half-width of the two-sided confidence interval, and (2) $t_{1-\alpha,df}$ is replaced by $t_{1-\alpha/2,df}$, which is the value of the Student's t distribution with n-1 degrees of freedom (df) such that the proportion of that distribution less than $t_{1-\alpha/2,df}$ is $1-\alpha/2$. Because n appears on both sides of the above equations (on the right side it appears in the degrees of freedom of the t distribution), the equation is solved iteratively using the method in Gilbert (1987, pg. 32).

Unbalanced RSS Design

For unbalanced RSS designs, the number of samples is computed using the "Adjusted Classical Formula" method in Perez and Lefante (1997, page 2789). The assumptions used in Version 2.0 for this design are that the distribution of the data is lognormal and the measurements are not correlated. The first step of this method is to compute an approximate sample size using the following formula:

$$n_{classic} = \left(\frac{Z_{1-\alpha/2}}{\pi} \right)^2 \left(GSD^{\ln(GSD)} - 1 \right)$$

where

$n_{classic}$ is the approximate recommended minimum number of samples

$Z_{1-\alpha/2}$ is the value of the standard normal distribution such that the proportion of the distribution less than $Z_{1-\alpha/2}$ is $1-\alpha/2$

GSD is the geometric standard deviation

ln is the natural logarithm

π is the maximum proportion difference that can be tolerated between the estimated mean and the true mean.

Next, the following linear regression equation is used to calculate the recommended minimum number of samples:

$$n_o = \beta_o + \beta_1\left(n_{classic}\right)$$

where

n_o is the recommended minimum number of samples needed for simple random sampling,

β_o is the Y-intercept of the regression line, and

β_1 is the slope of the regression line.

β_o is obtained from the following table (from Table III in Perez and Lefante, 1997, page 2791)

Confidence Level	GSD = 1.1	GSD = 1.5	GSD = 2.0	GSD = 2.5	GSD = 3.0	GSD = 3.5	GSD = 4.0
90%	2.9532	7.5249	11.3183	15.5638	20.1322	25.9327	30.3223
95%	3.3331	7.9237	14.0744	20.5406	27.1563	33.6865	40.1084
99%	4.9265	11.2470	20.5069	30.2478	40.1743	51.1945	60.6576

β_1 is obtained from the following table (from Table III in Perez and Lefante, 1997, page 2791)

Confidence Level	GSD = 1.1	GSD = 1.5	GSD = 2.0	GSD = 2.5	GSD = 3.0	GSD = 3.5	GSD = 4.0
90%	0.4714	0.6926	0.8509	0.8794	0.8499	0.7731	0.7033
95%	0.4726	0.8094	0.9046	0.9129	0.8731	0.8072	0.7288
99%	0.4740	0.8865	0.9808	0.9877	0.9444	0.8612	0.7796

Both of the above tables assume the data are lognormally distributed.

Step 2: Select a Value for the Set Size, m

The selected value of m is usually based on practical constraints in ranking locations in the field using

professional judgment or field screening measurements. It may be difficult to use professional judgment to accurately rank by eye more than m = 4 or 5 locations. Other constraints that may affect the size of m are time, staff, equipment, and other cost considerations. VSP limits m to 5 for judgment sampling and to 8 for field screening measurements.

Step 3: Determine the Relative Precision, RP, of RSS Compared to Simple Random Sampling

Balanced RSS Designs

For balanced designs, relative precision (RP) is found from the following table assuming the data are normally distributed [from Table 1 of Patil et. al. (1994, pg.176)]:

Set Size = 2	Set Size = 3	Set Size = 4	Set Size = 5
1.467	1.914	2.347	2.770

If the set size, m, is greater than 5, the RP is found using the following linear regression formula:

$$RP = 0.4342\ m + 0.6048$$

Unbalanced RSS Designs

For unbalanced designs, RP is found using a two-step process. First, a value for the RP is found from the following table, which assumes the data are lognormally distributed and that a balanced RSS design is being used [from Table 1 of Patil et. al. (1994, pg.177)]:

Set Size	CV* = 0.1	CV = 0.202	CV = 0.307	CV = 0.416	CV = 0.533	CV = 0.658	CV = 0.795	CV = 0.947	CV = 1.117	CV = 1.311
2	1.46	1.45	1.42	1.40	1.37	1.33	1.29	1.26	1.22	1.19
3	1.90	1.87	1.83	1.77	1.70	1.62	1.55	1.47	1.40	1.34
4	2.33	2.28	2.21	2.11	2.00	1.89	1.78	1.67	1.56	1.47
5	2.75	2.68	2.57	2.44	2.29	2.14	1.99	1.84	1.71	1.59
6	3.15	3.07	2.93	2.76	2.57	2.37	2.18	2.00	1.84	1.70
7	3.56	3.45	3.28	3.07	2.83	2.60	2.40	2.16	1.96	1.80
8	3.95	3.86	3.61	3.36	3.09	2.81	2.55	2.30	2.08	1.89

*CV is the *coefficient of variation*, which is defined to be $CV = \sqrt{e^{\ln(GSD)^2} - 1}$ when data are lognormally distributed.

Then, the RP from the above table is multiplied by a correction factor obtained by linear interpolation from the following table to obtain the approximate RP value for the unbalanced RSS design.

CV	0.1	0.3	0.5	0.8	1.3
Correction Factor	1.01	1.08	1.2	1.5	1.7

Step 4: Compute the Number of Cycles, r

Number of cycles, r, is computed using the following formula:

$$r = \frac{\left(n_o / m\right)}{RP}$$

Step 5: Compute the Total Number of Samples that Should be Collected.

For balanced designs, the total number of samples required is found using the following formula:

$$n = r \times m$$

For unbalanced designs, first the number of times, t, the largest rank needs to be sampled is found from the following table (from Table 8-5 in QA/G-5S, which was constructed from Figure 6 in Kaur et al 1995, page 14 and in Kaur et al. 1997):

CV	0.25	0.5	1.0	1.25	1.5	2.0	2.5	3.0	3.5	4.0
t	1	2	3	4	5	6	7	8	9	10

The total number of samples for unbalanced designs is then found using the following formula:

$$n = r \times (m + t - 1)$$

3.1.4.3 Adaptive Cluster Sampling (ACS)

Adaptive Cluster Sampling (ACS) uses a probability-based design such as simple random sampling to select an initial set of field units (locations) to sample. Then, additional neighboring samples are selected for observation when a characteristic of interest is present in an initial unit or when the initial unit has an observed value that meets some pre-specified condition (e.g., when a critical threshold value is exceeded).

ACS is appropriate in situations where the characteristic of interest is sparsely distributed but highly aggregated (clustered). Examples of such populations can be found in mineral investigations (unevenly distributed ore concentrations), animal and plant populations (rare and endangered species), pollution concentrations and hot spot investigations, and epidemiology of rare diseases. Possible environmental applications of ACS include soil remediation (investigating the extent of soil contamination while simultaneously estimating the mean), hazardous waste site characterizations, surveying Brownfields, and determining the extent of occurrence of effects of an airborne source of pollutant on nearby flora and fauna. Additional information on ACS is available in EPA (2001), Thompson and Seber (1996), Thompson

(1992, chapter 24) and Thompson (1990).

Two assumptions that underlie ACS are (1) the initial sampling locations are selected using simple random sampling, stratified sampling, grid sampling, or some other probability-based sampling design, and (2) the measurements obtained are statistically independent (not correlated).

ACS is implemented in Version 2.0 of VSP in four steps as now presented.

Step 1. Divide the Sample Area into a Grid of Sampling Units

VSP divides the selected sample area into square grid units of the specified size. The size of the grid unit is specified by the VSP user on the design dialog box. VSP assumes that a single measurement is associated with each grid unit.

Step 2. Define the Sample Design

2.1. Choose the initial set of sampling units

VSP computes the number of sampling units that should be sampled in the initial sample, n_1 ,by computing the number of samples required for computing a confidence interval for the mean. An assumption is made that either the measurements themselves have a normal distribution or the estimated mean has a normal distribution. VSP determines the sampling locations using simple random sampling.

For a one-sided confidence interval, the equation used by VSP to compute n_1 is given by Equation (3.16) for which d is the maximum desired width of the confidence interval. For a two-sided confidence interval, the equation used to calculate the number of initial samples is the same as Equation (3.16) except that d is the desired half-width of the desired confidence interval and $t_{1-\alpha,df}$ is replaced by $t_{1-\alpha/2,df}$, which is the value of the Student's t-distribution with n-1 degrees of freedom (df) such that the proportion of that distribution that is less than $t_{1-\alpha/2,df}$ is $1-\alpha/2$. Because n appears on both sides of these equations, (on the right side it appears in the degrees of freedom of the t distribution), the equation is solved iteratively using the method in Gilbert (1987, pg. 32).

2.2. Specify a rule or criterion for performing follow-up sampling

A criterion is needed to decide when follow-up sampling of neighboring grid units is needed. Version 2.0 of VSP only supports the use of a threshold criterion. That is, if a measured value for an initial unit exceeds some specified threshold value, then the grid units defined to be in the "neighborhood" of that unit need to be sampled.

2.3 Define the neighborhood of a sampling unit

VSP supports the use of two different neighborhoods: a 4-unit neighborhood and a 8-unit neighborhood. The 4-unit neighborhood consists of the 4 neighboring units that share an edge (side) with the original grid unit. The 8-unit neighborhood consists of the 8 neighboring units that share an edge (side) or touch a corner of the original grid unit. The VSP user selects which of these two types of neighborhoods is desired.

Step 3. Collect a Sample in Each of the Initial Units

Collect a sample in each of the n_1 initial units. The VSP user enters the measured value for each of these initial units into VSP. Then VSP determines if the units in the neighborhood of each initial unit needs to be sampled.

Step 4. Conduct Follow-Up Sampling

As in Step 3, collect a sample in the neighborhood units when VSP indicates that follow-up sampling is needed. Once these measurements are entered into VSP, the code applies the threshold criterion value to each of these new measurements to determine if units in the neighborhood of these new units need to be sampled. Follow-up sampling continues sequentially in this way until no more neighborhood units have measurements that exceed the threshold criterion value.

The final set of units sampled consists of the initial units for which there was no follow-up sampling conducted, plus the clusters of units centered on those initial units where follow-up sampling was conducted. Each cluster is surrounded by a set of measured units that do not exceed the threshold criterion value. These are called edge units. A cluster without its edge units is called a network. Any observed unit, including an edge unit, that does not exhibit the characteristic of interest is a network of size one. Hence, the final sample can be partitioned into non-overlapping networks.

The usual statistically-unbiased method of computing the sample mean and variance of the sample mean will be statistically biased when all the ACS measurements are used unless none of the n_1 initial units exceeded the threshold value. However, even when some of the initial units exceed the threshold value, statistically-unbiased estimates of the mean and standard error will be obtained if only the n_1 measurements for the initial units are used to compute the mean and standard error.

Thompson (1990) has developed statistically unbiased estimators of the mean and the variance of the estimated mean that make use of all the samples obtained using ACS. VSP computes the modified Horvitz-Thompson unbiased estimators when the initial set of n_1 units are placed on the study site by VSP using simple random sampling. The modified Horvitz-Thompson unbiased estimator of the mean is computed by VSP as follows:

$$\hat{\mu} = \frac{1}{N}\sum_{k=1}^{m}\frac{y_k^*}{\alpha_k}$$

and the variance of the estimated mean is computed by VSP as

$$\hat{Var}\left(\hat{\mu}\right) = \frac{1}{N^2}\left[\sum_{j=1}^{m}\sum_{k=1}^{m}\frac{y_j^* y_k^*}{\alpha_{jk}}\left(\frac{\alpha_{jk}}{\alpha_j \alpha_k} - 1\right)\right]$$

where

y_k^* is the sum of the values of the measurement of interest, y, for the k[th] network in the sample
N is the number of units in the population (study area)
m is the number of distinct networks (excluding edge units) in the final sample
α_k is the probability that the initial sample (set of n_1 initial units) intersects the k[th] network
α_{jk} is the probability that the initial sample intersects both the j[th] and the k[th] networks

Units in the initial sample whose measurements do not exceed the threshold value are included in the calculation as networks of size one, but edge units are excluded.

If there are n_1 units in the initial sample and x_k units in the k[th] network, then the intersection probabilities α_k and α_{jk} are calculated using combinatorial formulas as follows:

$$\alpha_k = 1 - \left[\binom{N - x_k}{n_1} \Big/ \binom{N}{n_1} \right]$$

$$\alpha_{jk} = 1 - \left[\binom{N - x_j}{n_1} + \binom{N - x_k}{n_1} - \binom{N - x_j - x_k}{n_1} \right] \Big/ \binom{N}{n_1}$$

where $\alpha_{jj} = \alpha_j$

3.1.5 Compare a Proportion to a Fixed Threshold Value

One-Sample Test for Proportions

The one-sample test for proportions, which is discussed in EPA (2000b, pp. 3-18, 3-19, 3-20), evaluates whether the true proportion of the population of measurements that exceeds some concentration limit (threshold) is greater than a specified fixed (standard) proportion. The assumptions needed for the test are that the data are representative and not correlated. The test is valid regardless of the shape of the data distribution. Hence, the data need not be normally distributed.

The equation used in VSP to compute the approximate number of samples required for the test, when the VSP user specifies that only r = 1 analytical replicate of each field sample will be measured, is

$$n = \frac{\left[Z_{1-\alpha} \sqrt{P_0 (1 - P_0)} + Z_{1-\beta} \sqrt{P_1 (1 - P_1)} \right]^2}{(P_1 - P_0)^2} \qquad (3.18)$$

where P_0 is the fixed (standard) proportion and P_1 is a specified value of the proportion different from P_0 that the VSP user indicates is important for the test to detect with probability $1 - \beta$. VSP asks the user to specify P_0 and Δ, where Δ is the absolute value of P_1 - P_0, i.e., the positive difference between P_1

and P_0 (the width of the "gray region"). Once P_0 and Δ are specified, then VSP computes P_1. VSP computes P_1 as $P_1 = P_0 - \Delta$ if the VSP user specifies the null hypothesis to be that "the site does not meet the standard." But, VSP computes P_1 as $P_1 = P_0 + \Delta$ if the null hypothesis is that "the site meets the standard." This difference in the way P_1 is computed can result in different sample sizes, depending on the null hypothesis selected.

The equation above is derived in Bowen and Bennett (1988, p. 190) and used in EPA (2000b, p. 3-19), EPA (1992, p. 5-40), and EPA (1989, p. 7-6). Desu and Raghavarao (1990, p. 16) compute n using a more complicated approximate formula.

3.1.6 Compare a Proportion to a Reference Proportion

Two-Sample Test for Proportions

The two-sample test for proportions, which is discussed in EPA (2000b, pp. 3-28, 3-29, and 3-30), evaluates whether the proportion of the site population that exceeds some concentration limit (threshold value) is greater than the proportion of the reference area population that exceeds that limit. The assumptions underlying this test are that the data are uncorrelated and representative of the populations. The data need not be normally distributed.

The equation used in Version 2.0 of VSP, as well as in EPA (2000b, p. 3-29), to compute the approximate number of samples required for this test, when the VSP user specifies r = 1, is

$$n = \frac{2(Z_{1-\alpha} + Z_{1-\beta})^2 \overline{P}(1 - \overline{P})}{(\delta_o - \delta_1)^2} \tag{3.19}$$

where

$$\overline{P} = (P_1 + P_2)/2$$

P_1 = the unknown true proportion at the study site greater than the limit

P_2 = the estimated proportion in the reference area (entered by the VSP user) that is greater than the limit

δ_0 = the specified difference in true proportions that defines the action level where the decision error rate α is specified

δ_1 = the user-selected difference in true proportions that defines the outer bound of the gray region where the decision error rate β is specified

Version 2.0 of VSP determines the value of P_1 as follows:

$$P_1 = P_2 \pm \Delta$$

where

Δ = the minimum difference in proportions P_1 and P_2 to be detected with power $1-\beta$ with $100(1-\alpha)$ percent confidence, and P_1 is chosen so that $\overline{P}$ is closest to 0.5 in order to maximize the value of n computed using Equation (3.35). The above approach for specifying P_1 and P_2 requires that the user enter P_1 and Δ to determine n. However, P_1 and P_2 also could be estimated using data obtained from previous studies at the study site and reference area.

Alternative methods for approximating n are available in the statistical literature. For example, Bowen and Bennett (1988, p. 193, Equation 4.98) provide a slightly different version of Equation (3.35).

3.1.7 Estimate a Proportion Using Stratified Sampling

Section 3.1.4.1 discussed the use of Stratified Sampling (StS) to estimate the mean of a study site. However, StS can also be used to estimate a proportion. For example, the study goal may be to estimate the proportion of the study area for which soil concentrations exceed a threshold (upper limit) value. Using StS to determine the locations where samples will be collected in the various subregions (strata) of the study area should provide a more precise estimate of the proportion for the entire study area than can be achieved if stratum boundaries are ignored and simple random sampling is used to determine sampling locations over the study area.

3.1.7.1 Methods for Determining the Total Number of Samples

Version 2.0 of VSP provides 3 methods for determining the total number of samples and 2 choices for determining the optimum allocation of those samples to the various strata. These formulas are very similar to those provided in Section 3.1.4.1 for estimating the mean using stratified sampling.

Method 1: Minimize Variance of the Sample Proportion for a Fixed Cost of Sampling and Analysis

VSP uses Method 1 to compute the total number of samples, n, such that the variance of the sample proportion for the total study area (all strata combined) is minimized (precision is maximized) when the dollar budget for sample collection and measurement is fixed at C - C_0. The formula used to calculate n is (Cochran 1977, pages 98 and 109) :

$$n = \frac{(C - c_o)\sum_{h=1}^{L} W_h \sqrt{P_h(1-P_h)/c_h}}{\sum_{h=1}^{L} W_h \sqrt{P_h(1-P_h)c_h}}$$

where

L is the number of strata

P_h is the estimated proportion of measurements in stratum h,

$W_h = N_h / N$ is the weight (relative size) of stratum h,

N_h is the size of stratum h, which may be interpreted to be the total number of possible sampling locations (units) in stratum h,

N is the size of the total study area (all strata combined), which may be interpreted to be the total number of possible sampling units in all strata combined,

$C = c_o + \sum_{h=1}^{L} c_h n_h$ is the total study cost (budget)

c_o is the fixed overhead cost (budget),

c_h is the cost of collecting and measuring a sample in stratum h, and

n_h is the number of samples collected in stratum h (n_h is computed below).

Method 2: Minimize Cost for a Specified Required Variance of the Sample Proportion

VSP uses Method 2 to compute the total number of samples, n, such that a pre-specified, fixed, value for the variance of the sample proportion is achieved without constraints on the total cost. The formula used to calculate the total number of samples is:

$$n = \frac{\left(\sum_{h=1}^{L} W_h \sqrt{P_h(1-P_h)} \sqrt{c_h} \right) \sum_{h=1}^{L} W_h \sqrt{P_h(1-P_h)} / \sqrt{c_h}}{V + \frac{1}{N} \sum_{h=1}^{L} W_h P_h (1-P_h)}$$

where V is the pre-specified variance. The other notation has been defined above.

Method 3: Predetermined Number of Total Samples, n

For Method 3, the VSP user supplies the value for the total number of samples, n. VSP does not compute n. There is no statistical basis for claiming that the value of n provided by the VSP user will result in an estimated proportion that achieves any desired level of precision.

3.1.7.2 Formulas to Determine the Allocation of n Samples to the Strata

The VSP user selects either Method 1 or Method 2 described below to determine the number of samples that should be collected in the h^{th} stratum. These methods assume that the following cost equation is applicable:

$$C = c_o + \sum_{h=1}^{L} c_h n_h$$

as defined above.

Method 1: Optimal Allocation to Minimize $Var(\hat{P})$ for Fixed Cost, C

VSP computes the number of samples that should be collected in the h[th] stratum, n_h , in order to minimize the variance of the estimated proportion, $\hat{P}$, for the entire study area for a fixed total budget (cost), C, using the following formula (Cochran 1977, page 109, equation 5.61):

$$n_h = n \frac{N_h \sqrt{P_h(1-P_h)/c_h}}{\sum_{h=1}^{L} N_h \sqrt{P_h(1-P_h)/c_h}}$$

The notation in this equation has been defined above.

Method 2: Optimum Allocation for Equal Stratum Costs (Neyman Allocation)

When the VSP user selects this allocation method, VSP computes the number of samples that should be collected from the h[th] stratum, n_h , as follows (Cochran 1977, page 108, equation 5.60):

$$n_h = n \frac{N_h \sqrt{P_h(1-P_h)}}{\sum_{h=1}^{L} N_h \sqrt{P_h(1-P_h)}}$$

The stratum costs do not appear in this equation because all the c_h , for h = 1, 2, ..., L, are equal and cancel out of the equation.

3.1.8 Looking for Areas of Elevated Concentrations ("Hot Spots")

Suppose a VSP user wants to determine the spacing of sampling locations at a study site that should be used to detect a hot spot of specified size. The VSP user can specify that the sampling locations should be laid out over the study area in a square, rectangular, or triangular pattern. VSP determines the optimum spacing between sampling locations using an implementation of a computer program called ELIPGRID-PC (Davidson 1995b). The events that led to the development of this program are now briefly summarized.

Singer and Wickman (1969) published an algorithm for calculating the probability of locating elliptical hot spots when sampling is done on a square, rectangular, or triangular grid pattern over space. Singer (1972) published a FORTRAN IV computer program, ELIPGRID, to automate the hot spot probability calculations. He also evaluated the efficiency of square and triangular grids in the search for elliptically-shaped hot spots (Singer 1975). Zirschky and Gilbert (1984) developed nomographs for answering the same questions addressed by ELIPGRID. These nomographs were published in Gilbert (1987, Chapter 10) along with examples of the calculations. Davidson (1995b) wrote and published ELIPGRID-PC for the personal computer, an upgraded and corrected version of the original ELIPGRID algorithm. ELIPGRID-PC was subsequently incorporated into VSP.

The assumptions that underlie the ELIPGRID-PC and the VSP implementation of that code are,

from Gilbert (1987, pp. 119-120):

1. The target (hot spot) is circular or elliptical. For subsurface targets, this assumption applies to the projection of the target to the ground surface.
2. Samples or measurements are taken on a square, rectangular, or triangular pattern.
3. The distance between grid points is much larger than the area sampled, measured, or cored at grid points; that is, a very small proportion of the area being studied can actually be measured.
4. The definition of "hot spot" is clear and unambiguous; the types of measurement and the levels of contamination that constitute a hot spot are clearly defined.
5. There are no measurement misclassification errors; that is, no errors are made in deciding when a hot spot has been found.

The computations conducted by VSP to determine the spacing of sampling locations are described in Singer (1972) and Davidson (1995b).

3.1.9 Finding a UXO Target Area

In some environmental sampling situations a detector or an array of detectors are moved across the study area in approximately parallel swaths (transects) or in a square grid of parallel swaths to search for target area that contains objects of interest or that exhibits concentrations of contaminants that are unusually large in magnitude. This type of design could, for example, be used with group-based or aircraft-based sensors to search for unexploded ordinance (UXO) or metal objects that are an indication that UXO may be present.

Version 2.0 of VSP can determine the spacing that can be allowed between the swaths while still attaining a specified probability of traversing and detecting a circular or elliptical target area of interest. First, the VSP user specifies (1) the size and shape of the target area of interest, (2) the width of the swath that will be used, (3) the required probability that one or more swaths cross the target area, (4) a trigger density (number per unit area) and a critical density of detectable objects of interest, (5) whether the density of objects within the target area is constant throughout the target area or varies as a bivariate normal distribution, and (6) the false negative detection error rate of the sensor being used in the survey. Then VSP computes the required transect spacing and the probability of both traversing and detecting the target area of interest. VSP can also compute the probability that a target area of interest exists even though none have been found using the design developed in VSP (using the method in Gilbert 1987, pages 128 and 129). This computation requires that the VSP user specify his/her a priori belief (probability) that a target area of concern exists.

VSP also has the capability to assess the performance of a meandering swath pattern over the study area to traverse and detect a target area. In this application, the VSP user supplies to VSP the actual swath path used. Then VSP determines locations in the study area that could contain a target area of a specified size and shape that had been missed by the swath pattern.

The computational methods in Version 2.0 for finding UXO target areas, although believed to be accurate, are none-the-less preliminary and subject to change. When those methods are finalized and thoroughly checked for accuracy, they will be documented in a separate PNNL report similar to this report.

3.1.10 Compliance Sampling

Version 2.0 of VSP has the capability of computing the number, n, of transects at the study site that should be surveyed in order to have a specified confidence, say 95%, that the fraction of transects that are defective (say, contain UXO) is less than a specified percent defective, say 1%. The method used to compute the required number of transects is from Shilling (1982, pp. 474-482). As Shilling's description of the method is very clear, the reader is referred to his book for details. An underlying assumption of the method is that simple random sampling is used to select the n transects that will be surveyed.

3.1.11 Non-Statistical Sampling Approach

Version 2.0 of VSP permits the user to simply specify the total number of samples, n, for a study area without going through the DQO planning process. Once n is specified, VSP determines the n sampling locations in the study area using either simple random sampling or a systematic grid pattern, as specified by the VSP user. This approach is non-statistical in the sense that no criterial (DQOs) are provided by the VSP user regarding the purpose of sampling or the quality of the decisions or estimated quantities that are obtained using the n measurements obtained.

3.2 Verification of VSP Computations and Outputs

The activities conducted to verify that Version 2.0 of VSP is computing, displaying, and describing VSP outputs correctly and accurately are documented in this section. Section 3.2.1 presents the verification results reported in Gilbert et al. (2001) for the simple random sampling option in VSP. These results also apply to the systematic sampling option in VSP because the equations used for that option are identical to those used for simple random sampling. Section 3.2.2 presents additional accuracy checking of Version 2.0 conducted by the Research Triangle Institute in 2002. Section 3.2.3 discusses verification results for the hot-spot detection option (from Gilbert et al. 2001). These "hot-spot" results are discussed more fully in Davidson (2001). Section 3.2.4 presents verification results for sequential sampling that uses the Sequential Probability Ratio Test (SPRT). The verification results for Barnard's sequential t-test are provided in Section 3.2.5. Tables 3.2 and 3.3 summarize the verification activities and sections of this report and previous reports where the verification results are documented.

Table 3.2 Summary and References to Verification of Computations in VSP Version 2.0

VSP Statistical Sampling Goal	Data Distribution Assumption	Sampling Design	Statistical Procedure	Section of this Report where the Statistical Procedure is Documented	Sections and Reports where Verification Computations are Found

Compare Average to Fixed Threshold	Normal	Random or Systematic (Grid)	One-Sample t-Test	3.1.1.1	3.2.1.1, 3.2.2.2, and 3.2.1.3 in this report and in Gilbert et al (2001); 3.2.2 in this report; 2.1.1 in Davidson (2001)
	Normal	Random	Sequential Probability Ratio Test and Barnard's Sequential t-Test	3.1.1.1	3.2.4 and 3.2.5 in this report
	Symmetric Distribution	Random or Systematic (Grid)	Wilcoxon Signed Ranks Test	3.1.1.2	3.2.1.2 and 3.2.1.3 in this report; 3.2.1 in Gilbert et al (2001); 2.2.1 in Davidson et al (2001)
	None	Random or Systematic (Grid)	MARSSIM Sign Test	3.1.1.2	3.2.1.2, 3.2.1.3 and 2.2.5 in this report; 3.2.1 in Gilbert el al (2001)

Table 3.2 Summary and References to Verification of Computations (cont.)

Compare Average to Reference Average	Normal	Random or Systematic (Grid)	Two-Sample t-Test	3.1.2.1	3.2.1.2 and 3.2.2 in this report; 3.2.1 in Gilbert et al (2001); 2.1.2 in Davidson (2001)

	None	Random or Systematic (Grid)	Wilcoxon Rank Sum Test and MARSSIM Wilcoxon Rank Sum Test	3.1.2.2	3.2.1.2 and 3.2.1.3 in this report; 3.2.1 in Gilbert et al (2001); 2.2.4 and 2.2.6 in Davidson (2001)
Confidence Interval for the Mean	Normal	Random or Systematic (Grid)	Confidence Interval computed using the t distribution	3.1.3	3.2.1.2 and 3.2.1.3 in this report; 3.2.1 in Gilbert et al (2001); 2.1.3 in Davidson (2001)
Estimate the Mean	None	Random or Systematic (Grid)	Stratified Sampling; Ranked Set Sampling; Adaptive Cluster Sampling	3.1.4.1 3.1.4.2 3.1.4.3	3.2.2 in this report
Compare Proportion to Fixed Threshold Proportion	None	Random or Systematic (Grid)	One-Sample Test for Proportion	3.1.5	3.2.1.2 and 3.2.2 in this report; 3.2.1 in Gilbert et al (2001); 2.2.2 in Davidson (2001)

Table 3.2 Summary and References to Verification of Computations

Compare Proportion to Reference Proportion	None	Random or Systematic (Grid)	Two-Sample Test for Proportion	3.1.6	3.2.1.2 and 3.2.1.3 and 3.2.2 in this report; 3.2.1 in Gilbert et al (2001); 2.2.3 in Davidson (2001)
Estimate a Proportion	None	Random or Systematic (Grid)	Stratified Sampling	3.1.7	3.2.2 in this report
Looking for a Local Area of Elevated Concentrations	None	Systematic (Grid)	Described in Gilbert (1987, chapter 10)	3.1.8	3.2.3 in this report; 3.2 in Davidson (2001)
Finding a UXO Target Area	Unchanging distribution or Bivariate Normal Distribution of Density of UXO (number per square meter) in the target area	Parallel transects or a square or rectangular grid of transects	Current method described in Gilbert et al (November 2001)	3.1.9; General Description; Detailed Formula and Verification to be accomplished in future.	In progress; verification has not been completed for this module within Version 2.0
Non-Statistical Method to Determine the Number of Samples	None	Random or Systematic (Grid)	Specified by VSP User	3.1.10	3.2.2 in this report

Table 3.3 Summary and References to Verification and Documentation of Non-Statistical Portions of VSP Version 2.0

Verification Feature	**Section in This Report**
Installation Success for Computer Platform	4.1
File Import, Export, and Removal of Sampling Locations	4.2

Table 3.3 Summary and References to Verification and Documentation (cont.)

Verification Feature	Section in This Report
Drawing Function	4.2 Appendix B Appendix C
Correspondence Between Dialogue Box Values (DQOs) and Values in View Window	4.4 Appendix D
Algorithm to Determine Sampling Locations	4.5
Random Number Generators	4.6
Largest Unsampled Spot Algorithm	4.7
Post Survey Target Detection Algorithm	4.8

3.2.1 Simple Random Sampling and Systematic Sampling

3.2.1.1 Comparisons of Sample Sizes Computed by VSP and S-PLUS when MQO Module is not Used

The sample sizes computed by both VSP and S-PLUS were compared for the one-sample t test, the Wilcoxon Signed Ranks test, the one-sample proportion test, the MARSSIM Sign test, the confidence interval for the mean, the two-sample t-test, the Wilcoxon Rank Sum test and the MARSSIM WRS test. These computations were conducted for when the VSP user chooses not to use the Measurements Quality Objectives (MQO) module, in which case VSP sets r = 1, where r is the desired number of analytical replicates per field sample (Davidson 2001). The null hypothesis used was "the site is dirty." For this r = 1 case, 15 test cases for each of the statistical tests were computer-generated. For each test case, a value of each parameter in the sample-size equation was generated from a uniform distribution. This process yielded 15 cases for each test that had a wide variety of parameter value combinations. Then the sample size for each test case was computed using both the VSP program and the S-PLUS program.

All VSP-calculated sample-sizes agreed with the sample sizes calculated by the S-PLUS program. The results for the one-sample t-test (Equation 3.1) are given in Table 3.4. The complete results for all the other statistical tests are provided in Davidson (2001). The S-PLUS test file and underlying computer code are given in Appendices A and B, respectively, of Davidson (2001).

Table 3.4. Tests of VSP Software for Computing the Sample Size for the One-Sample t-Test

Test Case	Alpha	Beta	Delta	Standard Deviation	Calculated Sample Sizes	
					VSP	S-PLUS
1	0.22	0.22	4.43	3.20	2	2
2	0.06	0.08	3.52	2.66	7	7
3	0.19	0.19	1.19	4.52	45	45
4	0.14	0.13	0.38	2.55	220	220
5	0.07	0.18	0.25	4.28	1677	1677
6	0.12	0.24	2.79	0.86	2	2
7	0.03	0.03	3.02	3.87	26	26
8	0.10	0.21	1.99	7.17	58	58
9	0.05	0.15	3.75	9.87	52	52
10	0.14	0.14	0.94	5.33	151	151
11	0.23	0.24	4.66	6.54	5	5
12	0.23	0.21	4.15	3.16	2	2
13	0.02	0.17	3.79	9.62	61	61
14	0.02	0.16	3.13	6.69	45	45
15	0.14	0.09	3.75	9.21	36	36

3.2.1.2 Comparisons of Hand Calculations with VSP and S-PLUS when MQO Module is not Used

The accuracy of the VSP and S-PLUS calculations was evaluated by also computing the sample-size equations by hand. The hand computations were conducted for 3 of the 15 test cases for each of the statistical tests. The 3 cases for each test were selected so that the computed values of n ranged from very small (less than 10) to very large (more than 1000). The results are provided in Table 3.5. In total, hand calculations were conducted for 27 of the 135 total test cases (15 test cases for each of the 9 statistical tests equals 135). For all 27 cases, the sample sizes computed by hand were exactly equal to those obtained using the VSP and S-PLUS computer programs.

Table 3.5. Hand Calculations to Test if VSP Software Accurately Computes Sample Sizes

Statistical Test	Case	Alpha	Beta	Delta [a]	Sd_{total}	Calculated Sample Size		
						VSP	S-PLUS	By Hand
One-sample t-test	1	0.22	0.22	4.43	3.20	2	2	2
	2	0.07	0.18	0.25	4.28	1677	1677	1677
	3	0.02	0.17	3.79	9.62	61	61	61
Wilcoxon signed ranks test	1	0.22	0.22	4.43	3.20	2	2	2
	2	0.07	0.18	0.25	4.28	1946	1946	1946
	3	0.02	0.17	3.79	9.62	71	71	71
One-sample proportion test	1	0.22	0.22	0.289	NA	2	2	2
	2	0.07	0.18	0.02	NA	3407	3407	3407
	3	0.02	0.17	0.37	NA	7	7	7
MARSSIM sign test	1	0.22	0.22	4.43	3.20	4	4	4
	2	0.07	0.18	0.25	4.28	2636	2636	2636
	3	0.02	0.17	3.79	9.62	117	117	117
Confidence interval for mean	1 [b]	0.07	NA	0.64 [c]	8.90	637	637	637
	2 [d]	0.23	NA	0.51 [e]	1.05	4	4	4
	3 [b]	0.27	NA	0.17 [c]	5.72	1379	1379	1379
Two-sample t-test	1	0.22	0.22	4.43	3.20	3	3	3
	2	0.07	0.18	0.25	4.28	3353	3353	3353
	3	0.02	0.17	3.79	9.62	118	118	118
Wilcoxon rank sum test	1	0.22	0.22	4.43	3.20	4	4	4
	2	0.07	0.18	0.25	4.28	3889	3889	3889
	3	0.02	0.17	3.79	9.62	137	137	137
MARSSIM WRS test	1	0.22	0.22	4.43	3.20	4	4	4
	2	0.07	0.18	0.25	4.28	3512	3512	3512
	3	0.02	0.17	3.79	9.62	126	126	126

Table 3.5. Hand Calculations (cont.)

Statistical Test	Case	Alpha	Beta	Delta [(a)]	Sd_{total}	Calculated Sample Size		
						VSP	S-PLUS	By Hand
Two-sample test for proportions	1	0.22	0.22	0.44	NA	7	7	7
	2	0.07	0.18	0.03	NA	3062	3062	3062
	3	0.02	0.17	0.38	NA	23	23	23

NA = Not applicable; [(a)] width of gray region of the Decision Performance Goal Diagram
[(b)] two-sided confidence interval was computed; [(c)] half-width of two-sided confidence interval
[(d)] one-sided confidence interval was computed; [(e)] width of one-sided confidence interval.

3.2.1.3 Comparisons of Hand Calculations with VSP and S-PLUS When the MQO Module is Used

The sample-size equations that apply when the MQO module was used, i.e., when the desired number of analytical replicates per field sample (r) was greater than 1, were hand-calculated for the following specific cases:

- one-sample t-test [Equation (3.2) for r = 1]
- Wilcoxon signed ranks test [Equation (3.4) for r = 1]
- confidence limit on the mean [Equation (3.17) for r = 2]
- MARSSIM sign test [Equations (3.5) and (3.7) for r = 1 and 2]
- two-sample t-test [Equation (3.9) for r = 2]
- Wilcoxon rank sum test [Equation (3.12) for r = 2]
- MARSSIM WRS test [Equations (3.13) and (3.15) for r = 2]

For each of the above cases, the hand calculation of the sample size agreed exactly with that computed by VSP. However, two problems were identified and corrected. The most important problem found (and subsequently corrected) was that in some cases VSP incorrectly displayed the DPGD specified by the VSP user for the MARSSIM sign test and the Wilcoxon rank sum test.

The one-sample and two-sample test for proportions are not included in the above list because the MQO module is not implemented in Version 2.0 of VSP. However, two test cases (hand calculations) were conducted for the one-sample test for proportions for the non-MQO module case. VSP correctly computed the sample size when the null hypothesis selected was "the site is dirty." However, VSP did not correctly compute the sample size when the null hypothesis "the site is clean" was selected. This problem was corrected.

3.2.2 Comparisons Conducted by the Research Triangle Institute (RTI)

Verifications of the accuracy of the sample-size calculations in Version 2.0 of VSP reported in this section were conducted by the Research Triangle Institute (RTI) in the spring of the year 2002. The results indicate that the sample-size equations in VSP that were evaluated by RTI are properly implemented and contain no rounding errors for the particular VSP inputs used in the comparisons. As the equations used in

Version 1.2 are also used in Version 2.0 of VSP, the sample-size equations are also properly and accurately implemented in Version 2.0.

Table 3.6 defines the terms used in Tables 3.7 through 3.15, which present the results of the comparisons conducted by RTI.

Table 3.6. Definition of Terms for the Comparisons Conducted by the RTI

Inputs	
mean vs. AL	"true mean ≥ action level" or "true mean ≤ action level"
alpha	false rejection rate (in %)
beta	false acceptance rate (in %)
delta	width of gray region
GSD	geometric standard deviation
stdev	estimated standard deviation
diff means vs. AL	"difference of true means ≥ action level" or "difference of true means ≤ action level"
dist	laboratory data distribution
rank	ranking method
err/cor	ranking error or analysis/screening correlation
m	set size in ranked set sampling
analy C	analytical cost per analysis
rank C	ranking cost
sided	one-sided or two-sided confidence interval
RP	relative precision
t	number of times top rank is sampled per cycle
conf	confidence level
md	maximum acceptable % difference between estimated and true mean
d	maximum acceptable width (or half-width) of confidence interval
P vs. P_0	"true proportion ≥ given proportion" or "true proportion ≤ given proportion"
P_0	given proportion (action level)
P_1 vs. **P_2**	"difference of proportions ≥ specified difference" or "difference of proportions ≤ specified difference"
P_2	estimated proportion in reference area
P_1	estimated proportion in survey unit
$delta_0$	specified difference of proportions (action level)
method	fixed cost or fixed variance
V	required variance of mean
C	total budget
C_0	initial fixed costs
meth of alloc	optimal allocation or optimal allocation assuming equal stratum sampling/measurement costs
L	number of strata
N_i	number of potential sample in stratum i
P_i	proportion in stratum i
C_i	cost per sample in strata i

Table 3.6. Definition of Terms for the Comparisons Conducted by the RTI

Inputs	
# of samples	number of samples
Outputs	
testn	sample size obtained from independent testing
VSPn	sample size obtained in VSP
agree?	"yes" indicates that the independent testing and VSP agree for that case
testnm	sample size for survey area and reference area obtained from independent testing
VSPnm	sample size for survey area and reference area obtained in VSP
testA/r	analyC / rankC obtained from independent testing
VSPA/r	analyC / rankC obtained in VSP
testn0	sample size under SRS obtained from independent testing
VSPn0	sample size under SRS obtained in VSP
testCr	cost ratio obtained from independent calculations
VSPCr	cost ratio obtained from VSP
testCe	independent calculations determined cost effective
VSPCe	VSP determined cost effective
$testn_i$	sample size in stratum i obtained from independent testing
$VSPn_i$	sample size in stratum i obtained in VSP

Table 3.7. Systematic Grid Sampling for Comparing an Average to a Fixed Threshold Using the One-Sample t-Test*

Case	**Alpha**	**Beta**	**Delta**	**Stdev**	**Testn**	**VSPn**	**Agree?**
1	14	22	4.43	9.21	16	16	yes
2	6	8	3.52	2.66	7	7	yes
3	0.5	0.5	1.19	4.52	387	387	yes
4	14	13	0.38	2.55	220	220	yes
5	7	18	0.25	4.28	1677	1677	yes
6	12	24	1.00	0.86	4	4	yes
7	75	3	3.02	3.87	3	3	yes
8	10	1	1.99	7.17	170	170	yes
9	5	15	3.75	9.87	52	52	yes

*The sample size equation verified is Equation (3.1) in the text.

Table 3.8 Simple Random Sampling for Comparing a Proportion to a Fixed Threshold*

Case	**P vs. P_0**	**Alpha**	**Beta**	**Delta**	**P_0**	**Testn**	**VSPn**	**Agree?**
1	≥ **	0.5	22	0.2	0.4	62	62	yes
2	≤	6	8	0.52	0.4	5	5	yes
3	≥	19	16	0.19	0.82	19	19	yes
4	≤	14	13	0.3	0.3	13	13	yes
5	≥	7	18	0.25	0.25	7	7	yes
6	≤	12	24	0.1	0.4	87	87	yes

Table 3.8 Simple Random Sampling for Comparing a Proportion to a Fixed Threshold* (cont.)

Case	P vs. P_0	Alpha	Beta	Delta	P_0	Testn	VSPn	Agree?
7	≥	3	3	0.02	0.1	2887	2887	yes
8	≤	10	0.5	0.9	0.1	1	1	yes
9	≥	5	15	0.75	0.9	2	2	yes

* The sample-size equation verified is Equation (3.18) in the text.

** The notation ≥ and ≤ refers to whether the null hypothesis states that the true average was "greater than or equal to" or "less than or equal to" the threshold value, respectively.

Table 3.9. Systematic Grid Sampling for Comparing a Proportion to a Fixed Threshold*

Case	P vs. P_0	Alpha	Beta	Delta	P_0	Testn	VSPn	Agree?
1	≥ **	0.5	22	0.2	0.4	62	62	yes
2	≤	6	8	0.52	0.4	5	5	yes
3	≥	19	16	0.19	0.82	19	19	yes
4	≤	14	13	0.3	0.3	13	13	yes
5	≥	7	18	0.25	0.25	7	7	yes
6	≤	12	24	0.1	0.4	87	87	yes
7	≥	3	3	0.02	0.1	2887	2887	yes
8	≤	10	0.5	0.9	0.1	1	1	yes
9	≥	5	15	0.75	0.9	2	2	yes

* The sample-size equation verified is Equation (3.18) in the text.

** The notation ≥ and ≤ refers to whether the null hypothesis states that the true average was "greater than or equal to" or "less than or equal to" the threshold value, respectively.

Table 3.10. Systematic Grid Sampling for Comparing an Average to a Reference Average*

Case	Alpha	Beta	Delta	Stdev	Testnm	VSPnm	Agree?
1	14	22	4.43	9.21	30	30	yes
2	6	8	3.52	2.66	11	11	yes
3	0.5	0.5	1.19	4.52	768	768	yes
4	14	13	0.38	2.55	439	439	yes
5	7	18	0.25	4.28	3353	3353	yes
6	12	24	1	0.86	6	6	yes
7	75	3	3.02	3.87	5	5	yes
8	10	1	1.99	7.17	339	339	yes
9	5	15	3.75	9.87	101	101	yes

*The sample size equation being verified is Equation (3.8) in the text.

Table 3.11. Simple Random Sampling for Comparing a Proportion to a Reference Proportion*

Case	P_1 vs. P_2	Alpha	Beta	P_2	P_1	Delta	$Delta_0$	Testn	VSPn	Agree?
1	≤ **	0.5	22	0.3	0.1	0.2	0.4	23	23	yes
2	≥	6	8	0.8	0.6	0.52	0.4	23	23	yes
3	≤	19	16	0.9	0.1	0.19	0.81	3	3	yes
4	≥	14	13	0.1	0.2	0.3	0.3	14	14	yes
5	≤	7	18	0.3	0.99	0.25	0.25	42	42	yes
6	≥	12	24	0.0	0.3	0.1	0.1	91	91	yes
7	≤	3	3	0.9	0.0	0.02	0.1	701	701	yes
8	≥	10	1	0.2	0.4	0.99	0.1	547	547	yes
9	≤	5	15	0.4	0.9	0.75	0.25	53	53	yes

*The sample size equation being verified is Equation (3.19) in the text.

** The notation ≥ and ≤ refers to whether the null hypothesis states that the true average was "greater than or equal to" or "less than or equal to" the threshold value, respectively.

Table 3.12. Systematic Grid Sampling for Comparing a Proportion to a Reference Proportion*

Case	P_1 vs. P_2	Alpha	Beta	P_2	P_1	Delta	$Delta_0$	Testn	VSPn	Agree?
1	≤ **	0.5	22	0.3	0.1	0.2	0.4	23	23	yes
2	≥	6	8	0.8	0.6	0.52	0.4	23	23	yes
3	≤	19	16	0.9	0.1	0.19	0.81	3	3	yes
4	≥	14	13	0.1	0.2	0.3	0.3	14	14	yes
5	≤	7	18	0.3	0.99	0.25	0.25	42	42	yes
6	≥	12	24	0.0	0.3	0.1	0.1	91	91	yes
7	≤	3	3	0.9	0.0	0.02	0.1	701	701	yes
8	≥	10	1	0.2	0.4	0.99	0.1	547	547	yes
9	≤	5	15	0.4	0.9	0.75	0.25	53	53	yes

*The sample size equation being verified is Equation (3.19) in the text.

** The notation ≥ and ≤ refers to whether the null hypothesis states that the true average was "greater than or equal to" or "less than or equal to" the threshold value, respectively.

Table 3.13. Ranked Set Sampling for Estimating the Mean

Symmetric Distribution Case

Case	Dist.	Rank	Err/ cor	M	Anal y C	Rank C	TestA/r	VSPA/r	Test Cr	VS PC r	Test Ce	VSPC e
1	sym	prof	min	2	400	75	5.3	5.3	4	4	yes	yes
2	sym	prof	sub	2	400	75	5.3	5.3	7	7	no	no
3	sym	field	1.0	2	400	75	5.3	5.3	5	5	yes	yes
4	sym	field	0.7	4	500	600	0.833	0.833	12	12	no	no
5	sym	field	0.8	4	500	50	10	10	8	8	yes	yes

Case	Sided	Conf	Stde v	D	Testn 0	VSP n0	Testn*	VSPn	Agree?
1	one	90	3.20	5	3	3	4	4	yes
2	two	90	3.2	5	3	3	4	4	yes
3	two	98	20	3	244	244	168	168	yes
4	one	83	30	1.8	254	254	112	112	yes
5	two	83	30	1.8	524	524	224	224	yes

Asymmetric Distribution Case

Case	Dist.	Rank	Err/ cor	M	analy C	Rank C	TestA/ r	VSP A/r	Test Cr	VSP Cr	Tes tCe	VSP Ce
1	asym	prof	min	2	400	75	5.3	5.3	6	6	no	no
2	asym	field	0.9	2	400	75	5.3	5.3	6	6	no	no
3	asym	prof	min	5	500	1000	0.5	0.5	4	4	no	no
4	asym	prof	sub	5	500	50	10	10	9	9	yes	yes
5	asym	field	0.7	8	500	50	10	10	16	16	no	no

Case	Sided	Conf	GSD	Md	Testn0	VSPn0	Testn*	VSP n	Agree?
1	n/a	95	2	50	23	23	24	24	yes
2	n/a	99	1.5	50	16	16	15	15	yes
3	n/a	90	3.0	80	29	29	30	30	yes
4	n/a	95	1.1	60	4	4	5	5	yes
5	n/a	99	3.5	30	293	293	156	156	yes

* In calculating $n_{classic}$ (an intermediate step in computing testn), the computed value of $n_{classic}$ was not rounded up to the next largest integer.

** In determining values from Table 1 in Patil et al (1994) and from Tables 8-5 and 8-6 in EPA (2001), the nearest value of CV that was $\geq$ to the CV independently calculated was used, except for Case (2) for which the CV was interpolated to be equal to 4

Table 3.14. Adaptive Cluster Sampling for Estimating the Mean

Case	Sided	Conf	Stdev	d	Testn	VSPn	Agree?
1	one	90	3.2	5	3	3	yes
2	two	90	3.2	5	3	3	yes
3	one	99	4	2	25	25	yes
4	two	99	4	2	30	30	yes
5	one	50	4.2	6	2	2	yes
6	two	50	4.2	6	2	2	yes
7	one	83	30	1.8	254	254	yes
8	two	83	30	1.8	524	524	yes
9	one	97	1.8	3	4	4	yes
10	two	97	1.8	3	5	5	yes

NOTE: In determining n, the independently calculated values in the iterative scheme were rounded (but not necessarily rounded up). This follows the way rounding was done in the example on page 32 of Gilbert (1987).

Table 3.15. Stratified Sampling for Estimating a Proportion for the Case of Fixed Cost, Fixed Variance, or Predetermined Number of Total Samples

Fixed Cost

Case	Method		C	C_0	Meth of Alloc	L	N_1	P_1	C_1
1	fixed cost		10,000	1,000	optimal	2	100	.7	300
2	fixed cost		10,000	1,000	equal strat cost	2	100	.7	300
3	fixed cost		75,000	10,000	optimal	2	672	.3	1,000
4	fixed cost		75,000	10,000	equal strat cost	2	672	.3	1,000
5	fixed cost		99,000	40,000	optimal	2	500	.5	50

Case	N_2	P_2	C_2	Testn	VSPn	Testn 1	$VSPn_1$	$Testn_2$	$VSPn_2$
1	200	.8	350	29	29	11	11	**18**	18
2	200	.8	350	29	29	11	11	18	18
3	700	.5	900	70	70	32	32	38	38
4	700	.5	900	70	70	33	33	37	37
5	500	.6	50	1181	1181	597	597	584	584

Fixed Variance

Case	Method	V			Meth of Alloc	L	N_1	P_1	C_1
1	fixed var	.004			equal strat cost	2	100	.7	300
2	fixed var	.004			optimal	2	100	.7	300
3	fixed var	.002			equal strat cost	2	672	.3	1,000
4	fixed var	.002			optimal	2	672	.3	1,000
5	fixed var	.009			equal strat cost	2	500	.5	50

Table 3.15. Stratified Sampling for Estimating a Proportion (cont.)

Case	N_2	P_2	C_2	Testn	VSPn	$Testn_1$	$VSPn_1$	$Testn_2$	$VSPn_2$	Agree?
1	200	.8	350	40	40	15	15	25	25	yes
2	200	.8	350	40	40	15	15	25	25	yes
3	700	.5	900	108	108	51	51	57	57	yes
4	700	.5	900	108	108	49	49	59	59	yes
5	500	.6	50	28	28	14	14	14	14	yes

Predetermined Number of Total Samples

Case	# of Samples	Meth of Alloc	L	N_1	P_1	C_1	N_2	P_2	C_2
1	100	optimal	2	632	.7	100	600	.5	89
2	1000	optimal	2	180	.2	75	250	.9	100
3	86	optimal	2	50	.5	300	50	.5	300
4	100	equal strat cost	2	632	.7	n/a	600	.5	n/a
5	1000	equal strat cost	2	180	.2	n/a	250	.9	n/a

Case	Testn	VSPn	$Testn_1$	$VSPn_1$	$Testn_2$	$VSPn_2$	Agree?
1	101	101	48	48	53	53	yes
2	1001	1001	526	526	475	475	yes
3	86	86	43	43	43	43	yes
4	101	101	50	50	51	51	yes
5	1001	1001	490	490	511	511	yes

3.2.3 Hot Spot Sampling

Davidson (1995a, 1995b) conducted Monte Carlo simulations to verify that the upgraded ELIPGRID-PC program was computing hot-spot detection probabilities correctly. He also compared the computed ELIPGRID-PC detection probabilities with those of 100 test cases published in Singer (1962). *Davidson (1995b) found that his upgraded ELIPGRID-PC program produced probabilities that exactly matched those of the 100 test cases.* Davidson (2001) applied this same testing procedure to the implementation of ELIPGRID-PC in the VSP software. *He found that the VSP program also gave complete agreement with the 100 test cases in Singer (1972).* These results provide confidence that the ELIPGRID computations in VSP are accurate. The input and output files of these 100 cases are listed in Appendices C and D, respectively, of Davidson (2001).

3.2.4 Sequential Sampling Using the Sequential Probability Ratio Test (SPRT)

Verification of the consistency and correct coding of the Sequential Probability Ratio Test (SPRT) in Version 2.0 of VSP was obtained by comparing results of the SPRT from Version 2.0 with the results of SPRT obtained from an independently written SAS (Statistical Analysis System) (Version 8.2) code.

The verification was conducted by first generating six sequential data sets with 50 observations each using the "RANNOR" function in SAS (Version 8.2). Three of the data sets were applied to the null and alternative hypothesis

H_{oc}: The site mean is less than the action limit (AL)
H_{ac}: The site mean is greater than or equal to the AL

The remaining three data sets were applied to the hypotheses

H_{od}: The site mean is equal to or greater than the AL
H_{ad}: The site mean is less than the AL

For all data sets the AL was equal to 10, the gray-region width was equal to 2, the known standard deviation was equal to 3, the maximum acceptable Type I decision error rate (probability of falsely rejecting the null hypothesis) was 0.05, and the maximum acceptable Type II decision error rate (probability of falsely accepting the null hypothesis) was 0.10. Among the three data sets used to test H_{oc} versus H_{ac}, one data set was randomly generated such that almost all the data were less than the AL (10), one data set had some values larger and some values smaller than the AL. The third data set had almost all data values greater than the AL. This approach was also used for the three data sets used to test H_{od} versus H_{ad}.

The consistency and correct coding of the Sequential Probability Ratio Test (SPRT) in Version 2.0 of VSP was verified on the basis of the absolute differences in value of the following three test parameters computed by both the SAS and VSP codes:

- Sample mean
- Upper decision boundary (UL_c if H_{oc} is used; UL_d if H_{od} is used)
- Lower decision boundary (LL_c if H_{oc} is used; LL_d if H_{od} is used)

(See Section 3.1.1.1 for the equations used to compute UL_c, UL_d, LL_c and LL_d.)

Among the 300 differences computed for each parameter, the maximum differences were:

- Sample mean: 1.00E-8
- Upper Decision Boundary: 0.00E+00
- Lower Decision Boundary: 0.00E+00

Clearly, there is almost exact agreement between the two codes, which indicates that Version 2.0 of VSP is consistently and correctly computing the SPRT.

3.2.5 Barnard's Sequential t-Test

Verification of the consistency and correct coding of Barnard's Sequential t-test in Version 2.0 of VSP was obtained by comparing VSP results with results of that test obtained using an independently written SAS (version 8.2) code. We note that this verification assumes that Barnard's Sequential t-test methodology as described in the scientific literature (Barnard 1952) is correct. The methodology of Barnard's test is discussed in more detail in Section 3.1.1.1 and Appendix A.

The verification was conducted by first generating six data sets of 50 observations each using the "RANNOR" function in SAS (version 8.2). Three of the data sets were applied to the hypotheses**:**

H_{oc}: The site mean is less than the action limit (AL)

H_{ac}: The site mean is greater than or equal to the AL

The remaining three data sets were applied to the hypotheses

H_{od}: The site mean is equal to or greater than the AL
H_{ad}: The site mean is less than the AL

All hypothesis tests were specified using an action level equal to 10 and a gray-region width equal to 2. Values required to implement Barnard's sequential t-test were then independently computed in SAS (ver. 8.2), which is referenced here as the 'test code.' The Version 2.0 VSP C++ source code is referenced here as the 'source code.' These independently generated values were compared on the basis of absolute and relative (%) differences in value between the test code and the source code. Relative % differences were formed by dividing the absolute difference by the value computed by the source code and multiplying by 100. The values compared were as follows:

1. X — Randomly generated value from a normal distribution
2. Xbar — The sample mean. This value is computed initially from the first 10 samples per data set, and is then updated in sequence starting with the 11th sample.
3. StdErrX — The sample standard error of the mean, based on the sample standard deviation. This value is computed initially from the first 10 samples per data set, and is then updated in sequence starting with the 11th sample.
4. Delta — The non-centrality parameter supplied to the non-central t PDF (probability density function). This value is computed as the width of the gray-region divided by StdErrX.
5. tstat — The t-statistic value used to compute both the non-central and central t PDF values
6. NCTPDF — The computed value from the non-central t PDF given the t statistic and delta value
7. CTPDF — The computed value from the central t PDF given the t statistic and setting the non-centrality parameter (Delta) to zero
8. Ln — The likelihood ratio test (LRT) statistic
9. logLn — The natural-log transformed LRT statistic

Table 3.16 summarized the verification results. The absolute differences for the randomly generated data (X) agree exactly. This comparison was made as a quality control measure to ensure that all further comparisons were appropriate and correct. Also in Table 3.16, absolute differences in values for Xbar, StdErrX, delta (the non-centrality parameter), and the t-statistic all agree to between 9 or 10 decimal places. Relative (%) differences for these same parameters agree to between 6 and 8 decimal places.

Table 3.16. Barnard's Sequential t-Test Verification Results

Parameters	Maximum Absolute Difference	Maximum Absolute Percent Difference
Random Normal Data Value (X)	00.00E+00	00.00E+00%
Sample Mean (Xbar)	5.45E-09	4.86E-08%

Table 3.16. Barnard's Sequential t-Test Verification Results (cont.)

Parameters	Maximum Absolute Difference	Maximum Absolute Percent Difference
Standard Error of Mean (StdErrX)	6.21E-10	1.82E-07%
Non-Centrality Parameter (Delta)	6.25E-10	6.83E-08%
t Statistic (tstat)	6.93E-10	1.04E-06%
Non-central t PDF (NCTPDF)	2.55E-08	1.00E+02%
Central t PDF (CTPDF)	4.94E-08	3.60E-01%
Likelihood Ratio Test (LRS) Statistic (Ln)	2.10E+07	1.00E+02%
Natural-log Transformed LRS statistic (logLn)	4.17E+01	1.80E+02%

Absolute differences between calculated values for NCTPDF agree to 8 decimal places. However, there were cases where the relative differences were on the order of 10E+2. These cases occurred when the values calculated by the source code from the NCTPDF were on the order of 10E-16, while those computed from the test code were on the order of 10E-33. The largest NCTPDF values, corresponding to the most likely X values under the Alternative hypothesis, are on the order of 10E-02, or 14 orders of magnitude larger than those computed by the source code. Similarly, absolute difference values calculated under the CTPDF, agree to 8 decimal places, while relative differences agree to at least 1 decimal place. Again these relatively large relative differences (~0.1%) occur between very small values, with CTPDF values on the order of 10E-5 as computed by both the source code and test code.

Discrepancies between the test code and source code in computation of the CTPDF and NCTPDF may be seen as a compounding or propagation of error from the small discrepancies between the t-statistic values and the non-centrality parameter (delta). Discrepancies between the test and source code are particularly sensitive to this error-propagation when computing PDF values at the extreme tails of the probability distribution. It should also be noted that evaluation of the CTPDF and NCTPDF can only be derived numerically. Thus there is further opportunity for discrepancies to occur when computing PDF values in the tails of the probability distribution.

Also in Table 3.16, the largest discrepancies in values are shown to occur in the LRT statistic (Ln). This statistic was formed by dividing NCTPDF by CTPDF. Again, propagation of error is likely the cause of much of these discrepancies. However, all cases where this discrepancy in value exceeds 10E-06 occur in cases where the relative likelihood of the data under the null hypothesis (indexed by the CTPDF), and the alternative hypothesis (indexed by the NCTPDF) differ greatly. These cases occurred with data that clearly favored one hypothesis over the other, and again involved computation of PDF values at the extreme tails of the distribution. Under such cases, a clear decision would have been indicated by Barnard's sequential t-test long before that data would have been collected.

Large discrepant values (i.e. > 10E-6) in Ln occur in data set 3 (obs. 44-50), and in data set 6 (obs. 30-50). Barnard's sequential t-test was carried out on data set 3 under the null hypothesis that the site's mean

lay below the action level. However, the X values in data set 3 clearly favor the alternative hypothesis. Thus the t-statistic values supplied to the CTPDF function were extremely unlikely (i.e., in one of the extreme tails of the CTPDF). Under this form of the sequential t-test as implemented in VSP, these extremely small values form the denominator of the LRT statistic, Ln, and are the source of this rather erratic behavior in Ln. These conditions in the Ln statistic value transfer directly to their natural log transformation.

A similar result occurs with the LRT statistic Ln with data set 6 (obs. 30-50). In this case Barnard's sequential t-test was carried out under the null hypothesis that the site's mean lay at or above the action level. Under this form of the sequential t-test as implemented in VSP, these extremely small values returned from the central t PDF again form the denominator of the LRT statistic (Ln) and produce a similar result.

In conclusion, the test code and source code demonstrate a level of agreement that is consistent with the intended uses of the VSP software. The cases where the absolute differences or relative differences seemed excessive occurred only where values were computed from the extreme tails of one of the central and non-central t PDFs. This is not unexpected as propagation of error is in effect, and these PDF values are numerically derived. Most importantly, it should be noted that the conditions leading to large discrepancies between the test code and source code output, still lead to the same correct conclusions with regard to the null and alternative hypotheses. Further, the conditions that produce any large discrepancies would be highly unlikely to encounter under any practical scenario, as they occurred only after a clear decision would have already been reached.

3.2.6 Other VSP Outputs

PNNL researchers also checked the accuracy of the four graphical displays of VSP outputs- the "Map View," the "Graph View," the "Report View," and the "Coordinate View" (see Figure 2.1). The Map View shows on the map of the study site the sampling locations determined by VSP. The Graph View shows the DPGD. The Report View documents the sampling design selected by the VSP user (e.g., simple random sampling) and other design details such as the size of the study site, the recommended minimum number of samples, and the total cost of sampling and measurement. The Coordinate View lists the geographical coordinates of the sampling locations.

It was discovered that VSP displayed an incorrect graph of the DPGD of the MARSSIM sign test and the Wilcoxon rank sum test in certain cases. The only other problem discovered was an incomplete listing of all the sample locations in the Coordinate View that occurred for one test case. These problems were corrected for Version 1.0 and, of course, for Version 2.0.

4.0 Verification and Documentation of Non-Statistical Portions of VSP

4.1 Installation Success for Various Computer Platforms

VSP was designed originally for the Microsoft (R) Windows (R) 95 operating system, but has been installed and run successfully on the following operating systems:

- Windows 95
- Windows 98
- Windows 98 Second Edition
- Windows Millennium Edition (ME)
- Windows 2000
- Windows XP
- Windows NT 4.0

4.2 Verification of File Import, Export, and Removal of Sampling Locations

One of the key functions of VSP is to overlay random sampling locations on a user's site map. The coordinates for these locations usually are generated by VSP but they also can be imported into a selected study area, exported (saved) to an ASCII text file, or removed from a selected study area. PNNL conducted three tests to verify the import, export, and removal of sampling locations.

A file of 100 random sampling locations (sample points) was generated for the following tests. See Appendix B for the file used. The test file had the following format:

Area 1

X Coord	Y Coord	Label	Value	Type	Historical
xxx.xx	*yyy.yy*	Test-*t*	*v.v*	Random	*h*

The first 2 lines are a literal header followed by 100 lines of the given format where:

xxx.xx is a random number between 0 and 100
yyy.yy is a random number between 0 and 100
t is the line number (1 2 3 …)
v is the line number divided by 10 (0.1 0.2 0.3 …)
h alternates between F and T for each line (F T F …)

Note: All values were separated by tab characters.

4.2.1 Import of Sampling Locations

The 100 samples were imported into a square sample area that ranged from (0,0) to (100,100). The attributes were copied from the coordinate view of VSP and pasted into a spreadsheet where all the attributes were compared to the original attributes from the test file. All the attributes matched exactly. The format of the Historical flag varied from the original because VSP copies 'T' or ' ' instead of 'T' or 'F', but the logic matched exactly.

4.2.2 Export of Sampling Locations

The 100 samples were exported to a text file. The attributes were compared to the original file of random points. All attributes matched. It was discovered that an extra carriage return character was added to the end of each line, resulting in double-spacing. This problem was corrected.

4.2.3 Removal of Sampling Locations

The remove function deletes sample points from the selected sample area if the coordinates match those of the ASCII text file. The remove function was run on the original file of random points. All sample points were removed from the sample area. The remove function was also run on the exported file of points. All sample points were removed.

4.2.4 Import of Swaths

Certain unexploded ordnance (UXO) designs use swaths or transect samples. In VSP these swaths are represented as a series of connected line segments. VSP supports loading these swath line segments from an ASCII text file.

A file of 40 points was generated for this test. See Appendix C for the file used. It contained 4 lines of 10 points each (9 line segments). The test file had the following format:

LINE
xxx.xx *yyy.yy*
xxx.xx *yyy.yy*
.
.
LINE
xxx.xx *yyy.yy*
xxx.xx *yyy.yy*

The literal string "LINE" indicates the start of a new series of connected points.
xxx.xx is the x coordinate of the point
yyy.yy is the y coordinate of the point

The swath file was imported and the resulting display on the map matched the expected pattern. The map was also exported to a DXF file and the coordinates exactly matched those in the original file.

4.3 Verification of Drawing Functions

Each drawing function in VSP allows input from the mouse or the keyboard. To test the accuracy of each method of input, researchers used the following procedure for each drawing function:

1. Inputs were made using the mouse, and the points were documented.
2. The same drawing function was performed using the keyboard.

3. The size of the resulting sample area was documented.
4. Visual comparison was made between the two figures.
5. The total size of the two sample areas was confirmed to have doubled.
6. The map was exported to a DXF file, and the coordinates of the figures were compared.

4.3.1 Polyline Drawing Function

The following points were used:

(-72.80,78.00)
(-13.60,26.80)
(20.00, 56.80)
(0,94.00)
(-72.80,78.00) to close
Single-sample area size: 3262.24 square feet
Two-sample area size: 6524.48 square feet
DXF file coordinates matched exactly.

4.3.2 Rectangle Drawing Function

The following corners were used:

(-38.00, 50.00)
(76.80, 93.20)
Single-sample area size: 4959.36 square feet
Two-sample area size: 9918.72 square feet
DXF file coordinates matched exactly.

4.3.3 Ellipse Drawing Function

The following bounding corners were used:

(15.60, 8.80)
(124.40, 76.40)
Single-sample area size: 5739.46 square feet
Two-sample area size: 11478.93 square feet
Note: In VSP, an ellipse is approximated with a series of line segments.
DXF file coordinates matched exactly.

4.3.4 Curve Drawing Function

The following bounding and control points were used:

(6.00, 82.00)
(36.80, 27.20)

(90.00, 78.80)
Single-sample area size: 1181.09 square feet
Two-sample area size: 2362.19 square feet
Notes: In VSP, a curve is approximated with a series of line segments.
Sample areas are not generated automatically from curves.
DXF file coordinates matched exactly.

4.3.5 MARSSIM Room Drawing Function

The following corner points and wall height were used:

(20.00, 4.00)
(80.00, 34.00)
(8.00)
Single-sample area size: 5040.00 square feet
Two-sample area size: 10080.00 square feet
DXF file coordinates matched exactly.

4.4 Verification of Correspondence Between Dialog Box Values and Values in View Windows

VSP provides the user with four views or windows on the problem at hand – the Report View, Graph View, Map View, and Coordinates View. For a variety of dialog-box input values and dialog-box calculated values, the four VSP views were found to have values that correspond to the dialog-box values.

VSP offers the user many dialog boxes for a variety of options, most of which are related in some way to one of the four views discussed above. An exhaustive test of all possible permutations of dialog box values and the four views was beyond the scope of PNNL's investigation. Instead, two of the most important options were chosen – the one-sample t-test and the two-sample t-test - as examples of VSP's correspondence between dialog box values and view values. A more exhaustive test could possibly be done in the future if a test method were developed that did not depend on the time-intensive process of manually entering the test data into the dialog boxes.

Six test cases were randomly generated using Excel for each of 22 different sampling designs. These random values were entered manually into the appropriate dialog boxes, and a check was made for exact correspondence between the dialog-box input values, the dialog-box calculated values, and the corresponding value(s) in one of the different view windows. In some cases, the random values were adjusted automatically by VSP because of constraints inherent to those sampling designs. In other cases, VSP rounds certain inputs to a certain number of significant digits. Each case was tested on the same basic site map having 2 elliptical sample areas of different sizes and shapes. The parameters tested are included in Appendix D. Some problems were found with the Hot Spot designs when using a different grid unit than the map unit. The problems were corrected and testing was conducted on the corrected version of VSP.

Table 4.1 Tests of View Values Compared with Dialog-Box Values

VSP Option	Map View	Graph View	Report View	Coordinate View
One-Sample t	OK	OK	OK	OK
One-Sample t (MQO)	OK	OK	OK	OK
Wilcoxon Signed Ranks	OK [1]	OK	OK	OK [1]
Wilcoxon Signed Ranks (MQO)	OK	OK	OK	OK
MARSSIM Sign Test	OK	OK	OK	OK
MARSSIM Sign Test (MQO)	OK [1]	OK	OK	OK [1]
Two-Sample t	OK	OK	OK	OK
Two-Sample t (MQO)	OK [1]	OK	OK	OK [1]
Wilcoxon Rank Sum	OK [1]	OK	OK	OK [1]
Wilcoxon Rank Sum (MQO)	OK [1]	OK	OK	OK [1]
MARSSIM WRS	OK [1]	OK	OK	OK [1]
MARSSIM WRS (MQO)	OK [1]	OK	OK	OK [1]
Confidence Interval	OK	OK	OK	OK
Confidence Interval (MQO)	OK [1]	OK	OK	OK [1]
One-Sample Proportion	OK	OK	OK	OK
Two-Sample Proportion	OK [1]	OK	NA	OK [1]
Hot Spot – Predetermined	OK	OK	OK	OK
Hot Spot – Cost	OK	OK	OK	OK
Hot Spot – Minimize Samples	OK	OK	OK	OK
Hot Spot – Minimize Spot Size	OK	OK	OK	OK
Predetermined Random	OK	NA	NA	OK
Predetermined Grid	OK	NA	NA	OK

[1] For one or more cases the number of samples was too great to manually count
NA View not available for the sampling design

4.5 Documentation of Algorithms to Determine Sampling Locations

The VSP user can choose from four methods that determine sampling locations. The three methods documented here produce sampling locations based on algorithms in VSP. The user also can manually select arbitrary sampling locations using the option Sampling Designs / Judgment Sampling / Manually Add Samples. Only the three computer-based methods are documented here.

4.5.1 Regular Simple Random Sampling Algorithm

Regular simple random samples are placed in the sample area by choosing one of the sampling designs listed on the Sampling Designs / Simple Random Sampling menu. The placement option should be set to Regular Random on the Options / Sample Placement menu (this is the default setting). Using this method, each sample is placed according to the following algorithm without regard to existing samples:

1. randomx and randomy values are obtained using the chosen random number generator (described above). These random values are fractions between 0 and 1.

2. The sample point coordinates are calculated by the following formulas:
x = XMin + (Xmax - XMin) * randomx

y = YMin + (Ymax - YMin) * randomy
where
x = x coordinate of sample
y = y coordinate of sample
XMin = minimum x extent of sample area
XMax = maximum x extent of sample area
YMin = minimum y extent of sample area
YMax = maximum y extent of sample area

3. The sample location (x, y) is checked to make sure it lies inside the sample area. If it does, a new sample point is added at the location. If not, another random location is tried.

4. For multiple Sample Areas, the number of samples assigned to each Sample Area is calculated by the following formula:

n = floor(samples * area / total)
where
n = number of samples to be assigned to the current Sample Area
floor() = function that returns the largest whole number less than or equal to a number
samples = total number of samples to be assigned to all Sample Areas
area = surface area of current Sample Area
total = total surface area of all Sample Areas

If the formula leaves unassigned samples (because of rounding down), one sample is added to each Sample Area in turn, until all remaining samples are placed.

4.5.2 Adaptive-Fill Simple Random Sampling Placement Algorithm

Adaptive fill samples are placed in the Sample Area by choosing one of the sampling designs listed on the Sampling Designs / Simple Random Sampling menu. The placement option must be set to Adaptive Fill on the Options / Sample Placement menu. Using this method, samples are placed according to the following algorithm:

1. An initial spacing value is calculated using the following equation:
   spacing = sqrt(area / samples)
   Where
   spacing = the initial spacing value
   sqrt() = the square root function
   area = the surface area of the Sample Area
   samples = the number of samples to place

   Then Steps 2 through 4 are performed once for each sample to be placed:

2. A sample location is obtained using the algorithm described in Regular Random.

3. The existing sample closest to the new sample location is found.

4. If the distance to the closest sample is greater than the spacing value, the sample is added.

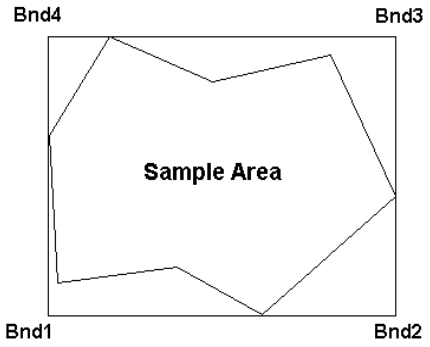
5. If not all samples were placed, then the spacing value is multiplied by 0.99 and Steps 2 through 4 are repeated for each remaining sample. If all samples were placed, then the algorithm is finished.

For multiple Sample Areas, a spacing value is calculated for each Sample Area, and the largest one is used as the initial spacing value. Each Sample Area is given an opportunity to place all the samples (Steps 2 through 4). The total number of samples placed is tracked, and the next Sample Area is given an opportunity to place the remaining samples. Processing continues with Step 5 after each Sample Area is given an opportunity to place samples.

4.5.3 Systematic Grid Sampling Location Algorithm

Grid samples are placed in the sample area by choosing one of the Systematic Grid sampling designs listed on the Sampling Goals menu. Using this method, samples are placed according to the following algorithm:

1. The bounding box of the Sample Area is defined by the following 4 points:



   Bnd1 = (minimum x extent of Sample Area, minimum y extent of Sample Area)
   Bnd2 = (maximum x extent of Sample Area, minimum y extent of Sample Area)
   Bnd3 = (maximum x extent of Sample Area, maximum y extent of Sample Area)
   Bnd4 = (minimum x extent of Sample Area, maximum y extent of Sample Area)

2. Find center of bounding box:

   Center = ((minimum x + maximum x)/2, (minimum y + maximum y)/2)

3. Tilt the bounding box by rotating each point around the center by the user-specified angle using the following formula:

   X' = Center x + cos(angle) * DX - sin(angle) * DY
   Y' = Center y + cos(angle) * DY + sin(angle) * DX

   where
   X' = x coordinate of rotated point
   Y' = y coordinate of rotated point
   Center x = x coordinate of sample area center

Center y = y coordinate of sample area center
angle = angle of tilt specified by user
DX = x coordinate of original point – x coordinate of sample area center
DY = y coordinate of original point – y coordinate of sample area center
cos() = cosine function
sin() = sine function

4. Find the length of the row of samples:

   Length = max rotated bounding point x coordinate – min rotated bounding point x coordinate
   + Grid Size * 2

5. Find the number of grid columns:

   Number of columns = Length / Grid Size + 2

6. Find the height of the column of samples:

   For rectangular grids:
   Height = max rotated bound point y coordinate – min rotated bounding point y coordinate
   + Short Side Grid Size * 2

   For other grids:
   Height = max rotated bound point y coordinate – min rotated bounding point y coordinate
   + Grid Size * 2

7. Find the number of grid rows:

   For rectangular grids:
   Number of rows = Height / Short Side Grid Size + 2

   For square grids:
   Number of rows = Height / Grid Size + 2

   For triangular grids:
   Number of rows = Height / (sin(PI/3.0) * Grid Size) + 2

8. Calculate Start Point (x, y):

   x = Center x – cos(PI/2 – angle) * (Height / 2)
   y = Center y – sin(PI/2 – angle) * (Height / 2)
   where
   PI = 3.1415926535

   If the user has specified a random start, then an offset point(x', y') is added to the starting point:

   For rectangular grids:
   x' = Random * Grid Size

y' = Random * Short Side Grid Size

For other grids:

x' = Random * Grid Size

y' = Random * Grid Size

where

Random is an independent, random fraction between 0 and 1

Note: the offset point is rotated around (0, 0) by "angle" before being added to start point

9. Calculate distance to move for each row

   For rectangular grids:

   DeltaXR = cos(PI/2 – angle) * Short Side Grid Size

   DeltaYR = sin(PI/2 – angle) * Short Side Grid Size

   For square grids:

   DeltaXR = cos(PI/2 – angle) * Grid Size

   DeltaYR = sin(PI/2 – angle) * Grid Size

   For triangular grids:

   DeltaXR = cos(PI/2 – angle) * sin(PI/3) * Grid Size

   DeltaYR = sin(PI/2 – angle) * sin(PI/3) * Grid Size

10. Calculate distance to move for each column:

    DeltaXC = cos(angle) * Grid Size

    DeltaYC = –sin(angle) * Grid Size

11. Calculate Current Point (xx, yy) based on Start Point(x, y):

    xx = x – length / 2 * cos(angle)

    yy = y + length / 2 * sin(angle)

12. For triangular grid, offset current point by half of grid on alternating rows:

    If row number is even:

    xx = xx + DeltaXC / 2

    yy = yy + DeltaYC / 2

13. If Current Point (xx, yy) is inside the sample area, add a sample point at that location

14. Move Current Point (xx, yy) to next column

    xx = xx + DeltaXC

    yy = yy + DeltaYC

15. Repeat Steps 13 and 14 for each column in Number of Columns

16. Move Start Point (x, y) to next row

    x = x + DeltaXR
    y = y + DeltaYR

17. Repeat Steps 11 through 16 for each row in Number of Rows

4.5.4 Ranked Set Sampling Placement Algorithm

Ranked set sampling field screening locations are created in groups called "sets". The user defines the set size during the design definition process. All the samples having the same set number are given a unique symbol shape to distinguish them from samples of other sets. All the samples belonging to one round of sampling (cycle) are given a unique color to distinguish them from samples of other cycles. If several Sample Areas are used in the design, the sample locations within each set need to be distributed uniformly among the Sample Areas. Sample placement is accomplished using the following algorithm.

1. The number of Sample Areas used in the design is determined and the surface area of each is calculated. Total sample surface area is also calculated.

2. For multiple Sample Areas, the number of samples assigned to each Sample Area is calculated by the following formula:

   N = floor(samples * area / total)

   where
   N = number of samples to be assigned to the current Sample Area
   floor() = function that returns the largest whole number less than or equal to a number
   samples = total number of samples to be assigned to all Sample Areas
   area = surface area of current Sample Area
   total = total surface area of all Sample Areas

3. If the sum of all "N"s is less than samples (because of rounding down), 1 is added to n for each Sample Area in turn, until all the sums of all "N"s = samples.

4. Cycle = 1

5. The color of Cycle is determined as follows:

   Color = RGB(Red, Green, Blue)

   where

   RGB() is a function that makes a composite color from the red, green, and blue elements each ranging in brightness from 0 (black) to 255 (brightest)

   Red, Green, Blue are interpolated from the following table using Color Index:

Index	0	1	2	3	4	5	6
Red	255	255	0	0	0	255	255
Green	0	255	255	255	0	0	0
Blue	0	0	0	255	255	255	0

Color Index = (Cycle - 1) * (6.0 / Total number of Cycles)

6. Set = 1

7. A symbol is chosen for the set:

Index	1	2	3	4	5	6	7	8
Symbol	Square	Triangle Up	Circle	Diamond	Plus	Triangle Down	X	Hour Glass

For balanced designs:
Set Index = Set
For unbalanced designs:
Set Index = Set, when Set < Set Size
Set Index = Set Size, when Set >= Set Size
Note: For unbalanced designs, when Set >= Set Size, a number (Set – Set Size + 1) is displayed along with the symbol.

8. Loc = 1

9. Choose a random location in the total area by the following formula:

   Location = Random * Total Surface Area

   where
   Random is an independent, random fraction between 0 and 1

10. Choose the Sample Area containing Location by the following algorithm:

    a. If Location <= Surface Area of this Sample Area, then this Sample Area is chosen
    b. Otherwise subtract Surface Area of this Sample Area from Location
    c. Repeat Steps a (and b if necessary) for next Sample Area

11. If N = 0 for the chosen Sample Area, then repeat Steps 9 and 10 to choose a different Sample Area

12. Subtract 1 from N for the chosen Sample Area

13. Choose a random location (x, y) within the Sample Area (using the method outlined in section 4.5.1 Steps 1 – 3)

14. Create a sample at (x, y) using Color and Symbol

15. Add 1 to Loc and repeat Steps 9 – 14 until Loc > Set Size

16. Add 1 to Set and repeat Steps 7 – 14 until:

    Set > Set Size (for balanced designs)
    Set > Set Size – 1 + t (for unbalanced designs)
    where:
    t = the number of times the top rank is sampled per cycle

17. Add 1 to Cycle and repeat Steps 5 – 14 until Cycle > Number of Cycles

4.5.5 Swath Sampling Location Algorithm

Swath samples are transect type samples represented by lines in parallel, square or rectangular patterns. They are placed in the sample area by choosing Sampling Goals / Finding a UXO Target Area / Parallel or grid swath sampling. Using this method, swaths are placed according to the following algorithm (which follows the algorithm used for systematic grid samples):

1. The bounding box of the Sample Area is defined by the following 4 points:

   Bnd1 = (minimum x extent of Sample Area, minimum y extent of Sample Area)
   Bnd2 = (maximum x extent of Sample Area, minimum y extent of Sample Area)
   Bnd3 = (maximum x extent of Sample Area, maximum y extent of Sample Area)
   Bnd4 = (minimum x extent of Sample Area, maximum y extent of Sample Area)

2. Find center of bounding box:

   Center = ((minimum x + maximum x)/2, (minimum y + maximum y)/2)

3. Tilt the bounding box by rotating each point around the center by the user-specified angle using the following formula:

   X' = Center x + cos(angle) * DX - sin(angle) * DY
   Y' = Center y + cos(angle) * DY + sin(angle) * DX

   where

   X' = x coordinate of rotated point
   Y' = y coordinate of rotated point
   Center x = x coordinate of sample area center
   Center y = y coordinate of sample area center
   angle = angle of tilt specified by user
   DX = x coordinate of original point – x coordinate of sample area center
   DY = y coordinate of original point – y coordinate of sample area center
   cos() = cosine function
   sin() = sine function

4. Find the length of the swaths:

Length = max rotated bounding point x coordinate – min rotated bounding point x coordinate
+ Extra
Where
For rectangular swath patterns:
Extra = 3 * ((Spacing * Ratio) + Width)
For other swath patterns:
Extra = 3 * (Spacing + Width)
Spacing = Spacing between swaths
Width = Width of swath
Ratio = Rectangle width to height ratio

5. Find the perpendicular distance of the swath coverage:

YDist = max rotated bounding point y coordinate – min rotated bounding point y coordinate
+ Extra

where
Extra = 3 * (Spacing + Width)

6. Calculate the number of swaths:

Number of swaths = floor(YDist / (Spacing + Width)) + 2

7. A random offset (xoff, yoff) is added to the Center point, and is calculated by:
yoff = (Spacing + Width) * Random
For rectangular swath patterns:
xoff = (Spacing + Width) * Ratio * Random
For other swath patterns:
xoff = (Spacing + Width) * Random

where

Random is an independent, random fraction between 0 and 1
Note: the offset point (xoff, yoff) is rotated around (0, 0) by "angle" before being added to Center

8. The current point (xstart, ystart) located at the center of the first swath is calculated as follows:

xcurrent = Center x – cos(PI / 2 – angle) * (YDist / 2)
ycurrent = Center y – sin(PI / 2 – angle) * (YDist / 2)

9. Calculate distance to move for each swath:

DeltaX = cos(PI / 2 – angle) * (Spacing + Width)
DeltaY = sin(PI / 2 – angle) * (Spacing + Width)

10. Calculate end points of swath:

Pnt1(xcurrent – Length / 2 * cos(angle), ycurrent + Length / 2 * sin(angle))

Pnt2(xcurrent + Length / 2 * cos(angle), ycurrent – Length / 2 * sin(angle))

11. Cut the line (Pnt1 : Pnt2) into segments where it crosses the edge of the sample area, discarding segments that are outside the sample area, defining segments inside the sample area as swaths.

12. Move current point (xcurrent, ycurrent) to center of next swath

    xcurrent = xcurrent + DeltaX
    ycurrent = ycurrent + DeltaY

13. Repeat Steps 10-12 for each swath in Number of Swaths

14. Continue with Step 15 for square and rectangular swath patterns, otherwise end.

15. Swap the values of YDist and Length

16. Calculate the number of swaths:

    For square patterns:
        Number of swaths = floor(YDist / (Spacing + Width)) + 2
    For rectangular patterns:
        Number of swaths = floor(YDist / (Spacing * Ratio + Width)) + 2

17. Calculate current point (xstart, ystart) located at the center of the first swath:
        xcurrent = Center x – cos(angle) * (YDist / 2)
        ycurrent = Center y – sin(angle) * (YDist / 2)

18. Calculate distance to move for each swath:

    For square patterns:
        DeltaX = cos(angle) * (Spacing + Width)
        DeltaY = –sin(angle) * (Spacing + Width)

    For rectangular patterns:
        DeltaX = cos(angle) * (Spacing * Ratio + Width)
        DeltaY = –sin(angle) * (Spacing * Ratio + Width)

19. Calculate end points of swath:

    Pnt1(xcurrent – Length / 2 * cos(PI/2 + angle), ycurrent + Length / 2 * sin(PI/2 + angle))
    Pnt2(xcurrent + Length / 2 * cos(PI/2 + angle), ycurrent – Length / 2 * sin(PI/2 + angle))

20. Cut the line (Pnt1 : Pnt2) into segments where it crosses the edge of the sample area, discarding segments that are outside the sample area, defining segments inside the sample area as swaths.

21. Move current point (xcurrent, ycurrent) to center of next swath

    xcurrent = xcurrent + DeltaX
    ycurrent = ycurrent + DeltaY

22. Repeat Steps 19-21 for each swath in Number of Swaths

4.5.6 Adaptive Cluster Sampling

Adaptive-cluster samplings involves:

- dividing the sample area into grid cells (sometimes called units)
- selecting several of those grid cells to be sampled initially
- selecting neighbor grid cells for follow-up sampling
- grouping cells into networks for statistical analysis

4.5.6.1 Grid Location Algorithm

For adaptive-cluster sampling, the sample areas need to be divided up into grid cells. This process follows the algorithm used for systematic grid sample using a square pattern.

1. The bounding box of the Sample Area is defined by the following 4 points:

   Bnd1 = (minimum x extent of Sample Area, minimum y extent of Sample Area)
   Bnd2 = (maximum x extent of Sample Area, minimum y extent of Sample Area)
   Bnd3 = (maximum x extent of Sample Area, maximum y extent of Sample Area)
   Bnd4 = (minimum x extent of Sample Area, maximum y extent of Sample Area)

2. Find center of bounding box:

   Center = ((minimum x + maximum x)/2, (minimum y + maximum y)/2)

3. Tilt the bounding box by rotating each point around the center by the user-specified angle using the following formula:

   X' = Center x + cos(angle) * DX - sin(angle) * DY
   Y' = Center y + cos(angle) * DY + sin(angle) * DX

   Where
   X' = x coordinate of rotated point
   Y' = y coordinate of rotated point
   Center x = x coordinate of sample area center
   Center y = y coordinate of sample area center
   angle = angle of tilt specified by user
   DX = x coordinate of original point – x coordinate of sample area center
   DY = y coordinate of original point – y coordinate of sample area center
   cos() = cosine function
   sin() = sine function

4. Find the length of the row of cells:

   Length = max rotated bounding point x coordinate – min rotated bounding point x coordinate

+ Grid Size * 2

5. Find the number of cell columns:

   Number of columns = Length / Grid Size + 2

6. Find the height of the column of cells:

   Height = max rotated bound point y coordinate – min rotated bounding point y coordinate
   + Grid Size * 2

7. Find the number of cell rows:

   Number of rows = Height / Grid Size + 2

8. Calculate Start Point (x, y):

   x = Center x – cos(PI / 2 – angle) * (Height / 2)
   y = Center y – sin(PI / 2 – angle) * (Height / 2)

   Where
   PI = 3.1415926535

9. Calculate distance to move for each row

   DeltaXR = cos(PI / 2 – angle) * Grid Size
   DeltaYR = sin(PI / 2 – angle) * Grid Size

10. Calculate distance to move for each column:

    DeltaXC = cos(angle) * Grid Size
    DeltaYC = –sin(angle) * Grid Size

11. Calculate Current Point (xx, yy) based on Start Point(x, y):

    xx = x – length / 2 * cos(angle)
    yy = y + length / 2 * sin(angle)

12. Calculate four corner points of grid cell

    CellPnt1 = (xx, yy)
    CellPnt2 = (xx+DeltaXC, yy+DeltaYC)
    CellPnt3 = (xx+DeltaXC+DeltaXR, yy+DeltaYC+DeltaYR)
    CellPnt4 = (xx+DeltaXR, yy+DeltaYR)

13. If the center of the grid cell is inside the sample area or if the distance from the center of the grid cell to the closest edge of the sample area is less than (Grid Size / 2.01), then the grid cell is added to the list of grid cells for the sample area. A label is attached to the grid cell with the format:

ACnn-rr-cc

where

nn is the sample area number
rr is the row number
cc is the column number

14. Move Current Point (xx, yy) to next column

    xx = xx + DeltaXC
    yy = yy + DeltaYC

15. Repeat Step 13-14 for each column in Number of Columns

16. Move Start Point (x, y) to next row

    x = x + DeltaXR
    y = y + DeltaYR

17. Repeat Steps 11-16 for each row in Number of Rows

4.5.6.2 Selection of Grid Cells for Initial Samples

First, VSP checks to see if there are enough available grid cells in the sample areas. If there are not enough, VSP puts up a warning and aborts the operation. If there are enough grid cells, selection proceeds as follows:

If pseudo-random sampling is chosen, selection of grid cells for the initial sampling proceeds as follows:

1. NumSelected is set to 0

2. A random index, i, is used to check one of the grid cells:

   i = Random * Total Number of Grid Cells

3. The i^{th} grid cell in the list is checked to see if it has already been selected. If it has not already been selected, it is marked as selected and NumSelected is increased by 1.

4. Steps 2-3 are repeated until NumSelected is equal to the number of samples that are needed.

If quasi-random sampling is chosen, selection of grid cells for the initial sampling proceeds as follows:

1. NumSelected is set to 0

2. A quasi-random pair (x, y) is generated.

3. The random point (x, y) is scaled to the number of rows and columns of the sample area:

   column = x * number of columns in sample area
   row = y * number of rows in sample area

4. The cell located at the specified column and row is located and checked to see if it has already been selected. If it has not already been selected, it is marked as selected and NumSelected is increased by 1.

5. Steps 2-4 are repeated until NumSelected is equal to the number of samples that are needed.

4.5.6.3 Selection of Neighbor Cells for Follow-Up Sampling

After entering the measured value of a grid cell and it is determined that follow-up sampling is needed for a grid cell, VSP searches the list of grid cells that are not already marked, and marks all cells that meet the following requirements:

If using the 4 neighbor rule:

If the row is the same as the original cell and the column is one more or one less
Or
If the column is the same as the original cell and the row is one more or one less

If using the 8 neighbor rule:

If the row is the same as the original cell and the column is one more or one less
Or
If the row is one more or one less and the column is the same or one more or one less

4.5.6.4 Network Assignment Algorithm

Each time that grid cells are selected for sampling (initial sampling assignment or assignment for follow-up sampling) the following method is used to assign network numbers to each assigned grid cell.

1. Network Number is initialized to 0.
2. Network Number is assigned to each grid cell.
3. For each grid cell, perform Step 4 and following.
4. If the grid cell is marked for sampling and its Network Number is zero, continue to Step 5. Otherwise, check next grid cell.
5. Continue with Step 6 if the grid cell value is less than the threshold value, otherwise proceed to Step 8.
6. The grid cell is a network of size one, so the Network Number is increased by one and is assigned to the grid cell.
7. Continue to Step 4 for the next grid cell.
8. The Network Number is increased by one.
9. Do the recursive algorithm (starting at Step 11) using the current row, column, and Network Number.
10. Continue to Step 4 for the next grid cell.

11. The grid cell at the given row and column is found.
12. If the grid cell is marked for sampling and its Network Number is zero, continue to Step 13. Otherwise, the recursive algorithm is complete.
13. If the grid cell is greater than or equal to the threshold value, proceed to Step 14. Otherwise, the recursive algorithm is complete.
14. Network Number is assigned to the grid cell.
15. The recursive algorithm (starting at Step 11) is performed for each neighbor cell.
    For example: (row-1, column), (row+1, column), (row, column-1), (row, column+1).

4.6 Documentation of Random Number Generators

VSP users can choose between two random number generators for placement of samples by making a selection from the Options / Random Numbers menu. This section documents the algorithms used in VSP for these important functions.

4.6.1 Pseudo-Random Number Generator

The pseudo-random number generator produces a sequence of numbers using the following algorithm:

1. value = previous seed * 16807.0
2. seed = value MODULO 2147483647.0
3. random number = seed / 2147483647.0

The initial seed is set to 1.0 when a sample area is created. The seed value is preserved after each random number is produced.

This algorithm is based on a minimal standard random number generator recommended by Park and Miller (1988) and used or the Monte Carlo simulation tests of the ELIPGRID-PC algorithm by Davidson (1995a).

4.6.2 Quasi-Random Number Generator

This algorithm is based on Halton's Sequence discussed in Press et al. (1992, p. 300). The algorithm produces pairs of numbers useful for 2-dimensional problems. The algorithm uses a sequence of numbers comprising all the integers that can be represented in 4 bytes (0 to 2^{32}-1). Because the sequence (0,0), (1,1), (2,2), … wouldn't appear to be very random, the integers are used in two non-linear sequences: one sequence for the x dimension and a different sequence for the y dimension. Digit reordering is used to progress through the integers in a non-linear fashion. All the integers are still represented, but in a non-linear order. For the x-dimension, a sequence of integers (0, 1, 2, 3…) is expressed as a base 2 (binary) number and the digits are reversed to create the non-linear sequence. For the y-dimension, the sequence (0, 1, 2, 3…) is expressed as a base 3 number and the digits are reversed to create a non-linear sequence of numbers. The integers are normalized to represent fractions between 0 and 1, which are more useful for general applications.

The pattern of number pairs tends to be uniformly spread across the 2-dimension plane. In contrast, a pseudo-random number generator is not constrained from generating lists of X, Y coordinates with locations very near other locations.

The quasi-random number generator produces a sequence of paired (X, Y) numbers using the following algorithm:

1. An integer value, iSeq, is used as the basis of the algorithm (e.g., 457)
2. iSeq is expressed as a 32-digit binary number (e.g., bSeq = 00000000000000000000000111001001)
3. The sequence of digits is reversed (e.g., rbSeq = 10010011100000000000000000000000)
4. The sequence of digits is again expressed as a decimal integer (e.g., xVal = 2474639360)
5. xVal is converted to a fraction by dividing by 2^{32}-1, or 4294967295 (e.g., X = 0.576171875)
6. iSeq is expressed as a 20-digit base 3 number (e.g., tSeq = 00000000000000121221)
7. The sequence of digits is reversed (e.g., rtSeq = 12212100000000000000)
8. The sequence of digits is again expressed as a decimal integer (e.g., yVal = 2271910275)
9. yVal is converted to a fraction by dividing by 2^{32}-1, or 4294967295 (e.g., Y = 0.52897033)
10. The values X and Y are used as the random pair.

iSeq is initially set to 1. It is incremented by 1, and the value is preserved after each usage.

4.7 Largest Unsampled Spot Location Algorithm

The Largest Unsampled Spot location algorithm finds the largest circle that can be centered inside the Sample Area without enclosing a sample point. The algorithm scans through all the possible centers at user-defined increments. This method is simple and has a predictable run-time, but its accuracy is limited to the size of the specified increment.

In addition to the increment size, the user determines:

- whether the spot can overlap the edge of the sample area
- whether the spot will be bounded by the vertices of the Sample Area in addition to the sampling locations

The algorithm proceeds as follows:

1. Largest is set to 0.
2. A list of constraining points is created from the sample locations.
3. Total Points is set equal to the number of constraining points.
4. If the user has chosen the spot to be bounded by Sample Area vertices, then add the number of vertices to Total Points.
5. If Total Points is greater than or equal to 20, then the faster method is used.
6. If the faster method is used, the constraining points are quick-sorted by their y-coordinates.
7. Current X is set to the minimum X extent of the Sample Area.
8. Current Y is set to the minimum Y extent of the Sample Area.
9. If the center point (Current X, Current Y) is outside the Sample Area, skip to Step 25.
10. If the faster method is not used, skip to Step 19.
11. Use a binary search on the sorted list of constraining points to find which constraining point has a Y Coordinate closest to Current Y. This constraining point will be the initial point.
12. MinDist is set to the distance from the center point to the initial point.
13. Check the next constraining point lower in the list. If the distance from the next point to the center point is less than MinDist, then set MinDist to that distance.
14. Move lower in the list and repeat Step 13 until there are no more points or the difference between the center point y coordinate and the constraining point y coordinate is greater than MinDist.

15. Start at the initial point again.
16. Check the next constraining point higher in the list. If the distance from the next point to the center point is less than MinDist, then set MinDist to that distance.
17. Move higher in the list and repeat Step 16 until there are no more points or the difference between the center point y coordinate and the constraining point y coordinate is greater than MinDist.
18. Continue with Step 20.
19. Slow method: Check each constraining point, setting MinDist to the smallest distance between the constraining point and the center point.
20. If the user has not chosen the spot to be bound by Sample Area vertices, skip to Step 22.
21. Check distance from each Sample Area vertex to the center point. If the distance is less than MinDist then set MinDist to that distance.
22. If the user has chosen to let the spot overlap the edge of the Sample Area, skip to Step 24.
23. If the circle defined by center point and MinDist overlaps the edge of the Sample Area, then set MinDist to –1.
24. If MinDist is greater than Largest then: set Largest = MinDist, set LargestX = Current X, set LargestY = Current Y
25. Add defined increment to Current X.
26. Repeat Step 9 and following until Current X is greater than the maximum x extent of the Sample Area.
27. Add defined increment to Current Y.
28. Repeat Step 9 and following until Current Y is greater than the maximum y extent of the Sample Area.
29. The largest unsampled spot is now centered at (LargestX, LargestY) and has a radius of Largest.

4.8 Post Survey Target Detection Algorithm

The Post Survey Target Detection algorithm determines the probability of traversing a circular or elliptical target area with a given set of transects. Transects consist of all the lines that traverse the sample area on the map, but exclude any lines that coincide with the boundaries of the sample area. The algorithm simulates the probability of traversing the target using the following Monte Carlo method:

1. Set HitCount = 0
2. Calculate the sum (TotalArea) of the surface area of all selected Sample Areas.
3. Steps 4-21 are performed for each selected Sample Area:
4. For the sake of efficiency, a list of polylines (connected points) traversing the Sample Area is compiled. (Polylines not traversing the Sample Area are excluded.)
5. The extent of each polyline is calculated and stored with the polyline.
6. Calculate the number of trials for this Sample Area:

    Trials = Total Trials * Surface Area of Sample Area / TotalArea

7. Steps 8-21 are performed for each Trial:
8. A random point (x, y) inside the Sample Area is selected.
9. If the user has specified an angle of orientation, then Angle is set to the given angle, otherwise a random Angle between 0 and PI is chosen. Note that Angle is counter-clockwise with respect to the positive x-axis.
10. Set Hit = False.

11. Steps 12-20 are performed for each polyline in the list, or until Hit is True:
12. Steps 13-20 are performed for each line segment in the polyline, or until Hit is True:
13. Calculate Distance as the shortest distance from the point (x, y) to the line segment.
14. If the target is circular, go to Step 15; if the target is elliptical go to Step 17.
15. If Distance is less than (Radius + Swath Width / 2), then set Hit = True.
16. Continue with Step 20.
17. Calculate SwathAngle as the angle between the x-axis and the line segment. Note that SwathAngle is clock-wise with respect to the positive x-axis.
18. Calculate height of ellipse (Y) with respect to the swath segment:

    Y = sqrt(R2 – D2 * Cos * Cos)
    Where:
    Cos = cosine(SwathAngle + Angle)
    R2 = (Length of semi-major axis of ellipse) * (Length of semi-major axis of ellipse)
    D2 = R2 – (Length of semi-minor axis of ellipse) * (Length of semi-minor axis of ellipse)

19. If Distance is less than (Y + Swath Width / 2), then set Hit = True.
20. If Hit is True, then add 1 to HitCount.
21. After all swaths have been checked and Hit is still False, a sample point is added to the sample area if the user has chosen that option.
22. After all Sample Areas have been checked, the Probability of Hit (P) is calculated:
    P = HitCount / Total Trials

Several tests of the algorithm were conducted on a square sample area. The extent of the sample area was (0,0) and (100,100). Four swaths were used, each passing through the point (50, 50). One swath was vertical, one swath was horizontal, one swath was oriented at 45 degrees, and the last swath was oriented at negative 45 degrees. 10,000 simulation points were used. The simulation points that missed the swaths were exported from VSP and imported into a spreadsheet where each point was checked against the 4 swaths to make sure that no missing point was closer to the swaths than expected. Results are listed in the following table.

Table 4.2 Checking the Target Detection Algorithm

Swath Width	Target Size and Shape	Target Angle	# Miss Points	Results
0 feet	10 foot radius Circle	Random	2658	All OK
10 feet	10 foot radius Circle	Random	730	All OK
0 feet	10 foot semi-major axis 0.5 Ellipse	Random	3969	All OK

Table 4.2 Checking the Target Detection Algorithm (cont.)

Swath Width	Target Size and Shape	Target Angle	# Miss Points	Results
10 feet	10 foot semi-major axis 0.5 Ellipse	Random	1541	All OK
0 feet	10 foot semi-major axis 0.5 Ellipse	0/	3924	All OK
10 feet	10 foot semi-major axis 0.5 Ellipse	0/	1480	All OK
0 feet	10 foot semi-major axis 0.5 Ellipse	45/	4015	All OK
10 feet	10 foot semi-major axis 0.5 Ellipse	45/	1616	All OK
0 feet	10 foot semi-major axis 0.5 Ellipse	90/	3932	All OK

10 feet	10 foot semi-major axis 0.5 Ellipse	90/	1496	All OK

5.0 References

Barnard, G.A. 1952. "The Frequency Justification of Certain Sequential Tests," *Biometricka* 39:144-150**.**

Bowen, W.M. and C.A. Bennett. 1988. *Statistical Methods for Nuclear Material Management*. NUREG/CR-4604, PNL-5849, prepared by Pacific Northwest National Laboratory for the U.S. Regulatory Commission, Washington, D.C.

Cochran, W.G. 1977. *Sampling Techniques*, 3rd edition, John Wiley & Sons, New York.

Conover, W.J. 1999. *Practical Nonparametric Statistics*, 3rd edition, John Wiley & Sons, Inc, New York.

Davidson, J.R. 1995a. *Monte Carlo Tests of the ELIPGRID-PC Algorithm*. ORNL/TM-12899, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Davidson, J.R. 1995b. *ELIPGRID-PC: Upgraded Version.* ORNL/TM-13103, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Davidson, J.R. 2001. *Verification of the Accuracy of Sample-Size Equation Calculations for Visual Sample Plan Version 0.9C.* PNNL-13434, Pacific Northwest National Laboratory, Richland, Washington.

Desu, M.M. and D. Raghavarao. 1990. *Sample Size Methodology.* Academic Press, New York.

EPA. 1989. *Methods for Evaluating the Attainment of Cleanup Standards, Volume 1: Soils and Solid Media*. EPA 230/02-89-042, Office of Policy, Planning, and Evaluation, U.S. Environmental Protection Agency, Washington, D.C.

EPA. 1992. *Methods for Evaluating the Attainment of Cleanup Standards*, *Volume 2: Ground Water*. EPA 230-R-92-014, Office of Policy, Planning, and Evaluation, U.S. Environmental Protection Agency, Washington, D.C.

EPA. 1997. *Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM)*. EPA 402-R-97-016, NUREG-1575, U.S. Environmental Protection Agency, Washington, D.C.

EPA. 2000a. *Guidance for the Data Quality Objectives Process*, (QA/G-4), EPA/600/R-96/055, Office of Environmental Information, U.S. Environmental Protection Agency, Washington, D.C.

EPA. 2000b. *Guidance for Data Quality Assessment – Practical Methods for Data Analysis – EPA QA/G-9, QA00 Update*. EPA/600/R-96/084, Office of Environmental Information, U.S. Environmental Protection Agency, Washington, D.C.

EPA. November 2001. *Guidance for Choosing a Sampling Design for Environmental Data Collection*, EPA QA/G-5S, Peer Review Draft, Quality Staff, Office of Environmental Information, U.S. Environmental Protection Agency, Washington, D.C.

Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring*. John Wiley & Sons, Inc., New York.

Gilbert, R.O. and J.R. Davidson, Jr., J.E. Wilson and B.A. Pulsipher. February 2001. *Visual Sample Plan (VSP) Models and Code Verification*, PNNL-13450, Pacific Northwest National Laboratory, Richland, WA**.**

Gilbert, R.O. and B.A. Pulsipher, D.K. Carlson, R.F. O'Brien, J.E. Wilson, D.J. Bates and G.A. Sandness. November 2001. *Designing UXO Sensor Surveys for Decision Making*, DRAFT, Pacific Northwest National Laboratory, Richland, WA

Gogolak, C.V., G.E. Powers and A.M. Huffert. 1997. *A Nonparametric Statistical Methodology for the Design and Analysis of Final Status Decommissioning Surveys*. NUREG-1505, U.S. Nuclear Regulatory Commission, Washington, D.C.

Guenther, W.C. 1977. *Sampling Inspection in Statistical Quality Control*. Griffin's Statistical Monographs and Courses, No. 37, London: Charles Griffin.

Guenther, W.C.. 1981. "Sample Size Formulas for Normal Theory T-Tests." *The American Statistician* 35(4):243-244.

Hahn, G.J. and W.Q. Meeker. 1991. *Statistical Intervals*. John Wiley & Sons, Inc., New York.

Iman, R.L. and W.J. Conover. 1983. *A Modern Approach to Statistics*. John Wiley & Sons, Inc., New York.

Johnson, N.L. and S. Kotz. 1970. *Continuous Univariate Distributions-2*, Houghton Mifflin Company, Boston.

Kaur, A., G.P. Patil and C. Taillie. 1995. Unequal Allocation Models for Ranked Set Sampling with Skew Distributions, Technical Report Number 94-0930, Pennsylvania State University, University Park, PA.

Kaur, A., G.P. Patil and C. Taillie. 1997. "Unequal Allocation Models for Ranked Set Sampling with Skew Distributions," *Biometrics* 53:123-130.

McIntyre, G.A. 1952. "A Method for Unbiased Selective Sampling Using Ranked Sets," *Australian Journal of Agricultural Research* 3:385-390.

Mode, N.A., L.C. Conquest, and D.A. Marker. 1999. "Ranked Set Sampling for Ecological Research: Accounting for the Total Costs of Sampling," *Environmetrics* 10:179-194.

Noether, G.E.. 1987. "Sample Size Determination for Some Common Nonparametric Tests." *Journal of the American Statistical Association* 82:645-647.

Owen, D.R. 1965. *Handbook of Statistical Tables*. Addison-Wesley, Reading, Massachusetts.

Park, S.K. and K.W. Miller. 1988. "Random number generators: Good ones are hard to find." *Communications of the ACM*, 31:1192-1201.

Patil, G.P., A.K. Sinha and C. Taillie. 1994. "Ranked Set Sampling," *Handbook of Statistics 12*, Environmental Statistics, North-Holland, New York, pp. 167-200.

Perez, A. and J.J. Lefante. 1997. "Sample Size Determination and the Effect of Censoring when Estimating the Arithmetic Mean of a Lognormal Distribution," *Communications in Statistics, Theory and Methods* 26(11):2779-2801.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. 1992. *Numerical Recipes in Fortran 77: the Art of Scientific Computing*. Cambridge University Press, Cambridge, United Kingdom.

Rushton, S. 1950. "On a Sequential t-Test," *Biometrika* 37:326-333.

Shilling, E.G. 1982. *Acceptance Sampling in Quality Control*, Marcel Dekker, Inc, New York.

Singer, D.A. 1972. "ELIPGRID, a FORTRAN IV Program for Calculating the Probability of Success in Locating Elliptical Targets with Square, Rectangular and Hexagonal Grids." *Geocom Programs* 4:1-16.

Singer, D.A. 1975. "Relative Efficiencies of Square and Triangular Grids in the Search for Elliptically Shaped Resource Targets." *Journal of Research of the U.S. Geological Survey* 3(2):163-167.

Singer, D.A. and F.E. Wickman. 1969. *Probability Tables for Locating Elliptical Targets with Square, Rectangular and Hexagonal Point Nets*. Pennsylvania State University, University Park, Pennsylvania.

Thompson, S.K. 1990. "Adaptive Cluster Sampling," *Journal of the American Statistical Association* 85:412.

Thompson, S.K. 1992. *Sampling*, John Wiley & Sons, New York.

Thompson, S.K. and G.A.F. Seber. 1996. *Adaptive Sampling*. John Wiley & Sons, New York.

Wald, A. 1947. *Sequential Analysis.* Dover, New York.

Wallis, W.A. 1947. "Uses of Variables in Acceptance Inspection for Percent Defective." *Techniques of Statistical Analysis*, McGraw-Hill, New York.

Wetherill, G.B. 1966. *Sequential Methods in Statistics*. Chapman and Hall, New York.

Zirschky, J. and R.O. Gilbert. 1984. "Detecting Hot Spots at Hazardous Waste Sites." *Chemical Engineer*, July 9, pp. 97-100.

Appendix A

Barnard's Sequential t -Test

The sequential t-test assumes that data are sampled sequentially from a normal distribution with unknown mean (m) and known or unknown standard deviation (s). Two versions of the sequential t-test procedure (s known, s unknown) are implemented in VSP. The basis of the sequential t-test lies in forming a sequence of likelihood ratio tests computed from the central and non-central t probability density function (pdf) (Equation A.1). The central t pdf is a special case of Equation A.1 when the non-centrality parameter, δ, is set to zero [Johnson and Kotz (1970, p.204); Barnard (1952); Rushton (1950)].

$$pdf(t|\nu,\delta) = \frac{\nu!}{2^{(\nu-1)/2}\sqrt{\pi\nu}\,\Gamma(\nu/2)} e^{-\frac{\nu\delta^2}{\nu+t^2}} \left(\frac{\nu}{\nu+t^2}\right)^{(\nu+1)/2} Hh_\nu\left(-\frac{\delta t}{\sqrt{\nu+t^2}}\right) \qquad \textbf{(A.1)}$$

where ν = degrees of freedom, and

$$Hh_\nu(x) = \frac{1}{\nu!}\int_0^\infty u^\nu e^{-(\nu+x)^2/2} du$$

In general, the non-centrality parameter (δ) has the effect of shifting the mode or central tendency of the central t probability distribution away from 0 and skewing its shape. This shift and skew occurs according to the sign of δ (Figure A.1). In Figure A.1 the central t distribution is seen to be symmetric and centered at 0. The top panel in the figure shows a non-central t distribution with positive non-centrality parameter. The resulting non-central t distribution is shifted to the right and skewed to the right. The bottom panel shows a non-central t distribution with negative non-centrality parameter. This distribution is shifted to the left and skewed to the left.

A typical application of the sequential t-test occurs when investigators or regulators wish to test whether the true mean contamination at a site lies above or below a threshold or *action-level* μ_{AL}. There are two ways of stating the null and alternative hypotheses (H_0 and H_A, respectively). First, it may be desired to assume that the site is dirty (until proven clean), in which case

$$H_0: \mu \geq \mu_{AL}$$
$$H_A: \mu < \mu_{AL}$$

Second, it may be desired to assume that the site is clean (until proven dirty), in which case

$$H_0: \mu \leq \mu_{AL}$$
$$H_A: \mu > \mu_{AL}$$

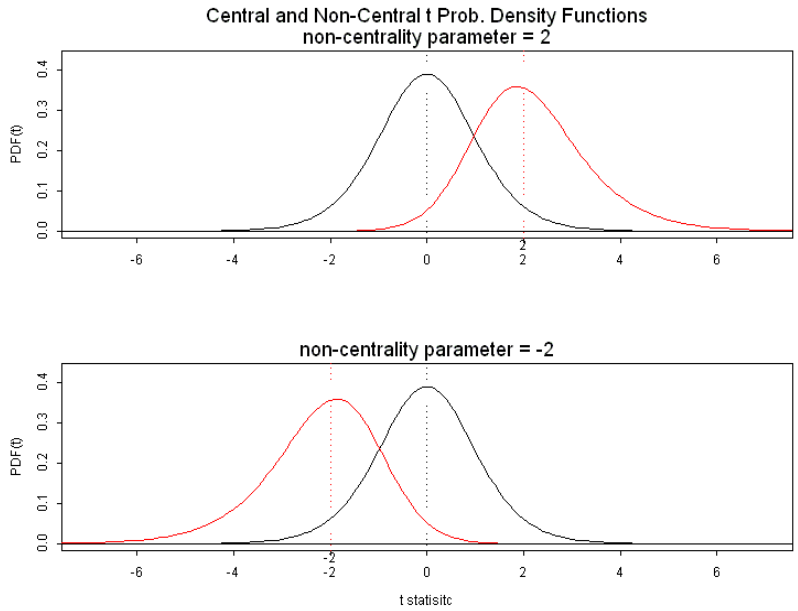
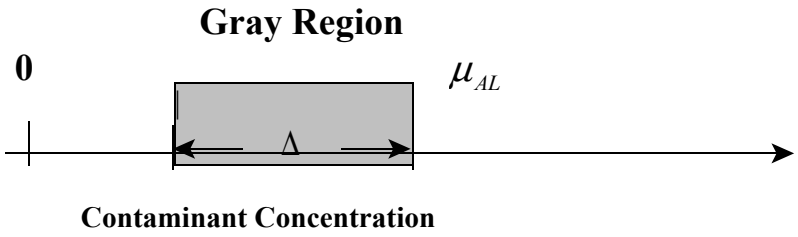


Figure A.1. Central and non-central t probability distributions with positive non-centrality parameter (top panel) and a negative non-centrality parameter (bottom panel).

Figure A.2 shows the "gray region" (EPA 2000a) used in the sequential t-test. The gray region has a width of absolute value $\Delta = \mu_A - \mu_0$ as shown in Figure A.2. Note that the action level, μ_{AL}, always lies at the upper end of the gray region. Also note that when the null hypothesis used is that the site is clean, then μ_0 is the true mean at the lower boundary of the gray region, whereas μ_0 is the true mean that lies at the upper end of the gray region if the null hypothesis used is that the site is dirty.



H_0: The site is clean. μ_0 μ_A

H_0 : The site is dirty. μ_A μ_0

Figure A.2. The Gray Region and the values of the true mean used at the ends of the gray region for the two versions of the null hypothesis.

Figure A.3 shows the gray region and the performance curves for the sequential t-test when the null hypothesis used is that the site is dirty (top panel of Figure A.3) and when the null hypothesis used is that the site is clean (bottom panel of Figure A.3). For either null hypothesis used, the upper end of the gray

region is set at the action level, μ_{AL}. The performance curves in Figure A.3 are for the case when the number of measurements used in the test is n = 21, the standard deviation of the measurements is 3, the action level is μ_{AL} = 10, and the lower end of the gray region is μ = 8. The red vertical line in the top and bottom portions of Figure A.3 indicates the value of μ_o, which is the value of the true mean assumed to be true in the null hypothesis. Hence, μ_o equals 10 in the top panel of Figure A.3 and μ_0 equals 8 in the bottom panel of Figure A.3, which is consistent with Figure A.2. Consequently, μ_A equals 8 in the top panel and equals 10 in the bottom panel, which is also consistent with Figure A.2.

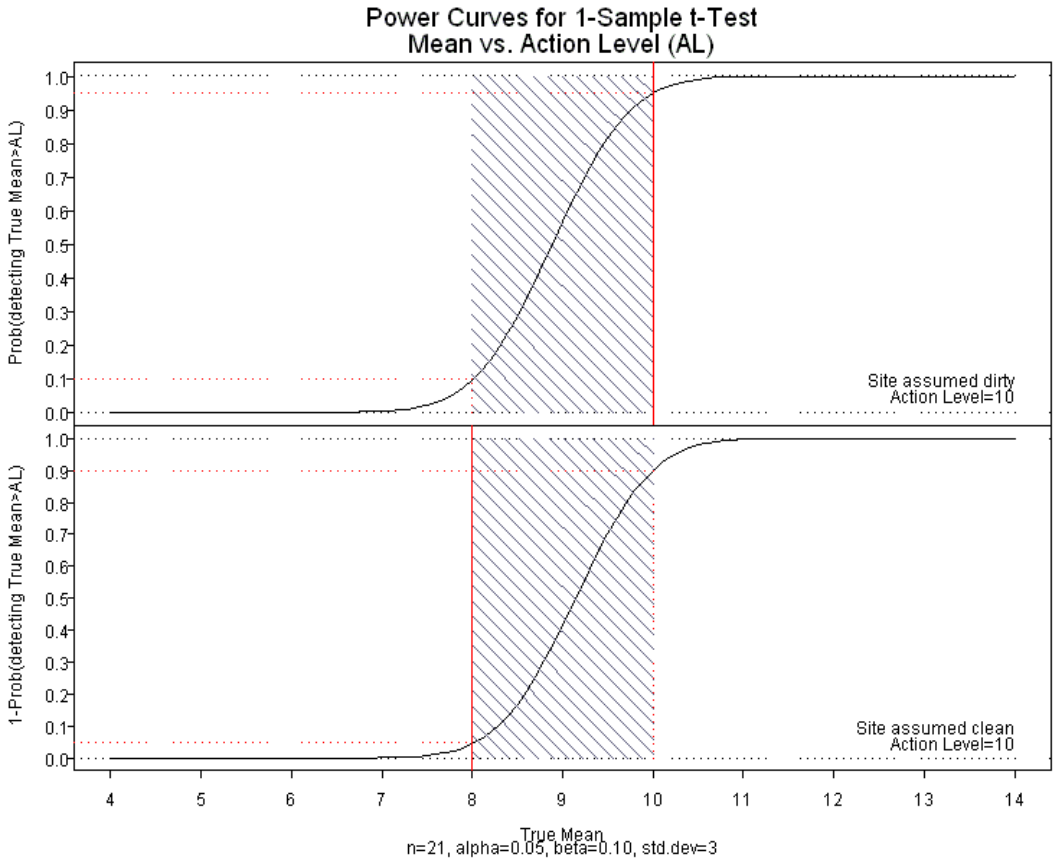


Figure A.3. Power curves for the 1- sample sequential t-test for when the null hypothesis is that the site is dirty (top panel) and when the null hypothesis is that the site is clean (bottom panel).

The testing procedure for the sequential t-test is as follows:

1. The action level (μ_{AL}) and the width of the gray area (Δ) are specified prior to sampling, along with acceptable levels of statistical confidence ($1-\alpha$) and power ($1-\beta$).
2. Samples are collected from a study site sequentially over time and analyzed for contaminate concentrations.

3. After the n measurements ($x_1, x_2, \ldots, x_n$) have been collected and measured, a t-statistic is computed from these data as the ratio of the mean deviation of the n measurements from the value assumed under the null hypothesis (μ_0) to the standard error of the sample mean:

$$t_n = \frac{\bar{x}_0}{SE_{\bar{x}_n}}$$

where

$\bar{x}_0 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)$ is the mean deviation of the data from the value assumed under the null hypothesis (μ_0)

$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean on the n measurements

$SE_{\bar{x}_n} = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}_n\right)^2}{n(n-1)}}$ is the standard error of the sample mean based on the n measurements.

Note that when σ is known, then $SE_{\bar{x}_n}$ is replaced by $\sigma / \sqrt{n}$. If σ is known, then the sequential probability ratio test (SPRT) discussed in Section 3.1.1.1 should be used.

4. The value of the non-centrality parameter δ is computed as the width of the gray region, Δ , scaled to the standard error of the sample mean:

$$\delta = \frac{\Delta}{SE_{\bar{x}_n}}$$

Again, when σ is known, $SE_{\bar{x}_n}$ is replaced by $\sigma / \sqrt{n}$ and the SPRT should be used.

5. A value from the non-central t pdf is then computed using Equation A.1, with t = t_n , degrees of freedom ν = n-1, and the computed non-centrality parameter δ. Denote this computed value as $pdf(t_n | \nu, \delta > 0)$, which denotes the likelihood of the alternative hypothesis being true given the n observed measurements made thus far.

6. A value from the central t distribution is similarly computed using Equation A.1, with t = t_n, the same degrees of freedom (ν = n - 1), but with non-centrality parameter (δ) set equal to zero. Denote this value as $pdf(t_n|\nu, \delta = 0)$, which denotes the likelihood of the null hypothesis being true given the n measurements made thus far.

7. If it is assumed that the site is clean, the hypotheses are stated as follows:

$H_0: \mu \leq \mu_{AL}$ that is, the true site mean is less than or equal to the action level

vs.

$H_A: \mu > \mu_{AL}$ that is, the true site mean greater than the action level (i.e., the non-centrality parameter δ is greater than 0)

For these hypotheses, the sequential t-test likelihood ratio test statistic (L_n) is computed as follows:

$$L_n = \frac{pdf(t_n|\nu, \delta > 0)}{pdf(t_n|\nu, \delta = 0)}$$

The decision whether to accept the null hypothesis or the alternative hypothesis is made using the following decision rules:

a. If $\ln(L_n) \leq \ln\left(\frac{\beta}{1-\alpha}\right)$, then conclude that the null hypothesis, H_0, is true, i.e. that the site is clean.

b. If $\ln(L_n) \geq \ln\left(\frac{1-\beta}{\alpha}\right)$, then conclude that the alternative hypothesis, H_A, is true, i.e. that the site is dirty.

c. If $\ln\left(\frac{\beta}{1-\alpha}\right) < \ln(L_n) < \ln\left(\frac{1-\beta}{\alpha}\right)$, then continue sampling and repeat the test after the additional measurements have been obtained.

8. If it is assumed that the site is dirty, the hypotheses are stated as follows:

$H_0: \mu \geq \mu_{AL}$ that is, the true mean is greater than or equal to the action level

vs.

$H_A: \mu < \mu_{AL}$ that is, the true mean is less than the action level (i.e., the non-centrality parameter δ is less than 0)

For these hypotheses, the sequential t-test likelihood ratio test statistic (L_n) is computed as follows:

$$L_n = \frac{pdf\left(t_n \middle| \nu, \delta = 0\right)}{pdf\left(t_n \middle| \nu, \delta < 0\right)}$$

The decision whether to accept the null hypothesis or the alternative hypothesis is made using the following decision rules:

a. If $\ln(L_n) \le \ln\left(\frac{\alpha}{1-\beta}\right)$, then conclude that the alternative hypothesis, H_A , is true (i.e. that the site is clean)

b. If $\ln(L_n) \ge \ln\left(\frac{1-\alpha}{\beta}\right)$, then conclude that the null hypothesis is true (i.e., that the site is dirty)

c. If $\ln\left(\frac{\alpha}{1-\beta}\right) < \ln(L_n) < \ln\left(\frac{1-\alpha}{\beta}\right)$, then continue sampling and repeat the test after the additional data are available.

Note that the VSP software requires that n = 10 sample be collected and measured before computing the first iteration of this sequential t-test. The first 10 samples are used to compute the statistic t_n, and the corresponding likelihood ratio statistic L_n. This approach helps prevent spurious hypothesis test conclusions based on insufficient data, which is considered to be n < 10. The test is repeated (updated) following the acquisition of additional measurements.

Also note that regardless of whether the null hypothesis is that the site is clean or dirty, the above decision rules are set up such that, if the lower decision boundary is broken, the test will conclude in favor of a "clean site." Correspondingly, if the upper decision boundary is broken, the test will conclude in favor of a "dirty site." This provides an intuitive link between the likelihood ratio statistic (L_n) and the contaminant concentration levels. Lower values of the ratio L_n correspond to lower contaminant concentrations, while higher values of the ratio L_n correspond to higher levels of contaminant concentration. This is accomplished by forming the test statistic L_n so that the numerator of the ratio always corresponds to the condition or hypothesis of a "dirty site," while the denominator of the ratio L_n always corresponds to the condition or hypothesis of a "clean site." Hence, the ratio L_n when the null hypothesis is that the site is clean is inverted from what it is when the null hypothesis is that the site is dirty. The decision boundaries are then necessarily inverted and the inequalities reversed to maintain consistency with the inversion of the likelihood ratio statistic L_n.

Appendix B

File Used to Verify Import, Export, and Removal of Sampling Locations

Area 1

X Coord	Y Coord	Label	Value	Type	Historical
67.55	49.51	Test-1	0.1	Random	F
60.73	61.80	Test-2	0.2	Random	T
41.87	66.58	Test-3	0.3	Random	F
20.57	96.18	Test-4	0.4	Random	T
89.94	91.95	Test-5	0.5	Random	F
32.76	9.11	Test-6	0.6	Random	T
95.69	54.42	Test-7	0.7	Random	F
59.33	35.90	Test-8	0.8	Random	T
16.93	60.72	Test-9	0.9	Random	F
68.36	87.72	Test-10	1	Random	T
14.96	46.52	Test-11	1.1	Random	F
10.96	97.50	Test-12	1.2	Random	T
30.00	14.18	Test-13	1.3	Random	F
98.46	59.51	Test-14	1.4	Random	T
79.30	43.48	Test-15	1.5	Random	F
6.35	57.39	Test-16	1.6	Random	T
42.54	75.57	Test-17	1.7	Random	F
77.89	54.85	Test-18	1.8	Random	T
4.31	49.90	Test-19	1.9	Random	F
58.33	48.59	Test-20	2	Random	T
68.80	29.95	Test-21	2.1	Random	F
14.46	52.31	Test-22	2.2	Random	T
72.16	8.68	Test-23	2.3	Random	F
99.16	6.56	Test-24	2.4	Random	T
43.73	85.37	Test-25	2.5	Random	F
91.53	79.30	Test-26	2.6	Random	T
24.46	77.87	Test-27	2.7	Random	F
20.09	6.05	Test-28	2.8	Random	T
9.63	53.00	Test-29	2.9	Random	F
24.69	87.64	Test-30	3	Random	T
32.43	62.19	Test-31	3.1	Random	F
36.05	17.46	Test-32	3.2	Random	T
23.97	84.71	Test-33	3.3	Random	F
29.35	38.39	Test-34	3.4	Random	T
98.84	19.98	Test-35	3.5	Random	F
1.39	45.68	Test-36	3.6	Random	T
9.79	79.90	Test-37	3.7	Random	F
26.48	38.86	Test-38	3.8	Random	T
3.86	88.80	Test-39	3.9	Random	F
31.67	4.63	Test-40	4	Random	T
27.95	39.75	Test-41	4.1	Random	F
32.92	74.04	Test-42	4.2	Random	T
60.99	20.79	Test-43	4.3	Random	F
47.92	75.87	Test-44	4.4	Random	T

12.41	56.53	Test-45	4.5	Random	F
59.21	88.42	Test-46	4.6	Random	T
6.18	61.43	Test-47	4.7	Random	F
59.01	66.79	Test-48	4.8	Random	T
8.92	57.16	Test-49	4.9	Random	F
88.16	55.38	Test-50	5	Random	T
16.60	35.13	Test-51	5.1	Random	F
4.94	82.73	Test-52	5.2	Random	T
18.06	53.95	Test-53	5.3	Random	F
78.37	0.58	Test-54	5.4	Random	T
71.35	16.71	Test-55	5.5	Random	F
82.77	78.38	Test-56	5.6	Random	T
50.56	15.89	Test-57	5.7	Random	F
23.21	89.01	Test-58	5.8	Random	T
84.82	5.57	Test-59	5.9	Random	F
65.09	56.48	Test-60	6	Random	T
7.74	52.64	Test-61	6.1	Random	F
8.87	26.81	Test-62	6.2	Random	T
92.75	60.25	Test-63	6.3	Random	F
10.06	14.48	Test-64	6.4	Random	T
32.58	53.67	Test-65	6.5	Random	F
46.17	11.57	Test-66	6.6	Random	T
4.60	29.32	Test-67	6.7	Random	F
34.05	90.53	Test-68	6.8	Random	T
68.23	48.88	Test-69	6.9	Random	F
5.56	18.59	Test-70	7	Random	T
17.08	49.84	Test-71	7.1	Random	F
7.95	71.59	Test-72	7.2	Random	T
34.62	2.94	Test-73	7.3	Random	F
35.32	62.99	Test-74	7.4	Random	T
48.51	1.53	Test-75	7.5	Random	F
82.52	79.11	Test-76	7.6	Random	T
7.56	5.16	Test-77	7.7	Random	F
13.15	34.66	Test-78	7.8	Random	T
33.54	34.11	Test-79	7.9	Random	F
28.80	93.59	Test-80	8	Random	T
73.34	20.85	Test-81	8.1	Random	F
0.13	29.68	Test-82	8.2	Random	T
6.02	66.21	Test-83	8.3	Random	F
85.02	45.15	Test-84	8.4	Random	T
22.39	20.52	Test-85	8.5	Random	F
45.78	30.19	Test-86	8.6	Random	T
68.21	38.84	Test-87	8.7	Random	F
8.86	8.85	Test-88	8.8	Random	T
76.58	65.52	Test-89	8.9	Random	F
2.27	60.88	Test-90	9	Random	T
96.76	22.33	Test-91	9.1	Random	F
40.70	92.19	Test-92	9.2	Random	T
72.81	47.28	Test-93	9.3	Random	F
26.07	1.82	Test-94	9.4	Random	T
21.30	60.89	Test-95	9.5	Random	F
47.60	55.87	Test-96	9.6	Random	T
15.85	15.06	Test-97	9.7	Random	F

89.37	43.27	Test-98	9.8	Random	T
9.23	25.24	Test-99	9.9	Random	F
54.65	55.60	Test-100	10	Random	T

Appendix C

Swath Import Test File

```
LINE
 0        40.00
 10       36.18
 20       26.18
 30       13.82
 40        3.82
 50        0.00
 60        3.82
 70       13.82
 80       26.18
 90       36.18
100       40.00

LINE
0         60.00
 10       56.18
 20       46.18
 30       33.82
 40       23.82
 50       20.00
 60       23.82
 70       33.82
 80       46.18
 90       56.18
100       60.00

LINE
 0        80.00
 10       76.18
 20       66.18
 30       53.82
 40       43.82
 50       40.00
 60       43.82
 70       53.82
 80       66.18
 90       76.18
100       80.00

LINE
 0       100.00
 10       96.18
 20       86.18
 30       73.82
 40       63.82
 50       60.00
 60       63.82
```

70	73.82
80	86.18
90	96.18
100	100.00

Appendix D

Design Test Parameters

One Sample t

Hypo	Alpha	Beta	Delta	AL	SD
0	15.61	29.78	2.93	22.90	1.48
0	20.36	26.89	1.31	59.29	3.73
0	17.85	12.58	7.63	70.11	5.80
1	4.60	15.60	2.69	27.56	0.70
0	9.14	17.19	8.41	82.87	0.46
1	14.06	10.21	0.56	94.82	3.00

One Sample t (MQO)

Hypo	Alpha	Beta	Delta	AL	SD-S	SD-A	Reps
1	5.07	29.83	2.35	11.48	6.43	0.07	1
0	2.88	4.94	2.56	17.78	2.35	0.50	3
1	10.66	22.40	3.09	52.04	9.78	3.49	2
1	4.21	1.75	3.90	38.70	6.45	6.94	2
1	29.44	5.42	4.15	10.34	1.68	1.32	1
1	3.29	5.99	3.48	27.18	4.62	5.21	1

Wilcoxon Signed Ranks

Hypo	Alpha	Beta	Delta	AL	SD
1	14.04	14.37	7.88	38.30	5.57
0	6.28	24.83	5.85	96.39	5.10
1	24.73	12.89	4.95	5.68	1.06
1	27.90	13.66	6.73	25.06	4.95
1	26.98	18.52	0.02	48.27	7.99
1	21.33	14.21	3.82	11.39	5.68

Wilcoxon Signed Ranks (MQO)

Hypo	Alpha	Beta	Delta	AL	SD-S	SD-A	Reps
0	24.32	11.57	1.83	36.28	4.26	2.07	2
1	6.29	21.80	8.42	92.03	3.56	6.90	0
0	12.84	5.70	6.15	58.51	2.80	0.17	3
1	1.54	5.73	8.46	60.58	6.45	1.26	0
0	10.00	27.01	6.40	1.08	5.45	7.62	0
1	17.63	19.36	8.98	66.70	0.08	8.75	2

MARSSIM Sign Test

Hypo	Alpha	Beta	Delta	AL	SD	Pct
1	22.56	3.18	1.68	57.41	3.09	26
1	26.68	14.46	9.58	39.63	1.06	1
1	4.63	14.45	6.18	90.91	7.53	24
0	24.48	13.77	1.45	24.16	0.50	34
0	21.53	27.55	1.89	66.85	5.44	32

0	4.24	28.33	6.35	37.29	4.86	23

MARSSIM Sign Test (MQO)

Hypo	Alpha	Beta	Delta	AL	SD-S	SD-A	Reps	Pct
1	29.95	27.14	0.43	53.16	5.97	7.75	3	18
1	17.23	17.66	9.84	53.34	1.98	9.97	0	12
1	6.59	20.80	2.37	12.53	2.75	7.93	1	28
0	1.40	1.43	3.70	44.36	6.32	9.40	3	32
1	17.47	27.34	9.76	29.91	8.67	8.13	3	12
1	27.95	28.54	0.24	16.18	4.81	1.23	2	21

Two Sample t

Hypo	Alpha	Beta	Delta	Diff	SD
1	14.69	3.66	4.47	15.65	1.31
1	12.60	27.48	4.76	17.18	4.43
1	13.07	25.61	2.31	9.48	8.21
0	3.77	11.30	9.95	2.05	9.21
1	10.52	18.44	0.92	5.25	1.30
1	26.60	15.73	8.66	5.01	5.81

Two Sample t (MQO)

Hypo	Alpha	Beta	Delta	Diff	SD-S	SD-A	Reps
0	4.72	15.69	4.01	5.73	4.65	5.82	3
0	23.15	20.88	6.18	7.78	4.83	2.77	0
1	0.79	3.04	4.18	11.48	4.02	4.73	3
1	1.82	14.28	0.88	3.01	8.12	3.65	1
0	14.88	26.96	4.44	5.90	2.96	7.59	0
0	4.01	22.28	7.18	18.85	0.01	9.87	3

Wilcoxon Rank Sum

Hypo	Alpha	Beta	Delta	Diff	SD
0	28.30	3.37	0.70	6.48	1.28
0	17.24	20.59	3.73	9.94	1.25
1	25.33	28.08	1.93	11.59	0.36
1	3.69	26.54	0.16	4.00	2.96
1	1.24	24.95	6.78	19.59	8.22
0	11.54	23.57	5.21	14.34	4.26

Wilcoxon Rank Sum (MQO)

Hypo	Alpha	Beta	Delta	Diff	SD-S	SD-A	Reps
0	25.37	21.49	1.51	17.79	7.54	0.94	0
1	28.95	19.66	4.81	0.78	5.35	9.81	1
0	10.23	25.52	8.16	8.80	0.99	9.71	2
0	13.92	19.48	1.91	3.21	8.62	9.75	1
0	24.72	9.50	0.27	3.72	6.17	8.69	0
1	26.02	26.61	7.22	4.49	9.57	0.72	2

MARSSIM WRS

Hypo	Alpha	Beta	Delta	Diff	SD	Pct
1	7.94	26.90	8.72	17.45	5.43	99
0	2.57	24.34	2.29	5.18	8.98	52
0	2.24	25.69	9.10	11.05	0.64	5
0	16.19	23.49	9.54	6.20	9.80	69
1	8.00	15.72	6.86	13.92	3.73	2
0	11.87	29.74	5.43	4.43	9.49	70

MARSSIM WRS (MQO)

Hypo	Alpha	Beta	Delta	Diff	SD-S	SD-A	Reps	Pct
0	17.87	8.59	4.81	7.50	1.97	9.62	0	98
0	22.27	7.23	4.57	3.20	3.03	3.17	1	3
0	10.45	28.39	2.06	6.87	6.59	5.43	2	78
1	17.30	3.93	6.05	18.75	7.09	6.05	0	82
1	23.02	6.68	1.27	14.68	0.12	3.15	1	89
1	14.28	14.62	9.32	19.14	9.16	9.80	0	4

Confidence Interval

Sided	Conf	Width	SD
2	56.02	11.27	15.71
1	94.03	21.01	11.11
1	94.67	3.79	10.56
2	68.39	10.65	29.41
1	99.37	20.57	11.83
2	51.21	14.16	18.06

Confidence Interval (MQO)

Sided	Conf	Width	SD	SD-A	Reps
1	94.06	4.57	3.41	5.83	2
1	78.74	0.20	4.10	8.58	0
1	84.99	5.07	5.43	4.73	3
2	80.00	0.64	7.24	1.58	1
1	79.75	6.27	0.78	8.55	1
1	97.58	4.20	6.41	2.08	0

One-Sample Proportion

Hypo	Alpha	Beta	Delta	AL
0	12.36	1.97	0.83	0.86
0	8.71	19.23	0.56	0.44
0	3.77	15.07	0.66	0.88
1	1.35	20.41	0.59	0.84
0	1.55	27.53	0.22	0.89
1	7.09	10.67	0.50	0.26

Two-Sample Proportion

Hypo	Alpha	Beta	P-R	P-S	AL	Delta
1	23.71	20.44	0.67	0.71	0.15	0.33
0	2.83	1.53	0.70	0.25	0.82	0.63
1	18.14	1.69	0.97	0.77	0.95	0.45

1	11.88	5.87	0.09	0.39	0.02	0.44
0	11.78	13.22	0.64	0.82	0.78	0.65
0	4.20	5.92	0.40	0.33	0.65	0.08

Hot Spot - Predetermined

Grid	Length	Shape	Axis		Random	Angle
0	8.99	0.67	18.04	1		
0	10.04	0.72	29.45	0	49.03	
1	11.93	0.53	20.79	1		
2	9.57	0.46	17.58	0	69.24	(Use 1:2 rectangle)
0	5.85	0.23	19.16	1		
0	4.11	0.51	8.14	0	62.51	

Hot Spot - Cost

ost	Grid	Shape	Axis	Random	Angle	Fixed	Field	Analy	
47727.69	2	0.54	25.33	0	60.56	647.64	120.01	488.55	(Use 1:2 rectangle)
35488.97	0	0.75	21.01	1	822.57	54.73	12.97		
33726.19	0	0.67	0.99	0	81.40	867.39	19.61	432.69	
6651.38	1	0.72	15.78	1	138.44	95.48	374.38		
4194.69	0	0.24	8.44	0	79.87	542.91	165.81	215.99	
35734.00	0	0.75	26.27	1	876.22	275.54	184.82		

Hot Spot - Minimum Samples

Prob	Grid	Shape	Axis	Random	Angle	
82.81	1	1.00	4.65	1		
69.77	0	0.65	11.43	0	74.72	
54.31	2	0.65	15.23	0	66.48	(Use 1:2 rectangle)
76.50	0	0.59	19.52	1		
92.11	2	0.24	2.57	0	59.74	(Use 1:2 rectangle)
61.39	2	0.57	29.37	1		(Use 1:2 rectangle)

Hot Spot - Find Hot Spot Size

Prob	Grid	Length	Shape	Random	Angle	
85.54	2	13.70	0.63	0	81.75	(Use 1:2 rectangle)
83.22	1	9.47	0.96	1		
88.96	2	17.42	0.69	1		(Use 1:2 rectangle)
87.92	0	20.32	0.23	0	9.07	
51.21	2	26.94	0.96	1		(Use 1:2 rectangle)
95.81	0	2.25	0.31	1		

Predetermined Random

Num
95
17
18
31
22
58

Predetermined Grid
Num
62
93
9
40
88
23

Distribution

No. of Copies

OFFSITE

5 **EPA**

M. Carter
Ariel Rios Building
1200 Pennsylvania Avenue, N.W.
Washington, DC 20460

D. Crumbling
Ariel Rios Building
1200 Pennsylvania Avenue, N.W.
Washington, DC 20460

T. Jover
Ariel Rios Building
1200 Pennsylvania Avenue, N.W.
Washington, DC 20460

J. Warren
US EPA / Quality Assurance Division
1300 Pennsylvania Avenue
Washington, DC 20460

L. Zaragoza
USEPA Headquarters
Ariel Rios Building, 5204G
1200 Pennsylvania Avenue, N.W.
Washington, DC 20460

DOE

D. Bottrell
19901 Germantown Rd.
Germantown, MD 20874-1290

No. of Copies

2 **DoD**

A. Andrews
SERDP Program Office, Suite 303
901 N. Stuart Street
Arlington, VA 22203

J. Sample
NAVSEA 04XQ
1661 Redbank Rd Suite 104
Goose Creek, SC 29445-6511

RTI

K. Morgan
1615 M Street NW, Suite 740
Washington, DC 20036

ONSITE

DOE, RL

E.M. Bowers	A2-15

16 **Pacific Northwest National Laboratory**

D.J. Bates	K5-12
D.K. Carlson	K5-12
R.O. Gilbert (5)	K5-12
N.L. Hassig	K5-12
C.A. McKinstry	K5-12
R.F. O'Brien	K5-12
B.A. Pulsipher (5)	K5-12
J.E. Wilson	K5-12