PNNL-14970

Pacific Northwest National Laboratory
Operated by Battelle for the U.S. Department of Energy

Visual Sample Plan Version 3.0 User's Guide

N. L. Hassig
J. E. Wilson
R. O. Gilbert
B. A. Pulsipher

December 2004

Prepared for the U.S. Department of Energy
and the U.S. Environmental Protection Agency
under Contract DE-AC05-76RL01830

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service,
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161
ph: (800) 553-6847
fax: (703) 605-6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/ordering.htm

This document was printed on recycled paper.

PNNL-14970

Visual Sample Plan
Version 3.0
User's Guide

N. L. Hassig R. O. Gilbert
J. E. Wilson B. A. Pulsipher

December 2004

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Abstract

This user's guide describes Visual Sample Plan (VSP) Version 3.0 and provides instructions for using the software. VSP selects the appropriate number and location of environmental samples to ensure that the results of statistical tests performed to provide input to environmental decisions have the required confidence and performance. VSP Version 3.0 provides sample-size equations or algorithms needed by specific statistical tests appropriate for specific environmental sampling objectives. The easy-to-use program is highly visual and graphic. VSP runs on personal computers with Microsoft Windows operating systems (95, 98, NT, 2000, Millennium Edition, and XP). Designed primarily for project managers and users without expertise in statistics, VSP is applicable to any two-dimensional geographical population to be sampled (e.g., surface soil, a defined layer of subsurface soil, building surfaces, water bodies, and other similar applications) for studies of environmental quality as well as for devising transect sampling plans for unexploded ordnance target area detection and delineation..

Acknowledgments

Many U.S. government sponsors have contributed to the development of VSP. We wish to thank Mike Carter, Office of Solid Waste and Emergency Response, Environmental Protection Agency for his continued support for documentation and quality assurance of VSP, especially for his support of the development of this Users Manual. We also thank George Detsis and Chip Lagdon, Department of Energy Office of Environmental Safety and Health for continued support with training and VSP enhancements. Special thanks to David Bottrell, U.S. Department of Energy, for his support of VSP developments in the early stages of its development. In addition, we wish to thank John Warren of the Quality Staff, U.S. Environmental Protection Agency Office of Environmental Information, for his support in adding several new statistical design methods to VSP. We also wish to acknowledge the excellent guidance and support of specialized sampling modules offered by several other VSP sponsors including Navy sponsors Bill Ingersol, Fred McLean, and the late Jackie Sample, SERDP/ESTCP Unexploded Ordnance Coordinator Anne Andrews, and Technical Support Working Group sponsor Rebecca Blackmon.

Many statisticians have contributed to VSP developments. Special thanks are extended also to individuals in the Statistical Sciences Group at Pacific Northwest National Laboratory: Debbie Carlson, Brett Matske, Bob O'Brien, Rick Bates, Craig McKinstry for their excellent statistical input. Lucille A. Walker for her project financial accounting support; and Mary H. Cliff for her assistance in preparing the final report.

Content

Abstract iii

Acknowledgements v

1.0 Introduction 1.1

1.1 What is Visual Sample Plan? 1.1

1.2 What's New in VSP 3.0? 1.2

1.3 Installation and System Requirements 1.2

1.4 Overview of VSP 1.3

1.5 How Do I Use VSP to Provide a Defensible Sampling Plan? 1.5

2.0 Mechanics of Running VSP 2.1

2.1 Getting Started and Navigational Aids 2.1

2.2 Setting Up a Map 2.3
2.2.1 Importing a Site Map from a File 2.4
2.2.2 Importing a Site Map File in the VSP Format 2.4
2.2.3 Draw Map Using VSP Drawing Tools 2.4
2.2.4 Working with Maps 2.6
2.2.5 Additional Map Features 2.8

2.3 Sample Areas in VSP 2.9
2.3.1 Creating a Sample Area 2.9
2.3.2 Selecting or Deselecting Sample Areas 2.12
2.3.3 Deleting Selected Sample Areas 2.13
2.3.4 Sample Area Parameters 2.13

2.3.5 Extended Sample Area Topics 2.16

2.4 Individual Samples 2.16
2.4.1 Importing Samples 2.17
2.4.2 Historical Samples 2.19
2.4.3 Exporting Sampling Locations 2.20
2.4.4 Removing Sampling Locations 2.20

2.5 Rooms and Buildings in VSP 2.21
2.5.1 Drawing a Room 2.21

2.5.2 Extended Room Features 2.26

2.6 Saving a VSP File 2.29

3.0 Sampling Plan Development Within VSP 3.1

3.1 Sampling Plan Type Selection 3.1
3.1.1 Defining the Purpose/Goal of Sampling 3.1
3.1.2 Selecting a Sampling Design 3.3

3.2 DQO Inputs and Sample Size 3.6
3.2.1 Compare Average to a Fixed Threshold 3.8
3.2.2 Compare Average to Reference Average 3.21
3.2.3 Estimate the Mean 3.28
3.2.4 Construct Confidence Interval on Mean 3.39
3.2.5 Compare Proportion to Fixed Threshold 3.40
3.2.6 Compare Proportion to Reference Proportion 3.40
3.2.7 Estimate the Proportion 3.41
3.2.8 Locating a Hot Spot 3.43
3.2.9 Find UXO Target Areas 3.45
3.2.10 Access Degree of Confidence in UXO Presence 3.50
3.2.11 Non-statistical Sampling Approach 3.52
3.2.12 Establish Boundary of Contamination 3.52

4.0 Assessment of Sampling Plans 4.1

4.1 Display of Sampling Design on the Map: MAP VIEW button or View > Map 4.1

4.2 Display of Cost of Design 4.1

4.3 Display of Performance of Design: GRAPH VIEW button or View > Graph 4.2
4.3.1 Performance of Design for Sampling Goal: Compare Average to a Fixed Threshold 4.3
4.3.2 Performance of Design for Sampling Goal: Construct Confidence Interval on the Mean 4.7
4.3.3 Performance of Design for Sampling Goal: Comparing a Proportion to a Fixed Threshold 4.8
4.3.4 Performance of Design for Sampling Goal: Compare Average to Reference Average 4.9
4.3.5 Performance of Design for Sampling Goal for Hot Spot Problem 4.11
4.3.6 Performance of Design for Sampling Goal of Compare Proportion to a Reference Proportion 4.13
4.3.7 Performance of Design for Sampling Goal of Establish Boundary of Contamination 4.14

4.4 Display of the Report 4.15

4.5 Display of Coordinates 4.18

4.6 Multiple Displays.. 4.19

5.0 Extended Features of VSP 5.1

5.1 Tools 5.1
5.1.1 Largest Unsampled Spot 5.1
5.1.2 Reset Sampling Design 5.1
5.1.3 Measure Distance 5.2
5.1.4 Make Sample Labels 5.3

5.2 Options.. 5.3
5.2.1 Random Numbers 5.3
5.2.2 Sample Placement 5.4
5.2.3 Graph 5.5
5.2.4 Measurement Quality Objectives (MQOs) 5.6
5.2.5 Sensitivity Analysis 5.13
5.2.6 Preferences 5.13

5.3 View Menu 5.13

5.4 The Cost Tab: Setting Up Sampling Costs – Inputs for the General Screen 5.17

5.5 Multiple Areas To Be Sampled... 5.17

6.0 Room Features in VSP.. 1

6.1 Room Creation and Manipulation... 6.1
6.1.1 Creating a Room from a Sample Area 6.1
6.1.2 Drawing a Room 6.2
6.1.3 Room Delineation Method.. 6.2
6.1.4 Room Manipulation 6.4

6.2 Room Display Options.. 6.5
6.2.1 Current Room.. 6.5
6.2.2 Room View Types 6.6
6.2.3 Room North Arrow 6.7
6.2.4 Perspective Ceiling 6.8

6.3 Room Objects 6.8
6.3.1 Doors.. 6.8
6.3.2 Windows 6.9
6.3.3 Notes 6.10

6.4 Other Room Features 6.10

6.4.1 Surface Labels 6.10
6.4.2 Local Coordinates and Room Origin 6.10
6.4.3 Room Label 6.12

7.0 References 7.1

Figures

1.1 Screen Capture from VSP Using Quad Window Option 1.4

2.1 VSP Welcome Screen with Version Selection Menu 2.1

2.2 VSP Welcome Screen with VSP Advisor 2.2

2.3 Main Menu Items and Buttons on the Toolbar 2.3

2.5 The Millsite.dxf File Opened in VSP, showing MAP Pull-down Menu 2.5

2.9 Example of an Open Boundary with an Arrow to Show the Direction the Soil Contamination would be Expected to Move 2.11

2.10a Map with Multiple Sample Areas Selected 2.12

2.11 Sample Information Dialog Box for a Sample Area 2.14

2.14 Sample Information Dialog Box for a sample in Example2.VSP 2.17

2.15 The OneAcre.VSP Project with Sampling Locations Added from Windows Clipboard 2.18

2.16 Example of Sample Information Box 2.19

2.19 Room Information Dialog Box 2.22

2.21 Three perspective views of a room 2.23

2.22a Room in Map View 2.24

2.23 Door Object Displayed Using Map View 2.26

2.24 Door Room Object with Object Information Dialog Box Displayed 2.27

2.25 Window Room Object with Object Information Dialog Box Displayed 2.28

3.1 Menu Options in VSP for Comparing an Average to a Fixed Threshold 3.3

3.2 Dialog Box for Entering Type of Grid Design 3.5

3.3 Judgment Sampling in Vsp 3.7

3.4 Input Boxes for Case 1 with Original Error Rates 3.10

3.5 Input Boxes for Case 1 with Increased Error Rates 3.11

3.6 Dialog for Sequential Sampling ... 3.12

3.7 Data Input Dialog for Sequential Probability Ratio Test and Results from First Round of Sampling ... 3.13

3.8 Graph View of Sequential Sampling ... 3.14

3.9 Dialog Box for Sequential Sampling ... 3.15

3.10 Input Boxes for MARSSIM Sign Test ... 3.17

3.11 Input Boxes for Case 4 with Original Error Rates ... 3.22

3.12 Input Boxes for Case 4 with Increased Error Rates ... 3.24

3.13 Input Boxes for Case 5 Using Nonparametric Wilcoxon Rank Sum Test ... 3.25

3.14 Input Boxes for Case 6 Using Nonparametric Wilcoxon Rank Sum Test ... 3.26

3.15 Input Boxes for Case 7 Using Nonparametric Wilcoxon Rank Sum Test ... 3.27

3.16 Input Boxes for Case 9 with Larger Standard Deviation ... 3.28

3.17 Dialog Box for Stratified Sampling for Estimating a Mean ... 3.30

3.18 Dialog Boxes for Ranked Set Sampling Design ... 3.32

3.19 Map of RSS Field Sample Locations for All Sets in Cycle 3, Along with RSS Toolbar ... 3.33

3.20 Map of RSS Field Sampling Locations along with Their Labels ... 3.34

3.21 Map of the Sample Area with Grid Cells Displayed ... 3.34

3.22 Map of Sample Area with Initial Samples for Adaptive Cluster Sampling Shown as Yellow Squares, Along with Dialog Box ... 3.35

3.23a Dialog Input Box for Entering Sample Measurement Values and Labels for Initial Samples in Adaptive Cluster Sampling ... 3.36

3.24 Dialog Input Box for Entering Grid Size and Follow-up Samples ... 3.36

3.24b Examples of Combinations of Initial and Follow-up Samples from Adaptive Cluster Sampling ... 3.37

3.24c Input Dialog Box for Collaborative Sampling for Estimating the Mean ... 3.38

3.25 Dialog Input Box for Calculating a Confidence Interval on the Mean ... 3.40

3.26 Design Dialog for Comparing a Proportion to a Fixed Threshold..3.41

3.27 Input Dialog for Comparing a Site Proportion to a Reference Proportion......................................3.41

3.28 Dialog Box for Estimating a Proportion Using Stratified Sampling 3.42

3.29 Input Boxes for Case 9 ....3.44

3.30 Dialog Input Box for Finding the Swath Spacing and Probability of Traversing and Detecting Target Area 3.46

3.31 Dialog Input Box for Probability that One or More Target Areas Exist even though None were found in the Swath Survey 3.46

3.32a Imported Meandering Swath..3.47

3.32b Imported Meandering Swath with width 3.47

3.33 Input Dialog Box for Post-Survey Probability of Detection Using Simulation 3.48

3.34 Map View of Simulation Results for Probability of Traversing and Detecting a Target Area............3.49

3.35 Dialog Input Box for Find Target Areas..3.49

3.36 Dialog Input Box for Compliance Sampling for UXO and Map of Sample Area with Transects Selected 3.50

3.37 Dialog Input Box for Bayesian Method for UXO and Map of Sample Area with Transects Selected 3.52

3.38 Judgment Sampling with Six Sampling Locations Added Manually 3.53

3.39 Information Dialog Box for Judgment Samples 3.54

3.40 List of Default Contaminants of Concern and their Action Levels 3.55

3.41 Dialog Box for Entering Design Inputs for Sampling an Enclosed Boundary.....................................3.56

3.42 Enclosed Boundary with Two Bumped-Out Segments..3.57

3.43 Sample Information Box for Entering Data into VSP, Duplicate Samples Required...........................3.57

3.44 An Enclosed Boundary Showing the Five Primary Sampling Locations for each of the 17 Segments...3.58

4.1 Display of Sampling Locations on Map 4.2

4.2 Decision Performance Goal Diagram for Null Hypothesis: True Mean >=Action Level for Comparing Mean vs. Action Level 4.3

4.3 Graph of Probability of Making Correct Decision 4.5

4.4 Decision Performance Goal Diagram for Null Hypothesis: True Mean <= Action Level for Comparing Mean vs. Action Level 4.6

4.5 Decision Performance Graph for One-Sided 95% Confidence Interval 4.8

4.6 Decision Performance Goal Diagram for Comparing a Proportion to a Fixed Threshold 4.9

4.7 Decision Performance Goal Diagram for Comparing a Sample Area Mean to a Reference Area Mean 4.10

4.8 Decision Performance Graph for Comparing a Sample Area Mean to a Reference Area Mean 4.11

4.9 Probability of Hitting a Hot Spot vs. Number of Samples 4.12

4.10 Decision Performance Goal Diagram for Comparing a Sample Area Proportion to a Reference Area Proportion 4.14

4.11 Curve of trade-off Between Primary Sampling Locations and Size of Hot Spot that can be Detected 4.15

4.12 Report View of the Sampling Goal: Compare Average to a Fixed Threshold, Normality Assumed, Simple Random Sampling 4.17

4.13 Dialog Box for Changing Variables Displayed, and Range for Variables Shown, in Sensitivity Table in Report View 4.18

4.14 Coordinates Display of Sampling Locations 4.19

4.15 Quad Display of Map, Graph, Report, and Coordinates on Same Screen 4.21

4.16 Combined Display of VSP Inputs and Outputs 4.22

5.1 Largest Unsampled Spot Displayed on Map 5.2

5.2 Information Box for Largest Circle that Would Fit Within an Unsampled Area 5.2

5.3 Information Box Showing Percentage of Circle Within the Sample Area 5.2

5.4 Measuring Tool in VSP 5.3

5.5 Dialog Box for Creating Sample 5.3

5.6 Menu for Selecting Type of Random Number Generator 5.4

5.7 Adaptive-Fill Option for Sample Placement 5.5

5.8 Sample Information Window Displayed When the User Right-Clicks on Selected Sample Points on Map 5.6

5.9 Sample Exported Text File of Sampling Locations 5.7

5.10 Graph Options 5.7

5.11 MQO Input Dialog Box with Default Values Displayed 5.8

5.12 MQO Input Dialog Box Showing Positive Value for Estimated Analytical Standard Deviation with 1 Analysis per Sample 5.9

5.13 MQO Input Dialog Showing Positive Value for Estimated Analytical Standard Deviation with Multiple Analysis per Sample 5.10

5.14 Cost Input Dialog Box for MQO Option 5.11

5.15 Display of Cost Comparison for Method A and Method B from MQO Module 5.12

5.16 MQO Method Comparison Chart 5.13

5.17 Sensitivity Analysis for 3 DQO Input Parameters 5.14

5.18 Sensitivity Analysis for 4 DQO Input Parameters 5.15

5.20 Screen for Entering Sampling Costs for a Sampling Design – Accessed through the Cost Tab 5.17

5.23 Proportional Allocation of Samples to Multiple Sample Areas 5.20

6.1 Sample Area Information Dialog 6.2

6.2 Room Delineation Mode 6.3

6.3 Room Manipulation 6.4

6.4 Changing Segment Length 6.5

6.5 Current Room 6.5

6.6 Room View Types 6.6

6.7 Room North Arrows 6.7

6.8 Room Objects 6.8

6.9 Object Information Dialog Box 6.9

6.10 Room Surface Labels 6.11

6.11 Room Surface Labels 6.11

Tables

1.1 List of Sampling Goals....1.1

4.1 Interactive Graph Features....4.4

4.2 Graph Options Menu Commands....4.7

4.3 Window Menu Commands....4.20

5.1 Preferences Menu Items....5.16

1.0 Introduction

1.1 What is Visual Sample Plan?

Visual Sample Plan (VSP) is a software tool for selecting the right number and location of environmental samples so that the results of statistical tests performed on the data collected via the sampling plan have the required confidence for decision making. More than 4000 users from every state as well as many in other countries have registered a downloaded copy of VSP. Users include employees of the federal government, state and local governments, and private industry. Sponsors of this public domain software include the U.S. Environmental Protection Agency (EPA), U.S. Department of Energy (DOE), U.S. Department of the Defense (DOD and U.S. Department of Homeland Security (DHS). VSP 3.0 is a major new release of the software and incorporates many new features.

VSP provides sample designs and sample-size equations needed by specific statistical tests appropriate for several types of environmental problems. Table 1.1 is a list of the sampling goals that can be addressed in VSP 3.0.

Table 1.1. List of Sampling Goals

Sampling Goal	User's Guide Section
Compare Average to Fixed Threshold	3.2.1
Compare Average to Reference Average	3.2.2
Estimate the Mean	3.2.3
Construct Confidence Interval on Mean	3.2.4
Compare Proportion to Fixed Threshold	3.2.5
Compare Proportion to Reference Proportion	3.2.6
Estimate the Proportion	3.2.7
Locating a Hot Spot	3.2.8
Find UXO Target Areas	3.2.9
Assess degree of confidence in UXO presence	3.2.10
Non-statistical sampling approach	3.2.11
Establish boundary of contamination	3.2.12

VSP is easy to use, highly visual, and graphic. It has extensive online help and tutorial guides. Reports produced by VSP can be pasted directly into a quality assurance project plan or a sampling and analysis plan. VSP can be used to implement EPA's systematic planning process (EPA 2000) for a variety of problems: selection between clearly defined alternatives [Step 7 of the Data Quality Objectives (DQO) process], studies where a confidence interval on an estimated parameter is needed, or determination of whether a hot spot or target exists. The user specifies the criteria for "how good" the answer has to be (Step 6 of the DQO Process), and VSP uses this as input to the formula for calculating the required sample size. VSP is unique in this regard.

VSP is designed primarily for project managers and users who are not statistical experts, although those individuals with statistical expertise also will find the code very 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. VSP 3.0 is also applicable to sampling building interiors and room surfaces.

1.2 What's New in VSP 3.0?

VSP 3.0 offers expanded sampling designs—more UXO designs, collaborative sampling and contaminant boundary delineation. The **Report View** that describes in words the sampling plan designed by VSP along with the assumptions and formulas used in sample size calculations is now customizable by the user. VSP 3.0 now includes a linear-type boundary useful for contaminant boundary delineation. A major improvement to VSP 3.0 is to support complex-shaped rooms and sample attributes to support locating samples on room surfaces. VSP 3.0 now supports user-defined parameters associated with sample areas and rooms.

The experienced user will notice many other changes and enhancements. The expanded **Help** function provides an exhaustive list of formulas, definitions, and references used within VSP. Users are encouraged to liberally use the **Help** function as they navigate the expanded functionality within VSP 3.0. Help is available on a topic basis when accessed from the **Help** menu item on the main screen, and on a tutorial basis when accessed from the **Help** button at the bottom of the input screen for each of the sampling designs.

Several sponsors are supporting the addition of new features to VSP. We look forward to including these new features in a new revised version of VSP in the near future.

1.3 Installation and System Requirements

VSP 3.0 runs on Microsoft Windows 95 and later… *Visual Sample Plan (VSP) Models and Code Verification* (Gilbert et al. 2001) documents the successful installation of VSP on personal computers operating with Windows~ 95, 98, Millennium Edition, and Windows 2000. VSP is developed and successfully runs under Windows XP.

VSP will not run on Windows 3.1 or earlier Windows operating systems. VSP currently does not run on Macintosh® or UNIX®/Linux systems. Any personal computer with sufficient hardware to run one of the supported operating systems should run VSP. The *minimum* hardware recommended is

- Pentium processor
- 64 MB RAM
- 20 MB of free space on the hard drive.

The current version of the VSP setup file is available from http://dqo.pnl.gov/VSP. After the setup file is downloaded, installation of VSP is almost automatic. Simply run the VSP setup file, VSP30.exe (or later version), and follow the on-screen instructions. The VSP program and auxiliary files will be copied by default to the C:\Program Files\Visual Sample Plan folder (subdirectory). However, you may specify a different location for the files.

Once installation is complete, you will start VSP using option **Start > Program Files > Visual Sample Plan > Visual Sample Plan**. Alternatively, you may place a VSP shortcut on the desktop by selecting **New > Shortcut** from the menu obtained by right-clicking the mouse on the desktop. The appropriate command line for the default folder is

"C:\Program Files\Visual Sample Plan\VSample.exe".

VSP may be uninstalled using the Control Panel icon labeled Add/Remove Programs. You may access this option using the **Start** button and **Settings > Control Panel.**

New versions of VSP are often released as prototypes for testing. These demonstration versions have expiration dates. After the expiration date has passed, you will be given the option of continuing with the current version or going to the VSP website to download the latest version. Version 3.0 is not a demonstration version and does not have an expiration date.

1.4 Overview of VSP

Sampling is the process of gaining information about a *population* from a portion of that population called a *sample*. A key goal of *sampling design* is to specify the sample size (number of samples) and sampling locations that will provide reliable information for a specific objective (called the **Sampling Goal)** at the least cost. VSP does these required calculations for sample size and sample location and outputs a sampling design that can be displayed in multiple formats. VSP does not address sample types to take, sample collection methods, or sample results. It does address the trade-off between repeated analytical measurements on a single sample to reduce overall sample result variability (MQOs) and provides a sensitivity table for comparing analytical methods of varying accuracy and cost.

VSP can be used to develop a new sampling design. It can also be used to compare alternative designs. VSP automates the mechanical details of calculating sample size, specifying random sampling locations, and comparing sample costs with decision error rates. These activities can be accomplished in the context of your own site map displayed onscreen with various sampling plans overlain on sample areas that you select.

The first thing you will do after opening the program is to import or construct a visual map of the study site. Next, you select the area or areas to be sampled. The **Sample Area** may be only a portion of the study site (see the circular sample area in Figure 1.1, upper left window).

Then, for the Sampling Goal that you select, VSP will lead you through the quantitative steps of the DQO process (Steps 6 and 7) so that the program has the information needed to compute the recommended minimum number of samples (sample size). You can enter sampling costs and test alternative designs against a fixed budget.

The locations of the samples over the Sample Area are determined by the specific sampling design (pattern) that you select. For some Sampling Goals, and for some assumptions about the population, only certain designs are allowed from a statistical theory perspective. For example, sequential sampling is appropriate only for the sampling goal of **Compare Average to a Fixed Threshold** when the population units

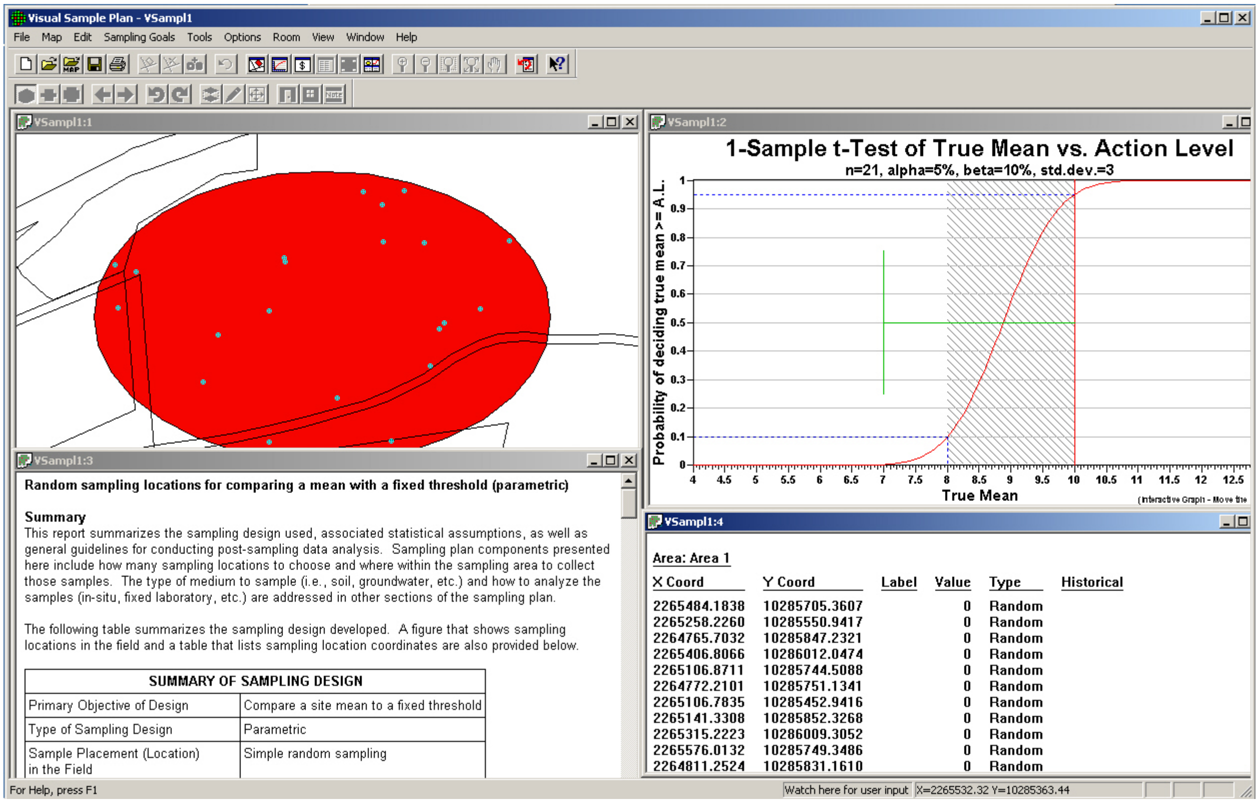


Figure 1.1. Screen Capture from VSP Using Quad Window Option

can be assumed to be distributed normally. When there is a choice of designs, VSP displays a dropdown menu of the allowable designs. VSP 3.0 has an expanded set of sampling designs from which to select.

On the site map, VSP displays the sample locations for easy visualization (see Figure 1.1, upper left window). VSP also lists the geographical coordinates of the sample locations (see Figure 1.1, lower right corner), which can be saved and exported as a Drawing Exchange Format (DXF) or SHP file for use in a geographical information system (GIS) or saved as a text file for use in global positioning system (GPS) software.

Two additional output formats for the design created in VSP are available: a **Graph View** of the design (see Figure 1.1, upper right window), and a **Report View** (see Figure 1.1, lower left window). The Graph View displays either a Decision Performance Goal Diagram for Sampling Goals that involved selecting between alternative actions, or a performance graph comparing number of samples to a design parameter for the other classes of sampling goals. The Report View is a text report that describes the sampling design in detail. The report contains the input values, the assumptions, the cost of the design, a technical description of the sample size formula used, and a sensitivity analysis table to assess what would happen if more or fewer samples are collected than the optimal number calculated by VSP.

New to VSP 3.0 is the **Room View** which allows the user to display and interact with sampling locations on the interior of building rooms. Chapter 6 discusses the new **Room View** in detail.

1.5 How Do I Use VSP to Provide a Defensible Sampling Plan?

To defend a sampling plan to a regulator concerned about safety ***and*** to a citizens' group concerned about saving taxpayer dollars requires balancing cost and risk. Defensible means that sufficient samples are taken, in a non-biased way, in order to make a decision, estimate a proportion, or declare an area free of UXO with a stated level of confidence. Additionally, once samples are taken and the results processed, someone needs to apply a statistical test to actually make a decision based on the data or calculate a confidence interval. VSP incorporates all this into the code it uses to calculate a sample size and sample locations. It asks the user to enter the assumptions, acceptable risk, and costs it needs for these calculations.

VSP follows the EPA-sanctioned planning approach for data collection and decision-making called the Data Quality Objectives (DQO) process. The DQO process achieves the user's limit on acceptable risk, at a minimum cost. See EPA (2000) for an extended discussion of the DQO process. There are 7 steps in the DQO process. Users must complete Steps 1 through 6 in order to have the inputs VSP needs. Then, using VSP, the user can complete Step 7, "Optimize the Design for Obtaining Data," because VSP can be used to try out different sampling designs and find the optimal design for the current problem.

Users familiar with the DQO process know that often a single site may have multiple sampling goals and multiple Sample Areas, each requiring its own set of DQO inputs and hence different sample requirements. VSP can help because it allows rapid prototyping and has many features that allow the overlay of designs and comparisons across designs.

2.0 Mechanics of Running VSP

2.1 Getting Started and Navigational Aids

Upon launching VSP, the first screen you will see is "Welcome to Visual Sample Plan" overlain with the initial navigational screen, **"Select"** (Figure 2.1).

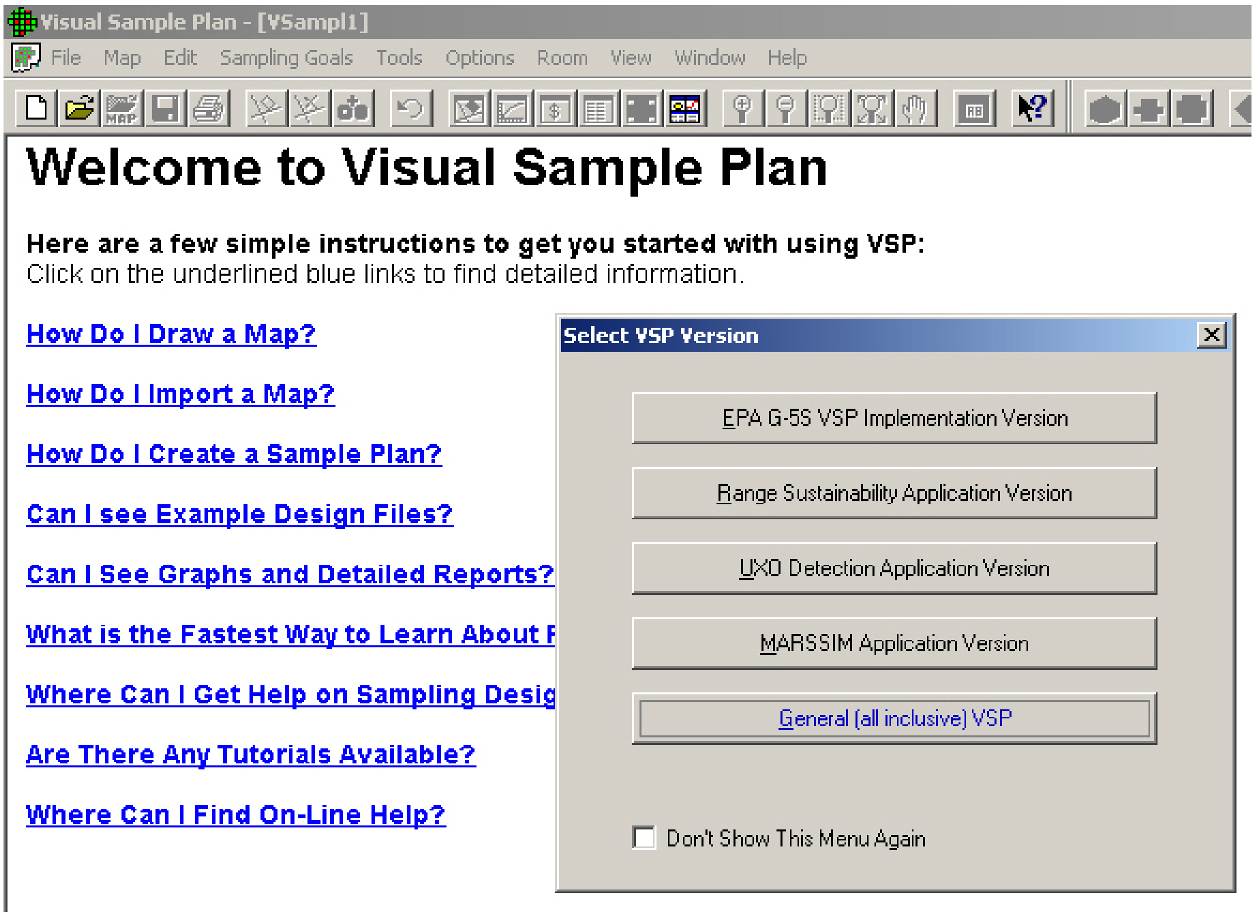


Figure 2.1. VSP Welcome Screen with Version Selection Menu

The choice of VSP versions is offered because different versions of VSP have been developed for different sponsors. Versions were designed to simplify the options presented to the single-purpose user as VSP became more complex. For example, users interested only in MARSSIM applications can select the MARSSIM version. That version contains menu items relating to only rooms and surfaces, and its statistical tests and sampling design options are limited to only those that are MARSSIM-approved. The **EPA G-5S VSP Implementation Version** implements only those sampling designs discussed in *Guidance for Choosing a Sampling Design for Environmental Data Collection* (EPA 2001). The **General (all inclusive) VSP** version provides access to all sampling designs and options.

Once a version has been selected, the second navigational menu pops up: **VSP Advisor**. This Help screen appears when VSP is first launched and again whenever **Advisor > Show Advisor** is selected from the main menu (Figure 2.2). Clicking any of the items under **VSP Advisor** will bring up a screen with brief explanation of the topic. Each topic description screen has a **Return to Main** that brings you back to **VSP Advisor**. You can close the **VSP Advisor** screen by either clicking in the "X" button on the top bar or selecting **Close** at the bottom of the screen.

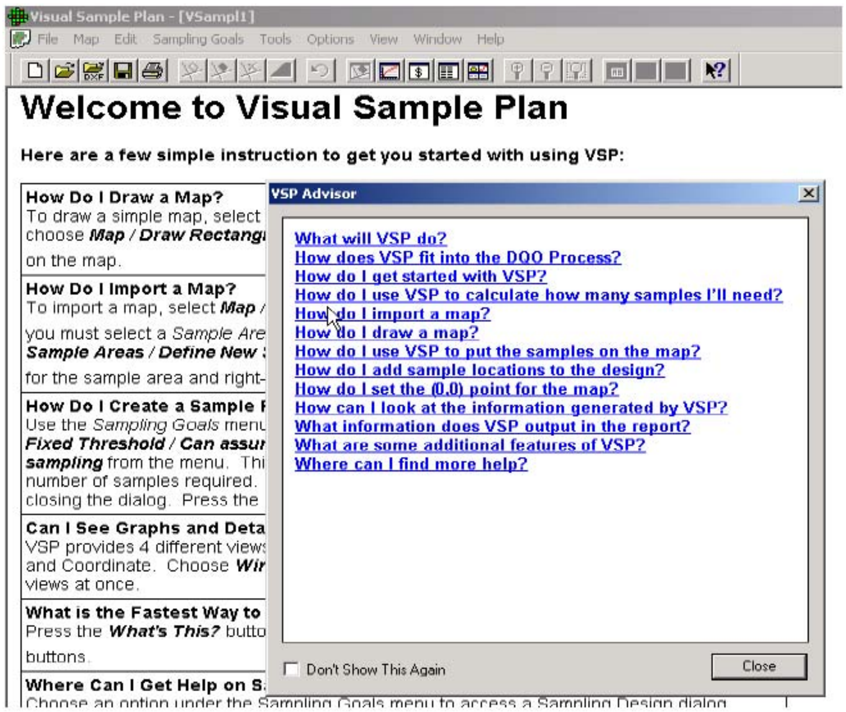


Figure 2.2. VSP Welcome Screen with VSP Advisor

Having closed **VSP Advisor**, you are now at the third navigational aid, the "Welcome to Visual Sample Plan" screen. The instructions on this page give answers to the most commonly asked questions from new VSP users. This screen will stay up until it is overlain with one of the View options, for example, when a map is loaded and you are in Map View.

You now are ready to begin using VSP after understanding one more piece of housekeeping. You have two ways to use VSP: pull-down menus from the top list of menu items, or the buttons on the main tool bar (select **View > Main Toolbar** to see the buttons). The pull-down menus offer a wider range of options. The buttons offer a quick one-click method for performing the primary VSP functions. Pull-down menus and buttons are shown in Figure 2.3. Holding the mouse over a button will reveal in text what that button does. For example, the Undo button is for undoing a key stroke during a map drawing session.

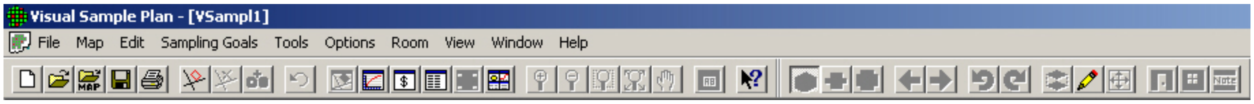


Figure 2.3. Main Menu Items (top row) and Buttons on the Toolbar (bottom row)

There are 4 separate toolbars: the main toolbar, map drawing toolbar, ranked set toolbar, and room toolbar. These toolbars can be moved around the screen by placing the mouse above the buttons on the toolbar and dragging to another place. Toolbars can be displayed or hidden depending on whether they are checked or not (see **View** pulldown list).

Starting with the **File** menu item on the top menu bar, the pull-down menu shows the various options for dealing with Projects.

VSP uses the term Project to refer to the map, report, sample information, and cost information associated with one sampling design. All this information is contained in the 'filename.VSP' created or selected by the user and is in a special VSP format file. Upon starting VSP, you either create a new project, **File** > **New Project**, or open an existing project, **File** > **Open Project** (Figure 2.4). If you are creating a new project, you will automatically be put into the "Welcome to Visual Sample Plan" screen after selecting **File** > **New Project**. If you are opening an existing project, you will be shown a list of existing VSP files and asked to select one.

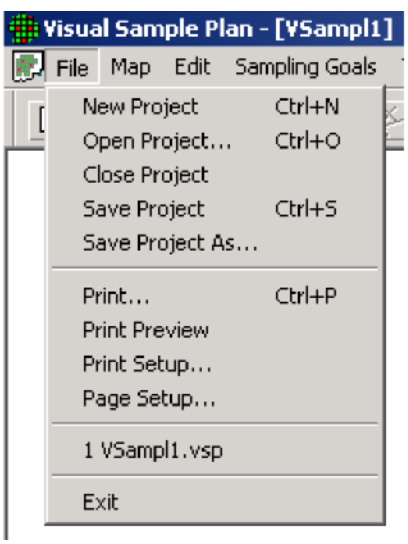


Figure 2.4. Pull-Down Menu Items Under File

2.2 Setting Up a Map

If you are starting a new project, you may obtain a map (drawing) of the site in any of three ways:

1. Import the site map from a drawing interchange format (DXF) file or ArcView SHP file. VSP supports the following DXF Objects: POLYLINE, LWPOLYLINE, LINE, ARC, CIRCLE, ELLIPSE, and TEXT. If you are having problems loading a DXF file into VSP, try converting your file to these types of objects.
2. Import the site map from a previous VSP project that was saved in VSP format (i.e., a .VSP file).
3. Draw the map or Sample Area using VSP's drawing tools.

These three methods are illustrated below. VSP uses the coordinate system associated with the imported map. Because neither DXF nor SHP files contain the distance units, VSP will assume your map is in feet until you change it to some other unit. This is done by selecting **Map** > **Set Map Extents** from the Main Menu. If you want to use a local origin in your design, use **Map** > **Set Origin** to click on the map at the location you want to become the new origin (0,0 point). You can also input the current location via the keyboard to become the new origin.

2.2.1 Importing a Site Map from a File

You can draw a complex site map in an architectural drawing program such as Autodesk Map~ AutoCAD~, or ArcView~ and save the drawing to a .DXF or SHP formatted file in that software package. The resulting file can be imported into VSP. The Millsite.dxf file is a sample DXF file provided with VSP. The following steps illustrate how to use this file in VSP:

1. From the main menu, select **Map > Load Map from File**. A quick alternative is to click on the **Load Map** button on the VSP toolbar.
2. A list of available files in the Visual Sample Plan folder is displayed. Select Millsite.dxf. You may change folder names to search for the desired file using standard Microsoft file-searching if the file is not in the Visual Sample Plan folder. Double-click on Millsite.dxf.
3. Choose whether or not you want to import the text embedded in the DXF file.

The site map should appear on your screen as illustrated in Figure 2.5.

2.2.2 Importing a Site Map File in the VSP Format

To open a VSP-formatted file, from the main menu select **File > Open Project** or use the Open button on the VSP toolbar. A list of available .VSP files is displayed. Double click on the .VSP file to be opened. Switch folders and/or directories if the desired file is in another folder or directory.

2.2.3 Draw Map Using VSP Drawing Tools

VSP provides a basic set of drawing tools for users who do not have a drawing program to create a site map. You can experiment with the drawing tools as follows:

- Create a new project by choosing **File > New Project** on the Main Menu or by clicking the New button on the main toolbar. To dismiss the “Welcome to Visual Sample Plan” displayed upon opening a new project, simply commence one of the drawing operations outlined below or **Map / Set Map Extents**. If the project window is not full screen, expand the project window by pressing the **Maximize** button on the upper right corner of the project window.
- Choose **View > Map Drawing Toolbar** from the Main Menu. This displays a toolbar used specifically for drawing a map. This toolbar also may be docked if you prefer to remove it from the project window. To dock the drawing toolbar, place the mouse cursor on the blue title bar and drag the drawing toolbar onto the VSP toolbar.

All the drawing functions described below also are available from the Main Menu option **Map**.

Draw Line. Click the **Draw Line** button on the toolbar. The cursor will become a cross, indicating that you are in drawing mode. Click a point on the map. You will now see a line between the cursor and point you clicked. Continue clicking points to make a complex polygon. If you make a mistake, click the Undo button on the VSP toolbar (or select **Edit > Undo** from the Main Menu or press Ctrl-Z on the keyboard). This will remove the last point you entered.

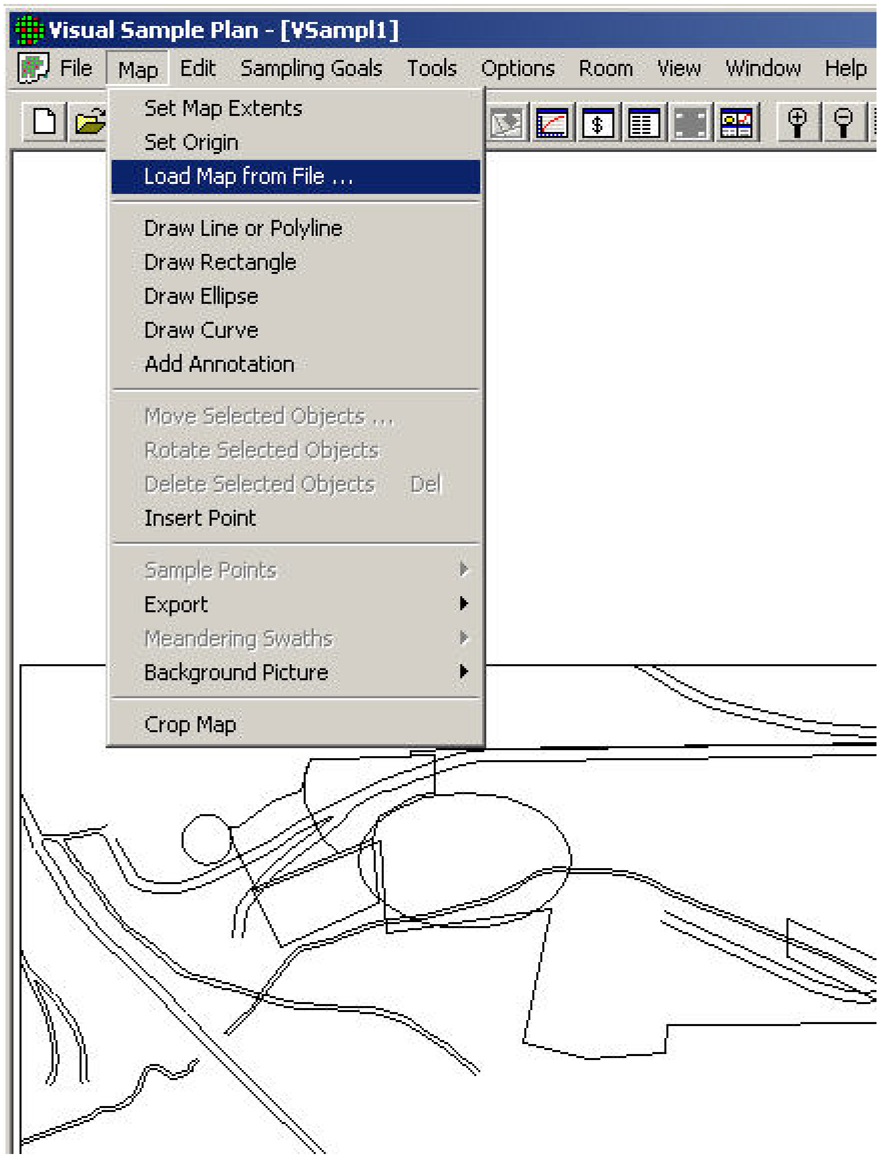


Figure 2.5. The Millsite.dxf File Opened in VSP, showing **MAP** Pull-down Menu

Points can be entered also on the keyboard. Just enter the x, y coordinates for each point (for example: type **32,48** and press the Enter key). You can see the coordinates that you are entering on the status bar at the bottom of the window. To connect a line to a point already entered (for example, to connect the last line to the first point to create a closed polygon), hold the Shift key while clicking with the mouse. Holding the Shift key can be used in most drawing operations to select the nearest point on the map without having to carefully position the cursor. Holding the Ctrl key while moving the mouse allows you to draw a horizontal or vertical line without having to be careful. To finish the line, right-click the mouse or click the **Draw Line** button on the toolbar again.

Draw Rectangle. Click the **Draw Rectangle** button on the toolbar. Click on a point on the map that you want to be one corner of a rectangle. Holding the Shift key while clicking causes that point to be attached to an existing point on the map. Move the cursor to the opposite corner of the rectangle and click the mouse button. Holding the Ctrl key while moving and clicking forces the rectangle to be a square. The x, y coordinates of the corner points can be entered on the keyboard also.

Draw Ellipse. Click the **Draw Ellipse** button on the toolbar. Drawing an ellipse is basically the same as drawing a rectangle. Holding the Ctrl key forces the ellipse to be a circle.

Draw Curve. Click the **Draw Curve** button on the toolbar. Click a point on the map. Click a second point on the map. A line is drawn between these first two points. As you move the cursor around the map, this line is stretched to become a curve. When the curve has the shape you want, click the mouse (this is the control point). The x, y coordinates for the three points also can be entered on the keyboard.

Add Annotation. Notes (also called note objects) can be added to maps using the annotation tool. Select **Map > Add Annotation**, and the cursor becomes a crosshair. Click on the map at the location where you want to add the note object. The location may also be entered on the keyboard. A default object containing the text "Right-Click Here" is added to the map.

After the default object is added, use the mouse to right-click on the note object. A **Map Label Information** dialog box pops up, as shown in Figure 2.6. You will be able to change the following parameters:

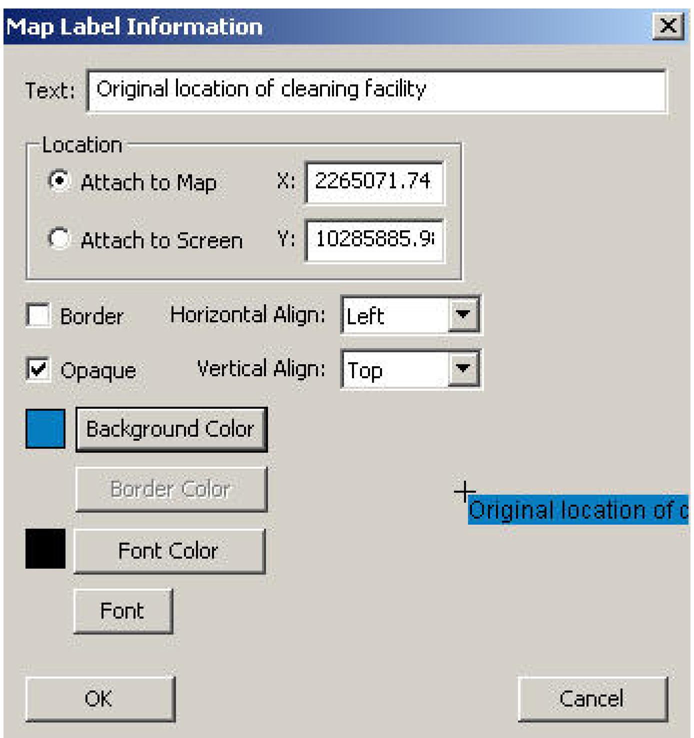


Figure 2.6. Map Label Information Dialog Box

- Note text
- Anchor point on map
- Anchor point to screen
- X and Y coordinates of anchor point (map or screen)
- Border, background color, and font
- Alignment (left/center/right, top/center/bottom)

2.2.4 Working with Maps

2.2.4.1 Selecting Lines and Notes on the Map

VSP imports DXF and SHP files and turns the objects into polylines or a series of connected points. Lines and note objects on the map can be selected by clicking on them with the mouse. When an object is selected, a small black box appears at each vertex, or point, on the object. Polylines appear as a series of vertices. Use the Ctrl key to toggle the selected status of a single line of note object.

Several lines may be selected by using a rectangular area. To do this, position the mouse at one corner of the rectangle then press and hold the left mouse button down while moving the mouse to the opposite corner of the rectangle. When the mouse button is released, all the lines and notes that pass through the rectangle will be selected. Use the Ctrl key to keep previously selected lines and note objects.

Move Selected Objects. Use this dialog to move the selected objects by the given offsets. Objects may also be moved by using the mouse. Position the mouse over a selected line or note object. Press and hold the left mouse button while moving the mouse to the new position. When the mouse button is released, the selected lines will be placed at the new position. This command applies to both map lines and annotation objects.

Rotate Selected Objects. After selecting this command from the menu, enter the pivot point by clicking with the mouse or entering the coordinates on the keyboard. Then enter the angle of rotation by moving the mouse and clicking or entering the degrees on the keyboard. This command affects all selected objects on the map.

Delete Selected Objects. This command deletes selected objects from the map. This action cannot be undone.

Insert Point. After selecting this command from the menu, click on the map with the mouse. A new vertex point is inserted in the nearest polyline on the map. If the polyline matches a sample area, then a matching vertex is also inserted into the sample area. After inserting a point into a polyline, it can be dragged with the mouse to a new location. See VSP's **Help > Help Topics > Map Menu > Insert Point >** <u>selecting and moving points and segments</u> for more information.

Deleting Segments of a Map. If you want to remove a segment from either an imported map or a user-drawn map, you may click on a segment and hit the Delete key on your keyboard. Right-clicking on any segment in a map displays the vertices of the polyline in a outline of bold squares. With the outline in bold squares displayed, hit the Delete key on your keyboard and that segment is removed.

Map Buttons on the Toolbar. The **Zoom In**, **Zoom Out**, and **Zoom Window**, **Zoom Max, and Pan** buttons in the middle of the VSP toolbar (and as pulldown items under Main Menu option **View**) provide methods to focus in on a Sample Area or other region of a site map. Press once on the **Zoom In** button and then click on the site map to make it grow larger. Turn off this mode by pressing the **Zoom In** button again. The **Zoom Out** button works the same way except that it makes the site map shrink. The location on the site map where you click determines the area of the new focus.

The **Zoom Window** button allows you to create an expanded rectangular window into the site map. For an example, press the **Zoom Window** button, drag the cursor across part of the screen, and release. The dashed lines illustrate the final window focus.

The **Zoom Max** button displays the map at the largest size that will fit the current view. VSP uses the current map extents to determine how the map will be positioned. Use **Map / Set Map Extents** to adjust the minimum / maximum x and y coordinates to use for this operation.

The **Pan** button repositions the map in the view window. Hold the left mouse button while dragging the map to a new location. When the left mouse button is released, the map will be redrawn at the new position.

2.2.5 Additional Map Features

The remaining pull-down menu items under **MAP** are specialty topics discussed in other sections of this manual. They will be briefly defined here.

Sample Points imports and exports sample points to text files. It is discussed in Section 2.4. **Export** exports a map and/or samples to various file formats. **Meandering Swaths** draws or imports meandering swaths from an ASCII text file to VSP. To draw meandering swaths on the map you must first enter the width of the swaths in the dialog that appears. Meandering swaths will only be added inside of existing sample areas. If you draw outside sample areas, the swaths will be clipped at the edge of the sample areas. This command works similar to the **Draw Line** command. Right-click with the mouse or re-select this command to stop drawing. This command is useful for creating swaths to be analyzed with **Sampling Goals > Find UXO Target Areas > Post-survey target detection evaluation**.

Figure 2.7. Background Picture VSPEx1 (JPEG Image) With Label Added

Map> Background Picture > Load from file loads a background picture from a graphics file into VSP. VSP comes with two sample pictures: VSPEx1 and VSPEx2. Once a picture file is loaded into VSP, sample areas can be located on the picture similar to how they are located on a map. **Map > Background Picture > Calibrate with Map** matches the background picture to the sampling map. **Map >Background Picture > Load World Map** can be used if a Picture World File is available for the background picture. VSP's **Help > Help Topics > Map Menu > Background Picture > Calibrate with Map** describes this process in detail. Figure 2.7 shows the background picture VSPEx1 loaded into VSP with a yellow sample area labeled "Sample area 1" placed on the picture.

Crop Map removes portions of a map that are not currently visible on the view window. It is useful for removing large amounts of extraneous map lines that tend to slow down the display and other map functions.

2.3 Sample Areas in VSP

2.3.1 Creating a Sample Area

Once a map is created, a Sample Area must be created. A Sample Area is a region in which to locate samples. While most sample areas are enclosed, one of the new Sampling Goals in VSP Version 3.0, Establish Boundary of Contamination, allows for an open type of sample area -- samples are located along a boundary. The user must identify the area to VSP in order to make sampling locations available. (Note: You can use any of the sampling designs except Judgement Sampling without a Sample Area defined, but they will not create sampling locations, only sample sizes.)

2.3.1.1 Define New (Closed) Sample Area

Press the **New Area** button on the VSP toolbar (or from the Main Menu select **Edit > Sample Areas > Define New Sample Area**). A **Color** dialog box appears. Use this dialog to choose the color of the Sample Area. After the color is selected, a tooltip box appears on the map to provide information on the selection method. Figure 2.8 shows a red Sample Area along with the dialog boxes for creating it. Repeat the operation to create a second Sample Area.

There are two basic ways in which to create the Sample Area:

Figure 2.8a. Map with a Single Sample Area

1. **One-Step Method**. Position the cursor inside one of the enclosed areas on the map and right-click with the mouse. The Sample Area is created, and a dialog box appears. This dialog box shows the size of the Sample Area and allows you to change the units of the map. Click the **OK** button on the dialog when done.
2. **Corner-Selection Method**. Position the cursor on each corner of the Sample Area and left-click with the mouse. If you hold down the Shift key while clicking, the nearest point on the map will be selected. If you make a mistake in choosing a corner, use the **Undo** feature. When you have finished defining the Sample Area, either click the **Finish Area** button on the VSP toolbar, select Main Menu option **Edit > Sample Areas > Finish New Sample Area**, or right-click the last segment in the corner selection method. The area dialog box appears, allowing you to change the map units. Note: A Sample Area cannot cross over itself. If this happens, an error message—"This area is invalid and will be removed"—appears.

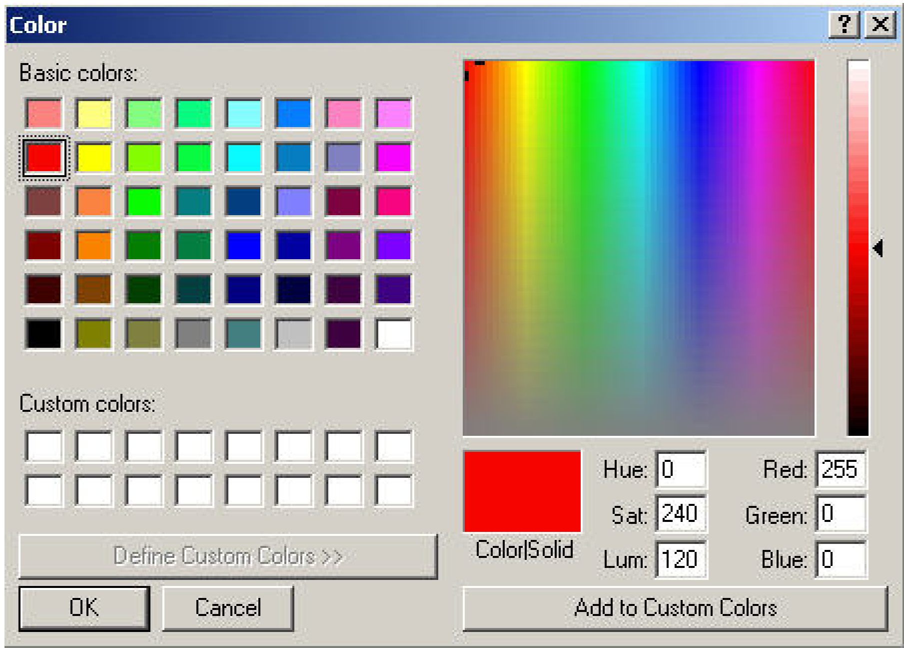


Figure 2.8b. (contd)

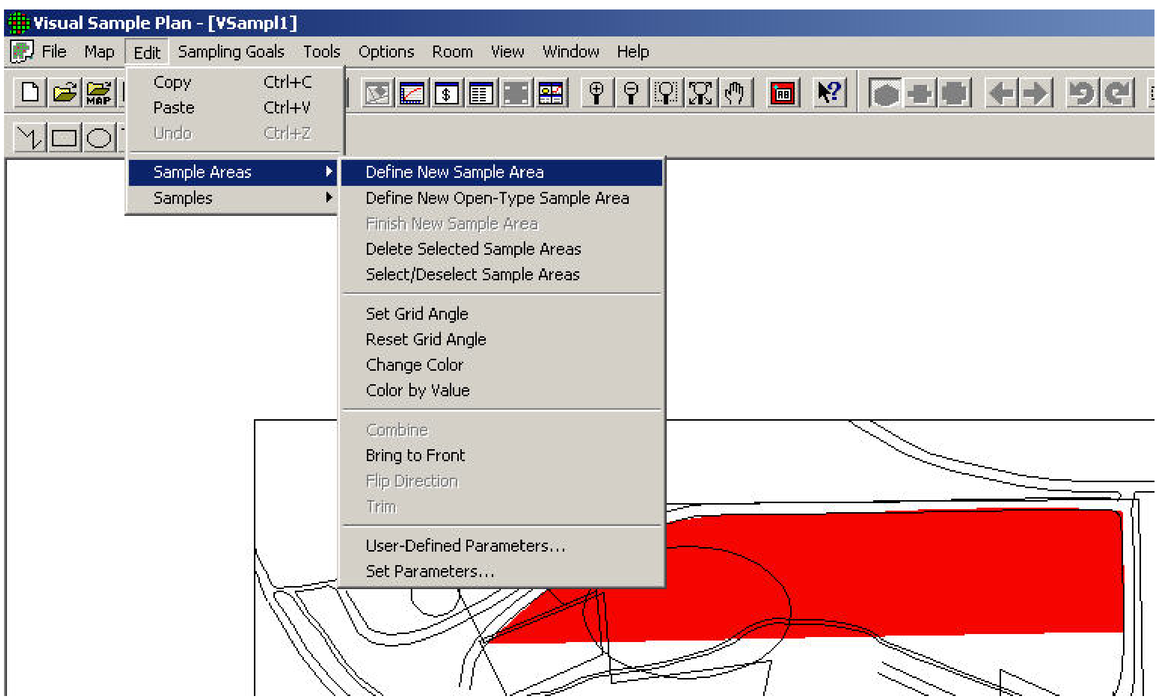


Figure 2.8c. (contd)

A map may contain a single Sample Area or multiple Sample Areas. For example, OneAcre.VSP (an example of a VSP file included with the program) is a single Sample Area, while Example1.VSP could have multiple Sample Areas because the map consists of several enclosed areas that could be selected as Sample Areas. When multiple Sample Areas are selected, samples located on the map by VSP are distributed across all the areas. When multiple samples area are combined using the VSP **Combine Areas** toolbar button or the **Edit > Sample Areas > Combine** menu selection, the combined area is treated as a single area (see Section 2.3.5 for a discussion of combining sample areas).

2.3.1.2 Define New Open-Type Sample Area

VSP provides sample design support for boundaries that do not completely surround a sample area. For instance, suppose it is reasonable to assume that the only portion of the boundary that could be breached by soil contamination is along the downhill side of the Sample Area. In that situation, the VSP user first clicks **Edit > Sample Areas > Define** New Open-Type Sample Area. Then the user places the cursor at the starting location of the desired partial boundary and clicks on each vertex along the boundary line until the end of the boundary of interest is reached. Then a click of the right mouse button finishes the creation of this partial boundary. An example of a partial boundary is shown as a red line in Figure 2.9.

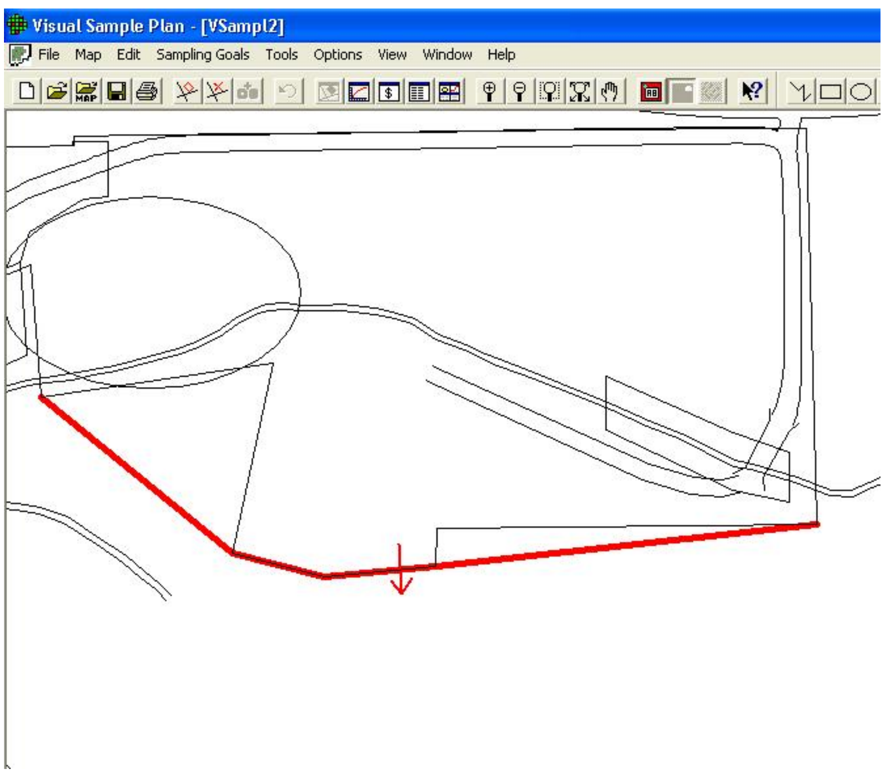


Figure 2.9. Example of an Open Boundary with an Arrow to Show the Direction the Soil Contamination would be Expected to Move (note that the arrow points toward the "clean" side)

Alternatively, the user may create an open-type sample area (partial boundary) using the single-click method. This is accomplished by selecting **Edit > Sample Areas > Define New Open-Type Sample Area** from the menu and then right clicking on an existing line on the map. The boundary may be shortened by selecting the **Edit > Sample Areas > Trim** from the menu and then clicking on two points on the boundary.

During the boundary selection process, VSP places an arrow on the boundary. This arrow points in the direction that contamination in soil may be expected to move, if such movement has or were to take place. If VSP points the arrow in the wrong direction, the direction of the arrow can be reversed by clicking **Edit > Sample Areas > Flip Direction**.

2.3.2 Selecting or Deselecting Sample Areas

VSP allows the user to control which Sample Areas are available for locating samples. Creating a Sample Area automatically "selects it" for locating samples. You know it is "selected" because it appears in a solid color on the map. "Deselected" Sample Areas appear with only the outline of the Sample Area in color and the interior blanked out. You may Select or Deselect a Sample Area in three ways: 1) left click within the Sample Area, 2) right-click on a sample area and change the **Selected** checkbox on the **Sample Area Information** dialog box, or 3) from the Main Menu select **Edit> Sample Areas > Select/Deselect Sample Areas**. The latter method brings up a dialog box that allows you to choose which areas to select or deselect. Figure 2.10 shows a VSP map with three areas selected and one area deselected, and the dialog box where the selections are made. Note that VSP automatically names the Sample Areas: Area 1, Area 2, Area 3, and Area 4 according to the sequence in which the areas were created. The names can be changed in the **Sample Area Information** dialog box discussed in Section 2.3.4.

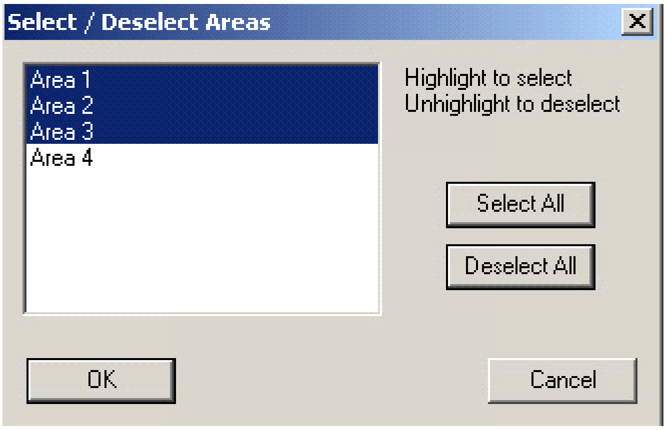


Figure 2.10a. Map with Multiple Sample Areas Selected

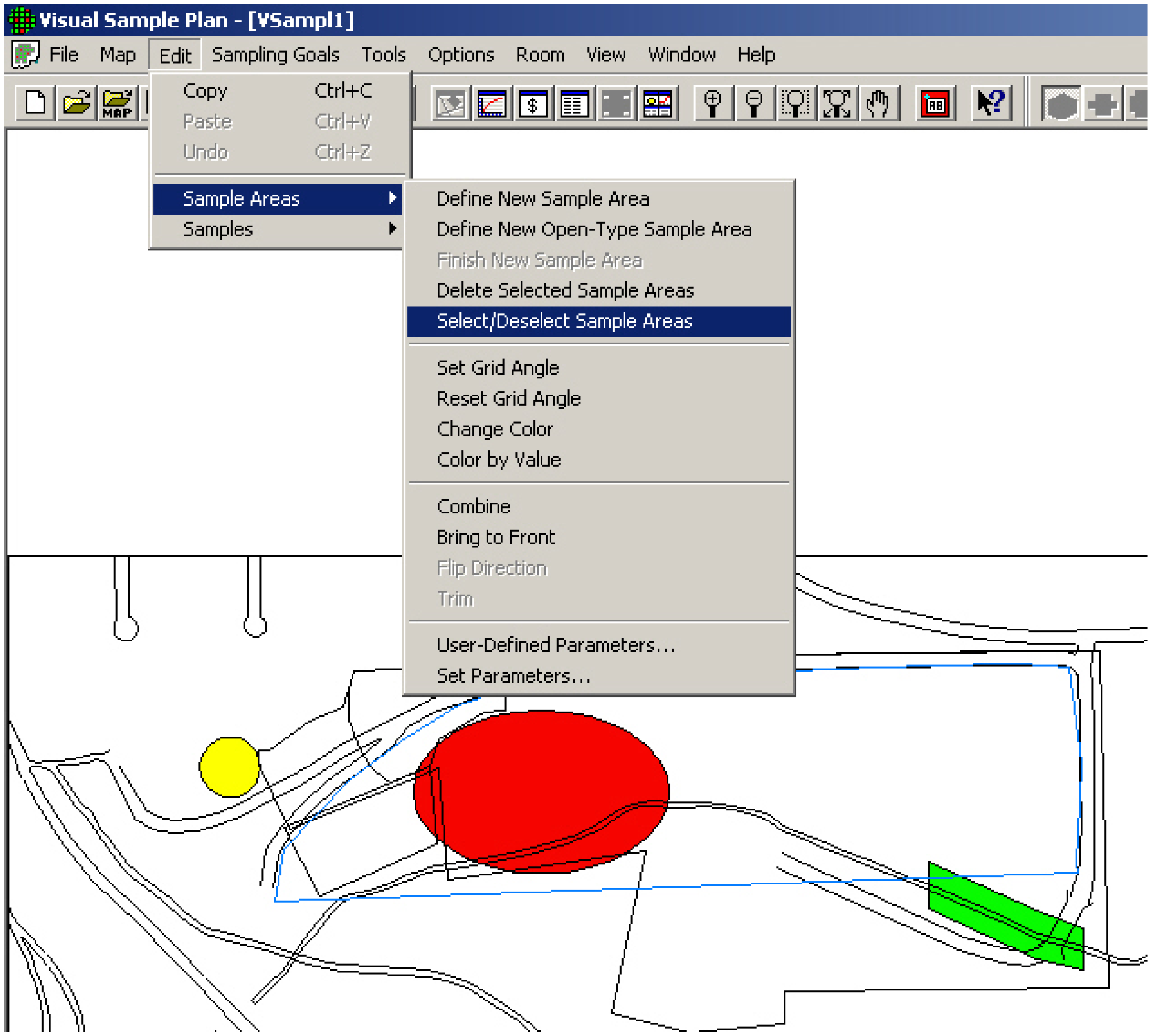


Figure 2.10b. (contd)

2.3.3 Deleting Selected Sample Areas

If you make a mistake, or just want to delete one or more of the Sample Areas you created, you must first make sure the Sample Area(s) is Selected (see above). Then, from the Main Menu, choose **Edit > Sample Areas > Delete Selected Sample Areas**. Be sure to deselect any sample areas that you want to save.

2.3.4 Sample Area Parameters

VSP automatically generates certain parameters for Sample Areas, such as the name, area and perimeter. This information can be accessed by right-clicking on the sample area on the map. The **Sample Area Information** dialog box for Area 3, the big red ellipse, is shown in Figure 2.11.

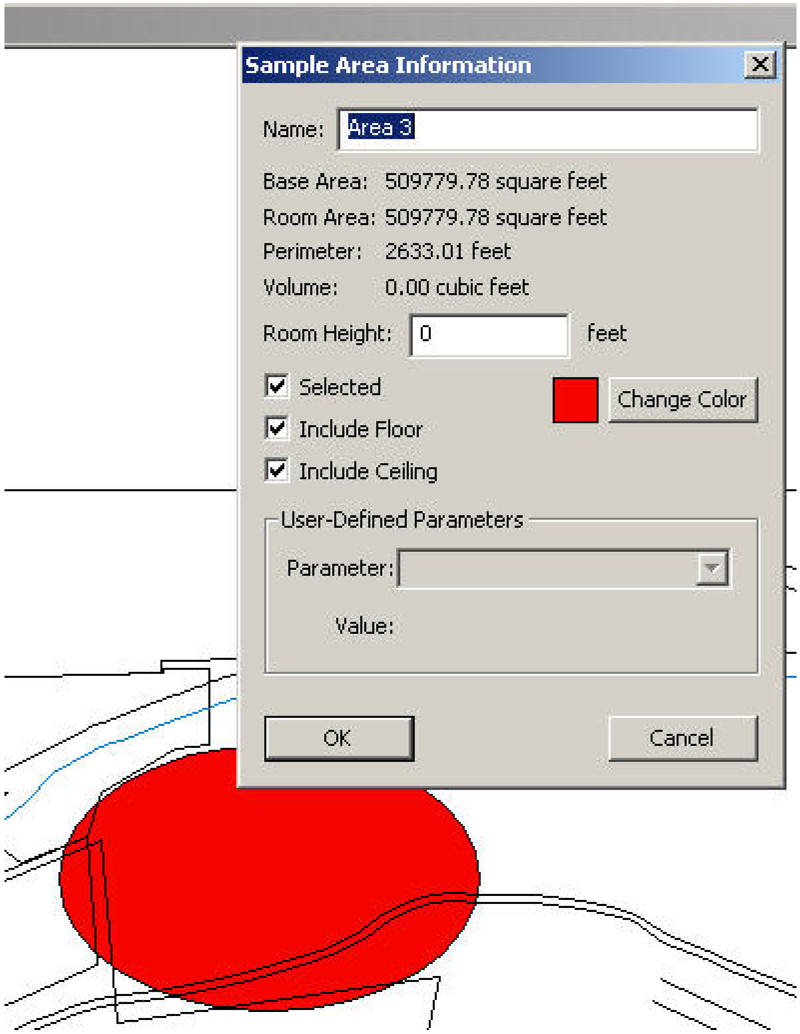


Figure 2.11. Sample Information Dialog Box for a Sample Area

Some parameters such as Name and Selected status can be changed in this dialog box. You will note parameters that refer to rooms in this dialog box. These will be discussed in Section 2.5. Briefly, Rooms are just Sample Areas with height greater than 0, so the same dialog box is used for both Sample Areas and Rooms.

VSP allows the user to define parameters for Sample Areas. These are called User-Defined Parameters. To create User-Defined Parameters, from the Main Menu select **Edit > Sample Areas > User-Defined Parameters**. A dialog box as shown in Figure 2.12 is displayed.

Press the **Insert New** button, and default values appear in the windows. Say you want to define a new parameter for Sample Areas and name it "Regulatory Status". Type "Regulatory Status" in the **Name** box. Say you select "Integer" for **Type**, and check the **List** option. **List** lets you limit the values assigned to Regulatory Status to those you supply. Figure 2.13 displays dialog box.

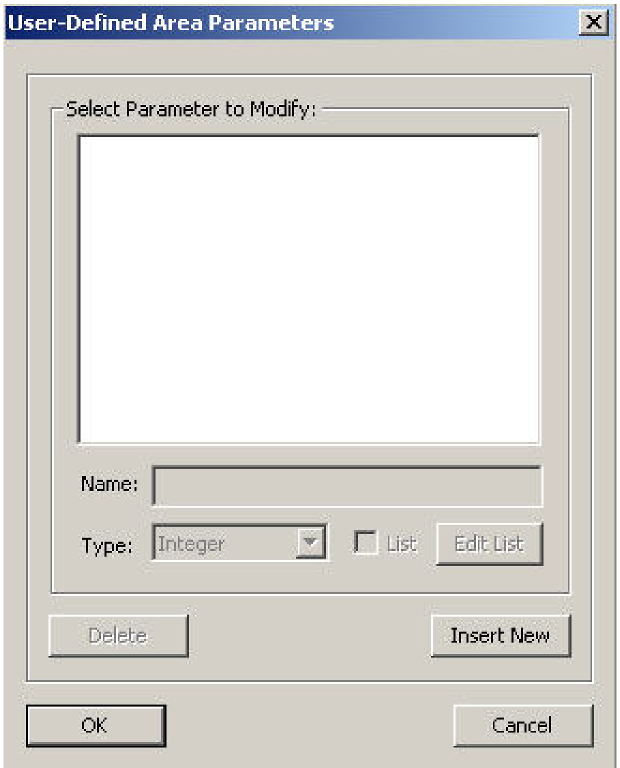


Figure 2.12. User-Defined Area Parameters Dialog Box

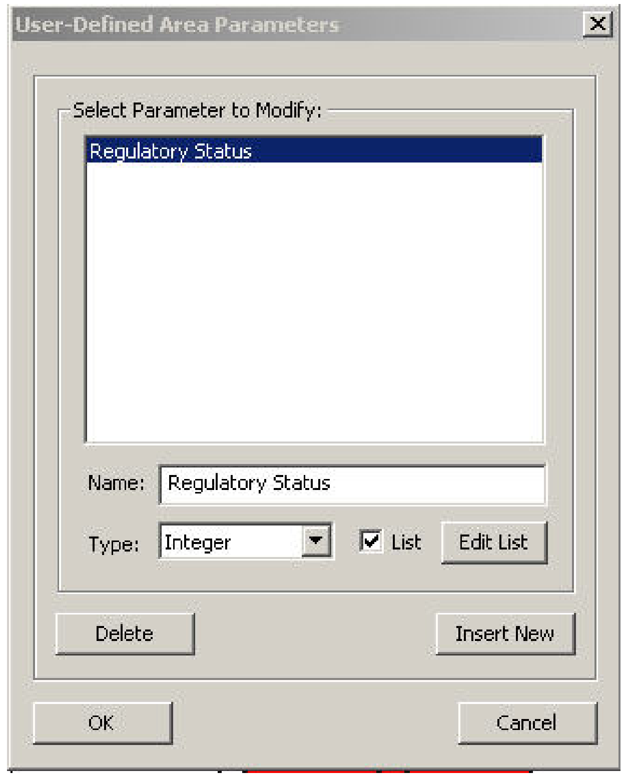


Figure 2.13. User-Defined Area Parameters Dialog Box with Edit List

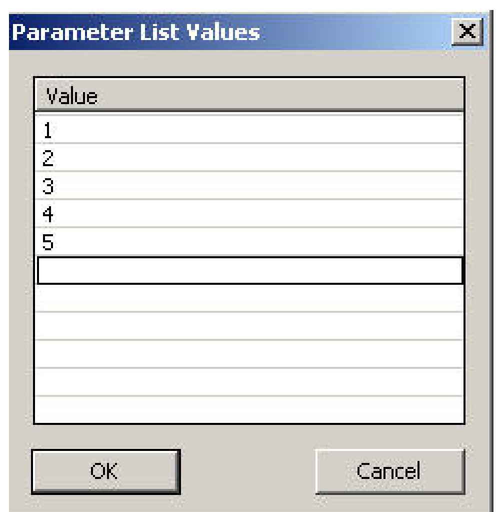


Figure 2.12. Parameter List Values Dialog Box

With a parameter highlighted in the Select Parameter to Modify window, hit the **Edit List** button. A new dialog box titled **Parameter List Values** comes up (see Figure 2.14). The user inputs values to this list by putting the mouse on an current value, hitting return, and typing the next value into the list. The list contains the valid arguments for that parameter.

Once a new parameter is defined, that parameter is attached to (or defined for) all Sample Areas. You may want some Sample Areas to have one value for the parameter, and other Sample Areas have another value. Parameter values may be set in the **Sample Area Information** dialog or by the **Edit > Sample Areas > Set Parameters** menu command. Shown in Figure 2.15, the **Set Parameters** dialog box allows the user to assign parameter values to Sample Areas based on a condition. The example shown in Figure 2.15 says to set the parameter Regulatory Status to the value 4 for Sample Areas that have a Base Area (one of the VSP-defined Sample Area parameters) greater than or equal to 100 square feet. The Set Parameters dialog box has many pull-down lists, making it easy for the user to quickly set parameter values for Sample Areas.

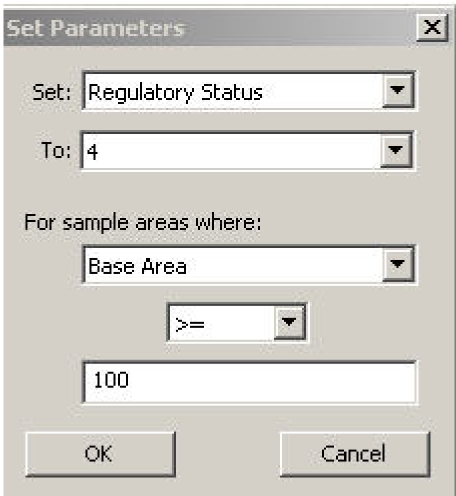


Figure 2.13. User-Defined Area Parameters Dialog Box with Edit List

2.3.5 Extended Sample Area Topics

There are several other features that deal with Sample Areas:

Set/Reset Grid Angles. Use this command to align gridded samples for selected sample areas. Left-Click with the mouse on one vertex of a sample area, then Left-Click on an adjacent vertex of a sample area. (Hold the shift key while clicking to select the exact point on the map.) Use the Reset Grid Angle command to change the grid angle back to its default setting.

Change Color. Use this command to change the color of all selected sample areas. The chosen color will also be the default color for subsequent new sample areas.

Combine. This command combines sample areas or creates interior holes in sample areas. To combine sample areas, select 2 or more non-overlapping sample areas on the map and use this command. The sample areas are combined so that they are treated as a single sample area. Clicking on one part of the sample area will select or deselect all the attached parts of the sample area. All the attached parts are treated as a single sample area for the purpose of sample placement. See **Help > Help Topics > Edit menu > Sample Areas > Combine** for an example.

Bring to Front. Use this command to bring the selected sample areas to the end of the draw list causing them to be drawn last. This will, in effect, cause them to appear in front of other sample areas. Use this command for sample areas that exist inside the hole of another sample area and cannot be seen. Note that this command changes the sample area numbers.

Flip Direction. This command switches the contaminated / uncontaminated side of the open-type sample area. The arrow points toward the uncontaminated side.

Trim. This command allows you to shorten the open-type sample area. After selecting this command, the cursor becomes a cross-hair. Use the mouse to click on two points on the open-type area. After the second point is selected, the area will be truncated at the two points.

2.4 Individual Samples (Importing, Exporting, Removing, and Labeling Them as Historical)

Individual samples have several attributes within VSP:

- location (x, y, z coordinates) and local coordinates (lx, ly)
- type (sampling design used to collect them)
- label (descriptive text)
- value (numerical value)
- Shape (marker symbol)
- historical sample indicator (true/false indicator).

Some of these attributes are relevant for only certain functions within VSP and are explained in future sections.

The primary way you will locate samples within a Sample Area is by pressing the **Apply** button from one of the dialogs once a **Sampling Goal** is selected from the Main Menu. This process is described in Section 3. Samples located in this way are automatically assigned Location, Type, and Shape. Samples that are imported and samples that are located manually do not have the same status as those located by VSP using a statistical approach. Imported samples and manually located samples must be assigned attributes by the user.

Sample attributes can be displayed using the **Sample Information** dialog box. With the map displayed, right-click on an individual sample. A Sample Information dialog box appears that displays current sample information. Information such as Label and Value can be assigned using this dialog box. In Figure 2.16, we see the VSP file Example 2, after right-clicking the right-most sample in the third row up from the bottom. We assigned that sample a Label of "A-24", and a value of "6.1". The fact that the **Historical** box is not checked means this sample originated from VSP by Applying one of the VSP Sampling Goals (rather being imported into VSP as part of an earlier sampling effort).

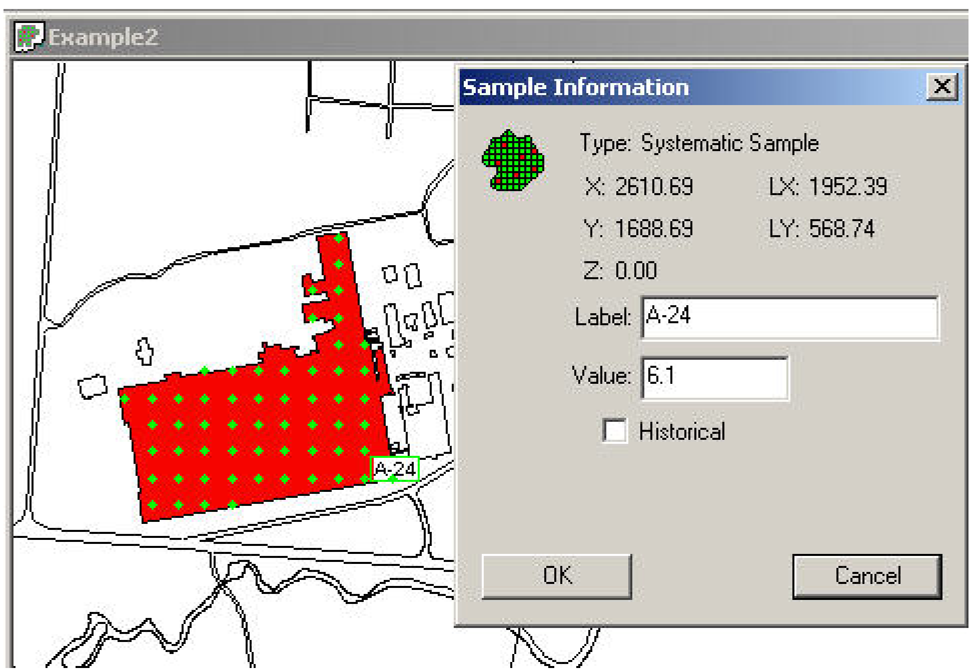


Figure 2.14. Sample Information Dialog Box for a sample in Example2.VSP

2.4.1 Importing Samples

There are two ways to import sampling locations:

1. Copy them from the Windows Clipboard. Edit the coordinates in a text editor, a word processor, or a spreadsheet. Each line (or row) represents a different sampling location. The first column is the x

coordinate; the second column is the y coordinate. The third column is the sample Type and is optional. Valid sample Types are Random, Systematic, Hotspot, Manual, Adaptive-Fill, or Unknown. The fourth column is the sample label and is optional. Spaces or tabs should separate columns. (Tabs are preferable.) The coordinates must lie inside a selected Sample Area.

Example: Type the following coordinates into a text editor such as Notepad:

10	10	Random
50	10	Systematic
10	50	Hotspot
95	60	Manual
99	99	Adaptive-Fill
150	150	Unknown

Now press Ctrl-A to select all the text and Ctrl-C to copy the text to the Windows Clipboard. Run VSP and load OneAcre.Vsp. Select the Main Menu option **View** > **Coordinates**. Paste the coordinates into VSP using either **Ctrl-V** or Main Menu option **Edit** > **Paste**. View the new sampling locations using Main Menu option **View > Map or Window** > **Quad Window**. Your map view should now look like Figure 2.17.

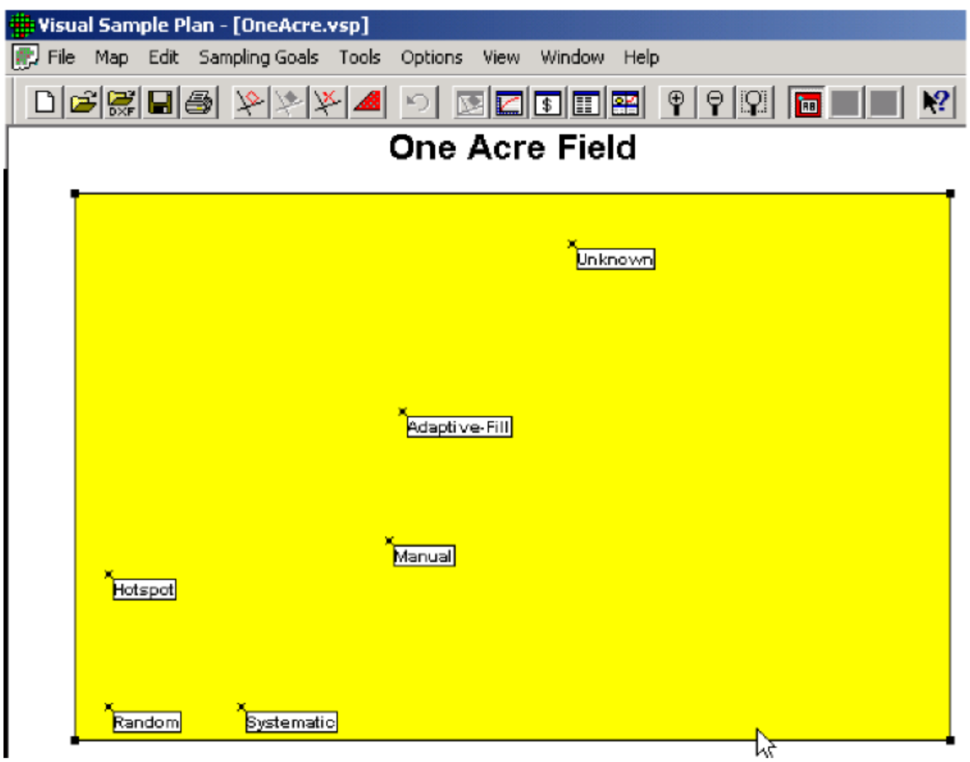


Figure 2.15. The OneAcre.VSP Project with Sampling Locations Added from Windows Clipboard

You can place the mouse on any sample point and right-click to see the attributes of the sample at that sample Location. Figure 2.18 shows the Sample Information VSP has for the sample near the arrow.

2. Import sampling locations from a text file. The text file must be formatted as described above. Choose Main Menu option **Map > Sample Points > Import** and enter the file name in the dialog box.

Samples that are imported are assigned Shapes depending on the Type attribute assigned. Sample Type can be edited by selecting **Edit > Samples > Shapes** from the Main Menu. The Dialog box that appears shows both the shapes assigned to valid Types (use the pull-down menu to select among valid Types), and gives a picture of the Shape. Figure 2.18 shows that when a sample was collected according to a **Random** design, it will be displayed with a **Small Cross** within a **Circle** Shape.

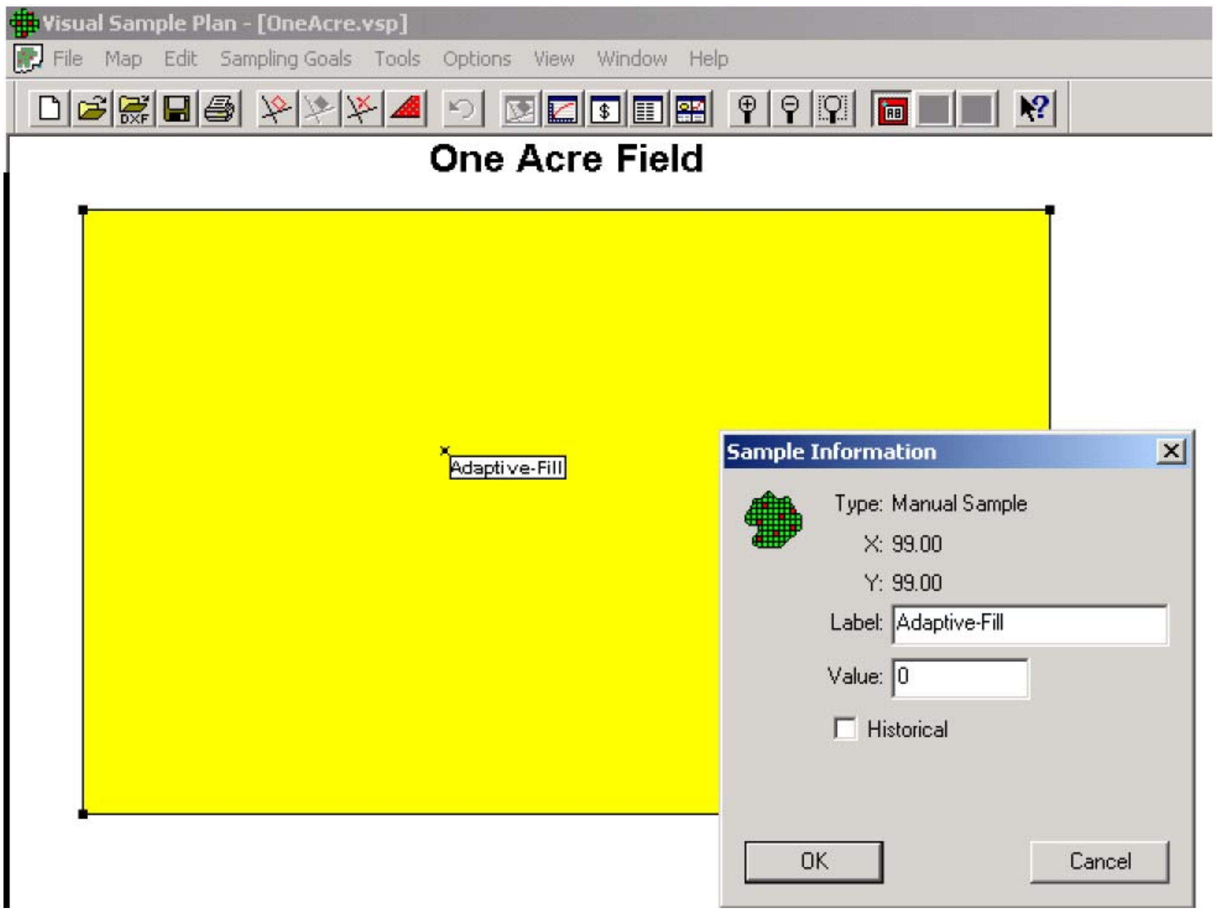


Figure 2.16. Example of Sample Information Box

2.4.2 Historical Samples

Sample locations with the Historical box checked (see Figure 2.18) have a unique role in VSP. VSP gives you "credit" for them in accounting for the total number of new samples needed. This is explained in Section 3.2. The important point to remember here is that if you import samples, manually add samples, or have a sampling design previously created within VSP, you can give specific samples a “Historical” status by placing your mouse over the sample location while in Map View and, in the dialog that comes up, checking the Historical box.

2.4.3 Exporting Sampling Locations

To export sampling locations to a text file (for example, to use the coordinates in a ground penetrating radar system),

1. Select the Sample Area as described above and develop the sampling design as described in Section 3.
2. Choose Main Menu option **Map** > **Sample Points** > **Export**. Provide a name for the text file and click Save.

2.4.4 Removing Sampling Locations

This option is best explained with an example:

1. Start VSP and open a new project using Main Menu option **File** > **New Project**
2. Open the Millsite.dxf file using Main Menu option **Map** > **Load Map from File.** ********formatting problems**********
3. Click the **New Area** button on the toolbar and, after choosing a color, select the large ellipse by rightclicking inside the oval. If you accidentally get some other area, click the **Remove Areas** button and start over. Place the cursor as far from other objects as possible but still inside the ellipse.

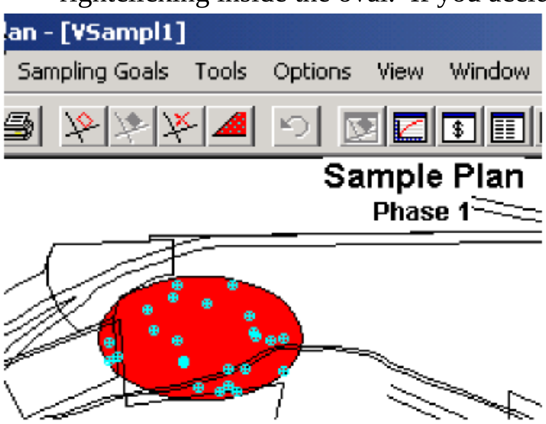


Figure 2.17. Example Sample Area with Sampling Locations

4. Choose the Main Menu; select **Sampling Goals** > **Compare Average to Fixed Threshold** > **Data not required to be normally distributed** > **Simple Random Sampling (Wilcoxon signed ranks test)**. Click the Apply button to place samples in the Sample Area. You should now have a Sample Area with 24 sampling locations similar to that shown in Figure 2.19.
5. Using the Main Menu option **Map** > **Sample Points** > **Export**, save all the sampling locations to a text file named ***Points.txt***.

6. Now we are ready to remove some of the sampling locations. First, delete the first 16 rows (sampling locations) from file ***Points.txt*** using a text editor like Notepad. Save the remaining 5 rows to a new file named ***Remove.txt***. These are the locations that will be *removed* from the Sample Area.
7. Finally, to remove the sampling locations listed in Remove.txt from the Sample Area, choose Main Menu option **Map > Sample Points > Remove**. Select the file ***Remove.txt*** and click the **Open** button. You will see in Figure 2.20 there now are only 16 sample points instead of the original 24 shown in Figure 2.19.

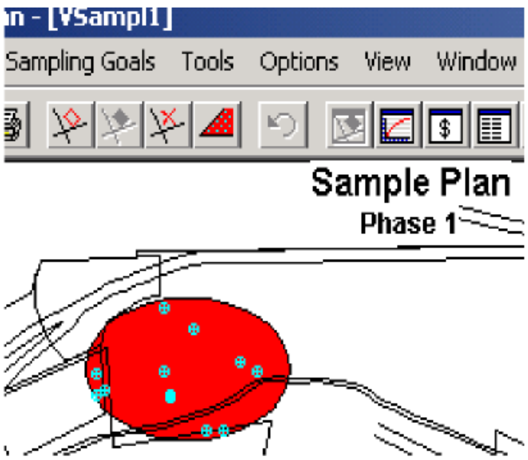


Figure 2.18. Example Study Area after Sampling

In other words, the coordinates in the ***Remove.txt*** file are the sampling locations that are deleted from the Sample Area. Just one location or all the locations can be removed.

2.5 Rooms and Buildings in VSP

One of the major enhancements in VSP version 3.0 is the ability to draw and apply sampling designs to rooms and hence, buildings. Rooms have height, spatial relationships with other rooms (i.e., can be assigned to floors within buildings), and a unique set of objects associated with them (e.g., doors and windows). They also have a unique set of parameters that are VSP-assigned (e.g., volume, floors, ceilings, walls) and user-assigned (e.g., zones, class, HVAC system, release point of threat agent, etc.).

The motivation for expanding the room functionality beyond that which was provided in release VSP 2.0 in the new interest in homeland security; threat assessment for buildings; sampling designs for floors, ceilings, and surfaces; and cleanup and release criteria for contaminated buildings.

At the most basic level, a Room is just an enclosed Sample Area with height greater than 0. As such, many of the VSP functions and screens associated with Sample Areas are the same for Rooms. However, there are also many new VSP functions for Rooms.

2.5.1 Drawing a Room

Start a new project using Main Menu option **File > New Project**. Click the **Draw Room** button on the drawing toolbar or select Main Menu option **Room > Draw**. (Use **View / Room Toolbar** to show the Room Toolbar.) A tooltip box displays the three ways to draw a room using this tool:

- Enter the room dimensions on the keyboard: LxWxH. (e.g., 12x10x8 <enter>)
- Enter the corners of the room on the keyboard: X, Y (e.g., 50, 50 <enter> 90,90 <enter>).

- Left-click the mouse at the upper-left hand corner of the room, and drag the mouse to the lower-right hand corner. Left-click the mouse to finish the room. This is similar to drawing a rectangle, except that VSP automatically sets a wall height of 8 ft. Room height can be changed using Main Menu item **Room > Set Room Height** or by clicking on the **Set Height** button on the drawing toolbar.

With the room displayed, right-click anywhere within the room. The **Sample Area Information** dialog box appears (Figure 2.21).

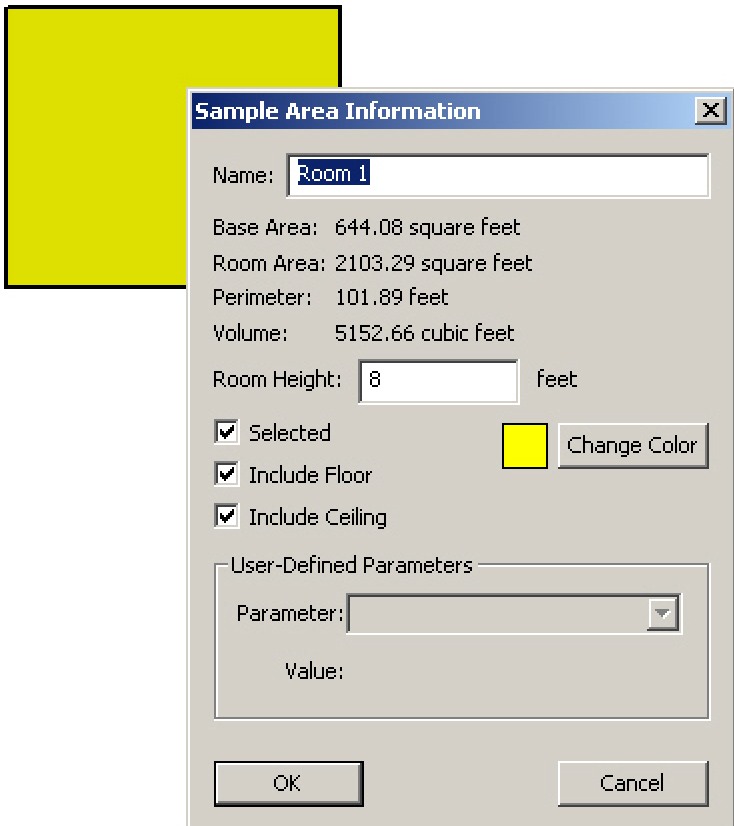


Figure 2.19. Room Information Dialog Box

This dialog box can be used to view parameters of the room such as base and room area, perimeter, and volume. The name of the room is set to “Room n” until changed by the user. The user can set and change other parameters such as room height, color, whether the floor and ceiling are included as part of the room (and hence samples will be applied to the floor and ceiling as well as the room walls). Any User-Defined Parameters set will be displayed in the pull-down list.. User-Defined Parameters for Rooms are set in the same way as User-Defined Parameters for Sample Areas (see Section 2.3.4). Note that you are in Map View when drawing the room.

The room can be modified in the Map View by inserting a point into a wall and then moving the wall section (see Figure 2.22). This is used to create L-shaped rooms, or irregular-shaped rooms.

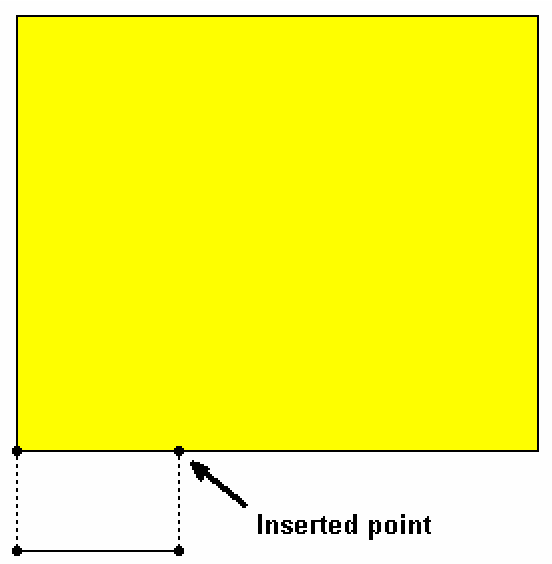


Figure 2.20. Room with Inserted Point

You can set the exact length of a line segment by right-clicking on it. First select the segment (hold the Shift key while clicking between two points on the map), then right-click on the selected segment. A dialog will appear that allows you to enter the exact length of the line segment. If the segment is attached to other segments at right angles, those other segments are moved or adjusted as well.

After the room is drawn (defined), it becomes the current room and can be displayed with Room view (**View** > **Room**). The current room is indicated on the map by a thick black outline and a darker shade. A room can be displayed in one of three view formats:

- Perspective
- Wall Strip
- Splayed

The display view format can be selected using one of the three buttons on the drawing toolbar, or by selecting one of the formats under the Main Menu item **Room** while in Room view. Figure 2.23 shows the three views of a room.

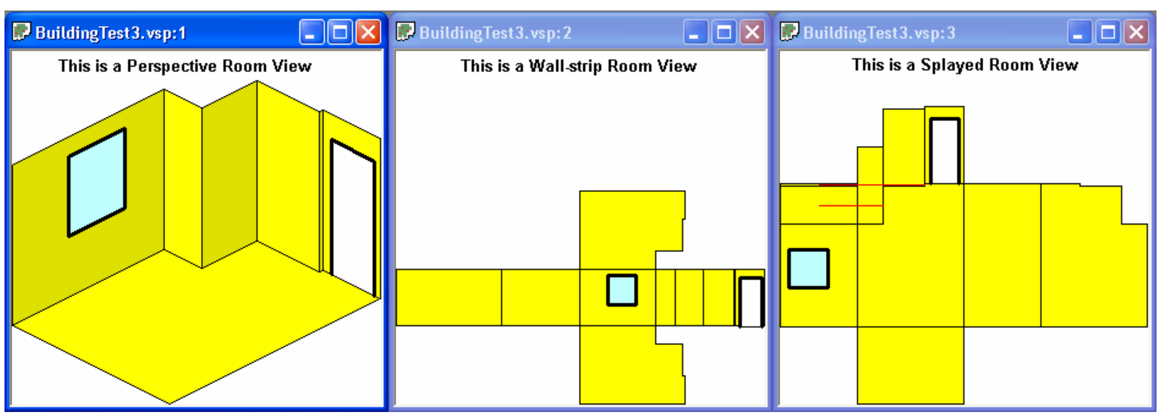


Figure 2.21. Three perspective views of a room

When rooms are defined using pictures or existing maps, the Delineate Rooms (Main Menu item **Room > Delineate Rooms**) mode allows the user to create rooms at right angles inside existing map shapes. Delineate Rooms is an on/off toggle switch. While in this mode (you must be in Map View), you can fill up the space inside an irregular-shaped area with individual rectangles. For more information on this tool, consult **Help > Help Topics (Contents) > Menus> Room> Delineate Rooms**.

Figure 2.24a is an imported CAD drawing of a floor layout with individual rooms delineated using the VSP **Delineate Rooms** tool. Figure 2.24a is the Map View of the room with Room Information displayed (right-click anywhere within the room to display information box). We also used the Main Menu item **Room > Insert Annotation** to put the label "Selected Room" on the map. Figure 2.24b is a Room View of the same room with **Perspective View** selected.

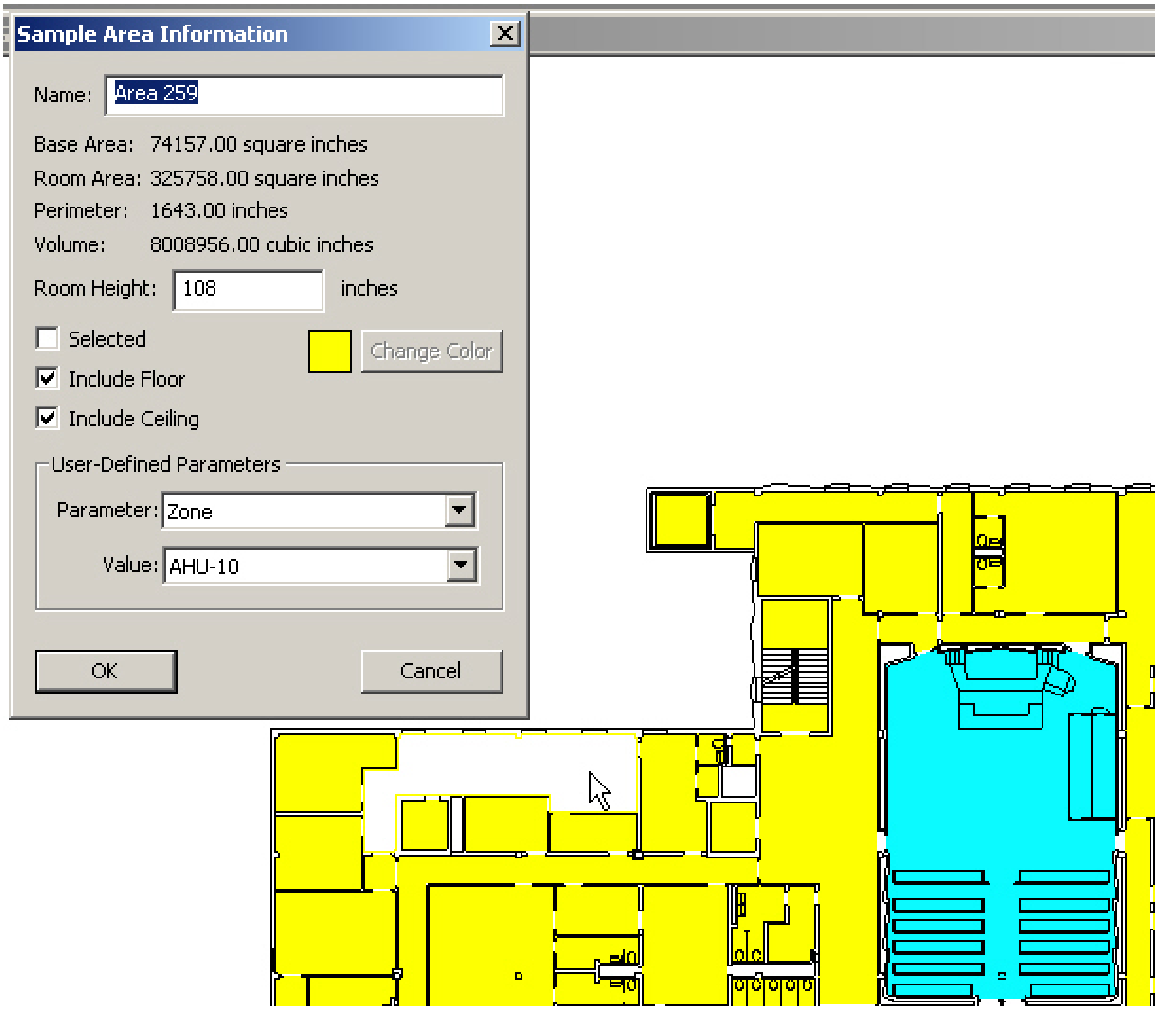


Figure 2.22a. Room in Map View

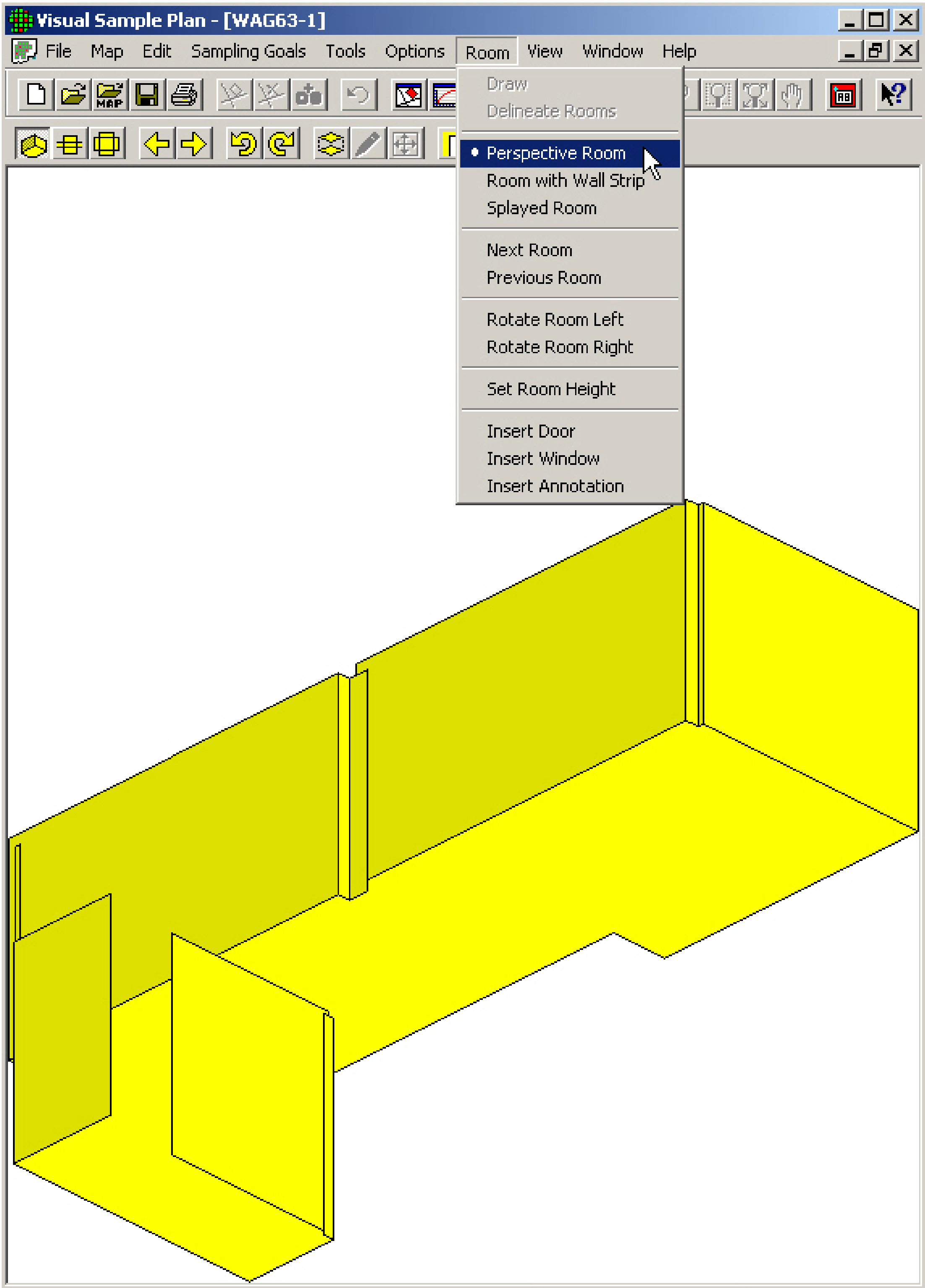


Figure 2.24b. Room in Room View

2.5.2 Extended Room Features

2.5.2.1 Room Objects

VSP Version 3.0 has two objects associated with rooms: doors and windows. There are two ways to view these objects: in Map View (see Figure 2.25) and in Room View (see Figure 2.26). Figure 2.25 is another section of the CAD drawing shown in Figure 2.24. A door in the CAD drawing was defined for VSP. Right-clicking on the Map where the arrow is pointing brings up the **Object Information** dialog box for the Door Object at that location.

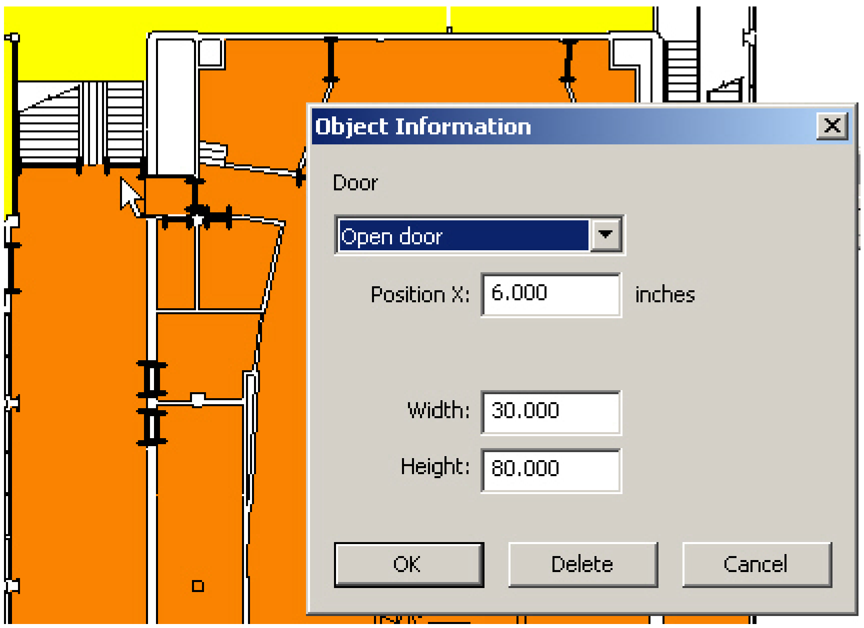


Figure 2.23. Door Object Displayed Using Map View

Doors and windows can be added in Room View using Main Menu item **Room > Insert Door/Window** or by clicking on the door or window button on the Room Toolbar. Once added, the properties of the room objects can be viewed by right-clicking on the object, which brings up the **Object Information** dialog box. This is shown for a door in Figure 2.26 and for a window in Figure 2.27.

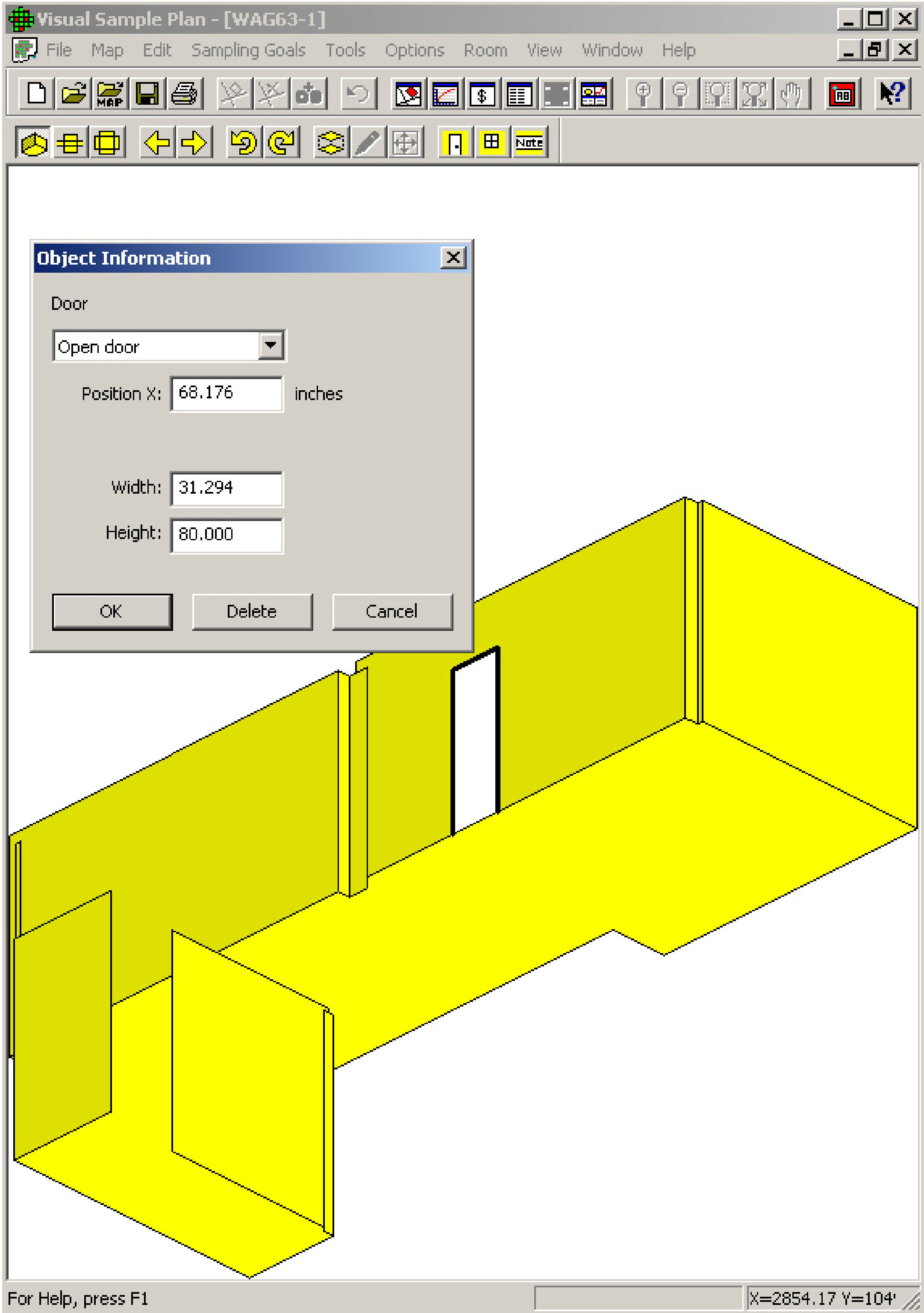


Figure 2.24. Door Room Object with Object Information Dialog Box Displayed

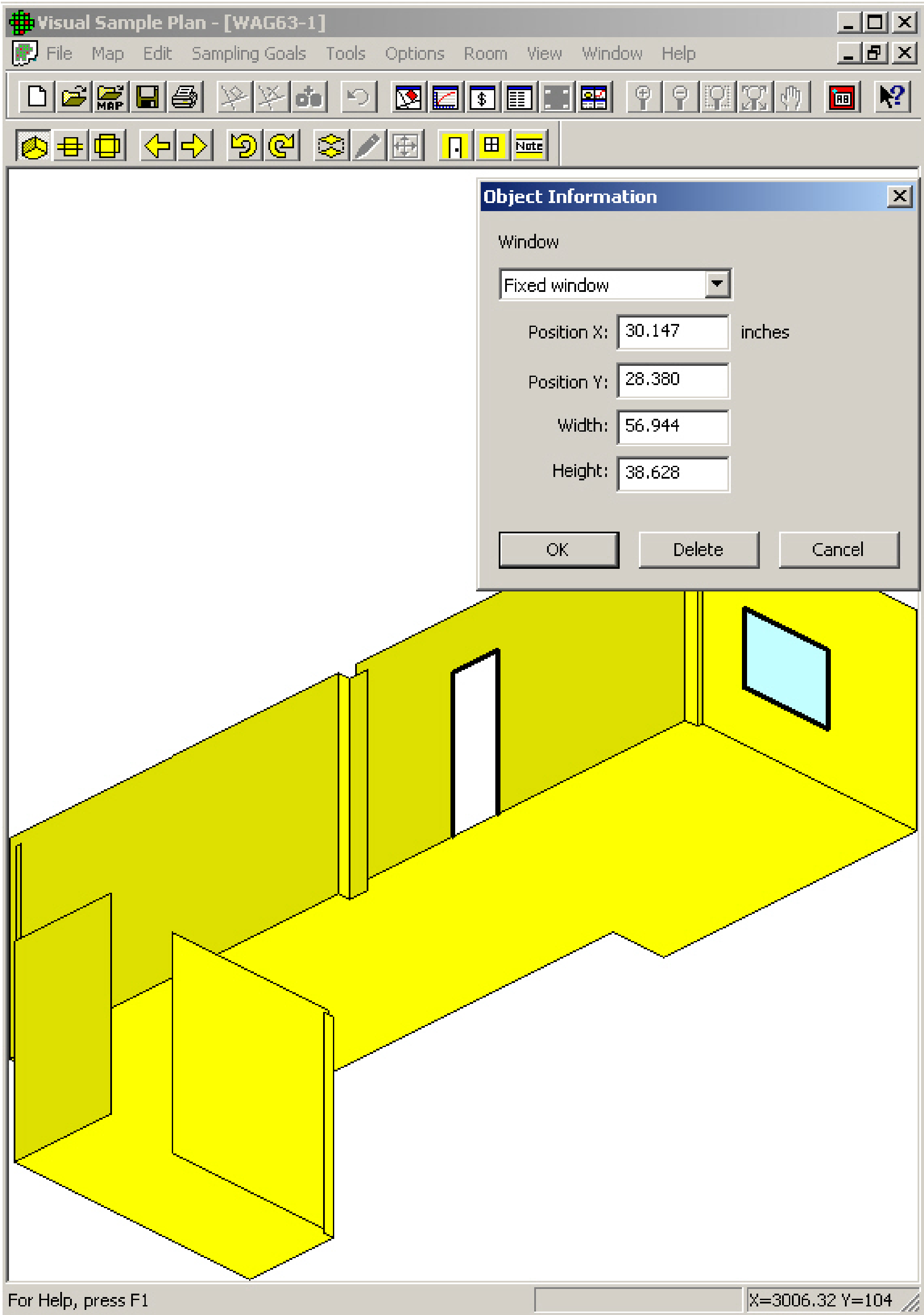


Figure 2.25. Window Room Object with Object Information Dialog Box Displayed

2.5.2.2 Room Parameters

Rooms can have VSP-defined parameters (e.g., Area, Perimeter, Volume, Name) and User-Defined Parameters (e.g., Zone, Class). The User-Defined Parameters are set in the **Sample Area Information** dialog box or with the Main Menu item **Edit > Sample Areas > Set Parameters**. This second option is a very powerful tool. Figure 2.24a shows that the room named "Area 259" is in Zone "AHU-10" (i.e., Air Handling Unit 10). All the rooms on AHU 10 are assigned to the same Zone so they are easily identifiable. In another example, rooms that exceed a release criteria could have a logical (yes/no) parameter assigned called "Contaminated", and coded red/green. Note that once a room parameter is set up using Main Menu item **Edit > Sample Areas > User-defined Parameters**, all rooms are given a default value for that parameter.

2.5.2.3 Room Color by Parameter

Rooms can be colored automatically based on the value of one of the built-in or user-defined parameters. Choosing the menu item **Edit > Sample Areas > Color by Value** displays the dialog shown in Figure 2.28. This dialog allows you to choose one of the predefined gradient or discrete color sets and one of the sample area parameters. Certain parameters can also be colored by the logarithm of the value. This dialog also allows the color by parameter function to be turned off or on.

2.5.2.4 Room Order

Clicking the button for **Next Room** from the Room Toolbar, or selecting **Room > Next Room** from the Main Menu changes the current room to the next selected room on the map. The current room is indicated by a thick black border and is a slightly darker hue in color-coded rooms. **Previous Room** changes the current room to the previously selected room on the map. The order of room selection is the order of creation. Note that the order for room and sample areas can be changed by the menu command **Edit / Sample Areas / Bring to Front**.

2.5.2.5 Room Rotation

Clicking the button for **Rotate Room Right** from the Room Toolbar, or selecting **Room > Rotate Room Right** from the Main Menu rotates the Perspective or Wall-Strip Room View clockwise 90 degrees. **Rotate Room Left** rotates the Perspective or Wall-Strip Room View counter-clockwise 90 degrees.

2.6 Saving a VSP File

No matter how you imported or created a site map or Sample Area for VSP, you can always save the information in VSP's own file format. From the Main Menu, select **File > Save Project** As and provide a name for the project. VSP will add the VSP file extension automatically. Alternatively, you can use the **Save** button with the disk icon on the VSP toolbar. After you have created a sampling design as discussed later in this guide, saving your project as a VSP file also will save the input data, cost data, and recommended sample sizes.

3.0 Sampling Plan Development Within VSP

3.1 Sampling Plan Type Selection

Sampling plan components consist of where to take samples, how many samples to take, what kind of samples (e.g., surface soil, air), and how to take samples and analyze them. We identified the general areas of where to take samples in Section 2.3, **Sample Areas in VSP**. In this section, we discuss where *within the Sampling Area* to locate the samples. We also discuss *how many* samples to take. The kind of samples to take—i.e., soil vs. groundwater, wet vs. dry, surface vs. core,—is determined during Step 3 of the DQO process (Define Inputs) and is not addressed directly in VSP. The Measurement Quality Objectives module in VSP (Section 5.4) deals with how the method selected for analytically measuring the sample relates to other components of the sampling plan.

3.1.1 Defining the Purpose/Goal of Sampling

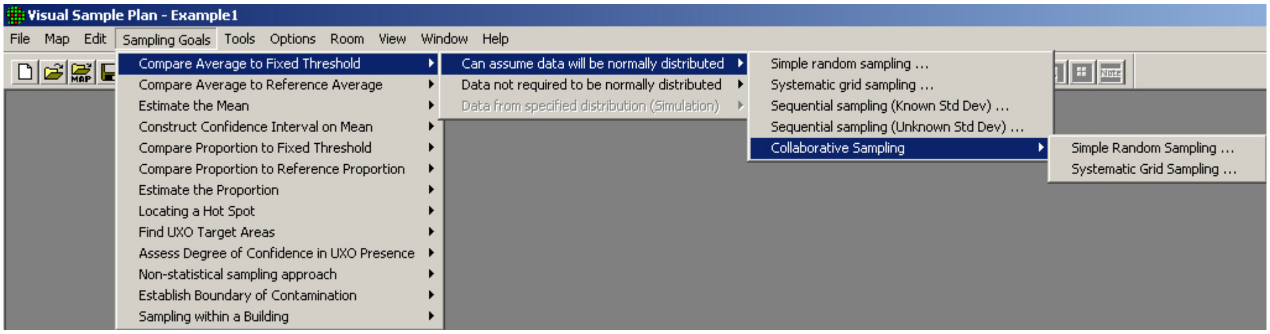
VSP follows the DQO planning process in directing users in the selection of the components of the sampling plan. The first thing you must do is to select the type of problem for which the current data collection effort will be used to resolve. In VSP, we call this the Sampling Goal. The following types of problems are addressed currently in VSP. Future versions will expand on this list:

This list of sampling goals available now in VSP reflects the targeted interests and specific problems of our current VSP sponsors. Therefore, the available sampling designs within VSP are not an exhaustive list of designs you might find in a commercial statistical sampling package. Future versions will work toward a complete set of sampling design offerings.

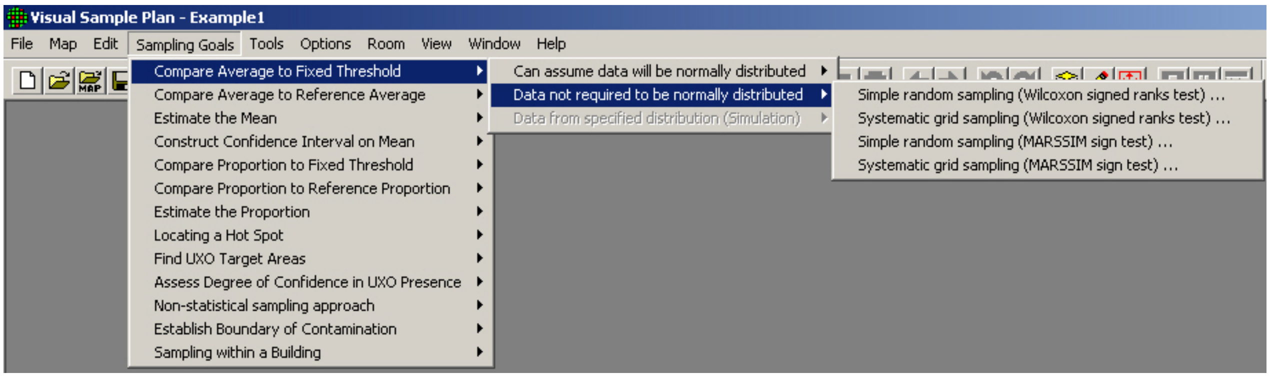
VSP lists “Non-statistical sampling approach” under Sampling Goals, but this is not really a goal. Under this category, VSP allows the user to specify a predetermined sample size and locate the samples judgmentally. Because VSP has no way of knowing how the sample size and sample locations were chosen, the sampling approach is considered to be “non-statistical” (i.e., no confidence can be assigned to the conclusions drawn from judgment samples).

To give you an idea of how VSP threads from Sampling Goal to selection of a sampling design, Figure 3.1 shows the sequence of pull-down menus for one of the goals, **Sampling Goal of Compare Average to a Fixed Threshold**. All endpoints from the Sampling Goal main menu result in a dialog box where the user provides inputs for the specific design selected. VSP allows only certain options and designs (e.g., simple random, systematic) under each goal. This is because VSP contains the algorithms for calculating sample number and locating samples for only certain *goal-assumptions-statistical test or method* sequences. Future versions of VSP will expand on the number and type of algorithms offered.

Sampling Goal	Description
Compare Average to Fixed Threshold	Calculates number of samples needed to compare a sample mean or median against a predetermined threshold and places them on the map. This is called a one-sample problem.
Compare Average to Reference Average	Calculates number of samples needed to compare a sample mean or median against a reference mean or median and places them on the map. This is typically used when a reference area has been selected (i.e., a background area) and the problem is to see if the study area is equal to, or greater than, the reference area. This is called a two-sample problem because the data from two sites are compared to each other.
Estimate the Mean	Calculates number of samples needed to estimate the population mean and places them on the map.
Construct Confidence Interval on Mean	Calculates number of samples needed to find a confidence interval on a mean and places them on the map.
Compare Proportion to Fixed Threshold	Calculates number of samples needed to compare a proportion to a given proportion and places them on the map
Compare Proportion to Reference Proportion	Calculates number of samples needed to compare two proportions and places them on the map.
Estimate the Proportion	Calculates number of samples needed to estimate the population proportion and places them on the map.
Locating a Hot Spot	Use systematic grid sampling to locate a Hot Spot (i.e., small pockets of contamination).
Find UXO Target Areas	Traverse and detect an elliptical target zone using swath sampling. Calculates spacing for swaths. Evaluates post-survey target detection.
Assess Degree of Confidence in UXO Presence	Assess degree of confidence in UXO presence.
Non-statistical sampling approach	Allows samples to be added to the map without the guidance of statistical methods.
Establish boundary of Contamination	Determine whether contamination has migrated across the boundary.
Sampling within a Building (Future VSP release)	Allows sampling within rooms, zones, floors, etc. for various contamination release scenarios and end goals.



(a)



(b)

Figure 3.1. Menu Options in VSP for Comparing an Average to a Fixed Threshold

3.1.2 Selecting a Sampling Design

VSP 3.0 offers several versions of the software (see Figure 2.1). Each version has a unique set of sampling designs available to the user – except **General (all inclusive) VSP** which contains all the designs. Some of the designs available under each of the Sampling Goal menu items are unique to that goal. Other designs are available under multiple goals. Thus, the Sampling Goal you select determines which sampling design(s) will be available to you.

If a user is new to VSP, and is not looking for a specific sample design but rather has a general definition of the problem to be resolved with sample data, a good discussion of how to select a sampling design is in EPA's *Guide for Choosing a Sampling Design for Environmental Data Collection* (EPA 2001) <http://www.epa.gov/quality/qa_docs.html>. See Table 3-1 on pages 23-24 in that source for examples of problem types that one may encounter and suggestions for sampling designs that are relevant for these problem types in particular situations. Another guidance document, *Multi-Agency Radiation Survey and Site Investigation Manual* (MARSSIM) (EPA 1997) <http://www.epa.gov/radiation/marssim/>, also provides insight into how to select a sample design.

One of the valuable ways to use VSP is to run through the various Goals and see what changes from one Goal to another, what sampling designs are available for each Goal, how designs perform, and what assumptions are required for each design. This trial and error approach is probably the best way to select a design that best fits your regulatory environment, unique site conditions, and goals.

An important point to keep in mind is the linkage between 1) the minimum number of samples that must be collected and where they are located, and 2) how you will analyze the sampling results to calculate summary values (on which you will base your decisions). The user must understand this linkage in order to select the appropriate design. Once the samples are collected and analyzed, the statistical tests and methods assumed in the sample size formulas and design must be used in the analysis phase, Data Quality Assessment (DQA).

Some of the designs in VSP contain limited DQA and require sample results to be input into VSP so tests can be executed and conclusions drawn based on the results. Adaptive Sampling is one design where results are input into VSP. There is a Data Analysis tab under the designs for **Compare Average to a Fixed Threshold**.

We cannot discuss all the technical background behind the designs here, but the technical documentation for VSP gives sample size formulas used in VSP and provides references. The VSP web site lists the technical documentation available, and allows download of the documents <http://dqop.pnl.gov/VSP/documentation>. The online help in VSP also provides technical help and references. Finally, the reports that are available within VSP are a good source for definitions, assumptions, sample size formulas, and technical justification for the design selected.

VSP allows both probability-based designs and judgmental sampling:

> Probability-based sampling designs apply sampling theory and involve random selection. An essential feature of a probability-based sample is that each member of the population from which the sample was selected has a known probability of selection. When a probability based design is used, statistical inferences may be made about the sampled population from the data obtained from the sampled units. Judgmental designs involve the selection of sampling units on the basis of expert knowledge or professional judgment (EPA 2001, pp. 9-10).

VSP currently deals only with two-dimensional spatial designs; however, with a little effort and repeated runs, it could be used to select designs in three dimensions or over a time domain. VSP allows you to select a design from the following list. All but the last are probability-based designs:

- simple random sampling – Sampling locations are selected based on random numbers, which are then mapped to the spatial locations.

- systematic grid sampling – Sampling locations are selected on a regular pattern (e.g., on a square grid, on a triangular grid, along a line) with the starting location randomly selected. Sampling is done only at the node points of the grid. The grid pattern is selected in the dialog box that appears once grid sampling is selected. Click the Grid tab on the dialog box to see this screen. Figure 3.2 shows this dialog box.

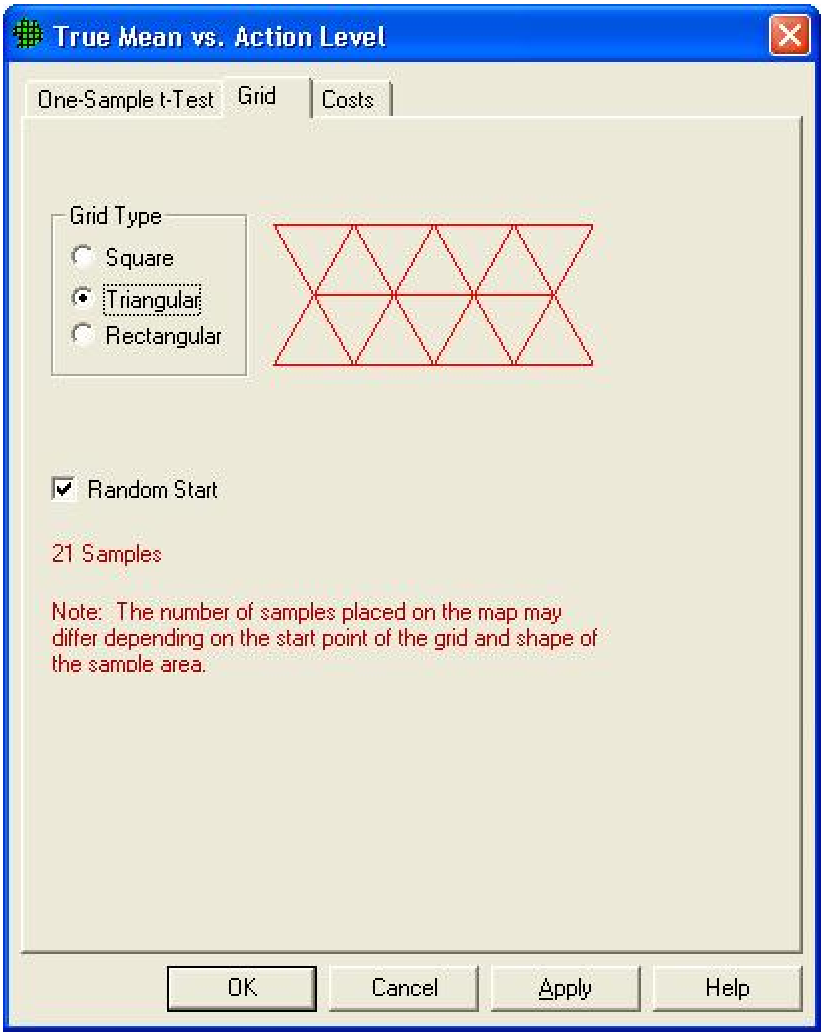


Figure 3.2. Dialog Box for Entering Type of Grid Design

You can see an example of the grid pattern selected in the right-hand side of the dialog box in red. You may specify **Random Start** or a fixed start for the initial grid point using the check box next to **Random Start**. Choosing **Random Start** will generate a new random starting location for the first grid location each time the **Apply** button is pushed. Once all selections have been made, press **Apply**.

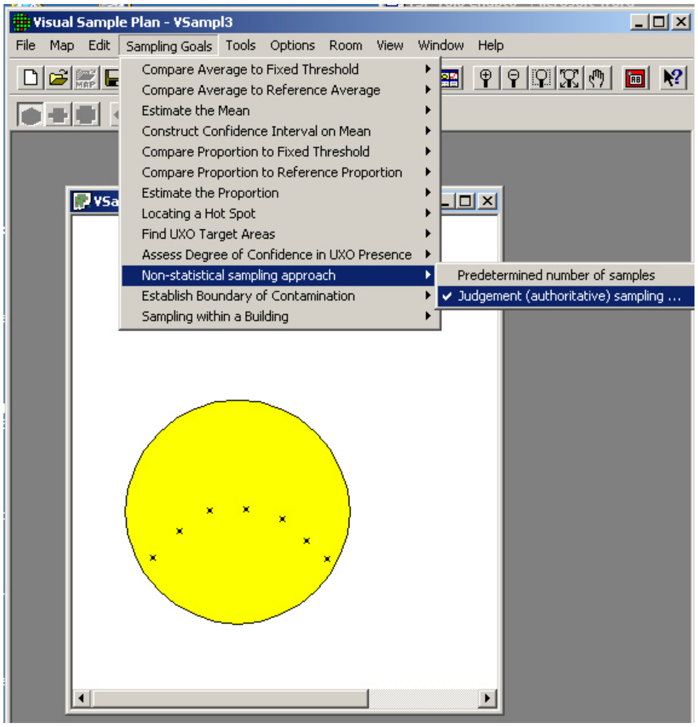
- stratified sampling – Strata or partitions of an area are made based on a set of criteria, such as homogeneity of contamination. Samples are drawn from each stratum according to a formula that accords more samples to more heterogeneous strata.
- adaptive cluster sampling – An initial n samples are selected randomly. Additional samples are taken at locations surrounding the initial samples where the measurements exceed some threshold value. Several rounds of sampling may be required. Selection probabilities are used to calculate unbiased estimates to compensate for oversampling in some areas.

- sequential sampling - Sequential designs are by their nature iterative, requiring the user to take a few samples (randomly placed) and enter the results into the program before determining whether further sampling is necessary to meet the sampling objectives.
- collaborative sampling - The Collaborative Sampling (CS) design , also called “double sampling”, uses two measurement techniques to obtain an estimate of the mean – one technique is the regular analysis method (usually more expensive), the other is inexpensive but less accurate. CS is not a type of sampling design but rather method for selecting which samples are analyzed by which measurement method.
- ranked set sampling – In this two-phased approach, sets of population units are selected and ranked according to some characteristic or feature of the units that is a good indicator of the relative amount of the variable or contaminant of interest that is present. Only the mth ranked unit is chosen from this set and measured. Another set is chosen, and the m-1th ranked unit is chosen and measured. This is repeated until the set with the unit ranked first is chosen and measured. The entire process is repeated for r cycles. Only the m X r samples are used to estimate an overall mean.
- sampling along a swath or transect – Continuous sampling is done along straight lines (swaths) of a certain width using geophysical sensors capable of continuous detection. Swath patterns can be parallel, square, or rectangular. The goal is to find circular or elliptical targets. This design contains the two elements of traversing the target and detecting the target. VSP application is for unexploded ordnance (UXO).
- sampling along a boundary - This design places samples along a boundary in segments, combines the samples for a segment, and analyzes each segment to see if contamination has spread beyond the boundary. If contamination has spread, VSP keeps extending the boundary until the sampling goals have been met.
- judgment sampling – You simply point and click anywhere in a sampling area. These sampling locations are based on the judgment of the user.

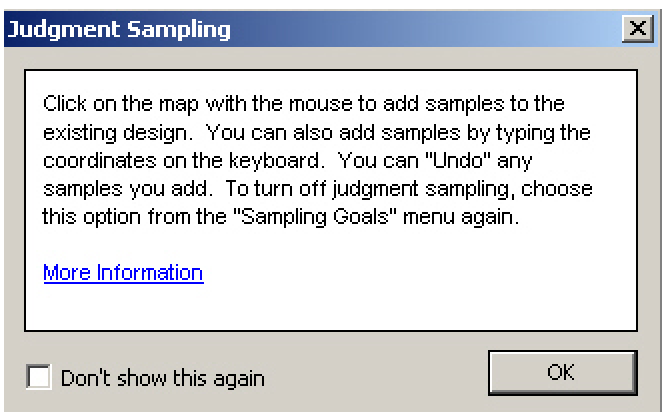
Because **Judgment Sampling** is not probability-based, users can bias the sampling results using this method. There is no basis in statistical theory for making confidence statements about conclusions drawn when samples are selected by judgment. However, some problem definitions might call for judgment sampling, such as looking in the most likely spot for evidence of contamination or taking samples at predefined locations. Figure 3.3 shows **Judgment Sampling** selected in VSP and six sampling locations selected manually.

3.2 DQO Inputs and Sample Size

The inputs needed for VSP’s sample-size calculations are decided upon during the DQO process. If you have not gone through the DQO process prior to entering this information, you can enter “best guess” values for each of the inputs and observe the resulting computed sample size. New inputs can be tried until a sample size that is feasible and/or within budget is obtained. This iterative method for using VSP is a valuable “what if” tool with which you can see the effect on sample size (and hence costs) of changing DQO inputs. However, be cautioned that all the DQO elements interact and have special meaning within the context of the problem. To be able to defend the sample size that VSP calculates, you



(a)



(b)

Figure 3.3. Judgment Sampling in VSP

must have a defensible basis for each of the inputs. There is no quick way to generate this defense other than going through Steps 1 through 6 of the DQO process.

The core set of DQO inputs that affect sample size for most of the designs are as follows:

- *Null Hypothesis Formulation* – The null hypothesis is the working hypothesis or baseline condition of the environment. There must be convincing evidence in the data to declare the baseline condition to be false. VSP uses a default of "Site is Dirty" as the working hypothesis that must be disproved with convincing evidence from the data.

- *Type I Error Rate (Alpha)* – This is called the false rejection rate in EPA's DQO guidance (EPA 2000a). This is the probability of rejecting a true null hypothesis. For the typical hypothesis test in which we assume the survey unit is dirty (above the action level), alpha is the chance a dirty site with a true mean equal to or greater than the Action Level will be released as clean to the public. In general, alpha is the maximum chance, assuming the DQO inputs are true, that a dirty site will be released as clean.

- *Type II Error Rate (Beta)* – This is called the *false acceptance* rate in EPA's DQO guidance. This is the probability of not rejecting (accepting) a false null hypothesis.

- For the typical hypothesis test in which we assume the survey unit is dirty, beta is the chance a specific clean site will be condemned as dirty. Specifically, beta is the chance that a clean site with a true mean equal to or less than the lower bound of the gray region will be condemned as dirty. In general, beta is the maximum chance, outside the gray region, that a clean site will be condemned as dirty.

- *Width of Gray Region (Delta)* – This is the distance from the Action Level to the outer bound of the gray region. For the typical hypothesis test in which we assume the survey unit is dirty, the gray region can be thought of as a range of true means where we are willing to decide that clean sites are dirty with high probability. Typically, these probabilities are 20% to 95%, i.e., from beta to 1 - alpha. If this region is reduced to a very small range, the sample size grows to be extremely large. Determining a reasonable value for the size of the gray region calls for professional judgment and cost/benefit evaluation.
- *Estimated Sampling Standard Deviation* – This is an estimate of the standard deviation expected between the multiple samples. This estimate could be obtained from previous studies, previous experience with similar sites and contaminants, or expert opinion. Note that this is the square root of the variance. In one form or another, all the designs require some type of user-input as to the variability of contamination expected in the study area. After all, it the area to be sampled was totally homogeneous, only one sample would be required to completely characterize the area.

Other inputs are required by some of the designs, and other inputs are required for design parameters other than sample size. For example, the stratified designs require the user to specify the desired number of strata and estimates of proportions or standard deviations for each of the stratum. The UXO (unexploded ordinance) modules use Bayesian methods and require the user to input their *belief* that the study area contains UXO. When simulations are used, as in the post-survey UXO target detection, the user must input assumptions about distribution of scrap or shrapnel in the target areas. In the discussions of the designs, we try to give an explanation of each input required of the user. If you are lost, use the VSP Help functions.

Note: The **Help Topics** function in VSP provides a description of each of the designs and its related inputs. You can also select the Help button on the toolbar, put the cursor on any of the designs on the menu and a description of the design and its inputs will appear in a **Help** window. In addition, pressing the Help button at the bottom of each design dialog will bring up a Help window that contains a complete explanation of the design. Finally, on each screen where input is required, highlight an item and press the *F1* key for a description of that input item.

The next section contains a discussion of the inputs required by most of the designs available in VSP 3.0. The designs are organized by the Sampling Goal under which they fall. Not all options for all designs are discussed. The **Help Topics** function describes all the options and is a good supplement to this User's Manual. Common design features (such as Costs, Historical Samples, MQO) that are found in multiple designs are not be discussed individually in this section but can be found in Section 5.0, **Common and Extended Features of VSP**.

3.2.1 Compare Average to a Fixed Threshold

Comparing the average to a fixed threshold is the most common problem confronted by environmental remediation engineers. We present different forms the problem might take and discuss how VSP can be used to address each problem formulation.

Continue where we left off in Section 2.3.3 with the Millsite.dxf map loaded. We selected a single Sample Area from the site. The Action Level for the contaminant of interest is 6 pCi/g in the top 6 in. of soil. Previous investigations indicate an estimated standard deviation of 2 pCi/g for the contaminant of interest. The null hypothesis for this problem is "Assume Site is Dirty" or H0: True mean >=AL.

We desire an alpha error rate of 1%. We also desire a beta error rate of 1%. According to EPA (2000a, pp. 6-10), 1% for both alpha and beta are the most stringent limits on decision errors typically encountered for environmental data. We tentatively decide to set the lower bound of the gray region at 5 pCi/g. We also decide that a systematic grid is preferable.

We will use VSP to determine the final width of the gray region and the number of samples required. Assume the fixed cost of planning and validation is $1,000, the field collection cost per sample is $100, and the laboratory analytical cost per sample is $400. We are told to plan on a maximum sampling budget of $20,000.

Case 1: We assume that the population from which we are sampling is approximately normal or that it is well-behaved enough that the Central Limit Theorem of statistics applies. In other words, the distribution of sample means drawn from the population is approximately normally distributed. We also decided that a systematic pattern for sample locations is better than a random pattern because we want complete coverage of the site.

VSP Solution 1: We start by choosing VSP Sampling Goal option of **Compare Average to Fixed Threshold > Can assume data will be normally distributed > Systematic grid sampling**. A grouping of the input dialogs is shown in Figure 3.4.

We see that for our inputs, using a one-sample t-test will require taking 90 samples at a cost of $46,000. Clearly, we need to relax our error tolerances or request more money.

For the sake of argument, suppose all the stakeholders agree that an alpha error rate of 5% and a beta error rate of 10% are acceptable. Figure 3.5 reveals that those changes lead to a significant reduction in the sampling cost, now $19,000 for n = 36 samples.

Are these new error rates justifiable? Only the specific context of each problem and the professional judgment of those involved can answer that question.

What about the assumption that we will be able to use a parametric test, the one-sample t-test? Unless the population from which we are sampling is quite skewed, our new sample size of n = 36 is probably large enough to justify using a parametric test. Of course, once we take the data, we will need to justify our assumptions as pointed out in *Guidance for Data Quality Assessment Practical Methods for Data Analysis* (EPA 2000b, pp. 3-5).

Case 2: We now decide that we want to look at designs that may offer us cost savings over the systematic design just presented. We have methods available for collecting and analyzing samples in the field making quick turnaround possible. We want to be efficient and cost-effective and take only enough samples to confidently say whether our site is clean or dirty. After all, if our first several samples exhibit

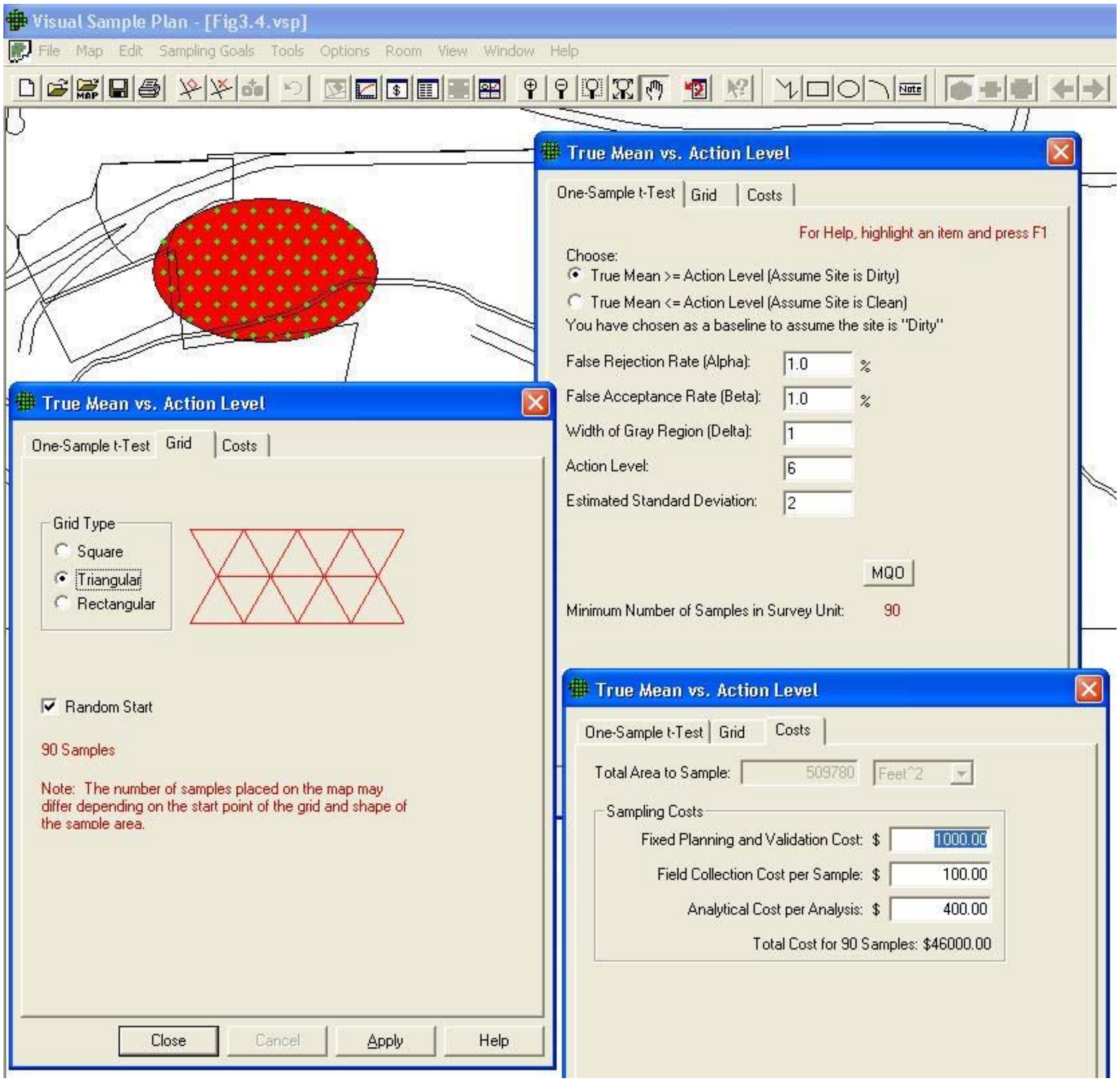


Figure 3.4. Input Boxes for Case 1 with Original Error Rates

levels of contamination so high that there is no possible scenario for the average to be less than a threshold, why continue to take more samples? We can make a decision right now that the site needs to be remediated. Sequential designs, and the tests associated with them, take previous sampling results into account and provide rules specifying when sampling can stop and a decision can be made.

VSP Solution 2a: From VSP's main menu, select **Sampling Goal of Compare Average to a Fixed Threshold > Can assume data will be normally distributed > Sequential Sampling (Known Standard Deviation)**. The dialog box in Figure 3.6 appears. We begin by entering the DQO parameters for Alpha, Beta, Action Level, etc. Next, enter the **Number of Samples Per Round**, shown here as **3**. This parameter indicates how many samples you want to take each time you mobilize into the field. Each time you press the **Apply** button, VSP places a pattern of this many sampling locations on the map.

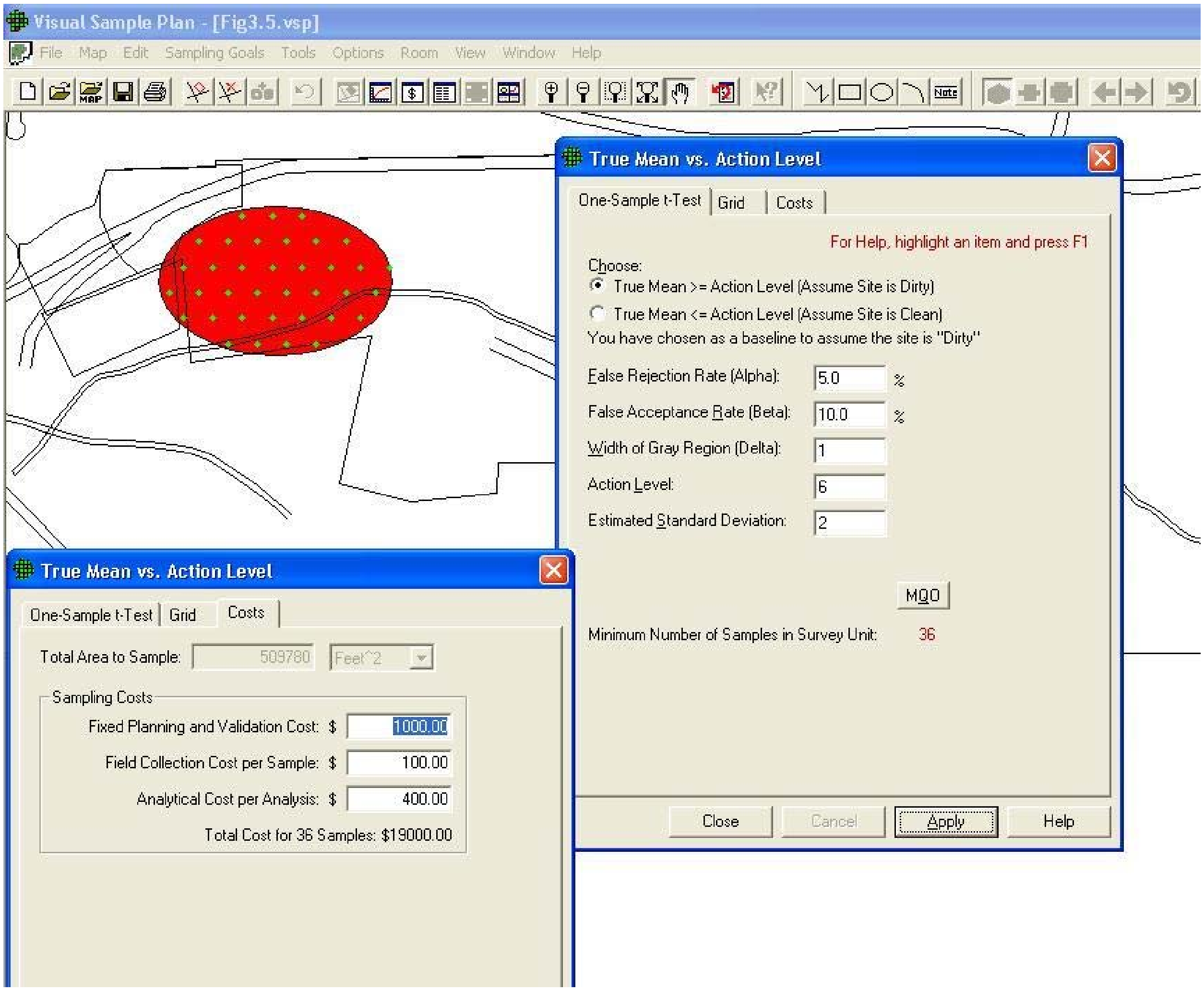


Figure 3.5. Input Boxes for Case 1 with Increased Error Rates

When you close this design dialog, this pattern of sampling locations is locked or "frozen." In Figure 3.6, we see the results of pressing Apply, and three locations are placed on the Map labeled "Seq-1, Seq-2, Seq-3".

You must now exit this dialog (close the display by clicking the X in the upper right-hand corner of the display), go out and take the samples, and analyze them. Once the sample results are available, re-open the Sequential Probability Ratio Test design dialog box (see Figure 3.7). The easiest way to re-open a sampling design is to use the menu item: **Sampling Goals > Last Design** or click the **Last Design** button on the main toolbar.

Note: This is the case with all VSP modules where the user must input sampling results – you must exit the dialog box, enter the results, and re-enter the dialog box.

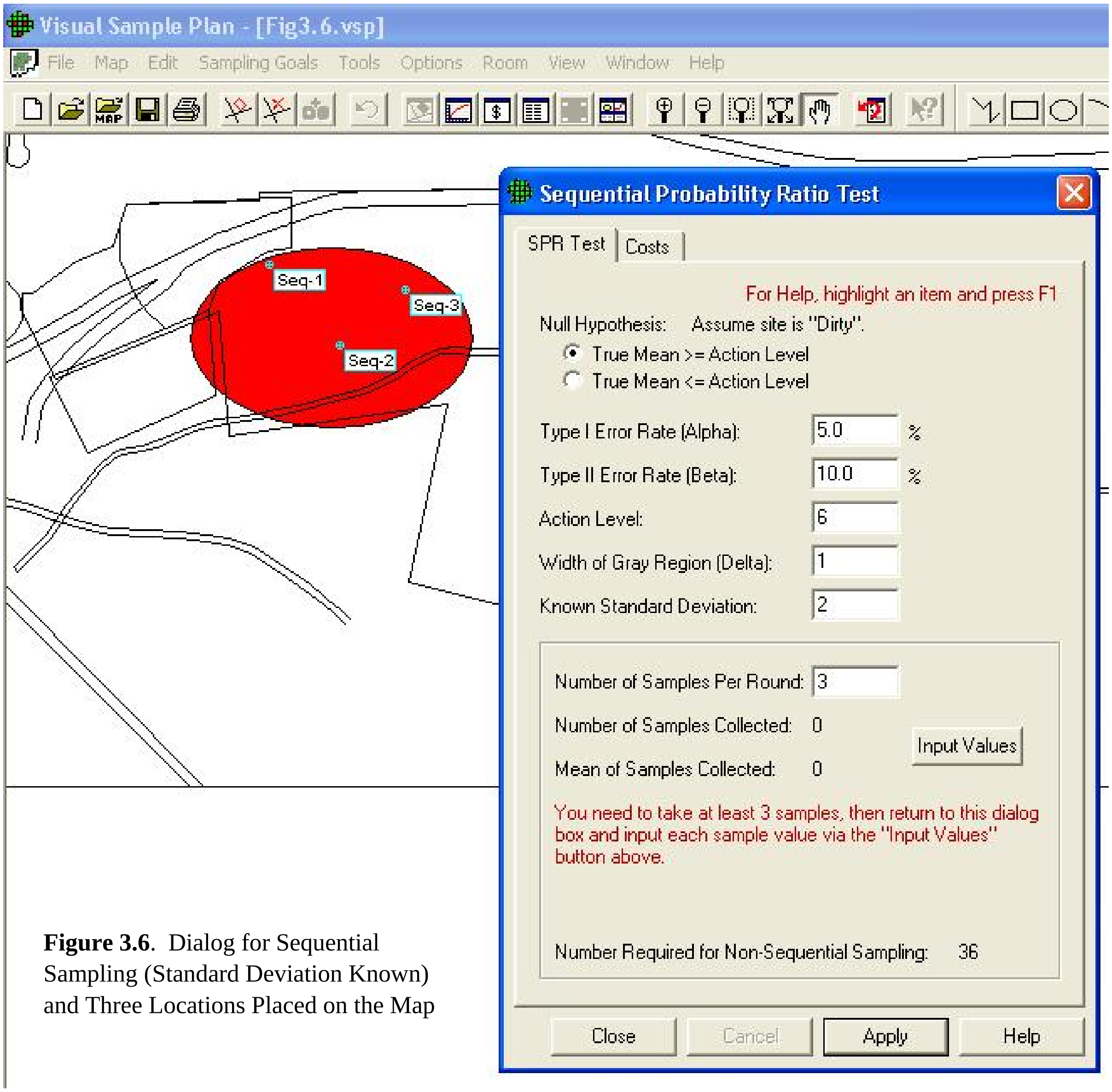


Figure 3.6. Dialog for Sequential Sampling (Standard Deviation Known) and Three Locations Placed on the Map

You now see the **Number of Samples Collected** as **3**. Press the **Input Values** button and enter the measurement values for those three samples into the grid on the data input dialog. We enter these values as **6.5, 8, and 5**. Press the **OK** button and VSP returns to the original SPRT dialog box. We now see that VSP calculated a mean of 6.5 and a standard deviation of 1.5 for the values we entered. VSP cannot accept or reject the null hypothesis within the error limits we specified based on these values and suggests that **9** additional samples are needed to make a decision.

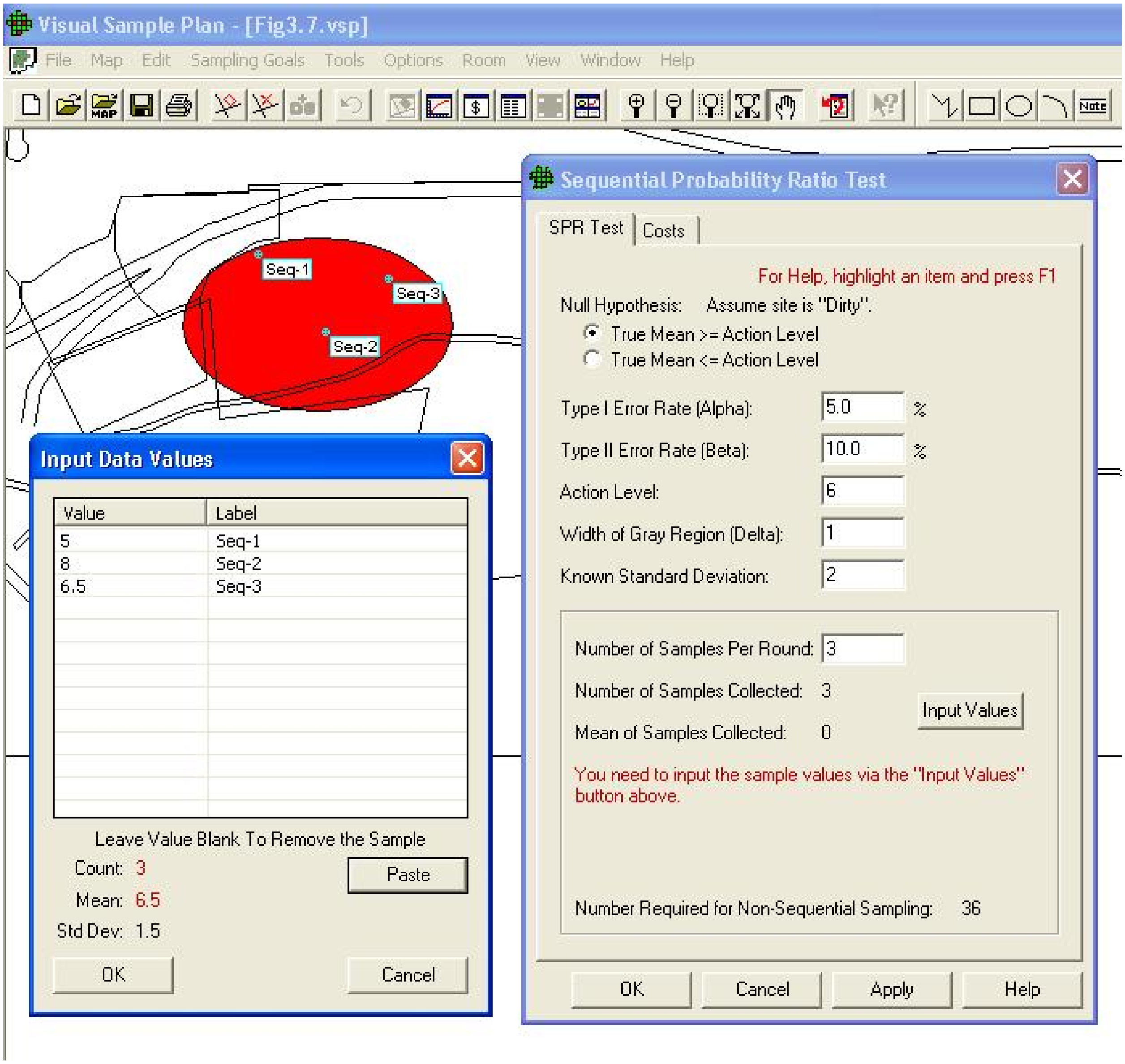


Figure 3.7. Data Input Dialog for Sequential Probability Ratio Test and Results from First Round of Sampling. Map View is shown in background.

Switching over to the Graph View in Figure 3.8, we can see that in order to accept the null hypothesis that the site is dirty by taking just three samples, we need a sample mean of approximately 9 (i.e., cross-hair set at 3 on the x-axis, intersects the lower boundary of the red area at 9 on the y-axis). If we move the cursor to the green area, the cross-hair intersects the upper boundary of the green area at approximately 2.5. In other words, we need a sample mean of 2.5 in order to reject the null hypothesis and declare the site clean with only three values. Remember, we told VSP that we *knew* the standard deviation to be 2, so that is what the program is using to make the projection of additional samples needed. If we are wrong, the VSP output will be misleading.

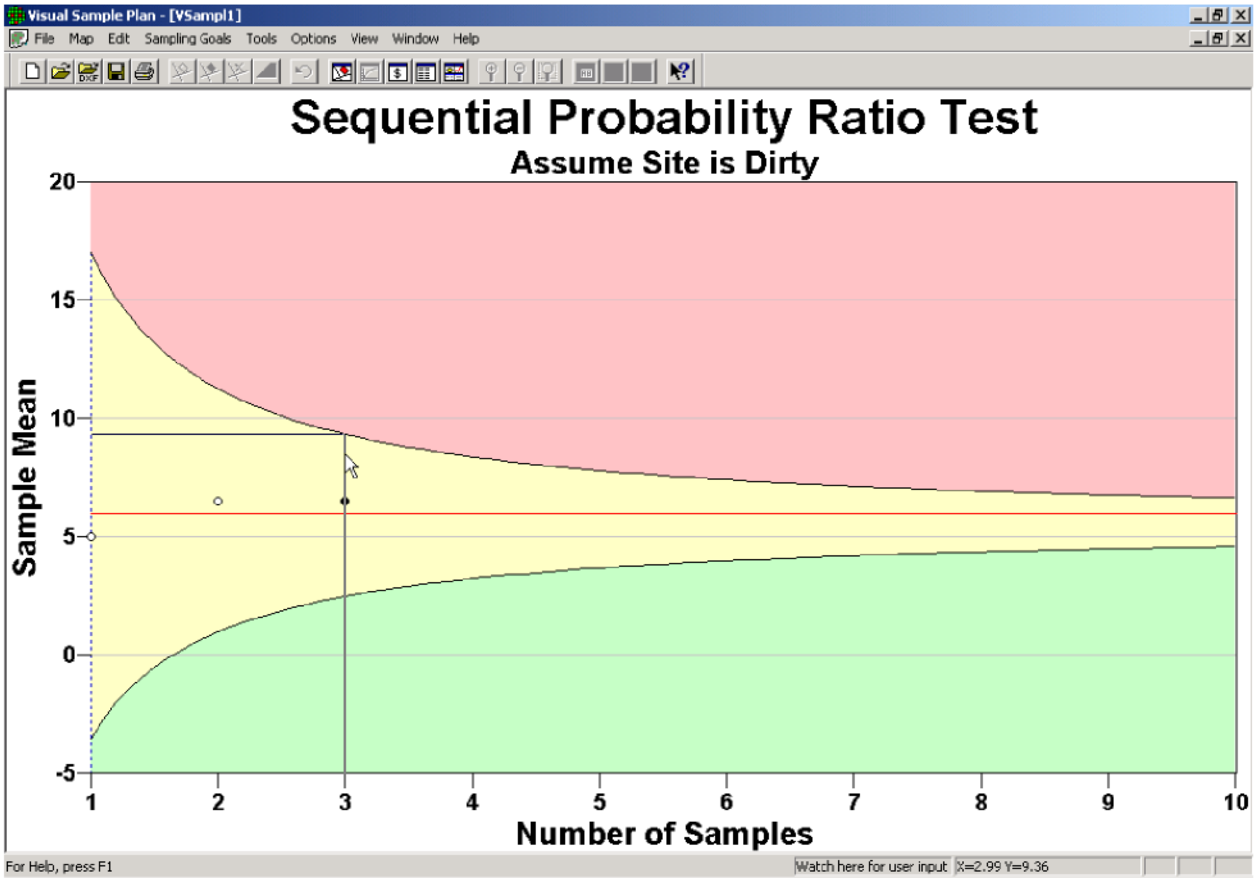
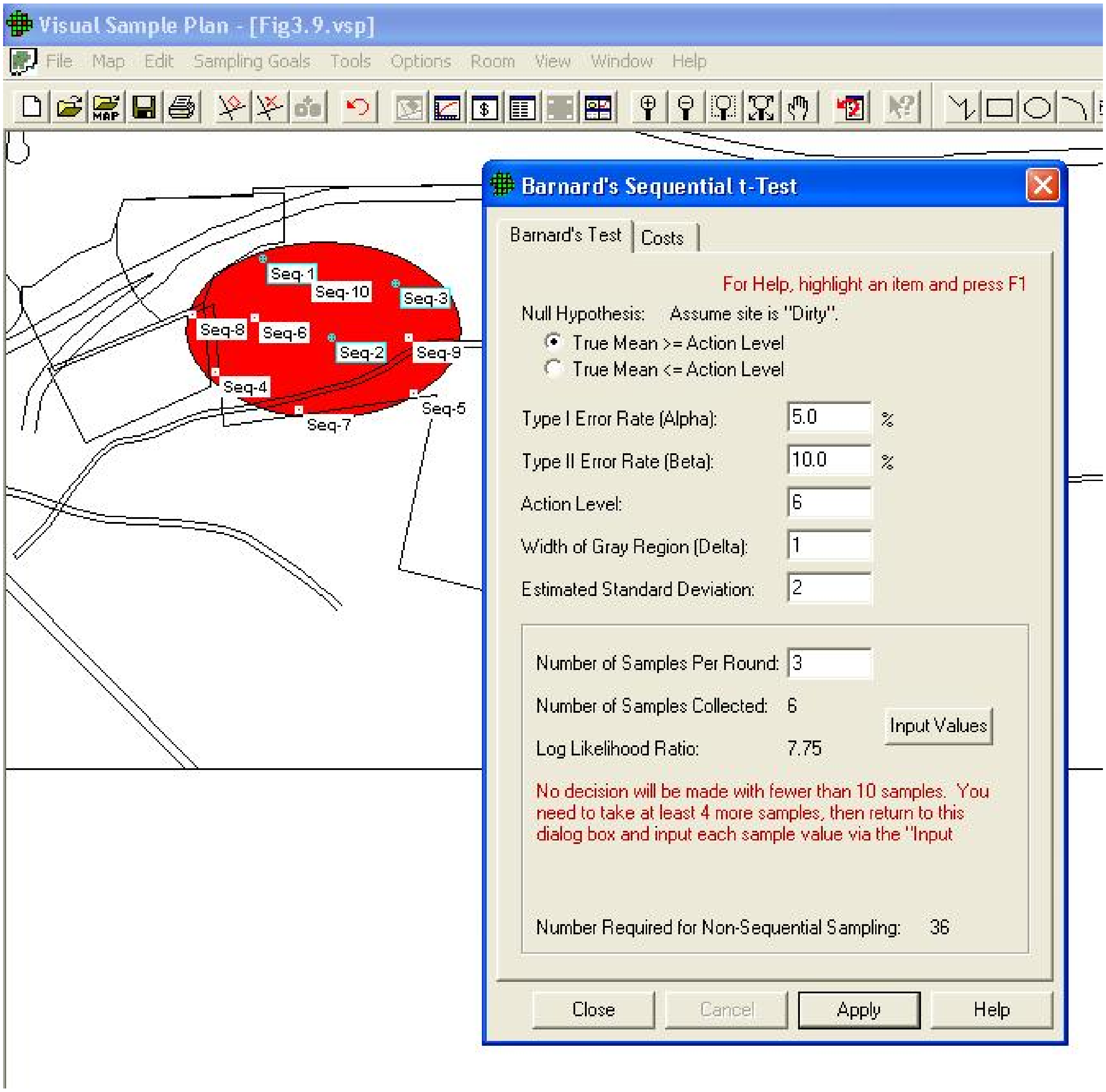


Figure 3.8. Graph View of Sequential Sampling

The two open circles in Figure 3.8 show the sample mean after the first two samples are collected. The closed circle shows the mean at the end of the third sample. We take the next set of three samples and get values **10, 5, and 12**. VSP now tells us that we can **Accept the Null Hypothesis** and conclude the site is dirty.

VSP Solution 2b: We now observe what happens when we do not know the true standard deviation of units in the population and select Sampling Goal of **Compare Average to a Fixed Threshold > Can assume data will be normally distributed > Sequential Sampling (Unknown Standard Deviation)**. The dialog box that appears is for a different test, Barnard's Sequential Test, shown in Figure 3.9. We enter the same DQO inputs of Alpha, Beta, Action Level, etc. We also say we will take 3 Samples Per Round. Using Barnard's test, VSP tells us that we need to take at least 10 samples in order to make a decision. VSP places these 10 samples on the map.

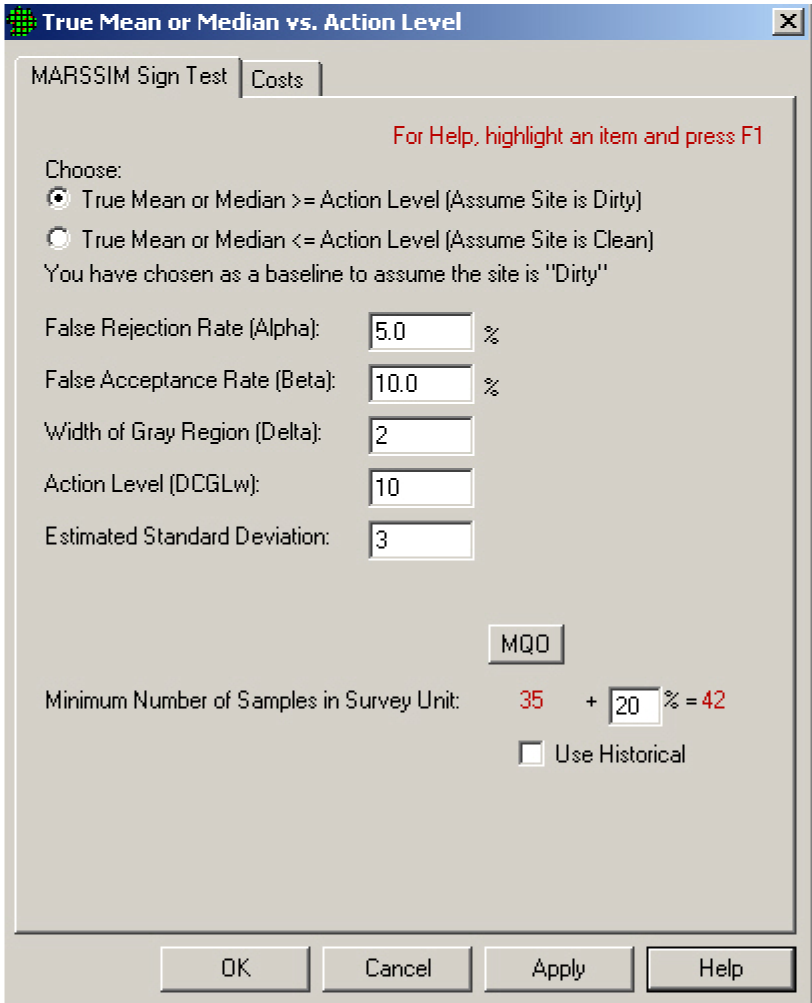
VSP Solution 2c: We have one other option for more cost-efficient sampling – reduce the analytic laboratory costs by taking advantage of measurement equipment that may be less accurate, but is less expensive. If we can still meet our DQOs (error levels, width of grey region) taking advantage of the less expensive equipment, we will save money.



(a)

Figure 3.9. Dialog Box for Sequential Sampling (Unknown Standard Deviation)

It works like this: At n’ field locations selected using simple random sampling or grid sampling, the inexpensive analysis method is used. Then, for each of n of the n’ locations, the expensive analysis method is also conducted. The data from these two analysis methods are used to estimate the mean and the standard error (SE: the standard deviation of the estimated mean). The method of estimating the mean and SE assumes there is a linear relationship between the inexpensive and expensive analysis methods. If the linear correlation between the two methods is sufficiently high (close to 1), and if the cost of the inexpensive analysis method is sufficiently less than that of the expensive analysis method, then CS



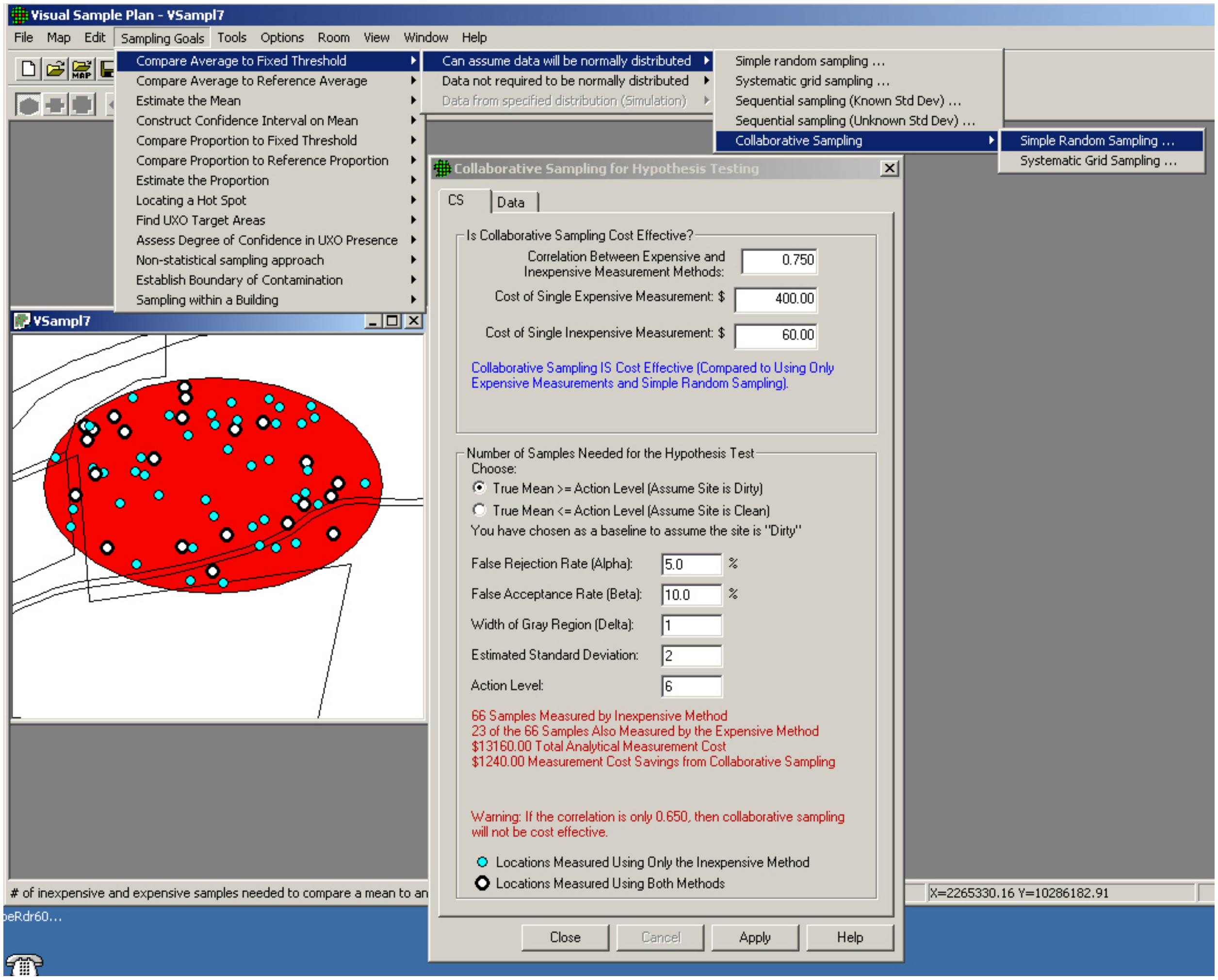
(b)

Figure 3.9. (contd)

is expected to be more cost effective at estimating the population mean than if the entire measurement budget was spent on obtaining only expensive analysis results at field locations selected using simple random sampling or grid sampling.

If Collaborative Sampling is chosen for the Sampling Goal of **Compare Average to a Fixed Threshold** the dialog box for input (with the CS tab selected) , and the resulting Map View of the applied CS samples on the Millsite.dxf map are all shown in Figure 3.10.

The first set of inputs requested in the Data Input Dialog Box for CS are those needed to determine whether CS sampling is more cost effective than using only expensive measurements and simple random sampling. The first input required is the correlation coefficient between expensive and inexpensive

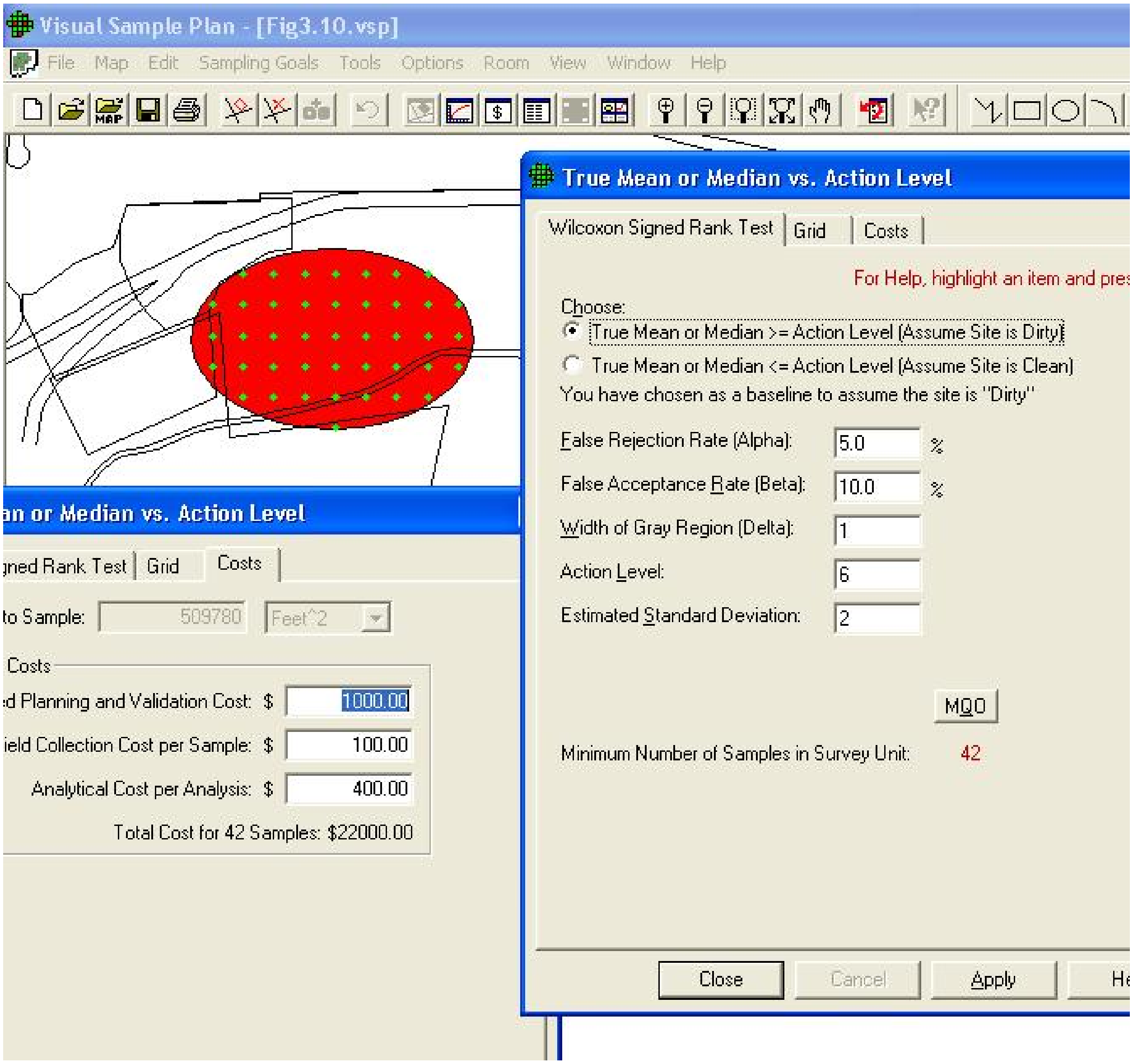


(a)

Figure 3.10. Input Boxes for MARSSIM Sign Test

measurements computed on the same set of samples. This is determined from data in prior studies or in a pilot study. The next two inputs are the cost estimates: the cost per unit of making a single expensive measurement, including the cost of sample collection, handling, preparation and measurement; and the cost per unit of making a single inexpensive measurement, including finding the field location and conducting the inexpensive analysis method.

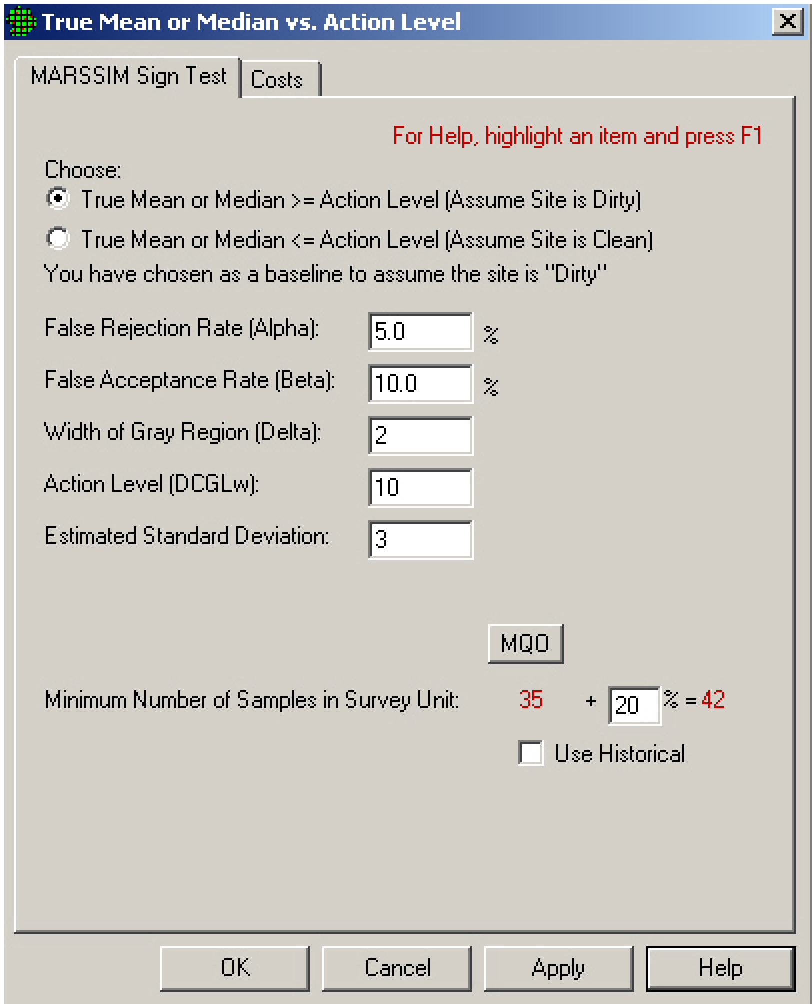
The next set of inputs comprises the DQOs for the problem. Notice that these are the same inputs we used for Case 1 with increased error rates (see Figure 3.5) when VSP calculated a required sample size of 36. If all those 36 samples were analyzed with the expensive method, the total cost would be 36 x $400 = $14,400. However, if we use CS and the same DQOs, VSP calculates we need to take 66 samples



(b)

Figure 3.10. (contd)

measured by the inexpensive method, and 23 of those 66 samples measured by the expensive method. This costs a total of 66 x $60 = 3960 plus 23 x $400 = $9200 for a total of $13,160. This represents a $1,240 cost savings over the $14,400 we were going to spend. And the best part is we can achieve this cost savings and still meet our required error rates (i.e., the stated DQOs). Note: If VSP determined that CS was not cost effective, it would not have computed the two samples n’ and n (66 and 23 samples, respectively) and reported only the number of samples that should be collected and analyzed using only the expensive method (36 samples).



(c)

Figure 3.10. (contd)

Once we hit the Apply button at the bottom of the Dialog Box, VSP places all 66 samples on the Sample Area on the map. VSP color codes those sample locations where both methods should be used vs. the sample locations where just the inexpensive measurement method should be used. The applied color-coded samples are shown in the **Map View** insert in Figure 3.10.

We now exit the Dialog Box by clicking on the X in the upper right-hand corner of the display. We take our samples, use the appropriate measurement method, and return to the sample Dialog Box to input the results from the lab. This time we select the tab labeled **Data** when we enter the Dialog Box. The data values can be entered by typing them into this input screen, or by importing the data from a file such as an Excel spreadsheet (see Section 2.4.1 Importing Samples). Figure 3.10a shows the Dialog Box for entering data.

Note that the values we entered result in a Standard Deviation of 2.24 – we estimated 2, and the two sets of sample values have a correlation of .769 – we estimated .75. We are well above the correlation limit of .650 in order for collaborative sampling to be cost effective. If we bring up the Graph View in a second window (**View > Graph**), we see that VSP has taken the data values we input and plotted the expensive measurements versus the inexpensive measurements. This plot can be used to assess whether the assumption of a linear relationship between the expensive and inexpensive measurements required for the use of CS is reasonable. Note that the calculated Rho = .769 (the correlation coefficient) is listed at the top of the graph. The regression line is the solid red line through the points. The dashed blue line represents the computed mean (x_{cs}). The horizontal red line represents the threshold value (Action Level). The bottom edge of the hashed red region represents the computed mean value below which the null hypothesis can be rejected.

VSP reports that based on the data values input, we can Accept the Null Hypothesis: Assume the Site is Dirty.

If we had chosen **Systematic Grid Sampling** rather than Simple Random Sampling, all the sample sizes would have been the same. The only difference would have been that the samples would have been placed on the Map in a grid pattern rather than randomly.

Case 3: We do *not* wish to assume that the population from which we are sampling is approximately normal or that the Central Limit Theorem applies. **In other words, we expect the possibility of a fairly skewed distribution**. We determine that a systematic grid is preferable.

VSP Solution 3a: We start by choosing VSP option **Compare Average to Fixed Threshold > Data not required to be normally distributed > Systematic grid sampling > Wilcoxon signed ranks test**. A grouping of the input dialogs is shown in Figure 3.10b.

For our inputs, and assuming that we will use a nonparametric Wilcoxon Signed Ranks test to analyze our data, VSP indicates that we are required to take 42 samples at a cost of $22,000. This is $3,000 more than the previous parametric case, given the same input parameters. Is the choice of a nonparametric test worth the extra $3,000 in sampling costs beyond what was required for the parametric one-sample t-test? VSP does not address that kind of question. Professional judgment is needed. You must make the decision based on the best available data, the consequences of decision errors, and legal and ethical considerations. If little pre-existing information is available, a pilot study to gain a better understanding of the characteristics of the population may be indicated.

VSP Solution 3b: The purpose of a MARSSIM sign test (Compare Average to Fixed Threshold > Data not required to be normally distributed > Simple Random sampling > Wilcoxon sign test is to test a hypothesis involving the true mean or median of a population against an Action Level. Note that using this test for the mean assumes the distribution of the target population is symmetrical. This assumption and the appropriate use of the Sign Test for final status surveys is discussed in Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) (EPA 2000). This document is currently available at: http://www.epa.gov/radiation/marssim/

The input for the MARSSIM Sign Test is shown in Figure 3.10c

3.2.2 Compare Average to Reference Average

We again start with the Millsite.dxf map from Section 2.3.3 with a single Sample Area defined. The Action Level for the contaminant of interest is 5 pCi/g above background in the top 6 in. of soil. Background is found by sampling an appropriate Reference Area. Previous investigations indicate an estimated standard deviation of 2 pCi/g for the contaminant of interest. The null hypothesis for this problem is “Assume Site is Dirty” or H0: Difference of True Means >= Action Level. In other words, the parameter of interest for this test is the *difference* of means, not an individual mean as was the case in the one-sample t-test.

We desire an alpha error rate of 1%. We also desire a beta error rate of 1%. We tentatively decide to set the lower bound of the gray region to 4 pCi/g above background, i.e., a *difference* of means of 4 pCi/g.

Using VSP, we will determine the final width of the gray region and the number of samples required. Assume that the fixed planning and validation cost is $1,000 for each area, and the field collection and measurement cost per sample is $100. The laboratory analytical cost per sample is $0 because we are able to justify the use of field measurements. We are told to plan on a maximum sampling budget of $20,000 for *both* the Reference Area and the Study Area.

Case 4: We assume that the populations we are sampling are approximately normal or that they are well-behaved enough so that the Central Limit Theorem of statistics applies. In other words, the distributions of sample means drawn from the two populations are approximately normally distributed. If that is the case, the distribution of the differences also will be approximately normally distributed. We also assume the standard deviations of both populations are approximately equal. In addition, we determine that a systematic grid sampling scheme is preferable.

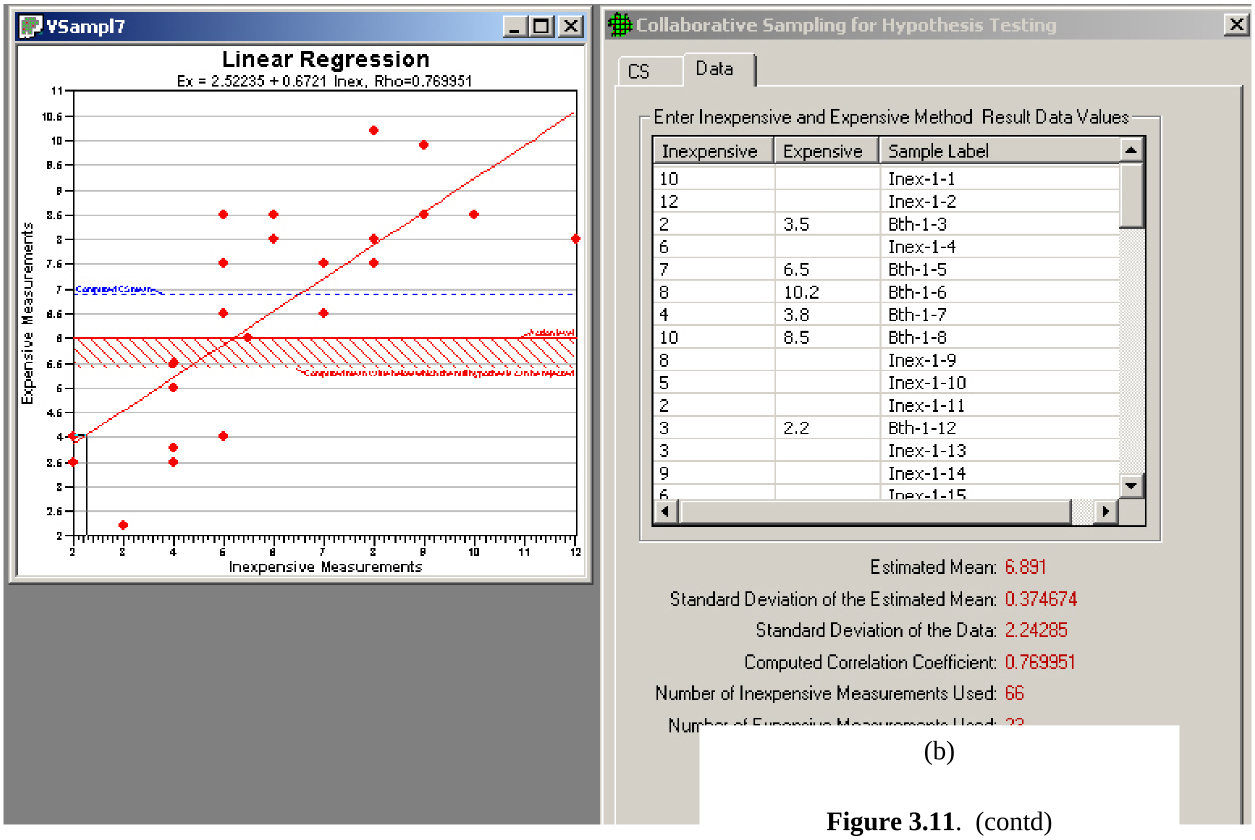
VSP Solution 4: We start by choosing from the main menu: **Sampling Goals > Compare Average to Reference Average > Can assume data will be normally distributed > Systematic grid sampling**. A grouping of the input dialogs is shown in Figure 3.11.

We see that for our inputs, using a two-sample t-test will require taking 175 field samples in the Sample Area at a cost of $18,500. The sampling cost for the Reference Area also will be $18,500. The combined sampling cost of $37,000 is significantly beyond our budget of $20,000. What will be the result if we relax the error rates somewhat?

In Figure 3.12, by increasing both the alpha error rate and the beta error rate to 5%, the sampling cost for one area has decreased to $9,800 based on n = 88 field samples. Thus, the new combined cost of $19,700 achieves our goal of no more than $20,000.

Can we justify these larger error rates? Again, only professional judgment using the best information related to the current problem can answer that question.

What about our planned use of a parametric test, the two-sample t-test? A sample size of 88 is large enough that we can probably safely assume the two-sample t-test will meet the assumption of normality for the differences of sample means. We should test this assumption after the data are collected.



(b)

Figure 3.11. (contd)

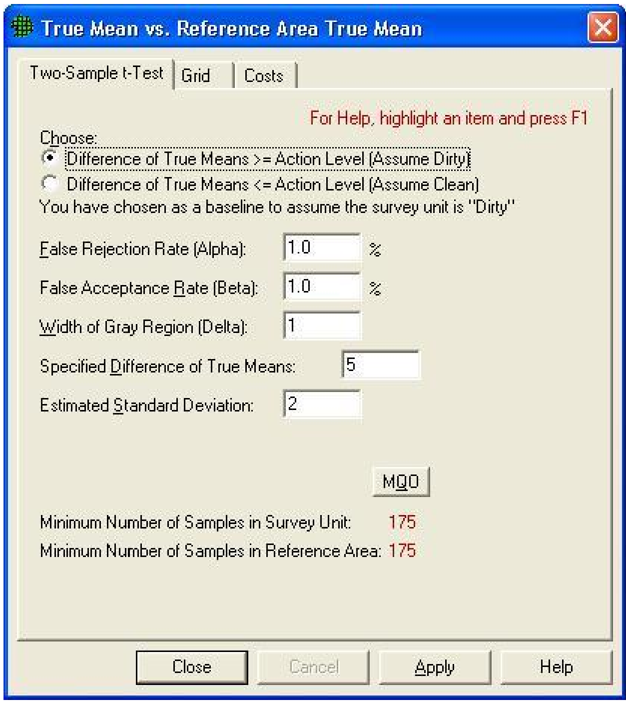
(a)

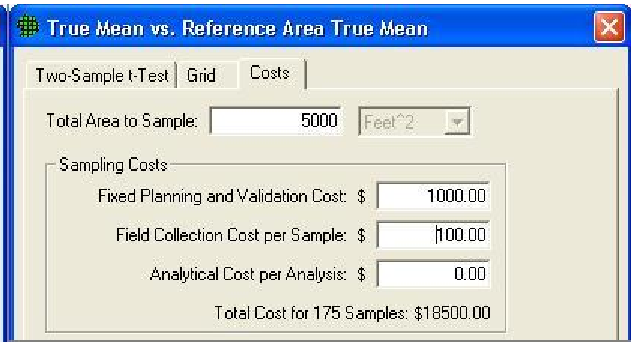
Figure 3.11. Input Boxes for Case 4 with Original Error Rates

What about the assumption of approximately equal standard deviations for the measurements in the Sample and Reference Areas? When we collect the data, we will need to check that assumption. See *Guidance for Data Quality Assessment Practical Methods for Data Analysis* (EPA 2000b, pp. 3-26) for the use of Satterthwaite's t-test when the standard deviations (or variances) of the two areas are not approximately equal.

Case 5: We now look at the case in which the nonparametric Wilcoxon rank sum test is planned for the data analysis phase of the project.

The Wilcoxon rank sum test is discussed in *Guidance for Data Quality Assessment* (EPA 2000b, pp. 3-31 – 3-34). The document can be downloaded from the EPA at: http://www.epa.gov/quality/qa_docs.html. It tests a shift in the distributions of two populations. The two distributions are assumed to have the same shape and dispersion, so that one distribution differs by some fixed amount from the other distribution. The user can structure the null and alternative hypothesis to reflect the amount of shift of concern and the direction of the shift.


True Mean vs. Reference Area True Mean
Two-Sample t-Test
Grid
Costs
For Help, highlight an item and press F1
Choose:
Difference of True Means >= Action Level (Assume Dirty)
Difference of True Means <= Action Level (Assume Clean)
You have chosen as a baseline to assume the survey unit is "Dirty"
False Rejection Rate (Alpha): 1.0 %
False Acceptance Rate (Beta): 1.0 %
Width of Gray Region (Delta): 1
Specified Difference of True Means: 5
Estimated Standard Deviation: 2
MQO
Minimum Number of Samples in Survey Unit: 175
Minimum Number of Samples in Reference Area: 175
Close
Cancel
Apply
Help


True Mean vs. Reference Area True Mean
Two-Sample t-Test
Grid
Costs
Total Area to Sample: 5000
Feet^2
Sampling Costs
Fixed Planning and Validation Cost: $ 1000.00
Field Collection Cost per Sample: $ 100.00
Analytical Cost per Analysis: $ 0.00
Total Cost for 175 Samples: $18500.00

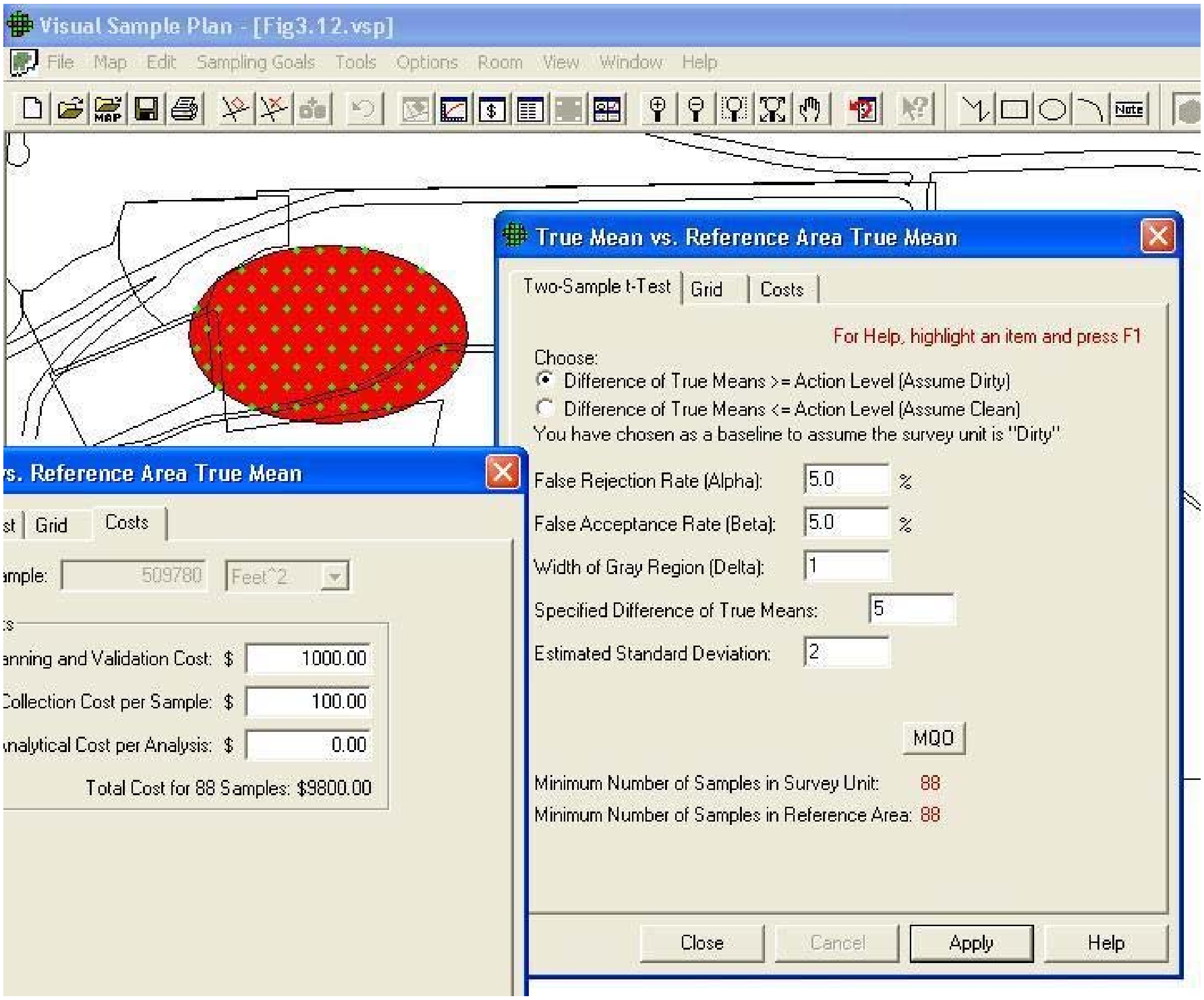


Figure 3.12. Input Boxes for Case 4 with Increased Error Rates

VSP Solution 5: We start by choosing from VSP's main menu **Sampling Goals > Compare Average to Reference Average > Data not required to be normally distributed > Systematic grid sampling (Wilcoxon Rank Sum Test)**. A grouping of the input dialogs is shown in Figure 3.13.

In Figure 3.13, you can see that the sample size increases to 102 for each sampling area, and the cost per area is now $11,200. Is the larger sample size of 102 instead of the previous sample size of 88 justified? Probably not. Again, professional judgment is needed.

Case 6: Next, assume that the population from which we will be sampling is definitely skewed and we again desire to use a nonparametric Wilcoxon rank sum test. However, we are limited to a total sampling budget for *both* areas of $10,000. By using VSP iteratively, we will adjust the various DQO input parameters and try to discover a sampling plan that will meet the new goals.

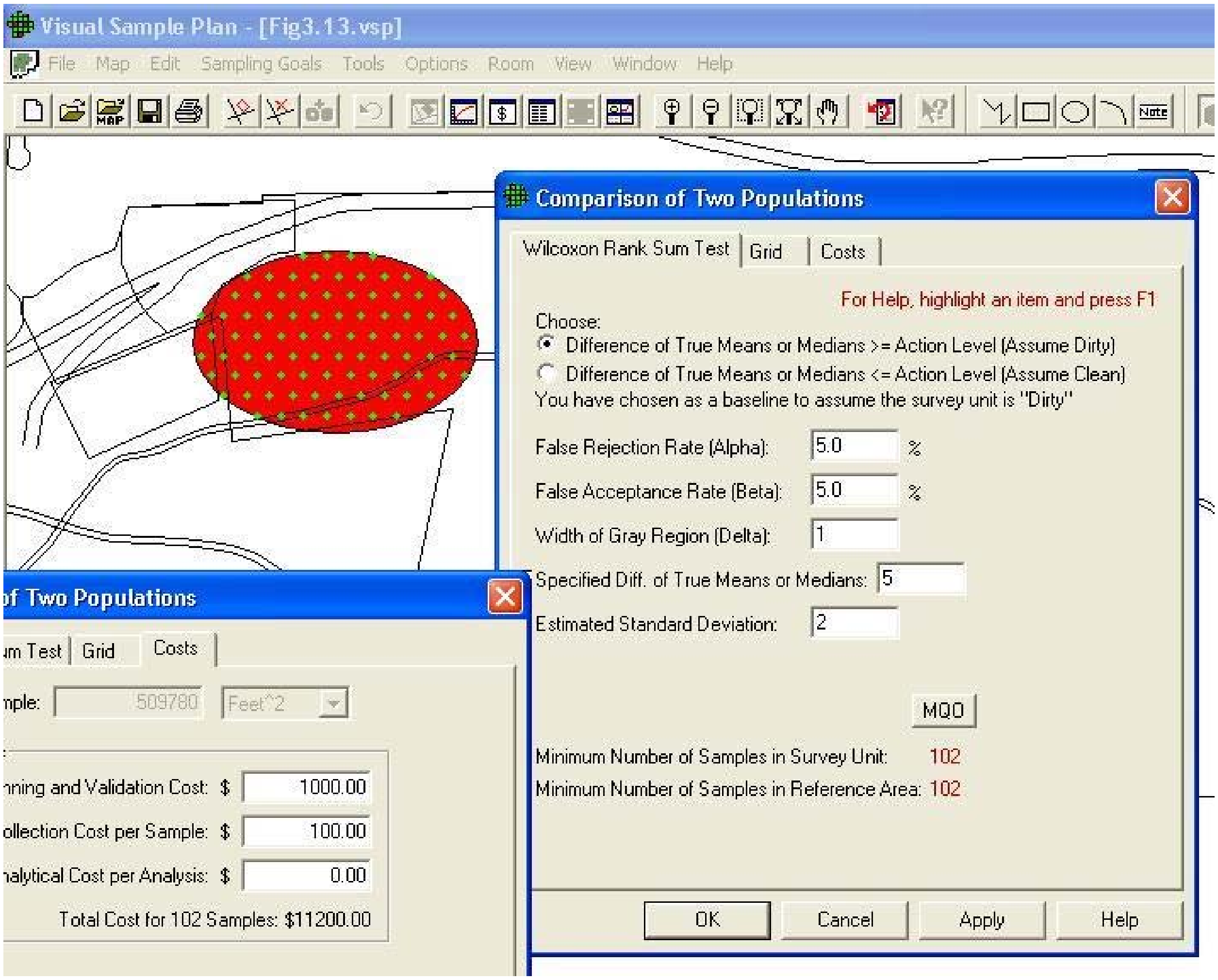


Figure 3.13. Input Boxes for Case 5 Using Nonparametric Wilcoxon Rank Sum Test

VSP Solution 6: Figure 3.14 shows that with an alpha of 5%, a beta of 20%, and a lower bound of the gray region of 3.75, the number of samples per area drops to 38. With a sampling cost of $4,800 for each sampling area, we now have a combined cost of $9,600 and thus meet our goal of $10,000.

Will relaxing the error tolerances and increasing the width of the gray region to meet the requirements of the smaller sampling budget be acceptable to all stakeholders in the DQO process? Again, it depends on the objectives and judgment of those involved in the process.

Case 7: Suppose our combined sampling budget is reduced to $5,000. Can VSP provide a sampling design that meets that goal?

VSP Solution 7: Figure 3.15 shows a design with just 14 samples per sampling area that meets the new sparse budget. We reduced the combined sampling cost, now $4,800, by increasing the width of the gray region to 2.1 pCi/g (lower bound of the gray region is 2.9 pCi/g).

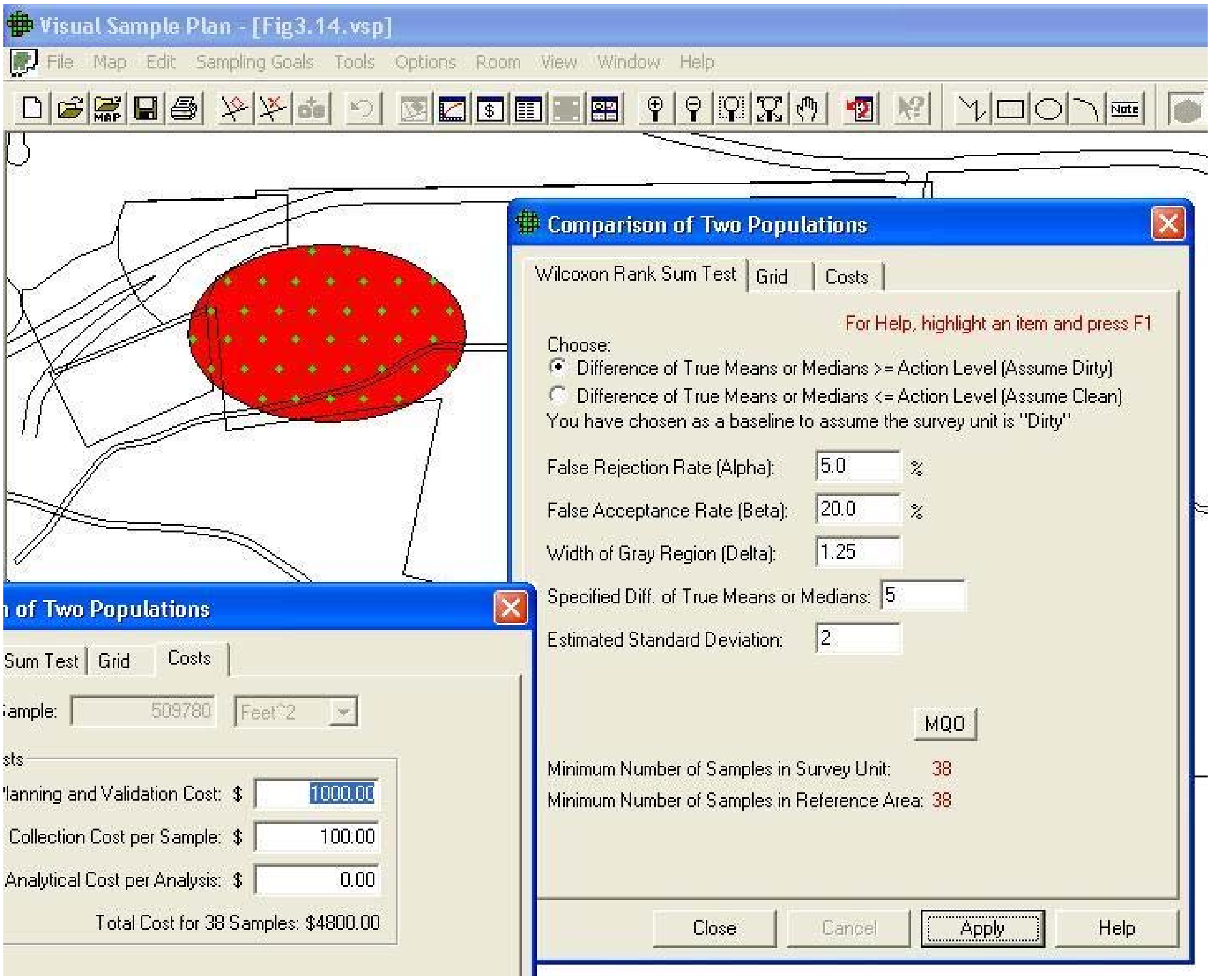


Figure 3.14. Input Boxes for Case 6 Using Nonparametric Wilcoxon Rank Sum Test

There are definite consequences of reducing sampling requirements to fit a budget. The consequences could include a greater chance of concluding that a dirty site is clean or a clean site is dirty. There is also a larger area of the gray region where you say you will not control (i.e., limit) the false acceptance error rate.

Is it justifiable to keep reducing the sampling budget in the above manner? Again, the answer depends on the specific problem. VSP, like most software, suffers from GIGO - Garbage In, Garbage Out. However, a responsible DQO process can provide valid information to VSP that overcomes GIGO and lets VSP help solve the current problem in an efficient manner.

Case 8: Now we assume we have seriously underestimated the standard deviation. Suppose that instead of 2 pCi/g, it is really 4 pCi/g. Now how many samples should we be taking?

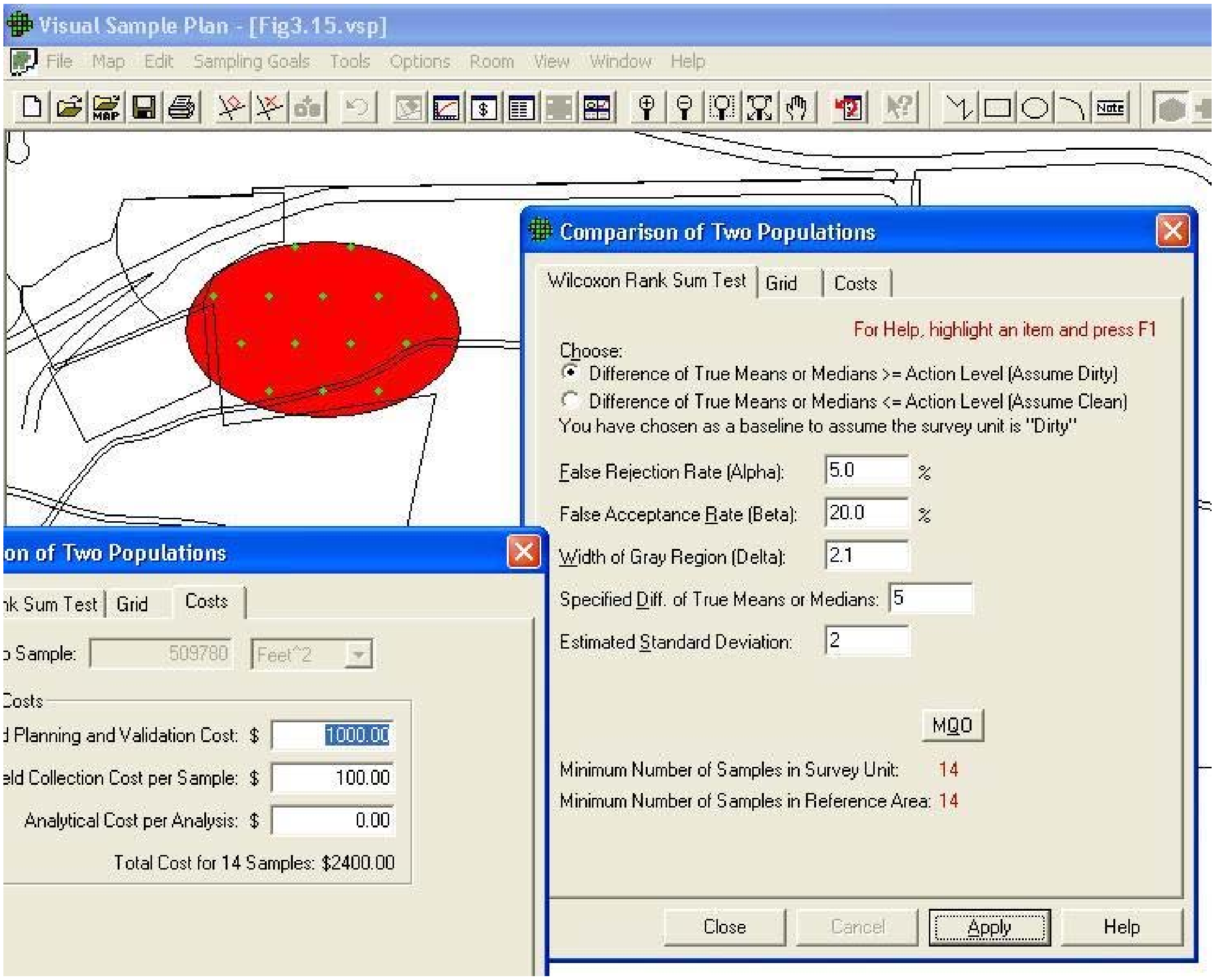


Figure 3.15. Input Boxes for Case 7 Using Nonparametric Wilcoxon Rank Sum Test

VSP Solution 8: Figure 3.16 shows the new sample size has jumped to 53, almost a four-fold increase over the 14 samples used in VSP Solution 7. For many sample-size equations, the number of required samples is proportional to the square of the standard deviation, i.e., the variance. Thus, an underestimate of the standard deviation can lead to a serious underestimate of the required sample size.

If we seriously underestimate the standard deviation of the measurements, what will be the practical implications of taking too few samples? Remember that we have as a null hypothesis "Site is Dirty." If the site is really clean, taking too few measurements means we may have little chance of rejecting the null hypothesis of a dirty site. This is because we simply do not collect enough evidence to "make the case," statistically speaking.

Case 9: When the assumptions of **data are not required to be normally distributed** is made, we can see from the pull-down menu lists under Sampling Goals that two non-parametric statistical tests are proposed – the Wilcoxon rank sum test and the MARSSIM WRS (Wilcoxon rank sum) test.

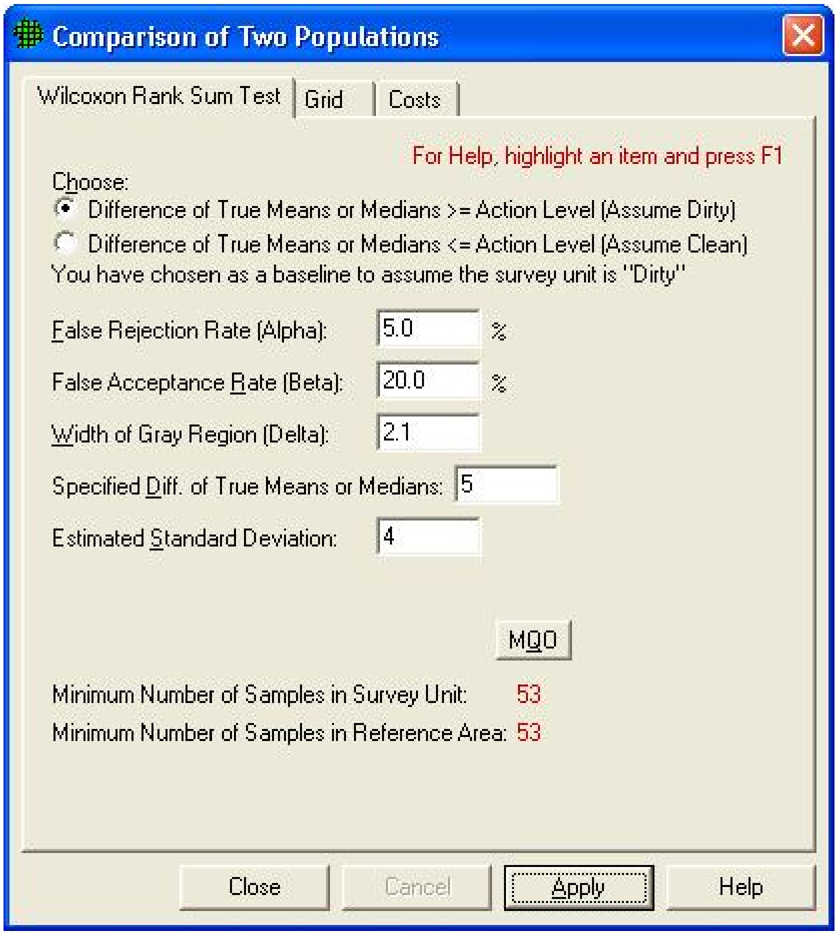


Figure 3.16a. Input Boxes for Case 9 with Larger Standard Deviation

The MARSSIM WRS test is used when the Sample Area population is symmetrical, the contaminant of concern in the SampleArea is present also in the background (Reference Area), and the contamination is uniformly present throughout the Sample Area. The *Multi-Agency Radiation Survey and Site Investigation Manual* (EPA 1997) contains a discussion of the MARSSIM WRS test. The manual is available from the EPA Internet site at: http://www.epa.gov/radiation/marssim/.

Shown in Figure 3.16b, the input dialog for the MARSSIM WRS test allows the user to supply a percent overage to apply to the sample size calculation. MARSSIM suggests that the number of samples should be increased by at least 20% to account for missing or unusable data and for uncertainty in the calculated values of Sample Size, (MARSSIM, p. 5-29). With the extra 20%, the sample size now becomes 53 samples required in both the Sample Area (i.e., Survey Unit or Study Area) and Reference Area.

The verification testing done on VSP shows that the Wilcoxon rank sum test requires slightly higher sample sizes than the MARSSIM WRS test for the same set of inputs, assuming all the appropriate assumptions for each test are met.

3.2.3 Estimate the Mean

When the Sampling Goal is to Estimate the Mean > Data not required to be normally distributed, four design options are more cost-effective (require fewer samples) than simple random sampling or systematic sampling. None of the four requires the assumption of normality as the underlying distribution of units in the population. The four options are:

- stratified sampling

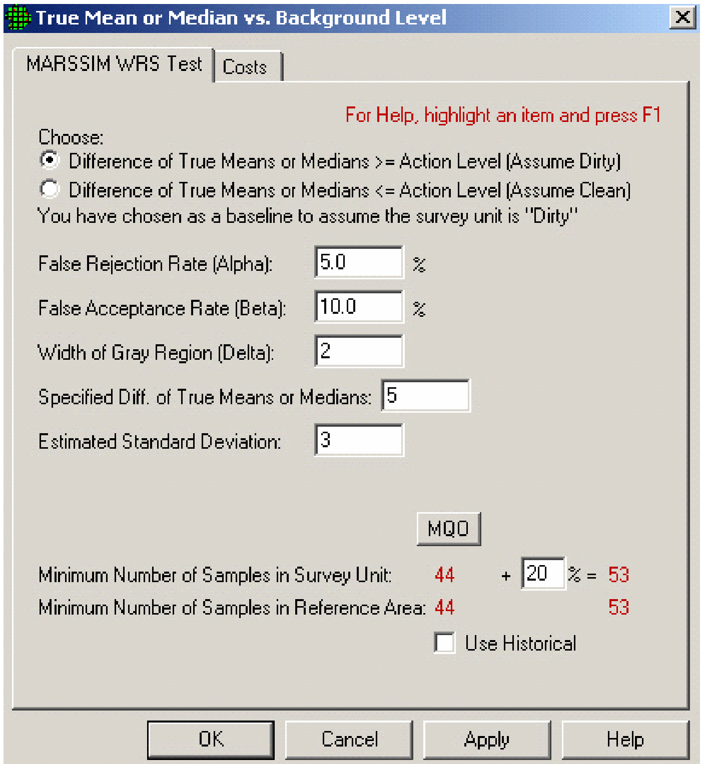


Figure 3.16b. (contd)

- ranked set sampling
- adaptive cluster sampling
- collaborative sampling

3.2.3.1 Stratified Sampling

In Figure 3.17, we see the dialog box for entering parameters for stratified sampling. Prior to running VSP to calculate sample sizes for the strata, the user must have pre-existing information to divide the site into non-overlapping strata that are expected to be more homogeneous internally than for the entire site (i.e., all strata). They must be homogeneous in the variable of interest for which we want to calculate a mean. The strata are the individual user-selected Sample Areas and can be seen using Map View.

With the Sample Areas selected (VSP shows total number of areas in **Numbers of Strata**), the dialog shows the initial values VSP assigns to the various inputs. The number of potential samples in each stratum is initially set at the number of 1-square-foot (or whatever units are used) units available to be sampled or approximately the **area of the Sample Area** (shown when the area is first selected). If the sample support is not a 1-square-foot volume, the user should change this to the correct value. The initial standard deviation between individual units in the stratum is assigned the value **1**. It is in the same units as the mean. This is a critical value in the sample size calculation, so the user should make sure this is a good estimate. The sampling and measurement costs per sample in each stratum and the fixed costs are input in dollars. After entering the values for stratum 1, the user selects the next stratum from the drop-down list under **Stratum #**.

VSP allows simple random sampling or systematic within the strata. This is selected using the pull-down menu under **Specify Sampling Design in Stratum n**.

The other inputs required by VSP pertain to the method the user wants to use for determining 1) the total number or samples in all strata and 2) the allocation of samples to strata. Methods are selected from the drop-down lists. VSP Help offers some insight into why one method might be selected over another, but the user should use the DQO process to flush out the site-specific conditions and project goals that will determine these inputs. Different inputs are required depending on which method is selected for determining the total number of samples. After you press **Apply**, the dialog shows in red the total number of samples and the number of samples in each stratum (use the pull-down **Stratum #** to switch between strata). You can see the placement of samples within strata by going to **Map View**.

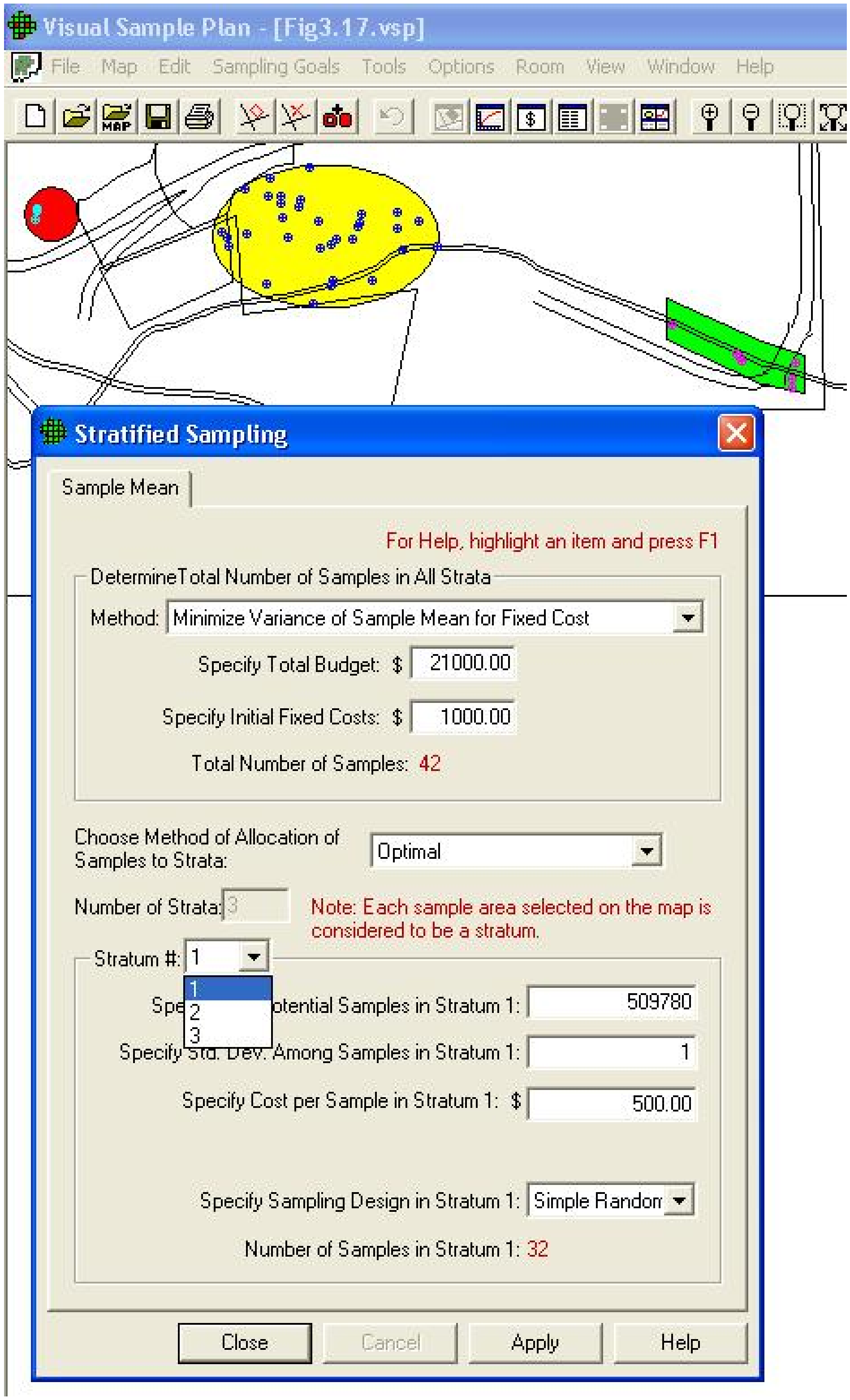


Figure 3.17. Dialog Box for Stratified Sampling for Estimating a Mean

3.2.3.2 Ranked Set Sampling

Ranked set sampling (RSS) is the second option for the Sampling Goal: **Estimate the Mean > Data not required to be normally distributed**. The number of inputs required for RSS is the most of any of the designs available in VSP. However, RSS may offer significant cost savings, making the effort to evaluate the design worthwhile. The VSP Help, the VSP technical report (Gilbert et al. 2002), and EPA (2001, pp. 79–111) are good resources for understanding what is required and how VSP uses the input to create a sampling design.

A simple example given here will explain the various input options. The user should have gone through the DQO process prior to encountering this screen because it provides a basis for inputs.

Under the tab **Ranked Set Sampling**, the first set of inputs deals with whether this design has any cost advantages over simple random sampling or systematic sampling where every unit that is sampled is measured and analyzed.

We select **Symmetric** for the distribution of lab data, thus telling VSP we think the lab data is distributed normally so VSP should use a balanced design. A balanced design has the same number of field locations, say $r = 4$, sampled for each of the say $m = 3$ ranks. That is, a sample is collected at each of the four locations expected to have a relatively small value of the variable of interest, as well as at the four locations expected to have a mid-range value, and at four locations expected to have a relatively large value. An unbalanced design has more samples collected at locations expected to have large values. EPA says that a balanced design should be used if the underlying distribution of the population is symmetric (EPA 2001, p. 86).

We select **Professional Judgement** as the ranking method. This selection requires us to say whether we think there is Minimal or Substantial error in our ranking ability. We select **Minimal**. Note: if we had chosen to use some type of Field Screening device to do the ranking, we would need to provide an estimate of the correlation between the field screening measurements and accurate analytical lab measurements. We choose a set size of 3 from the pull-down menu. The set size we select is based on practical constraints on either our judgment or the field screening equipment available.

Note: VSP uses set size to calculate the factor by which the cost of ranking field locations must be less than lab measurement costs in order to make RSS cost-effective. For our example, VSP tells us this factor must be at least **3** times.

The next set of inputs required for RSS is information required to calculate the number of samples needed for simple random sampling. This value, along with cost information, is used to calculate the number of cycles, *r*. We say we want a **one-sided confidence interval** (we want a tight upper bound on the mean and are not concerned about both over- and underestimates of the sample mean), we want that interval to contain **95%** of the possible estimates we might make of the sample mean, we want that interval width to be no greater than 1 (in units of how the sample mean is measured), and we estimate the standard deviation between individual units in the population to be **3** (in units of how the sample mean is measured). VSP tells us that if we have these specifications, we would need **26** samples if we were to take them randomly and measure each one in an analytical lab.

The box in the lower right corner of this dialog gives us VSP's recommendations for our RSS design: we need to rank a total of **45** locations. However, we need to send only **15** of those off to a lab for accurate measurement. This is quite a savings over the **26** required for simple random sampling. There will be *r* = **5** cycles required.

Note: If we had chosen an unbalanced design, VSP would tell us how many times the top ranked location needed to be sampled per cycle. Also, the inputs for the confidence interval would change slightly for the unbalanced design.

All costs (fixed, field collection per sample, analytical cost for sending a sample to the lab, and ranking cost per location) are entered on the dialog box that appears when the Cost tab is selected. In Figure 3.18, we see the two dialog boxes for RSS.

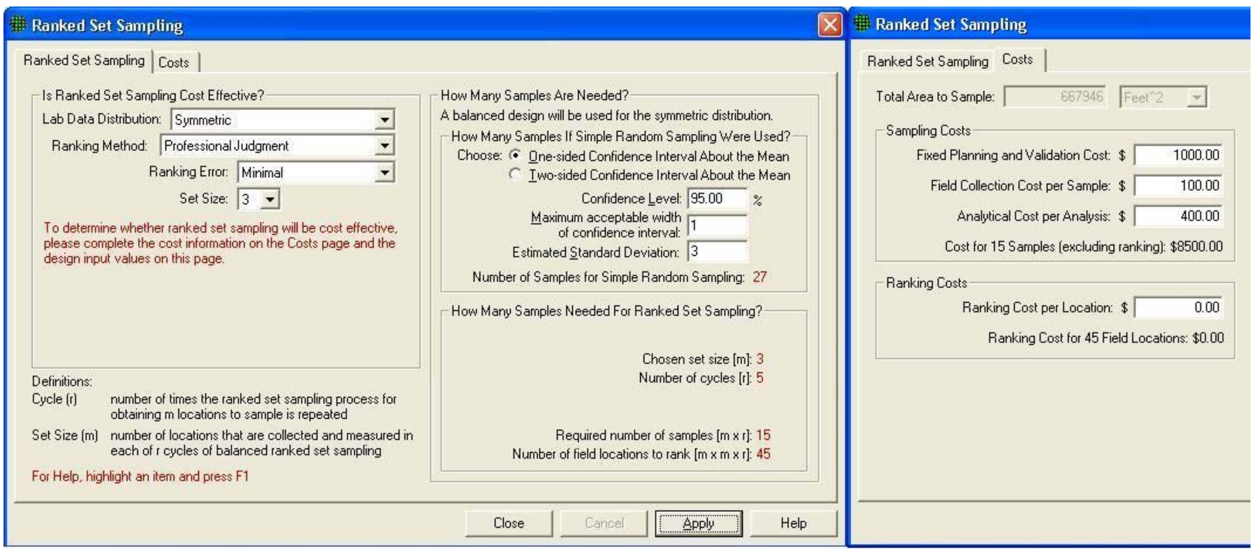


Figure 3.18. Dialog Boxes for Ranked Set Sampling Design

Once we press **Apply**, the RSS toolbar appears on our screen. The RSS toolbar lets us explore the locations to be ranked and the locations to be sampled and measured under **Map View**. VSP produces sample markers on the map that have different shapes and colors. The color of the marker indicates its cycle. The cycle colors start at red and go through the spectrum to violet. Selecting one of the cycles on the pull-down menu displays only the field locations for that cycle. In Figure 3.19, all the green field locations for **Cycle 3** are shown. The shape of the marker indicates its set. Field sample locations for the first set are marked with squares, locations for the second set are marked with triangles, and so on. We show **All Sets** in Figure 3.19. For unbalanced designs, the top set is sampled several times, so a number accompanies those markers. Our example is for a balanced design so we do not see numbers.

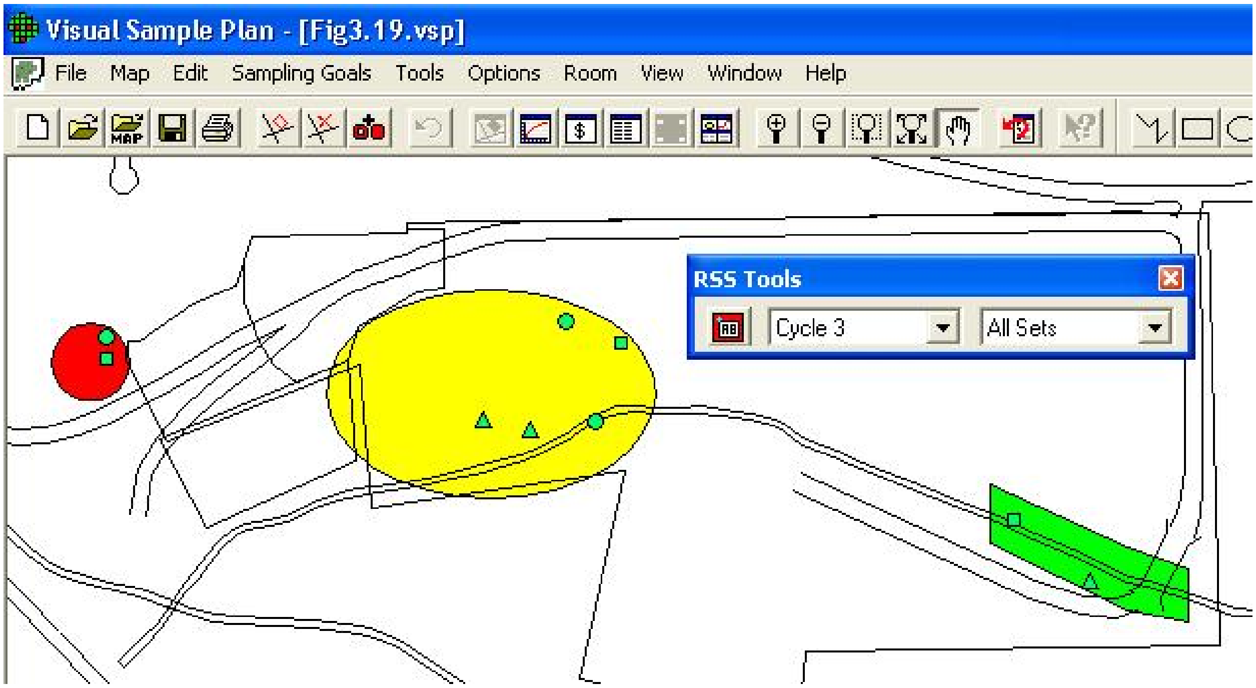


Figure 3.19. Map of RSS Field Sample Locations for All Sets in Cycle 3, Along with RSS Toolbar

Ranked set field sampling locations are generated with a label having the following format: RSS-c-s-i

where c = the cycle number

s = the set number (the unbalanced design for this number is also incremented for each iteration of the top set)

i = a unique identifier within the set.

Use **View > Labels > Labels** on the main menu or the **AB** button on the main toolbar (button also on the RSS toolbar) to show or hide the labels for the field sample locations. Figure 3.20 shows the labels on the map for field sample locations associated with **Cycle 3, All Sets**.

3.2.3.3 Adaptive Cluster Sampling

Adaptive cluster sampling is the third option for the Sampling Goal: **Estimate the Mean > Data not required to be normally distributed**. Because adaptive designs change as the results of previous sampling become available, adaptive cluster sampling is one of the two VSP designs that require the user to enter sample values while planning a sampling plan. (The other design that requires entering results of previous sampling is sequential sampling; see Section 3.2.1). The VSP Help, the VSP technical report (Gilbert et al. 2002), and the EPA (2001, pp. 105-112) are good resources for understanding what is required and how VSP uses the input to create a sampling design. A simple example here will explain the various input options. The user should have gone through the DQO process prior to encountering this screen because it provides a basis for inputs.

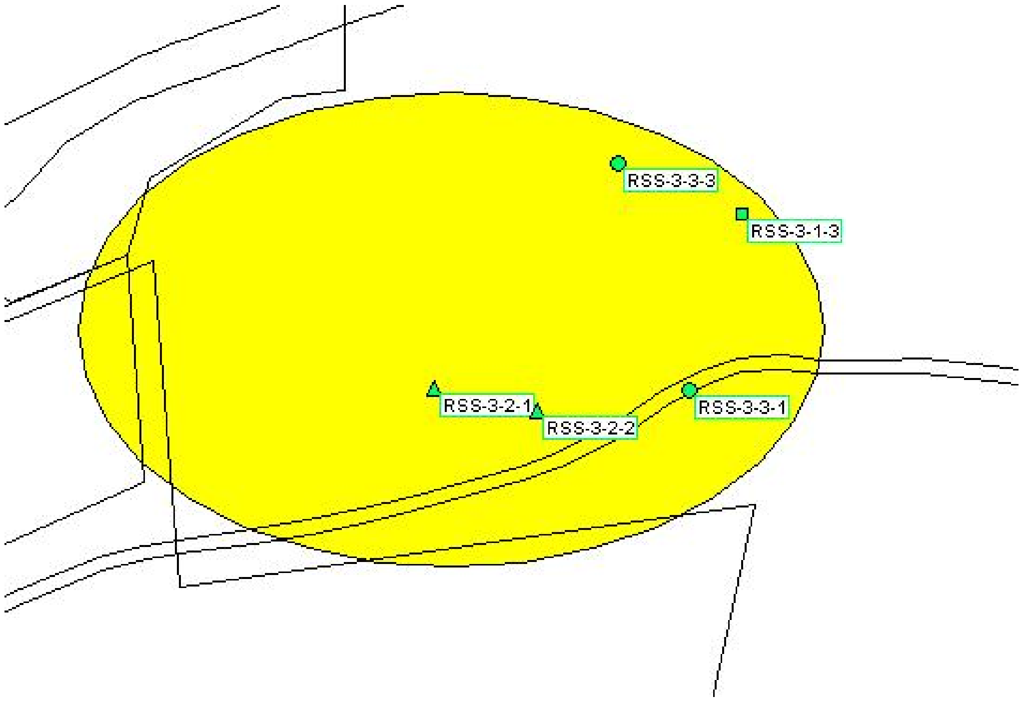


Figure 3.20. Map of RSS Field Sampling Locations along with their Labels

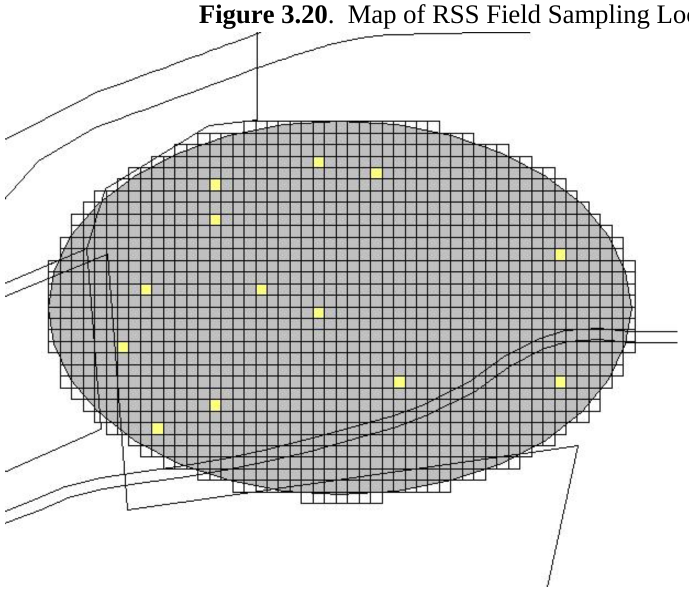


Figure 3.21. Map of the Sample Area with Grid Cells Displayed

The screen for entering values in the dialog box is displayed by selecting the tab **Number of Initial Samples**. Adaptive cluster sampling begins by using a probability-based design such as simple random sampling to select an initial set of field units (locations) to sample. To determine this initial sample number, either a one-sided or two-sided confidence interval is selected. We select **One-sided Confidence Interval** and enter that we want a **95%** confidence that the true value of the mean is within this interval. We want an interval width of at least 1 and we estimate the standard deviation between individual units in the population to be **2** (units of measure for interval width and standard deviation is same as that of individual sample values). VSP returns a value of **13** as the minimum number of initial samples we must take in the Sample Area.

In Figure 3.22, we can see the 13 initial samples as yellow squares on the map.

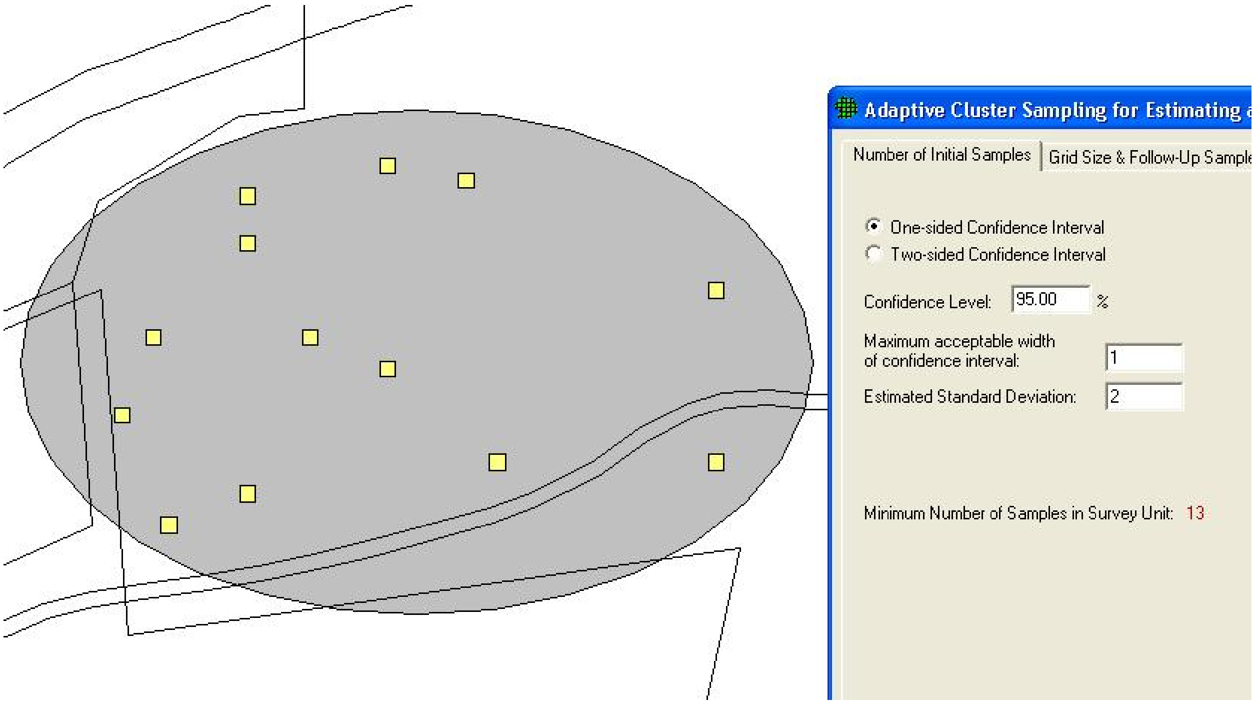


Figure 3.22. Map of Sample Area with Initial Samples for Adaptive Cluster Sampling Shown as Yellow Squares, Along with Dialog Box

The user now enters the analytical measurement results for the initial 13 sampling units. (Adaptive cluster sampling is most useful when quick turnaround of analytical results is possible, e.g., use of field measurement technology.) Place the mouse directly over each sample and right-click. An input box appears as shown in Figure 3.23. Enter a **measurement** value (shown here as 8) and, if desired, a **label** (shown here as **AC1-25-62**). Press **OK**. Enter another sample value and continue until all 13 sample values have been entered.

Select tab **Grid Size & Follow-Up Samples** on the Adaptive Cluster for Estimating a Mean dialog box. Enter the desired Grid Size for Samples, shown here as **20 ft**, and an upper threshold measurement value that, if exceeded, triggers additional sampling. We chose **10** as the threshold. We have a choice of how to expand sampling once the threshold is exceeded: 4 nearest neighbors or 8 nearest neighbors. We choose **4**. The dialog box is shown as the insert in Figure 3.24a. The grid units can be orientated at different angles by s**electing Edit > Sample Areas > Set Grid Angle and Edit > Sample Areas > Reset Grid Angle** from the main menu.

Once **Measurement** values have been entered, the yellow squares turn to either green, indicating the sample did not exceed the threshold, or red, indicating the sample exceeded the threshold. The red samples are surrounded with additional yellow squares that now must be sampled. This process continues until there are no more yellow grid cells. In Figure 3.24b, we see examples of green, single yellow, red surrounded by yellow, and red surrounded by green. Sampling and measurement continues until all the initial samples are green or red and all the added samples are green or red.

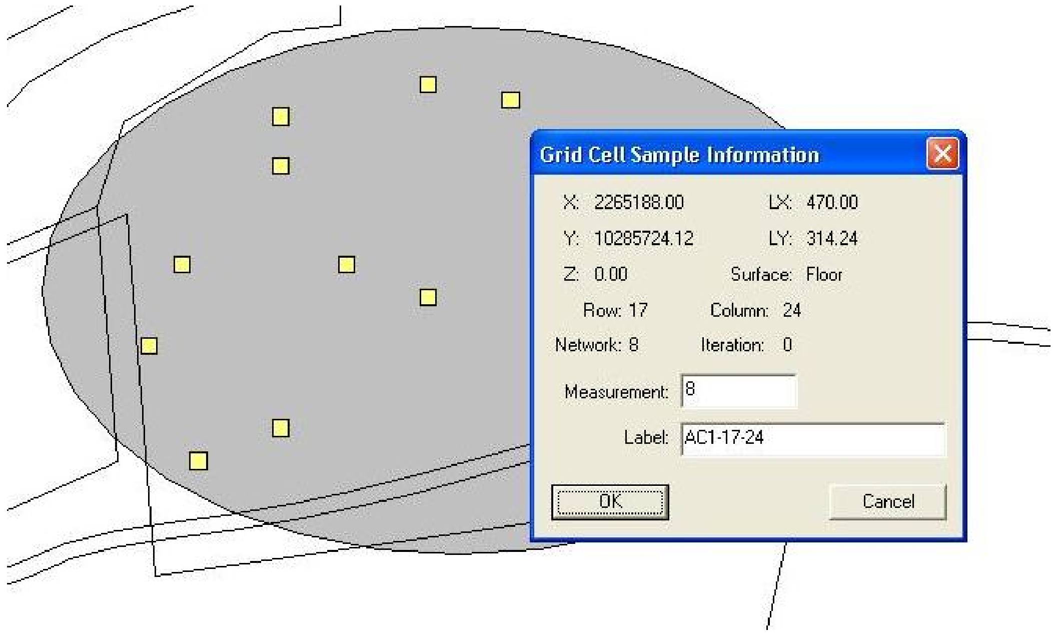


Figure 3.23. Dialog Input Box for Entering Sample Measurement Values and Labels for Initial Samples in Adaptive Cluster Sampling

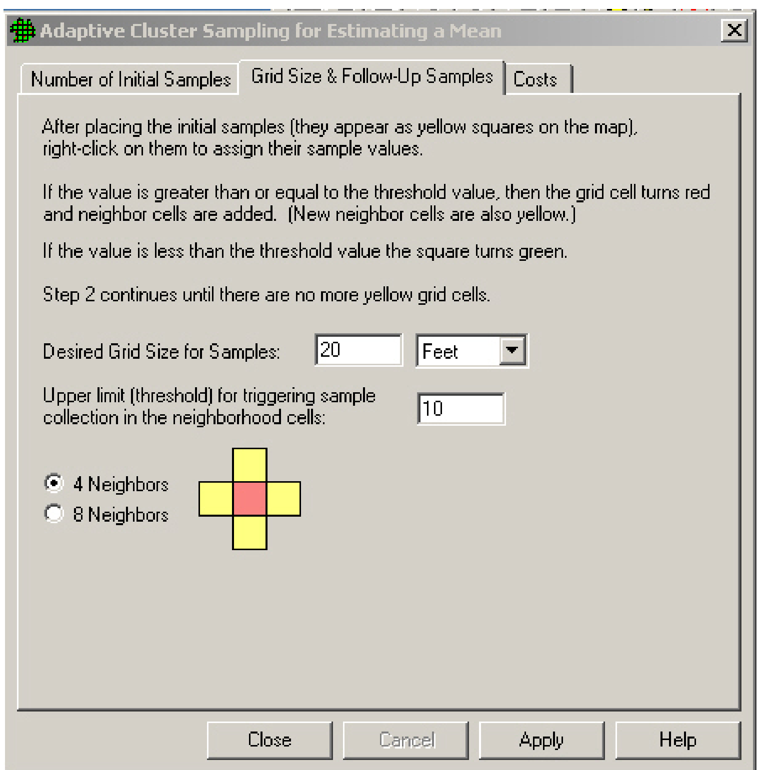


Figure 3.24a. Dialog Input Box for Entering Grid Size and Follow-up Samples

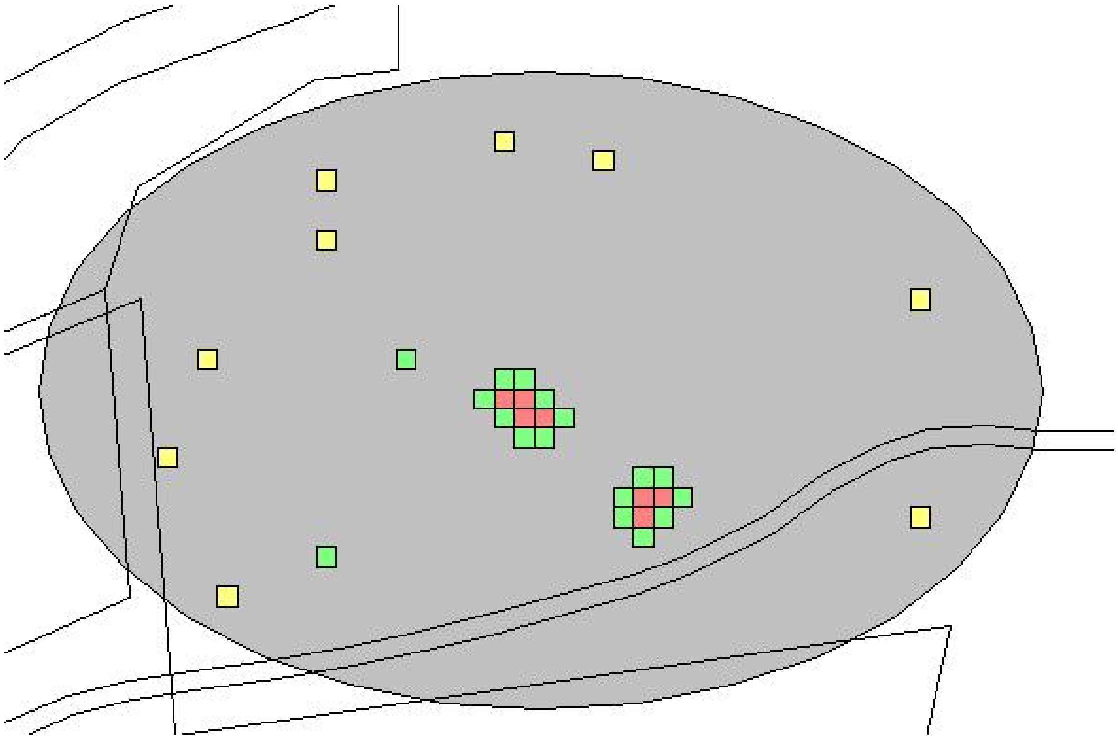


Figure 3.24b. Examples of Combinations of Initial and Follow-up Samples from Adaptive Cluster Sampling

Costs are entered using the **Cost** tab on the dialog box. The **Report** for adaptive cluster sampling shows the total cost for all the initial samples plus follow-up samples and provides an (unbiased) estimate of the mean and its standard error. Refer to VSP's **Help** for a complete discussion of adaptive cluster sampling.

3.2.3.4 Collaborative Sampling for Estimating the Mean

The final design we discuss for a cost-effective option for estimating the mean is Collaborative Sampling (CS) – sometimes called Double Sampling. This design is applicable where two or more techniques are available for measuring the amount of pollutant in an environmental sample, for example a field method (inexpensive, less accurate) and a fixed lab method (expensive, more accurate). The approach is to use both techniques on a small number of samples, and supplement this information with a larger of number of samples measured only by the more expensive method. This approach will be cost-effective if the linear correlation between measurements obtained by both techniques on the same samples is sufficiently near 1 and if the less accurate method is substantially less costly than the more accurate method.

Collaborative Sampling works like this: At n' field locations selected using simple random sampling or grid sampling, the inexpensive analysis method is used. Then, for each of n of the n' locations, the expensive analysis method is also conducted. The data from these two analysis methods are used to estimate the mean and the standard error (SE: the standard deviation of the estimated mean). The method of estimating the mean and SE assumes there is a linear relationship between the inexpensive and expensive analysis methods.

VSP has an extensive discussion of CS in the **On-Line Help**. CS is also discussed in Gilbert (1987), Chapter 9, where you can find an actual Case Study using CS. In Figure 3.24c we show the input screen for Collaborative Sampling.

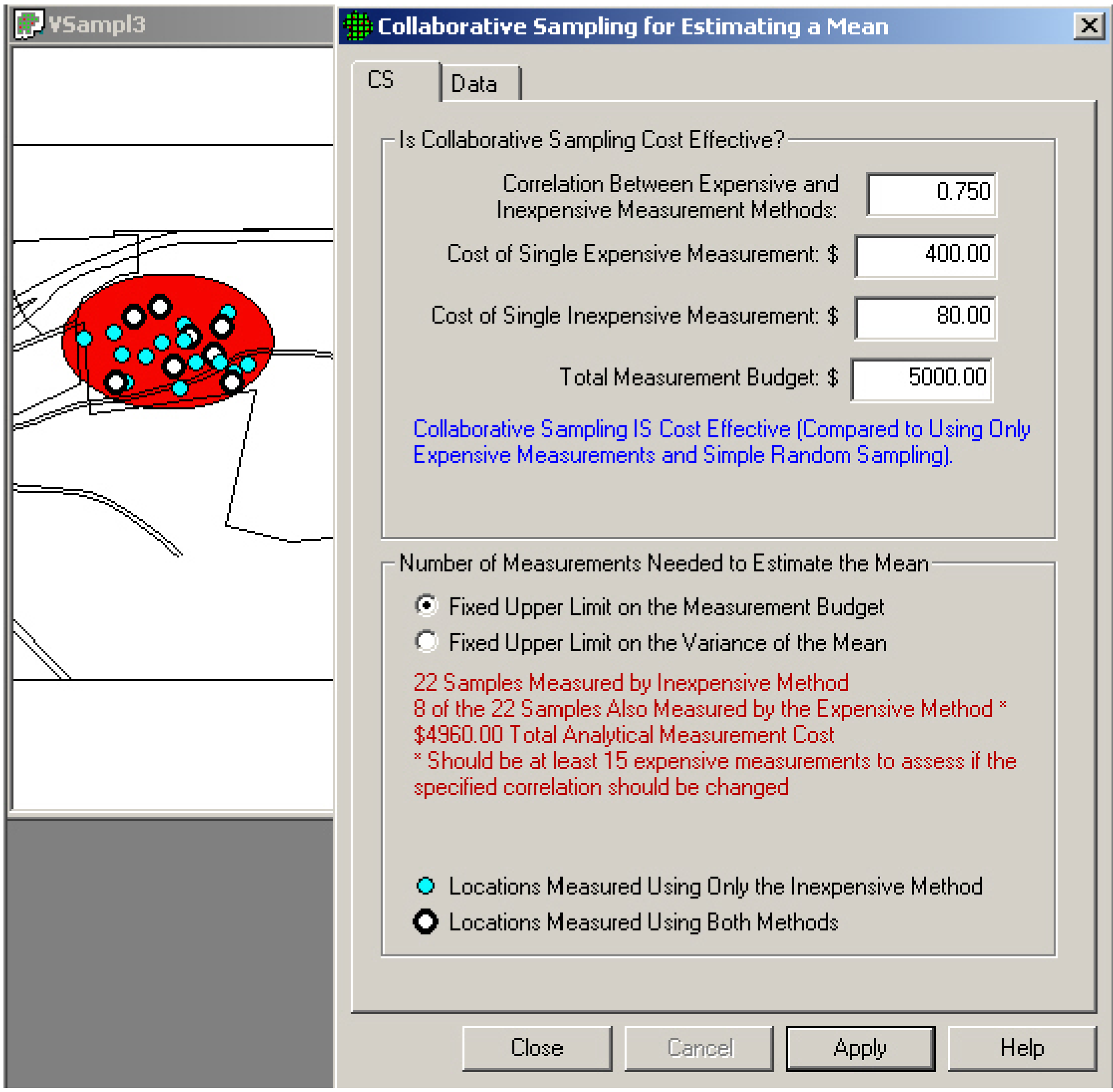


Figure 3.24c. Input Dialog Box for Collaborative Sampling for Estimating the Mean

For this example, we applied CS samples to an area on the Millsite map. After inputting the costs of each measurement technique, the total budget, and an estimate of the correlation between the two methods, VSP informs you whether or not CS is cost effective. For the vales we input, we see that it is cost effective. Then VSP uses the formulas discussed in the On-Line Help and the Report view to calculate two sample sizes, n' (22), and n (8). There are two options for optimizing the values of n' and n that the VSP user must select from:

- estimate the mean with the lowest possible standard error (SE: the standard deviation of the estimated mean) under the restriction that there is a limit on the total budget, or
- estimate the mean under the restriction that the variance of the estimated mean (square of the SE) does not exceed the variance of the mean that would be achieved if the entire budget were devoted to doing only expensive analyses.

We select the first option. VSP calculates that we need to take 22 samples and measure them with the inexpensive method, 8 of which are also measured using the more expensive methods. However, we get a warning message that we should be taking at least 15 measurements where we use both methods in order for VSP to assess whether our initial estimate of a 0.75 linear correlation coefficient is correct. Note that after we hit the **Apply** button, we see the sampling locations placed on the Sample Area we selected (Millsite.dxf used for this example).

As with Collaborative Sampling for Hypothesis Testing discussed in Section 3.2.1, VSP requires us to input the results of the sampling to verify that the computed correlation coefficient is close to the estimated correlation coefficient used to calculate the sample sizes. Data Results are input in the dialog box that appears after selecting the **Data** tab (see Fig 3.11). VSP calculates the estimated mean and standard deviation of the estimated mean once the data values are input.

3.2.4 Construct Confidence Interval on Mean

If the VSP wants a confidence interval on the true value of the mean, not just a point estimate of the mean as calculated in Section 3.2.3, the user selects **Sampling Goal > Construct Confidence Interval on the Mean**. Currently VSP has algorithms for only the case where the data can assumed to be normally distributed. Within that category, the user can choose a sampling design where the sample size is calculated assuming only a single type of sample is taken, **Simple Random Sampling** or **Systematic Grid Sampling**. For these two designs, three DQO inputs are required for the confidence interval sampling goal:

- the confidence you want to have that the interval does indeed contain the true value of the mean,
- how large that interval should be (width of confidence interval), and
- an estimate of the standard deviation between individual units of the population.

The user must also decide whether the confidence interval should be bounded on just one side of the mean (one-sided confidence interval) or on both sides of the mean (two-sided confidence interval). The two-sided confidence interval, smaller interval width sizes, and larger variation generally require more samples. In Figure 3.25, we see an example of the design dialog for the Confidence Interval on the Mean sampling goal, along with the recommended sample size of 38 that VSP calculated.

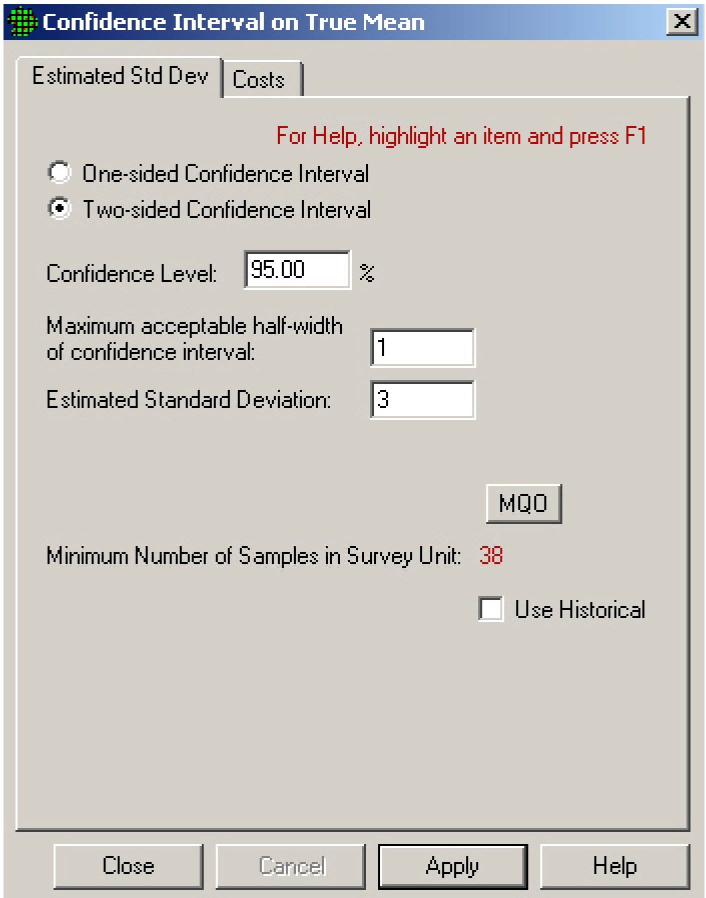


Figure 3.25. Dialog Input Box for Calculating a Confidence Interval on the Mean (Single Measurement Method Available)

If the user has more than one type of sample measurement method available, **Collaborative Sampling** should be explored to see if cost savings are available. Though not shown here, the inputs for Collaborative Sampling for Confidence Interval are similar to those in Figure 3.25, with the added cost inputs required to determine if Collaborative Sampling is cost effective (see discussion of Collaborative Sampling in Section 3.2.3.4).

3.2.5 Compare Proportion to Fixed Threshold

For comparing a proportion to a threshold (i.e., a given proportion), the designs available in VSP do not require the normality assumption. A one-sample proportion test is the basis for calculating sample size. The inputs required to calculate sample size are shown in the design dialog in Figure 3.26. The DQO inputs are similar to those for comparing an average to a fixed threshold, but since the variable of interest is a proportion (percentage of values that meet a certain criterion or fall into a certain class) rather a measurement, the action level is stated as a value from 0.01 to 0.99. Based on the inputs shown in Figure 3.26, VSP calculates that a sample size of 23 is required.

3.2.6 Compare Proportion to Reference Proportion

VSP formulates this problem as an environmental cleanup problem in which we have the proportion of contamination within a survey unit (Population 1) and we want to see if the difference between it and a reference area (Population 2) is greater (or less than) a specified difference. This specified difference becomes the action level. If we select the first formulation of the problem (P1 – P2 >= specified difference), we must enter a lower bound for the gray region. If we select the second formulation (P1 – P2 <= specified difference), we must enter an upper bound for the gray region. We must also enter our best guess of what we think the proportion of contamination is in both the survey unit and the reference unit. These two values are required to estimate the standard deviation of the proportions, which are then used as inputs to the sample size formula.

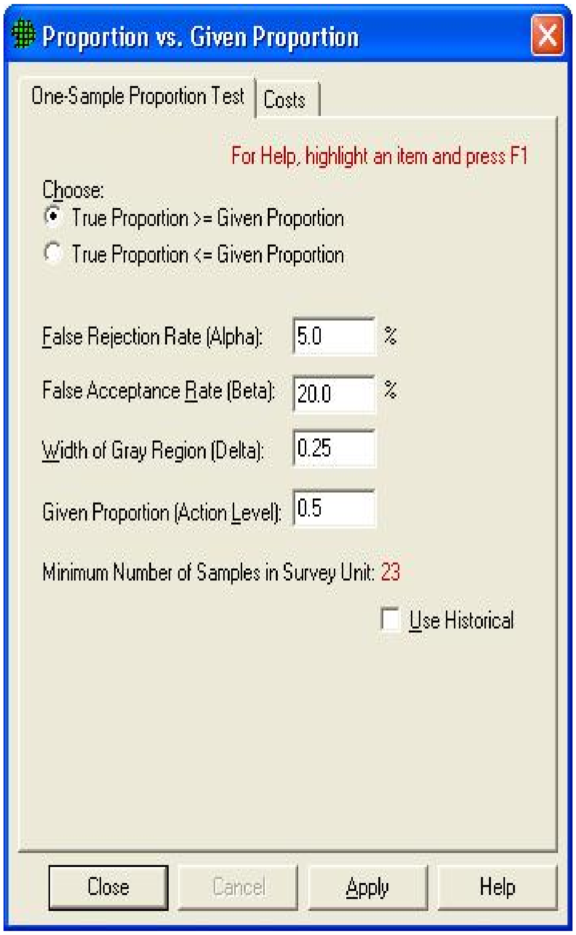


Figure 3.26. Design Dialog for Comparing a Proportion to a Fixed Threshold

Note that if the proportion of interest is the proportion of positive units in the environment, say the proportion of one-acre lots within a development area that have trees, then we need to select the null hypothesis that affords us the greatest protection against a false acceptance. In Figure 3.27, we see an example of the design dialog for this sampling goal. VSP calculates that we need 49 samples in the survey unit and 49 samples in the reference area for this set of inputs.

If no previous information is available on which to estimate the proportions in the survey unit or reference area, use 0.5 because at that value the sample sizes are the largest (i.e., the most conservative).

3.2.7 Estimate the Proportion

Similar to the designs available for estimating the mean, we see that VSP offers stratified sampling for the sampling goal of estimate the proportion because a stratified design may be more efficient than either simple random sampling or systematic sampling. Designs and sample size formulas for a simple random selection of samples are not in version 3.0 of VSP but can be found in standard statistics textbooks.

Prior to running VSP to calculate sample sizes for the strata, the user must have pre-existing information to use as the basis for dividing the site into non-overlapping strata. The strata should be more homogeneous internally than for the entire site (i.e., all strata). They must be homogeneous in the proportion of units that fall into one classification or another. The strata are the individual Sample Areas that the user selected and can be seen using Map View.

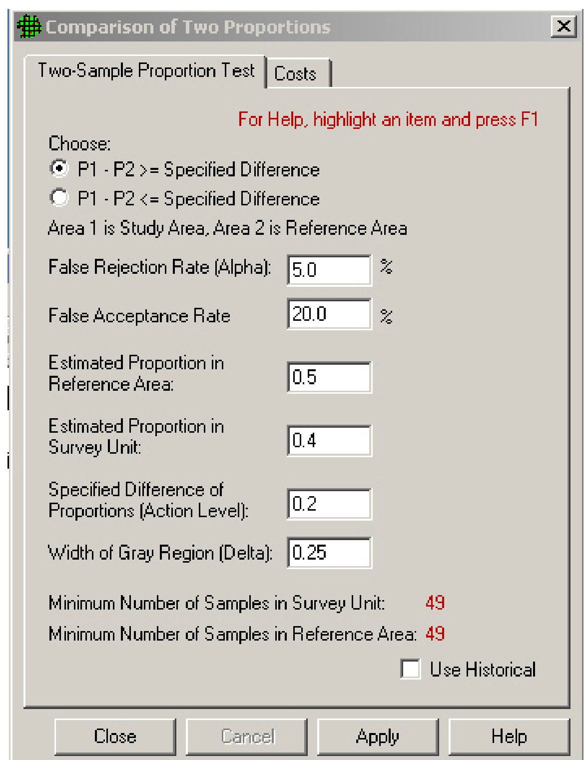


Figure 3.27. Input Dialog for Comparing a Site Proportion to a Reference Proportion

With the Sample Areas selected (VSP shows total number of areas in **Numbers of Strata**), the user may now open the dialog box. Note: Opening the dialog box prior to having Sample Areas selected will result in errors. Figure 3.28 shows the dialog box for one set of inputs.

The dialog box is separated into two blocks: the top deals with total number of samples in all strata; the bottom deals with allocation of total samples to individual strata. The user must select a method from the pull-down menu in each box. Different input is required depending on which method is selected. The **Help** function describes the various inputs required, why one method might be selected over another, and how they are used to calculate sample size. For methods where an estimate of the variance is required, the initial variance between individual units in the stratum is assigned the value **1**. It is in the same units as the mean. This is a critical value in the sample size calculation so the user should make sure this is a good estimate. In the bottom box, once values for stratum 1 are entered, the user selects the next stratum from the drop-down list under **Stratum #**.

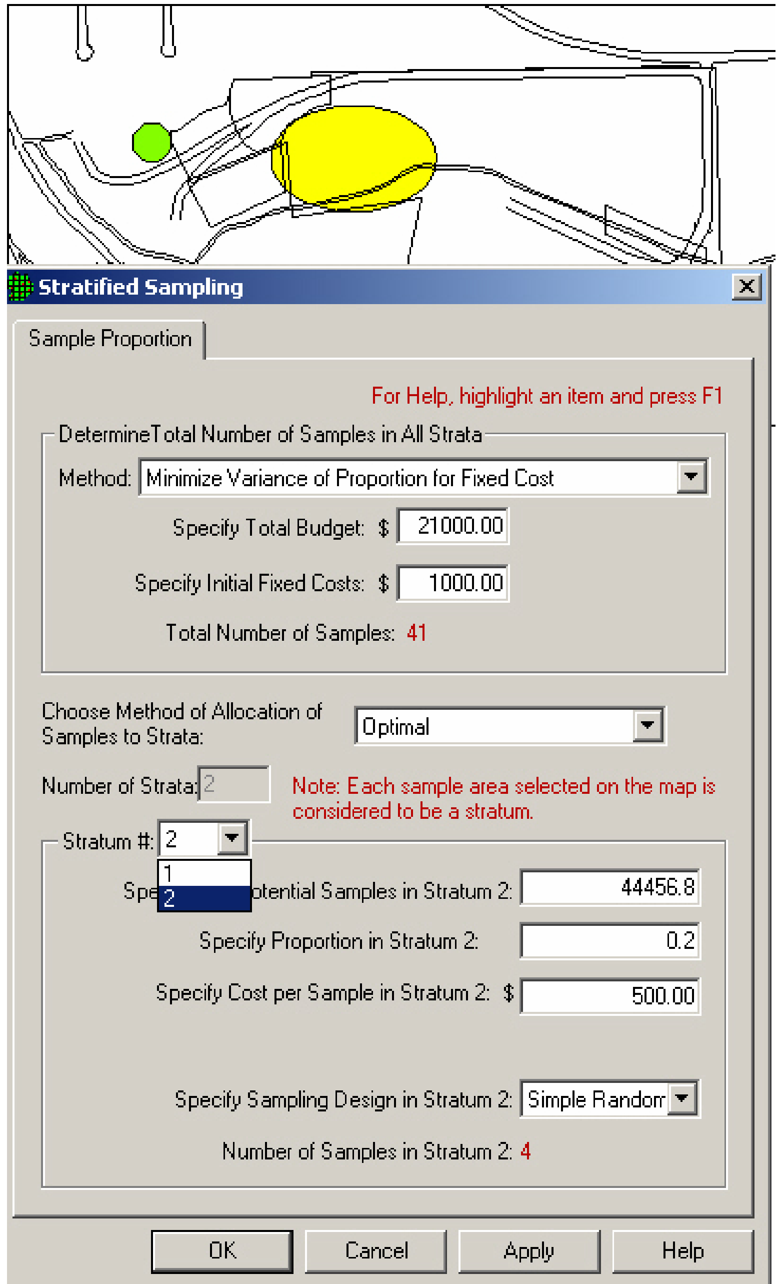


Figure 3.28. Dialog Box for Estimating a Proportion Using Stratified Sampling

VSP allows simple random sampling or systematic grid sampling within the strata. This is selected using the pull-down menu under **Specify Sampling Design in Stratum n**.

After supplying the required input, press **Apply** and the dialog shows in red the total number of samples and the number of samples required in each stratum (use the pull-down **Stratum #** to switch between strata). You can see the placement of samples within strata by going to **Map View**.

3.2.8 Locating a Hot Spot

The locating hot spots sampling goal is not the same type of problem and does not have the same type of algorithms for calculating sample size as the other sampling goals in VSP.

For one thing, the hot spot problem may be formatted in multiple ways:

- For a predetermined grid spacing, VSP will calculate the probability of finding a hot spot of a certain size.
- For a predetermined fixed cost, you can divide the area to be sampled into grids based on the budget, and VSP calculates the probability of finding a hot spot of a certain size.
- For a given probability and a given hot spot size, VSP calculates the minimum number of samples (maximum grid spacing) required to hit the hot spot.
- For a given probability and a given grid spacing, VSP calculates the smallest hot spot that can be detected.

The basic structure for these problems is that there are four variables (grid spacing, size of hot spot, probability of hitting a hot spot, and cost). You can fix any three of them and solve for the remaining variable.

The other unique feature of the hot spot problem is that there is only one type of error—the false negative or alpha error. VSP asks for only one probability for some formulations of the problem—the limit you want to place on missing a hot spot if it does indeed exist. The other error, saying a hot spot exists when it doesn't, cannot occur because we assume that if we do get a "hit" at one of the nodes, it is unambiguous (we hit a hot spot). We define hot spots as having a certain fixed size and shape, i.e., no amorphous, contouring hot spots are allowed. The hot spot problem is not a test of a hypothesis. Rather, it is a geometry problem of how likely it is that you could have a hot spot of a certain size and shape fitted within a grid, and none of the nodes fall upon the hot spot.

All the input dialog boxes for of the Hot Spot problem will not be shown in this user's manual. Rather, we demonstrate a common formulation of the problem: find the minimum number of samples to find a hot spot of a certain size, with specified confidence of hitting the hot spot. VPS's **Help**, and the textbook *Statistical Methods for Environmental Pollution Monitoring* (Gilbert 1997) are good resources for a complete discussion of the Hot Spot problem.

Problem Statement: A site has one Sample Area of one acre (43,560 square feet). We wish to determine the triangular grid spacing necessary to locate a potential circular pocket of contamination with a radius of 15 feet. We desire the probability of detecting such a hot spot, if it exists, to be at least 95%. The fixed planning and validation cost is $1,000. The field collection cost per sample is $50, and the laboratory analytical cost per sample is $100. Assume that the budget will be provided to support the sampling design determined from these requirements.

Case 9: We assume that the assumptions listed in Gilbert (1987, p. 119) are valid for our problem. We specify a hit probability of 95%, a shape of 1.0 (circular), and a radius (Length of Semi-Major Axis) of 15 feet. We will let VSP calculate the length of the side of the equilateral triangular grid needed for these inputs.

VSP Solution 9: First, open the file *OneAcre.vsp* using VSP Main Menu option **File > Open Project**. This is a VSP-formatted project file and it contains a previously defined Sample Area of the entire acre. Next, from the VSP Main Menu select **Sampling Goals > Locating a Hot Spot > Systematic Grid Sampling > Minimum Number of Samples**. A grouping of the input dialogs for the three tabs: **Find Grid, Hot Spot,** and **Costs** is shown in Figure 3.29.

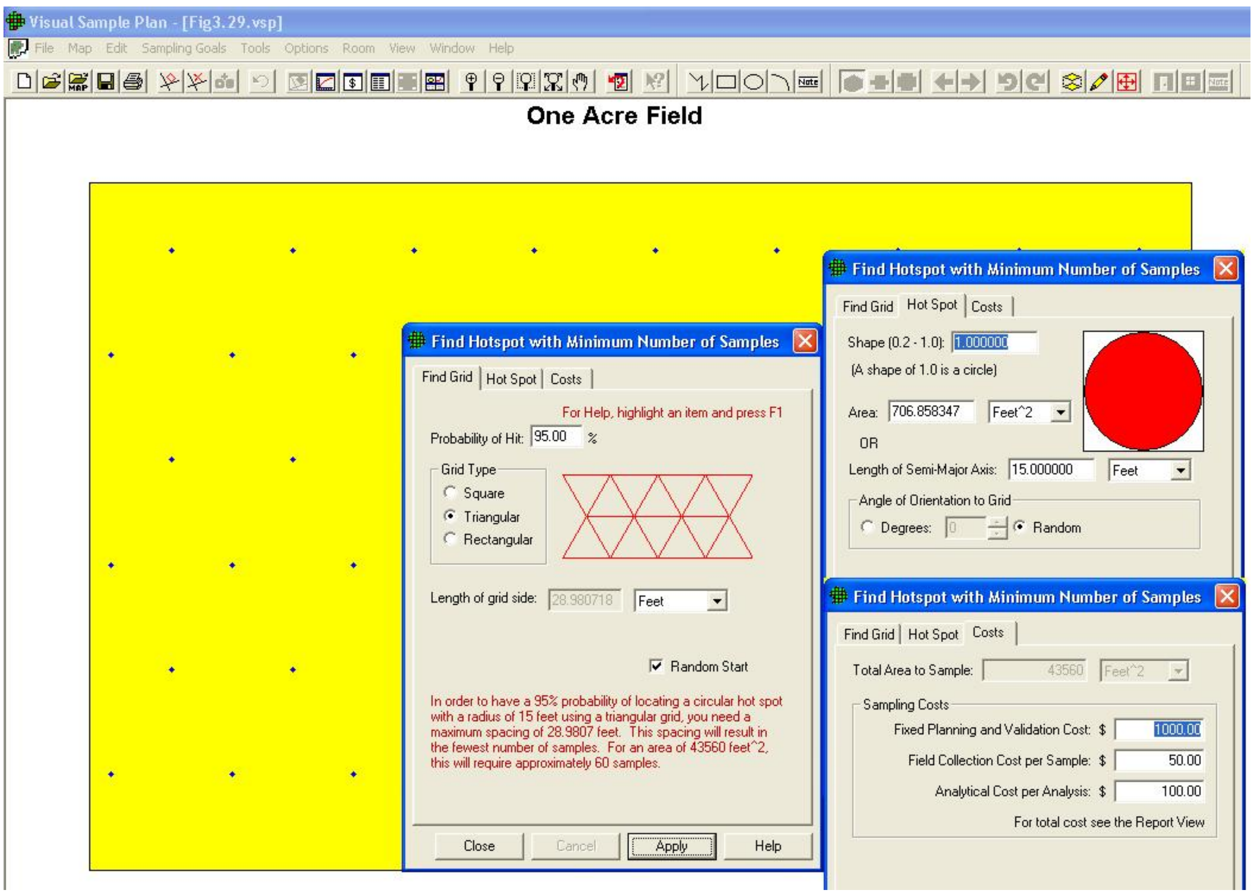


Figure 3.29. Input Boxes for Case 9

The recommended length of grid side is shown in the dialog box with the **Find Grid** tab. It is about 28.98 feet or, rounding up, a 30-foot triangular grid.

Note: For this set of inputs, VSP will always give the length of the triangular grid as 28.98 feet. The *Calculated total number of samples* in the **Report View** is always 60 for this set of inputs. However, the *Number of samples on the map* changes as you repeatedly press the **Apply** button. This occurs whenever the **Random Start** check box in the dialog box tabbed **Find Grid** is checked. Because the starting point of the grid is random, the way in which the grid will fit inside the Study Area can change with each new random-start location. More or fewer sampling locations will occur with the same grid size, depending on how the sampling locations fall with respect to the Sample Area's outside edges.

The input dialog boxes and report for the hot spot problem have some unique features:

- Placing the cursor in the **Length of Semi-Major Axis** on the **Hot Spot** tab and right-clicking displays a black line on the picture of the circle for the radius.
- **Shape** controls how "circular" the hot spot is. Smaller values (0.2) result in a more elliptical shape; 1.0 is a perfect circle.
- The user can specify the **Area** of the hot spot or the L**ength of the Semi-Major Axis**. Both fields have pull-down menus for selecting the unit of measurement.
- The Report provides additional information on the design such as the number of samples (both "on the map" and "calculated") and grid area.

The Hot Spot Sampling Goal takes into account the Total Area to Sample (see this field on the Cost tab) when calculating total number of samples. Many of the other designs use the standard deviation to control sample size.

3.2.9 Find UXO Target Areas

The next two Sampling Goals originated from specific unexploded ordinance (UXO) problems faced by the Department of Defense. The sampling designs the VSP developers came up with to address these problems are somewhat specialized. A separate User's Manual is available for the UXO Modules *(Version 2.0 Visual Sample Plan: UXO Module Code Description and Verification, PNNL -14267, Gilbert et al., 2003*). We present in this VSP 3.0 User's Manual a brief summary of the UXO designs.

At military ranges used for target practice, UXO could be expected to lie within a circular or elliptical "target area" of debris where ordnance missed the exact target and explosions scattered debris objects. The goal is to find these "target areas" by conducting geophysical surveys along transects (swaths) and detecting areas where metallic or explosive debris objects are more densely located than is expected in the (non target area) background. The swath width is the width of the geophysical sensor or sensor arrays; the spacing between swaths is calculated to achieve a specified level of confidence of finding circular or elliptical target areas. The problem separates into three distinct Sampling Goals:

- Find a swath spacing design that will locate elliptical or circular target areas
- In a post-survey situation, find the probability that a given set of transects will indeed transect and detect a target area of a specified size and shape
- Find areas of elevated anomaly densities in a set of pre-loaded swaths with anomalies placed on top of the swaths

Again, we point the user to the extensive VSP On-Line Help for these Sampling Goals, and the PNNL report referenced above for details not included in the brief discussion below.

3.2.9.1 Find Swath Spacing

The VSP menu selection **Sampling Goals > Find UXO Target Areas > Parallel or Grid Sampling**… brings up the dialog box shown in Figure 3.30. There are two parts to the problem: not only must the geophysical detector *traverse* (pass over) the target area (this is similar to the Hot Spot problem), it must also *detect* that a target area has been traversed (this is a simple false negative detection problem). Inputs for addressing these two parts to the problem are shown in the top and bottom boxes in Figure 3.30. In the top box, the desired probability of traversing the target, the swath pattern desired, and swath width of the detector are specified. VSP calculates the swath spacing required to detect the target defined under the **Target Zone** tab. In this example, VSP calculates a swath spacing of **23.33** ft is required, assuming a circular target of 10 ft radius, and a 10 ft swath width -- making the swath spacing 23.33 + 10 = 33.33 ft on centers.

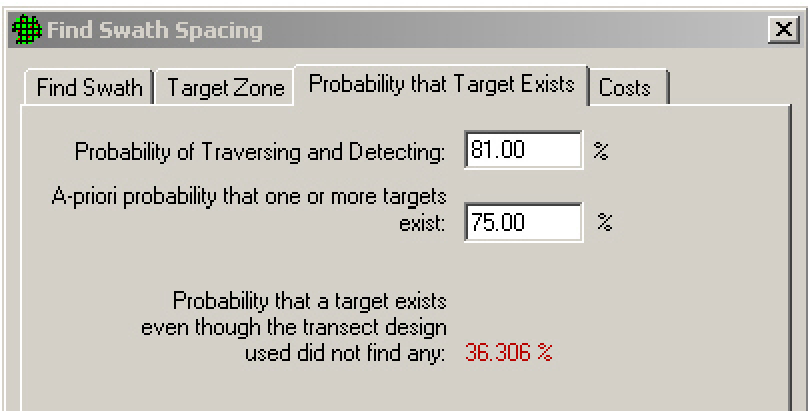


Figure 3.30. Dialog Input Box for Finding the Swath Spacing and Probability of Traversing and Detecting Target Area

In the bottom part of the dialog box, labeled **Target Detection**, the user inputs information about the density of objects or UXO expected within the target area and the false negative detection rate of the geophysical detector. Hitting the **Input** button brings up a separate dialog box for more completely describing the Target Area object density. The user must input the false negative detection error rate, which is the probability the geophysical sensor or sensor array does not detect an object when the sensor crosses over that object. Once all inputs are provided, the user hits the **Calculate** button and VSP computes the probability of both traversing *and* detecting the Target Area is **81%**.

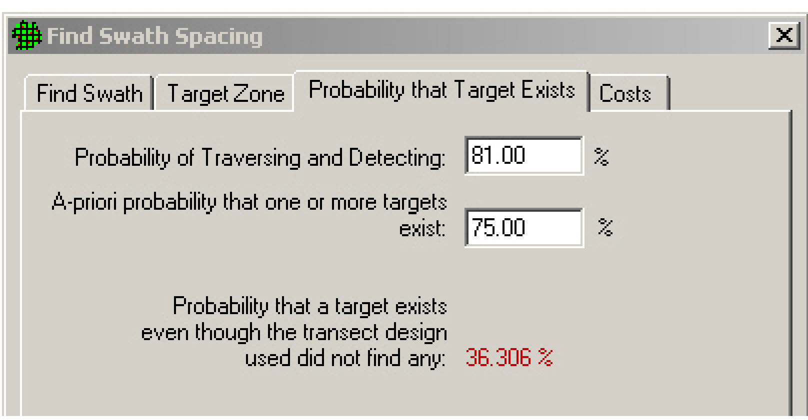


Figure 3.31. Dialog Input Box for Probability that One or more Target Areas Exist even though None were found in the Swath Survey

The **Probability that Target Exists** tab brings up a dialog box for calculating the probability that one or more target areas exists even though the transect design just specified did not find any. The input required for this is the Probability of Traversing and Detecting a target (calculated in dialog box shown in Figure 3.30 as 81%), and a user-input a priori probability that one or more targets exist (shown as 75% in this example). In Figure 3.31 we see that VSP calculates a probability of **36.3%** that a target exists even though the transect design did not find any. Note that the **Cost** tab for swath sampling requires slightly different input than the costs for individual samples.

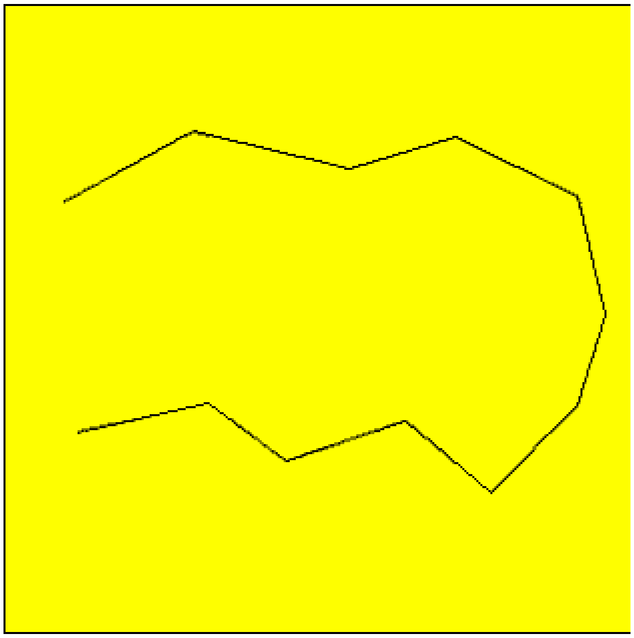


Figure 3.32a. Imported Meandering Swath (series of connected lines)

3.2.9.2 Post-survey Target Detection Evaluation

Menu selection **Sampling Goals > Find UXO Target Areas > Post-survey target detection evaluation** allows the VSP user to evaluate a user-input, actual set of transects/swaths and find the probability one or more targets areas would be traversed and detected by that set of transects.

We saw in Section 3.2.9.1 how to design a regular set of transects/swaths for detecting a circular or elliptical target area in which UXO or anomalous objects are suspected to exist. In practice, actual transect (swath) sampling plans won't follow regular parallel or rectangular patterns laid out by VSP. Field crews might find obstacles, dense vegetation, and other factors making it impossible to follow straight lines. To address this, VSP allows the user to input a set of "meandering swaths" (**Map > Meandering Swaths**), and then through simulation, calculates the probability of *traversing*, and *traversing and detecting* a user-specified Target Zone/Area. Figure 3.32a shows a **Map View** of a user-input meandering swath. Figure 3.32b shows a meandering swath with width reflecting the width of the detection equipment as it traverses the ground. The **Help** function describes this feature in more detail. Meandering swaths are loaded only if they fall inside a **Selected Sample Area**.

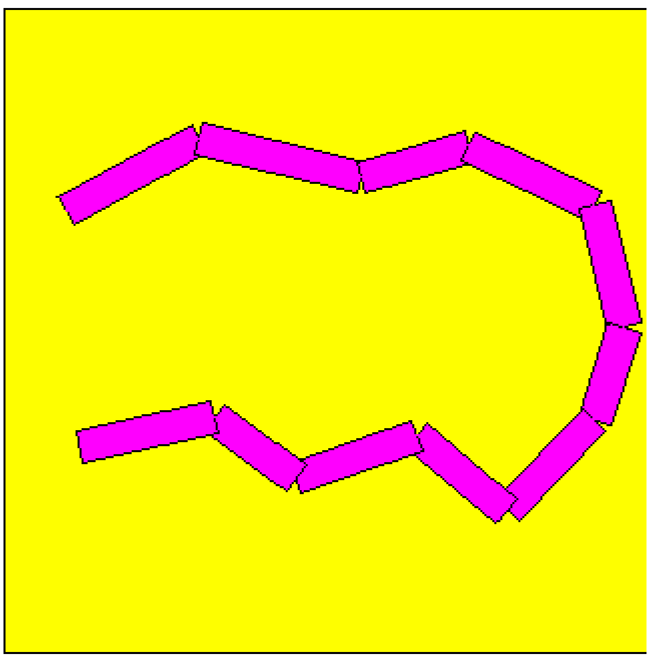


Figure 3.32b. Imported Meandering Swath with width (series of connected rectangles)

Menu selection **Sampling Goals > Find UXO Target Areas > Post-survey target detection evaluation** brings up the dialog box for inputting the information VSP needs to perform a simulation of the performance of a given set of transects/swaths. The dialog box is shown in Figure 3.33. Swath must be imported or created within a Sample Area before running the simulations.

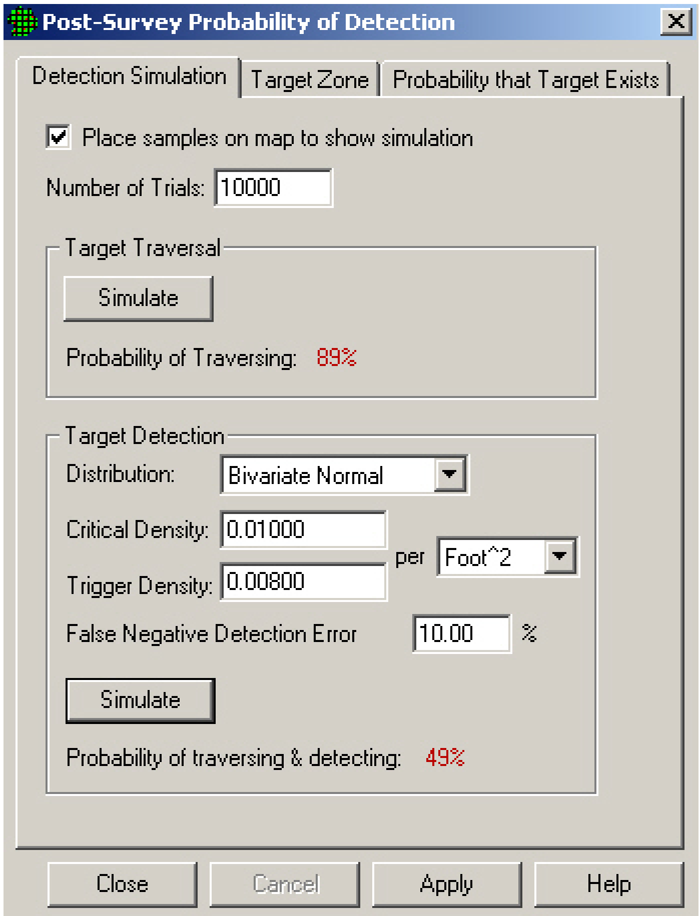


Figure 3.33. Input Dialog Box for Post-Survey Probability of Detection Using Simulation

The **Detection Simulation** tab has inputs required for the two simulations: Target Traversal, and Target Detection. The user selects the number of trials for the simulation, shown here as **10,000**. When the Target Transversal "**Simulate**" button is pushed, VSP throws down the 10,000 trial target areas and counts the number that touch one of the swath segments. A "hit" occurs when a trial target area touches a swath segment. The Probability of Traversing is computed as the number of hits divided by the number of trial targets, computed as **89%** for this example.

The inputs for the Target Detection are similar to those required for Find Swath Spacing (Section 3.2.9.1). When the Target Detection "**Simulate**" button is pushed, VSP throws down the number of trial target areas and determines the probability that each target area will be traversed and detected, making use of the critical density of anomalies in the target area, the trigger density at which the user is concerned about finding, and the false negative error rate of the sensor for actually detecting anomalies. For the inputs shown, this probability is **49%**.

If the "Place samples on map to show simulation" check-box is selected, VSP places a point on the map where any trial target area was not traversed. Figure 3.34 shows the results of using the shown meandering path to traverse a target of shape s=0.5 (ellipse) and angle of orientation to the map = 45 degrees. The dark area in the figure is the dense cluster of locations where target areas exist without being traversed by any portion of the meandering path. All targets in the clear (yellow) area would be traversed by one or more meandering swath segments.

As with Find Swath Spacing, prior to running the simulations the user must input the size and orientation of the Target Area using the dialog box under the **Target Zone** tab. Also similar to Find Swath Spacing, VSP will calculate the probability a target exists even though the meandering swath transect design did not find any using the inputs supplied under the **Probability that Target Exists** tab.

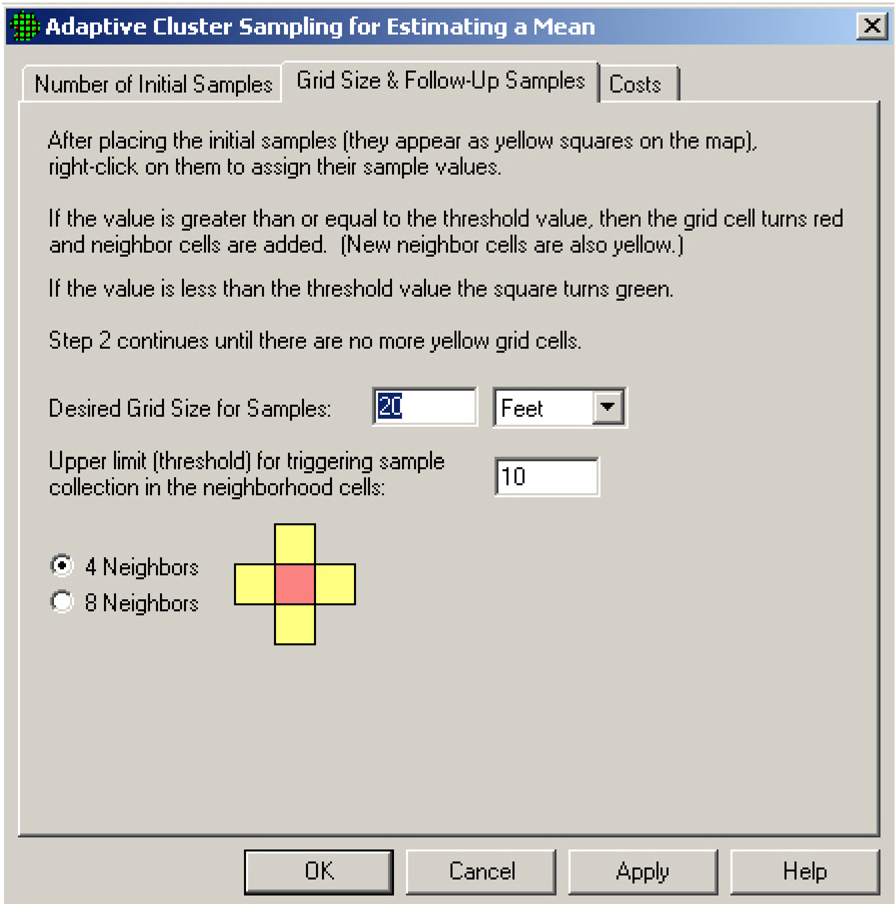


Figure 3.34. Map View of Simulation Results for Probability of Traversing and Detecting a Target Area

Menu selection Sampling Goals > Find UXO Target Areas > Find Target Areas brings up the dialog box shown in Figure 3.35 for finding areas of elevated densities. In order to use this tool, it is necessary to first load the anomalies as sample points on top of the swaths. Paste the anomaly coordinates into the Coordinate View, or use Map > Sample Points > Import.

When the Find Targets button is pressed, the tool searches along all the swaths checking each "window" (a linear or circular area) for elevated anomaly density and marking the locations on the Map based on the parameters set in this dialog. When the search is complete, a dialog is presented which displays a histogram of the frequency of windows at various densities. The Help function for this dialog box explains the inputs and Map display of the swaths and the targets.

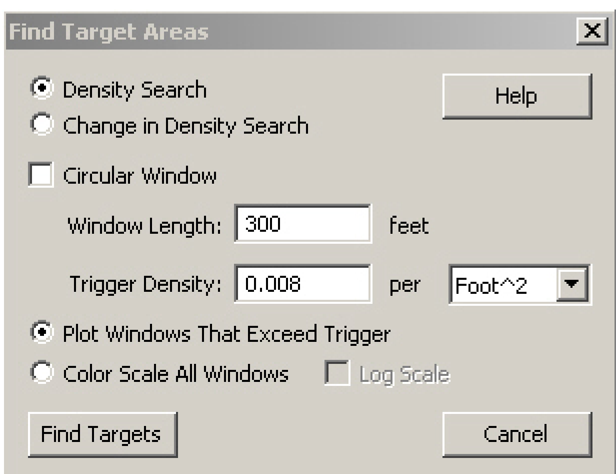


Figure 3.35. Dialog Input Box for Find Target Areas

3.2.10 Access Degree of Confidence in UXO Presence

The previous Sampling Goal dealt with large military ranges where remedial managers needed to find elliptical or circular target *areas* of multiple UXO or anomalous objects. In this Sampling Goal, if there are UXO, they are *single, isolated* objects not in any particular pattern. Two scenarios of interest are: a small UXO-remediated area where a survey is required to support a no further action (NFA) decision, and; a very large un-remediated area not expected to contain any UXO. In both scenarios, we assume the area of interest has a well-defined boundary. The area is divided into N non-overlapping, parallel transects with transect width equal to the width of the geophysical survey equipment. If N is large and only n of these transects can be surveyed, how should n be determined? This is similar to a compliance sampling problem in an industrial quality control setting where the goal is determine if there are any defects before releasing a product.

VSP addresses two variants of the problem:

- n must be of sufficient size to achieve a high confidence that *few* transects contain UXO, and
- n must be of sufficient size to achieve a high confidence that *no* transects contain UXO.

3.2.10.1 Achieve high confidence that *few* transects contain UXO

Menu selection Sampling Goals > Assess Degree of Confidence in UXO Presence > Achieve high confidence few transects contain UXO brings up the dialog box in Figure 3.36. With our familiar

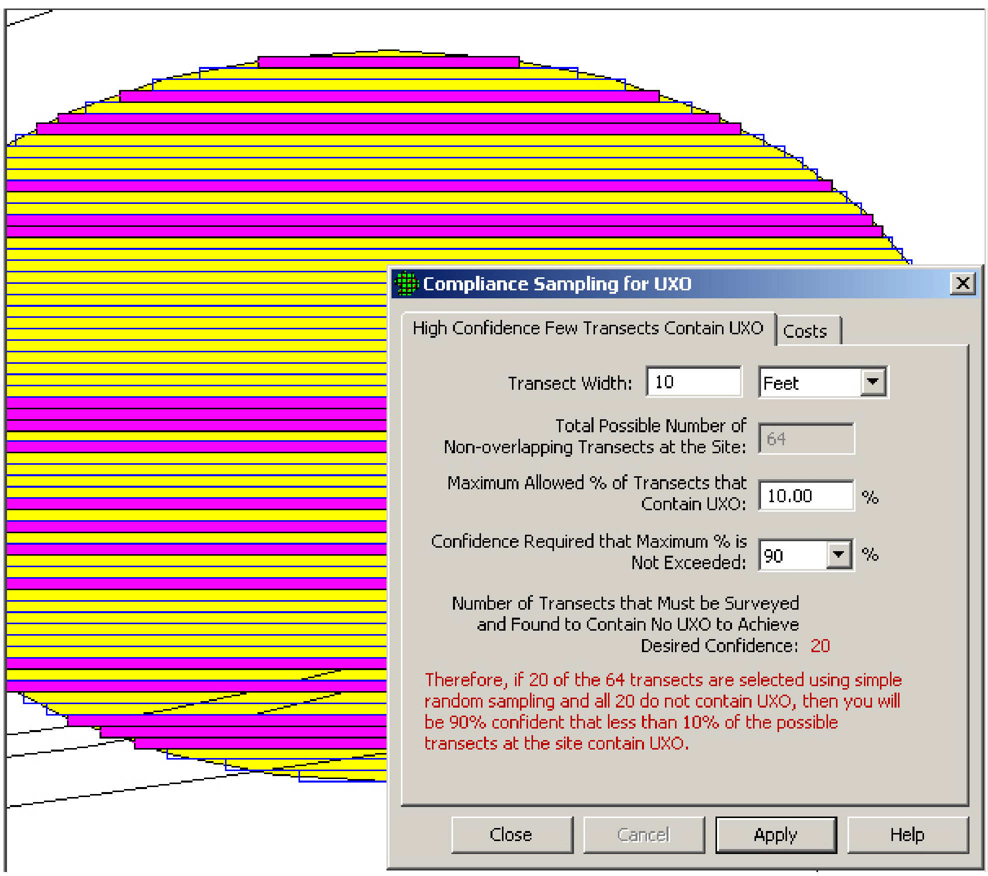


Figure 3.36. Dialog Input Box for Compliance Sampling for UXO and Map of Sample Area with Transects Selected

Millsite.dxf map file loaded, and the large elliptical area in the center of the map selected as our Sample Area, we input a transect width of 10 ft. VSP informs us that there are N= 64 total non-overlapping transects within this Sample Area. The 64 transects are shown in yellow in Figure 3.36. The VSP user inputs 10% as an upper limit on the maximum allowed percent of transects that could contain one or more UXO (DQO input), and 90% as the confidence that this maximum percent is not exceeded.

VSP calculates that 20 of the 64 transects must be selected, using simple random sampling, and none of these transects can contain UXO in order to be "90% confident that less than 10% of all possible transects contain UXO". The n=20 randomly selected transects that VSP placed on the Map when the Apply button is pushed are shown in pink in Figure 3.36. The algorithm for calculating sample size n programmed into VSP for this problem is a method developed by E.G. Schilling called "compliance sampling". For extensive discussion of this problem and compliance sampling methods consult VSP's Help and Gilbert, et al (The Schilling and Wright/Grieve Methods for Assessing Compliance with USO Requirements, 2003), or Gilbert, et al. (Version 2.0 Visual Sample Plan (VSP): UXO Module Code Description and Verification, PNNL-14267, 2003; Chapter 6).

3.2.10.2 Achieve high confidence that *no* transects contain UXO

Menu selection Sampling Goals > Assess Degree of Confidence in UXO Presence > Achieve high confidence no transects contain UXO bring up the dialog box in Figure 3.37. A sampling method for locating rare defects when there is strong prior evidence was developed by A.P Grieve and T. Wright. This method for finding n is programmed into VSP. It is a Bayesian method, meaning that stakeholders provide a quantitative measure of their *belief* that the study Sample Area contains UXO. This belief should be based on all information and data collected about the study area and the conceptual site model developed for the area.

With the Millsite.dxf map file loaded, and the large elliptical area in the center of the map selected as our Sample Area, we input a transect width of 10 ft. We require a 95% confidence that none of the possible 64 transects contain UXO. We say our prior belief that the percent of transects that contain UXO is Unknown Percentage. VSP interprets to mean that our prior belief on the percent of transects with UXO is 50%. If we had selected Extremely Low Percentage from the pull-down menu, VSP would have input our prior belief as 0.1%. If we had selected Extremely High Percentage, VSP would have input our prior belief as 99.9%.

VSP informs us that, we need to sample n= 61 of the 64 transects, and none of the 61 may contain UXO in order to say "we are 95% confident that none of the remaining (3) transects contain UXO". The 61 randomly selected transects are colored pink on the map in Figure 3.37. The sample size n is sensitive to the prior belief percentage:

- If the prior belief is 0.1% (Extremely Low Percentage) transects with UXO, n= 11.
- If the prior belief is 1% (Very Low Percentage) transects with UXO, n= 56
- By the time the prior belief is 90% (High Percentage) transects with UXO, n=64, all transects must be surveyed and you are 100% confident no transects contain UXO.

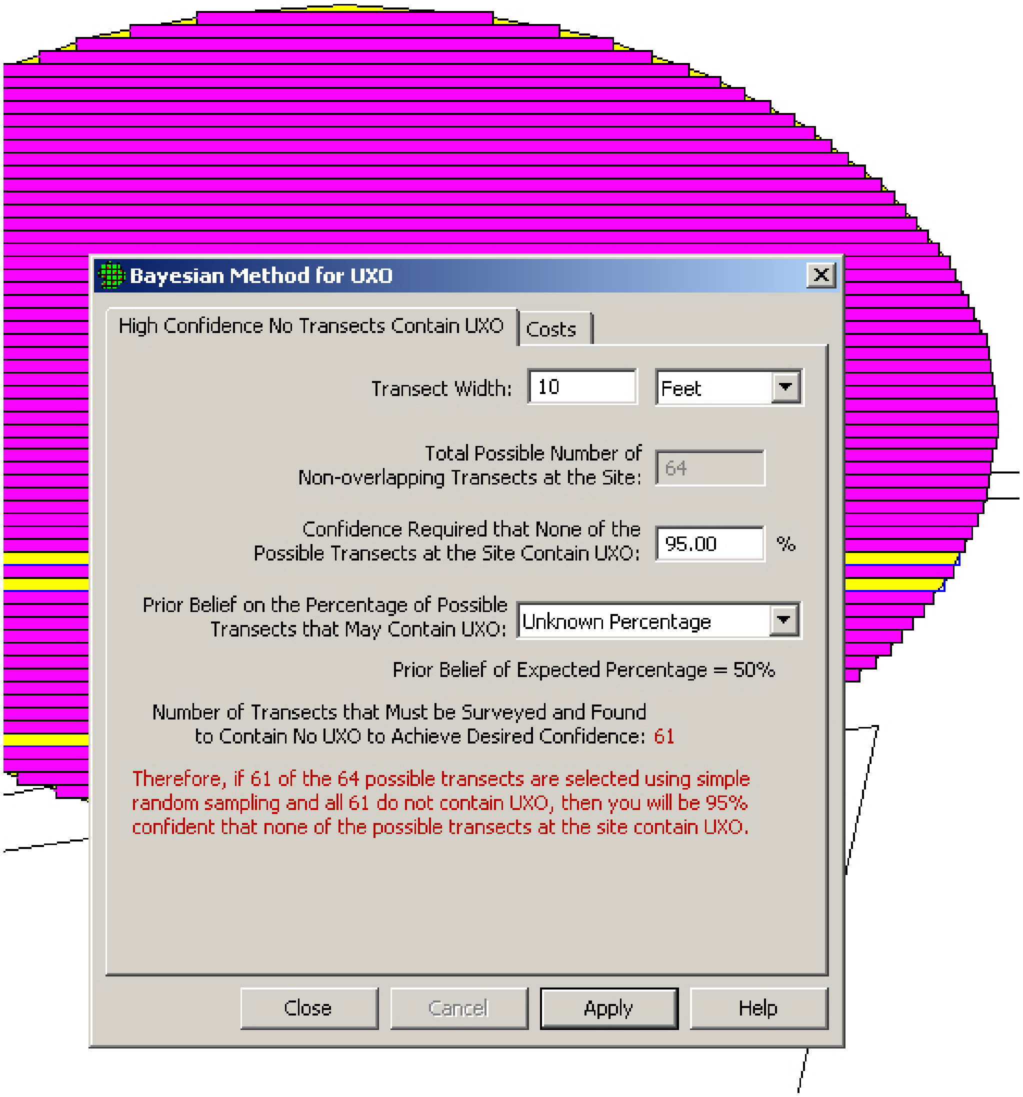


Figure 3.37. Dialog Input Box for Bayesian Method for UXO and Map of Sample Area with Transects Selected

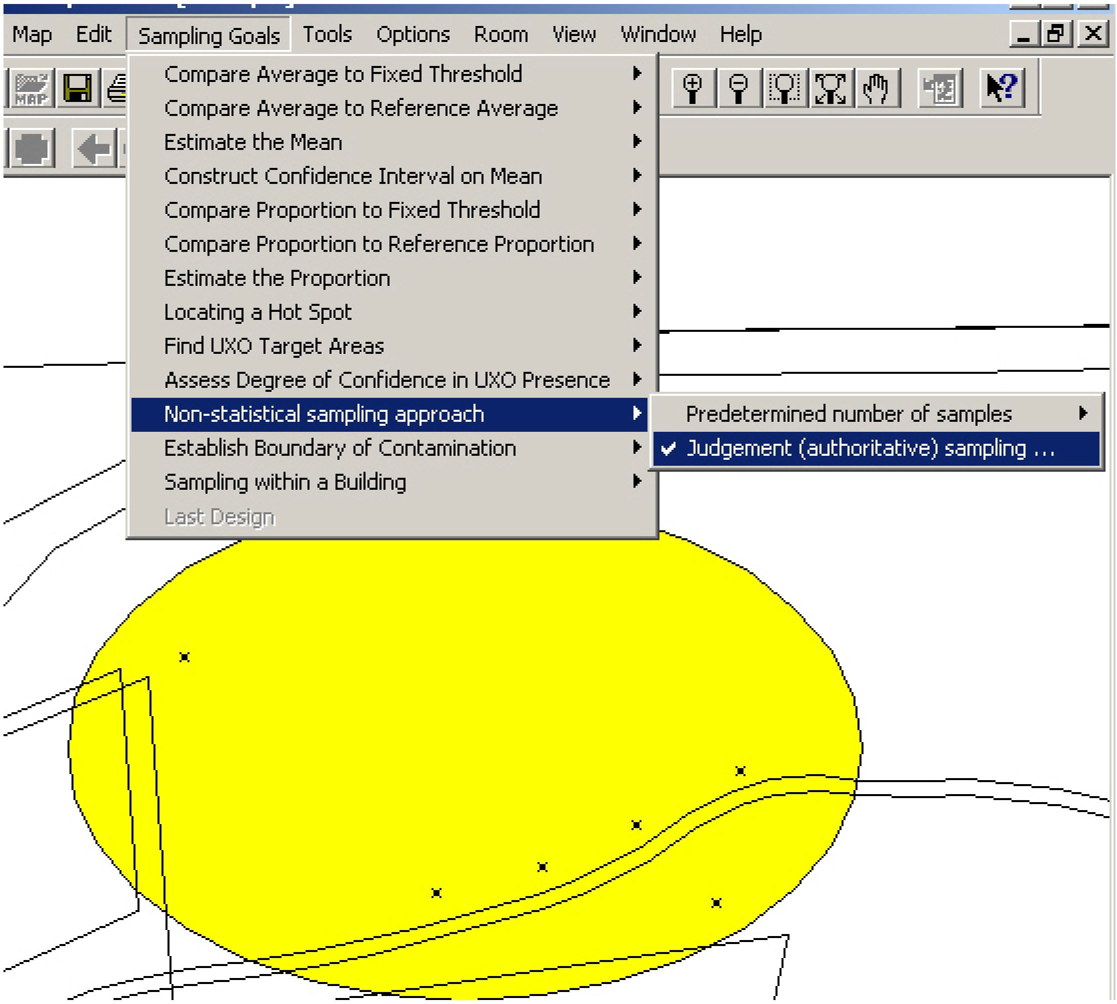


Figure 3.38. Judgment Sampling with Six Sampling Locations Added Manually

3.2.11 Non-statistical Sampling Approach

VSP allows the user to directly place samples in a Sample Area without going through the Sampling Goals and the DQO Process. If the user has a pre-determined number of samples, possibly obtained from a prior DQO study, VSP allows the user to input a sample size and place the samples within the Sample Area using either a random design or a systematic design. Menu selection Sampling Goals > Non-Statistical Sampling Approach > Predetermined Number of Samples brings up a simple dialog box where the user can input any value for Number of Samples, and by hitting the Apply button, the samples are placed in the Sample Area according to the design specified (random or systematic).

VSP allows user to manually place samples on a Map within a selected Sample Area using menu selection: Sampling Goals > Non-Statistical Sampling Approach > Judgment (authoritative) Sampling. This option is available only if View>Map is selected and a Sample Area defined. Judgment Sampling is a toggle switch. When it is turned on, any time the user clicks on the map, a sample marker is placed at that location. Judgment samples can be added to a blank Map or to an existing design. The Type is

"Manual" (see View > Coordinates). Manual samples may also be added by typing the coordinates (x, y) on the keyboard.

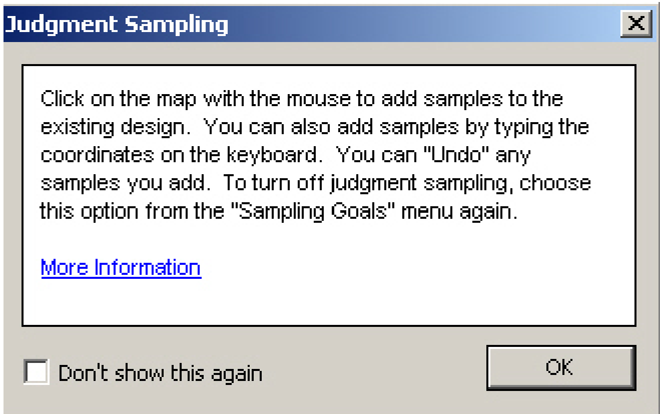


Figure 3.39. Information Dialog Box for Judgment Samples

In Figure 3.39, 6 samples have manually been added using Judgment Sampling. In Figure 3.39 we see an information dialog box for judgment sampling

3.2.12 Establish Boundary of Contamination

Finding the boundary of contamination is a problem faced by Department of Defense remediation managers. Training ranges or areas where the soil is known to contain explosive residues (or other contaminants of concern) may have boundaries that completely or partially enclose the contaminated area. Sampling is required to determine whether contamination has breached a known boundary line and if so, determine the correct boundary line. VSP has a special module for this sampling problem. The problem and the VSP solution are described in *Visual Sample Plan User's Guide for Establishing the Boundary of Contamination*, R.O. Gilbert, et al, PNNL-XXXX, 2004. In this Version 3.0 User's Guide we will provide a summary description of the VSP boundary module.

The VSP sampling design for this problem involves taking a representative sample (called a multiple increment or MI) for each segment along the known, user-input boundary. If the one or more samples show contamination, extend or "bump out" the boundary, and take more samples. The boundary continues to be bumped out until all samples taken along the new boundary line are "clean".

In Section 2.X, we described how to define enclosing and partial boundaries in VSP using Edit >Sample Areas > Define New Sample Area, and Edit > Sample Areas > Define New Open-Type Sample Area, respectively. VSP determines the number of segments using the length of the boundary and the specified width of a contaminant plume (hot spot) that would be of concern if it is present at the boundary or extends beyond the boundary line. VSP calculates the optimum segment length (OSL) along the current boundary, where all segments have the same length. One or two MI samples are collected per segment. VSP assumes that each MI sample collected in a segment consists of 25 small soil samples (increments) that have been collected in sets of 5 small samples clustered around each of 5 equally spaced Primary Sampling Locations along the segment. The spacing of the five segments depends on the specified width of the hot spot of concern at the boundary. The OLS is calculated as 5 times the user-specified width of the contamination plume (hot spot) of concern.

VSP provides two versions of the design: one for enclosing boundaries and one for partial (open-type) boundaries. Partial boundaries represent a dividing line, with contamination on one side and no contamination on the other side. VSP provides special tools for creating and manipulating open-type sample areas.

3.2.12.1 Enclosing Boundary

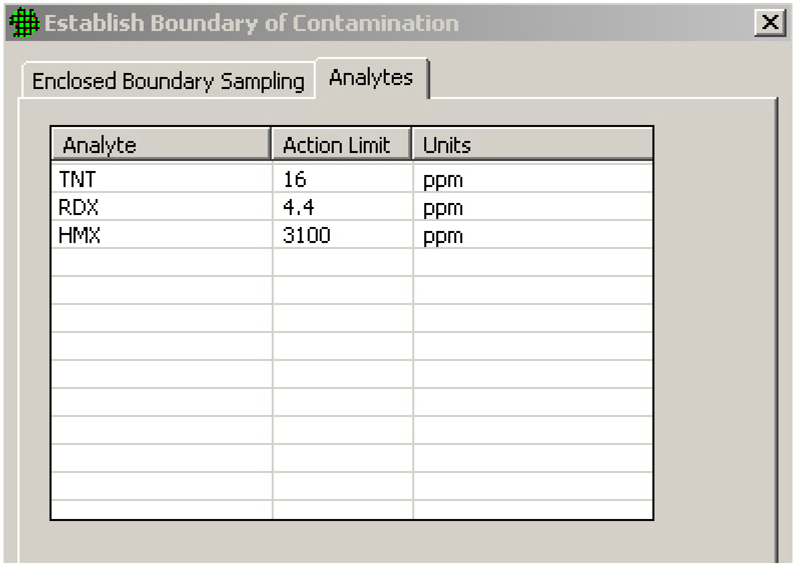


Figure 3.40. List of Default Contaminants of Concern and their Action Levels

Menu selection Sampling Goals > Establish Boundary of Contamination > Enclosed Boundary brings up the dialog box in Figure 3.40 for tab Enclosed Boundary Sampling. The first input required is the confidence needed that the mean calculated from limited sample data is indeed less than the action limit. For this example, that confidence level is 95%. The diameter of the area of contamination (i.e., the hot spot) that the user wants to be sure is detected at the boundary is input as 45 ft. The next box, labeled Duplicate Requirements, has to do with how many of the segments need duplicate MI samples to be collected. VSP requires that: at least 5 segments; or at least 10% of the segments, need duplicates. The user may select which requirement is used. While 10% is the minimum, the user may input any percentage for duplicates.

Note: The purpose of duplicate MI samples is to estimate the relative standard deviation of the data so that an Upper Confidence Limit (UCL) test can be conducted for each segment. See VSP Help for more information.

If the boundary of the site is very irregular, e.g., has various indentations, the VSP user can specify in the dialogue box that VSP should change the boundary to a convex hull. This has the effect of smoothing out the boundary irregularities, but it also enlarges the area enclosed by the initial boundary. In practice, the VSP user can try this option and view the resulting initial boundary to see if the new boundary is acceptable. In Figure 3.40 we leave this box unchecked.

The user now must input the contaminates of concern and the threshold (action level) at which we want VSP to trigger extending the contamination boundary line. The dialog box for tab Analytes is shown in Figure 3.41. VSP provides a default list of contaminants of concern (TNT, RDX, and HMX) and a default list of upper limit values (Action Limit) for each (16ppm, 4.4ppm, and 3100ppm). To remove a contaminant from the list, erase the name and the limit. To add a contaminant, enter its name and threshold value in the blank lint below the last contaminant.

For the Millsite.dxf map file selected, and the central ellipse in the center of the map selected as the Sample Area with an enclosing boundary, we see in Figure 3.42 that after clicking the Apply button in the previous screen, VSP divides the boundary into 17 segments. Shown are the 5 equally-spaced Primary

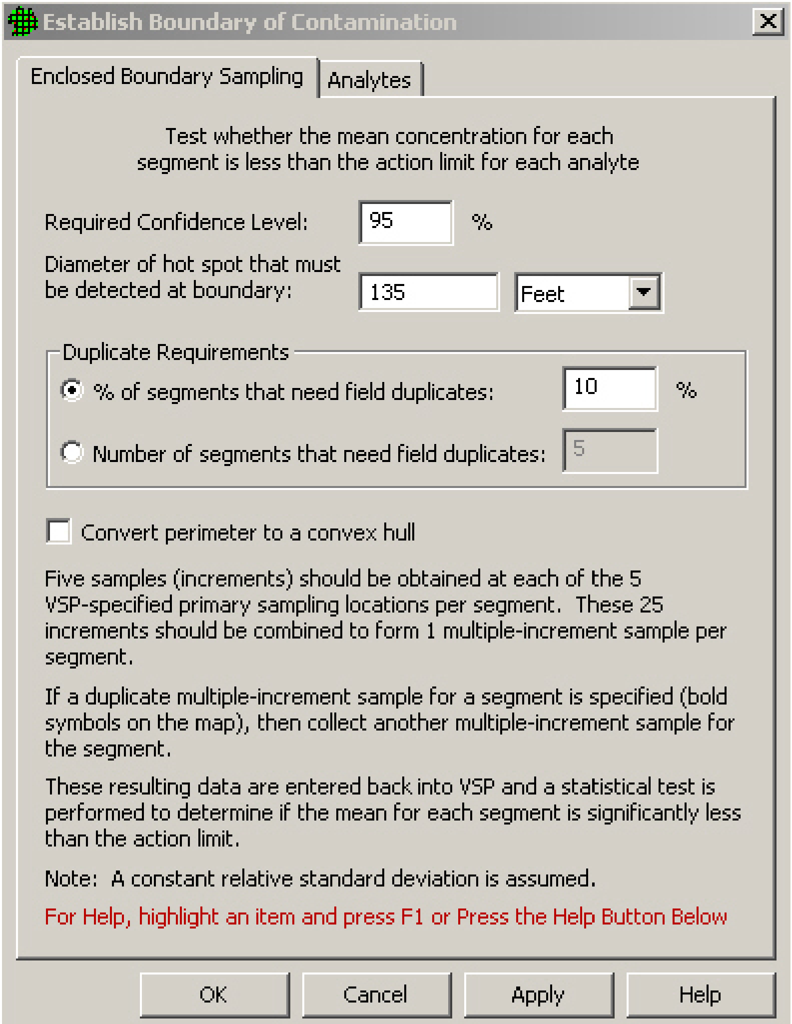


Figure 3.41. Dialog Box for Entering Design Inputs for Sampling an Enclosed Boundary

Sampling Locations in each of the segments. The segments for which the Primary Sample Locations are in bold type will have duplicate samples taken at each location. Different symbols are assigned to each segment to differentiate the segments visually. For each segment, VSP assumes the user will form the MI sample for that segment by mixing 5 small soil samples collected from each of the 5 Primary Sampling Locations. Hence, each MI sample is formed from the 25 small soil samples.

The user now collects the samples, mixes the samples to form a representative MI sample for each segment, and measures each MI sample. The results are input into VSP using the Sample Information box that appears when the cursor is placed over one of the Primary Sample Locations, and right-click the mouse. Use the keyboard to enter the measurement value into the appropriate row in the column labeled “Value” in the Segment Sample Results sub-box. Use the down arrow button on the keyboard to move between rows within the sub-box. Figure 3.43 shows the Sample Information Box.

We happened to click on a segment for which two MI samples are required. Thus, we will need to input two sets of measurements, one for each of the 3 analytes, making 6 input values required. Click the OK button on the dialog box to close the Sample Information box for that segment. Repeat the above process for each of the segments to enter all the measurement values. The Segment Sample Results box has a column headed “UCL”. VSP will fill in this box with the Upper Confidence Limit on the mean once all the measurement values for the segment are input. The UCL is used to test whether the mean exceeds the action level for that segment.

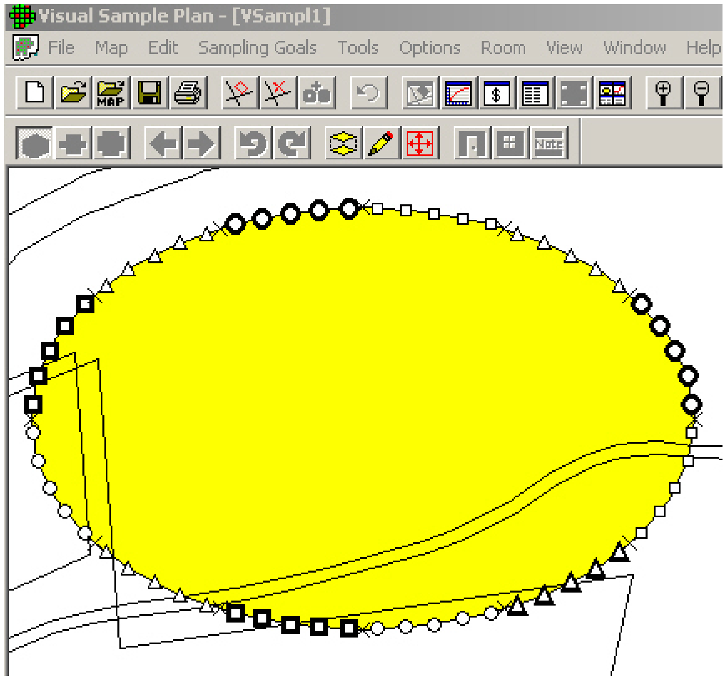


Figure 3.42. Enclosed Boundary with Two Bumped-Out Segments

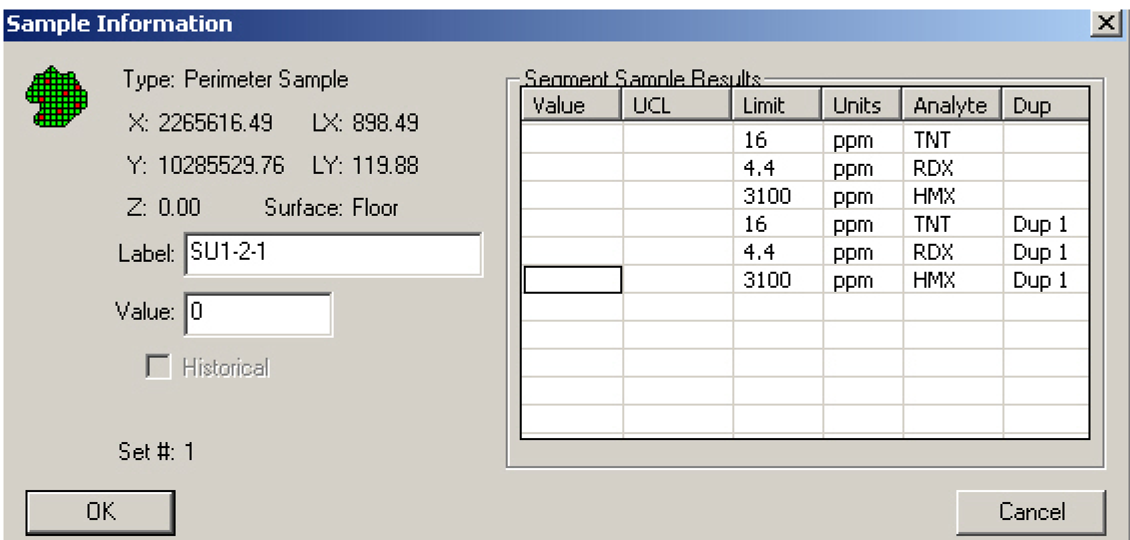


Figure 3.43. Sample Information Box for Entering Data into VSP, Duplicate Samples Required

Sample results can be entered into VSP using software such as a spreadsheet. Consult VSP's Help for instructions on this process.

VSP now tests whether each boundary segment should be enlarged (bumped out). This is described in Appendix to the report PNNL-XXXX referenced above. In Figure 3.44 we see an example of two

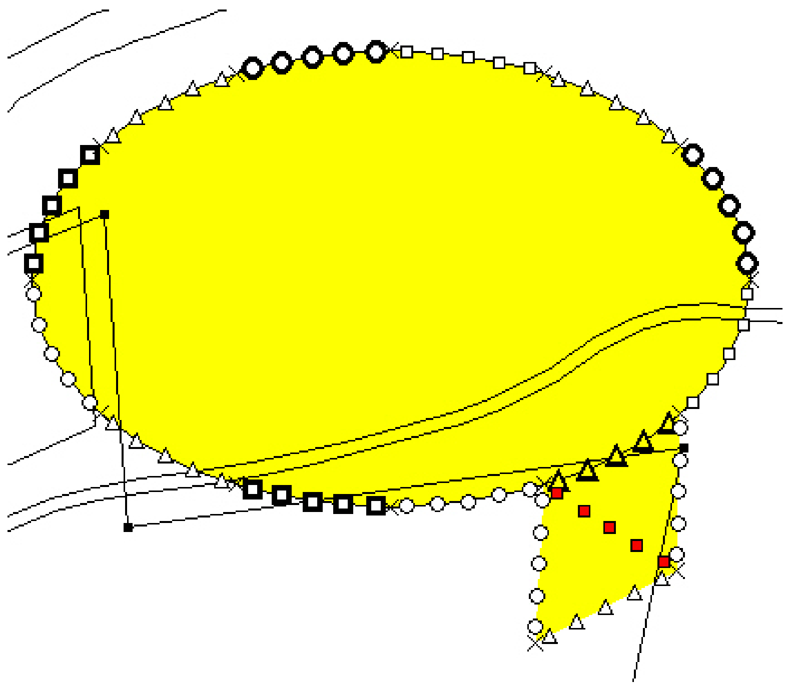


Figure 3.44. An Enclosing Boundary Showing the Five Primary Sampling Locations for each of the 17 Segments

expanded boundaries. Note the red colored Primary Location Segments indicate that segment did not pass the UCL test and hence had to be “bumped out”.

3.2.12.2 Partial Boundaries

The input screens, the dialog boxes, and the maps for Partial Boundaries problems are similar to those for the Enclosed Boundaries and will not be shown here. For a discussion of the Partial Boundaries problem consult the VSP Help function.

4.0 Assessment of Sampling Plans

VSP provides multiple displays for allowing you to assess the sampling plan that has been designed/ selected. VSP calls the displays **Views.** You can view a representation of the sampling locations on the map entered into VSP, view a graph of the performance of the design, look at a report that summarizes the key components of the design (such as number of samples, size of sampling area, cost, probabilities associated with the problem, assumptions, and technical justification), or see a listing of the coordinates of each sampling location. This section describes each of these views and discusses how you can use the views to assess the VSP sampling plan.

There are two ways to select/change views:

- Press one of the display buttons in the middle of the tool bar **(MAP VIEW, GRAPH VIEW, REPORT VIEW, COORDINATE VIEW)**
- From the main menu select View > Map (or Graph, Report, Coordinate)

4.1 Display of Sampling Design on the Map: MAP VIEW button or View > Map

In Section 2.2, we described how to set up a **Map.** In Section 2.3, we described how to set up a **Sample Area.** In Section 3.1, we described how to select a **Type of Sampling Plan.** In this section, we find out how to view the results of the sampling design we have just developed, displayed on the map.

In Figure 4.1, we see the display of a simple triangle we drew as our map and selected the entire triangle as our Sample Area. This is available in the file GridSize.vsp, which is included with the VSP program. We then selected from the main menu **Sampling goals > Locating a Hot Spot > Systematic grid sampling > Minimize number of samples.** We selected the Probability of Hit to be **90%** and selected a **Square** grid. We entered **4.0** feet for the **Length of Semi-Major Axis** and indicated that we wanted to detect a circular hot spot by selecting a **Shape** of **1.0.** We press **Apply,** and when we return to the map (**View > Map**) we see the **22** samples VSP calculated as required to meet the sampling goal displayed. Each time we press **Apply,** we refresh the map display with a new set of random-start sampling locations.

4.2 Display of Cost of Design

In Section 5.X, we describe how to enter costs. For most sampling designs, Total Cost (per unit plus fixed costs) is tallied and displayed on the same screen where we enter the per-unit costs—under the **Costs** tab on the dialog box used for entering design parameters. However, for the Hot Spot Sampling Goal, Total Cost is displayed only in the Report View (from the main menu, select **View > Report).** Reports are discussed in detail in Section 4.4.

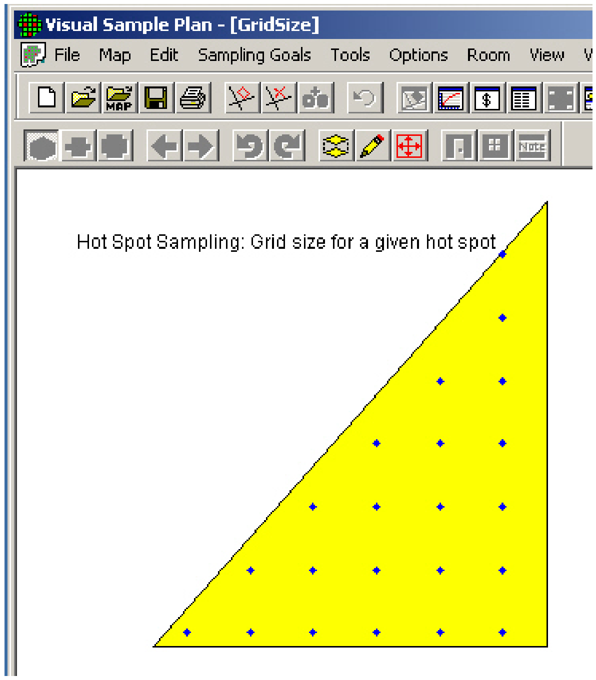


Figure 4.1. Display of Sampling Locations on Map

4.3 Display of Performance of Design: GRAPH VIEW button or View > Graph

VSP provides a display of the Performance of the Design for all of the sampling plans that result from sampling goals where a quantitative decision criterion is supplied. For some sampling goals, such as "Estimate a Mean", "Estimate a Proportion", "Find UXO Target Area", or "Assess Degree of Confidence in UXO Presence", where the only criterion the plan must meet is to minimize the variance of the estimate, or to minimize cost of the estimate, or to calculate a probability, there is little to graph in terms of the performance of the design. For such sampling goals, selecting **View** > **Graph** brings up a blank graph titled "**No Graph**".

For the sampling goals that *do* specify decision error rates or have confidence bounds on the estimates, VSP provides a graph of the performance of the sampling design that has just been created. Each sampling goal, or problem type, has a performance display tailored to it. Each graph tries to show the relationship between some parameter of the sampling design and how effective that design is at achieving the decision criteria. Once a **Sampling Goal** has been selected, the **DQO inputs** entered on the dialog box input screen, and the **Apply** button pressed to apply the design to the Sample Area, the display of the performance can be seen by pressing the **GRAPH VIEW** button on the tool bar or selecting **View** > **Graph** from VSP's main menu.

The following sections describe the major displays available for various types of problems. Displays not described are variants of those presented. Some of the graphs associated with unique sampling designs, such as Sequential Sampling, have been described in earlier sections, e.g., Graph View of Sequential Sampling (Fig 3.8), found in Section 3.

4.3.1 Performance of Design for Sampling Goal: Compare Average to a Fixed Threshold

The display for goal of comparing an average to a fixed threshold (i.e., the Action Level) is a graph of the probability of deciding the true mean of the sample area is greater than or equal to the Action Level on the vertical (y) axis as opposed to a range of possible true mean values on the horizontal (x) axis. Figure 4.2 is the Decision Performance Goal Diagram (DPGD) described in EPA's QA/G-4 guidance (EPA 2000a, pp. 6-7

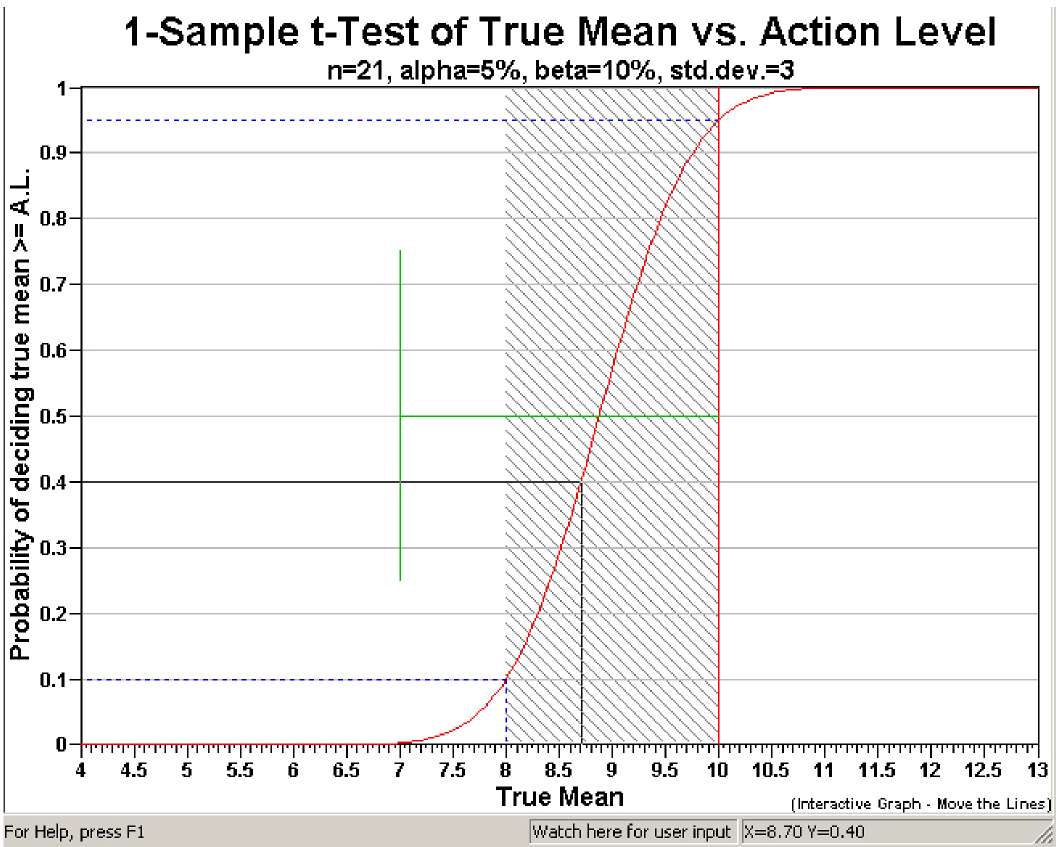


Figure 4.2. Decision Performance Goal Diagram for *Null Hypothesis: True Mean >=Action Level* for Comparing Mean vs. Action Level

through 6-11). The document can be downloaded from EPA at: http://www.epa.gov/quality/qa_docs.html. Notice how the graph changes as we alternate the null hypothesis between "Assume Site Dirty" to "Assume Site Clean."

The solid vertical red line is positioned at a true mean value of **10,** which corresponds to the Action Level. The area in gray hash marks is the gray region shown here from 8 to 10 and input as a delta (width) of **2.** The two dashed blue lines that extend from the y-axis to the x-axis mark the two types of decision error rates, alpha, set here at **5%,** and beta, set here at **10%.** Recall that Alpha is the probability of rejecting the null hypothesis when it is true (called a false rejection decision error), and beta is the probability of accepting the null hypothesis when it is false (called a false acceptance decision error). The error rates along with the user-supplied standard deviation of **3** and the VSP-calculated sample size **n=21** is shown on second row of the title. We also see in the title that we are using the sample size formula for the 1sample t-test.

The green vertical line marks off one standard deviation (3) from the action level. This mark allows the user to visually compare the width of the gray region to how variable, on average, we expect individual observations to be about the mean (definition of standard deviation). The sliding black lines (cross hairs) that move on the graph when the mouse is moved are provided to facilitate reading the x, y values off the graph. This cross-hair feature can be turned off or on by choosing **Options > Graph > Cross Hairs.**

Most of the parameters displayed on the DPGD can be changed interactively by moving the lines on the graph, rather than having to change the values in the input dialog box. Table 4.1 describes the interactive features.

Table 4.1. Interactive Graph Features

To Change	Do the Following
Alpha	Drag the horizontal blue dashed line up or down
Beta	Drag the horizontal blue dashed line up or down
Delta (and LBGR, or UBGR)	Drag the vertical edge of the shaded gray area to the left or right
Standard Deviation	Drag the vertical section of the green line left or right
Action Level	Drag the vertical red line left or right
Null Hypothesis	Click on the y-axis title

As you change these parameters, you can see the new value of the parameter on the bottom status bar after “watch here for user input.” You will notice that changing these values on the graph also changes their values on the other displays: the sampling design is modified in the report view, new samples are placed on the map view, and updated sample location information is list in the coordinate view.

Right-clicking anywhere in the graph (see Table 4.2 below) men brings up a pop-up menu, shown in Fig. 4.2a. This pop-up menu provides quick access to the menu choices available from the main menu under **Options > Graph**. One of the options is to view the complement of the Decision Performance Goal Diagram, which is referred to as the “Operating Characteristic” Curve. We select this option and the OC curve is shown in Figure 4.3. Note that the Y-axis is now labeled, “Probability of making the correct decision”.

When the null hypothesis is stated as **H_0: True Mean >= Action Level** (Site is Dirty), the gray region is on the left side of the Action Level. However, when the null hypothesis is stated as **H_0: True Mean <= Action Level** (Site is Clean), the gray region is on the right side of the Action Level. In practical terms, when we assume a site is dirty, the majority of the decision errors will occur for clean sites with true means just below the Action Level. On the other hand, when we assume a site is clean, the majority of decision errors will occur for dirty sites with true means just above the Action Level.

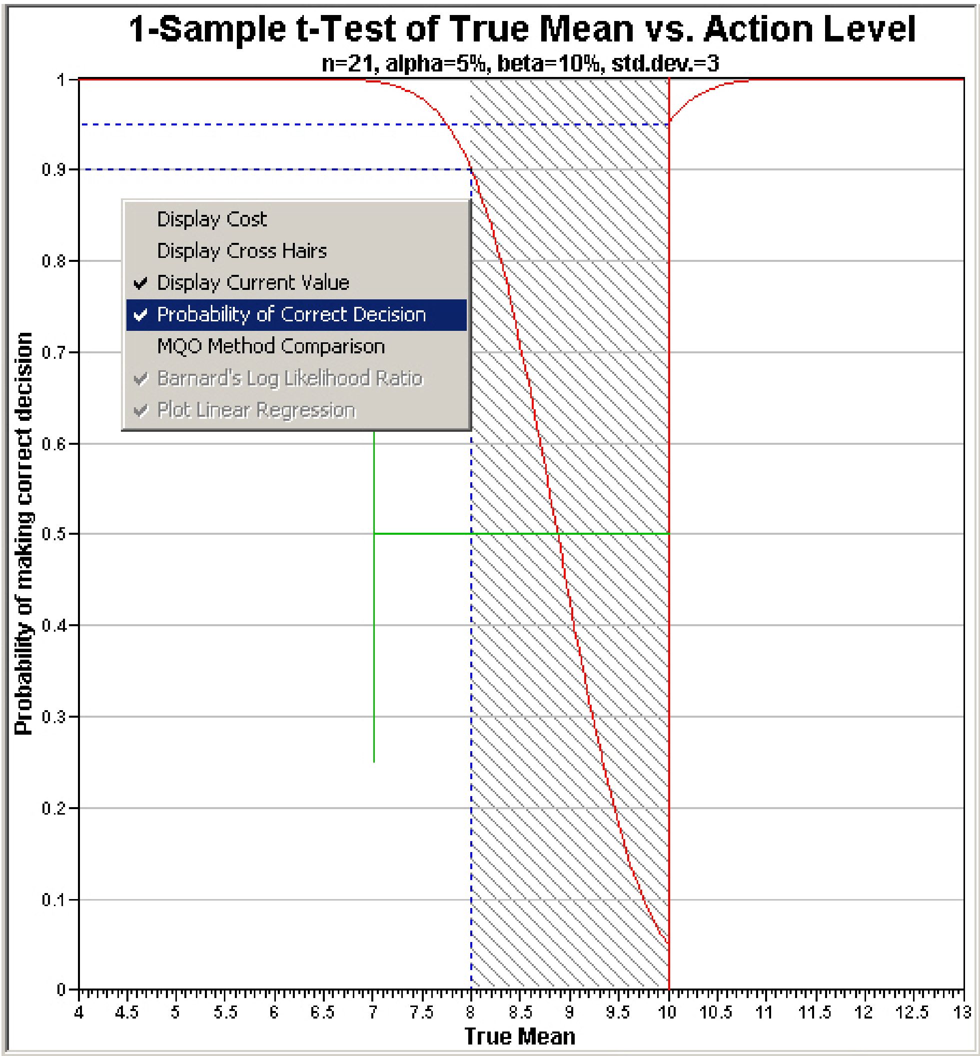


Figure 4.3. Graph of Probability of Making Correct Decision

The DPGD graph in Figure 4.2 is telling us that for the "Site is Dirty" null hypothesis,

- Very clean sites will almost always result in sets of random sampling data that lead to the decision "Site is Clean."
- Very dirty sites will almost always result in sets of random sampling data that lead to the decision "Site is Dirty."

What we may not know intuitively is how our choice of the null hypothesis affects decisions near the Action Level. The graph in Figure 4.2 also is telling us:

- Clean sites with true means just *below* the Action Level will lead to mostly *incorrect* decisions.
- Dirty sites with true means just *above* the Action Level will lead to mostly *correct* decisions.

However, when we reverse the null hypothesis and state it as **H0: True Mean <= Action Level, i.e.,** assume "Site is Clean," we see in Figure 4.4 that the gray region where the majority of decision errors occur shifts to the right side of the Action Level. Sites that are dirty now lead to the majority of decision errors. Also note that alpha is now defined for values less than the action level, while beta is defined for values above the upper bound of the gray region.

You should carefully study EPA's QA/G-4 guidance document (EPA 2000a, especially pp. 6-1 to 7-6) to better understand how to use VSP to balance the choice of null hypothesis, decision error rates, width of the gray region, total sampling costs, and costs of incorrect decisions.

An option available in Graph View is the pop-up menu shown in Figure 4.2. In Graph View, right click anywhere within the graph and the pop-up menu appears. Table 4.2 briefly describes the options. Options that are "grayed out" are unique to graphs associated with other Sampling Goals.

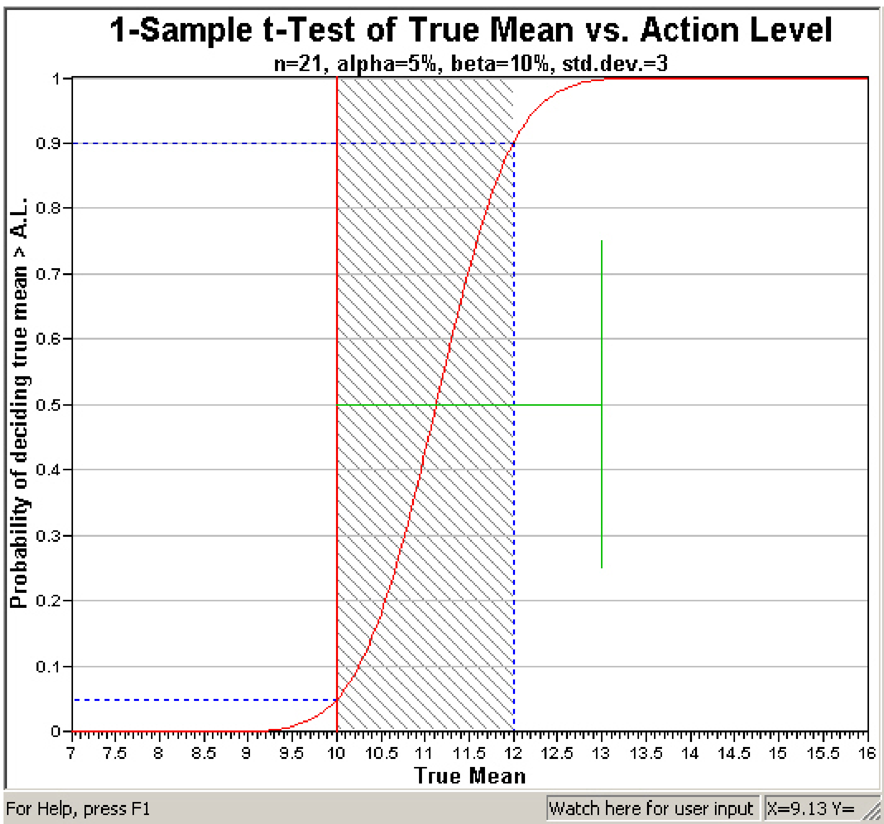


Figure 4.4. Decision Performance Goal Diagram for *Null Hypothesis: True Mean <= Action Level* for Comparing Mean vs. Action Level

Table 4.2. Graph Options Menu Commands

Display Cost	Display cost on any graph axis that would otherwise display the number of samples.
Display Cross Hairs	Display an interactive cross-hair that allows the user to see the X and Y values for any point on the graph
Display Current Value	Displays a cross-hair that corresponds to the current X or Y value produced by the current sampling design.
Probability of Correct Decision	Displays the probability of a correct decision in place of a decision performance goal diagram
MQO Method Comparison	Displays a bar graph that shows the relative costs of MQO sampling design alternatives.
Barnard's Log Likelihood Ratio	Display's a plot of Barnard's Log Likelihood Ratio (which is the test statistic used in Barnard's sequential t-Test)
Plot Linear Regression	Displays the linear regression plot (available for collaborative sampling designs).

4.3.2 Performance of Design for Sampling Goal: Construct Confidence Interval on the Mean

The display for assessing a confidence interval for a mean differs somewhat from that for comparing an average to a threshold because this is an estimation problem, not a testing problem. As such, there is only one type of decision error rate, alpha. Shown in Figure 4.4 is the Performance Design for a problem where the user specified the width of the confidence interval as **1.0,** the standard deviation as **3,** and a desired **95% one-sided** confidence interval on the mean. We are using a one-sided confidence interval (vs. a two-sided) because we are concerned only about values that exceed the upper bound of the confidence interval, not values both above the upper bound and below the lower bound. This is consistent with problems in which the mean to be estimated is average contamination, so we are not concerned about values below the lower bound of the confidence interval.

VSP calculated that a sample size of **26** was required. The performance graph is a plot of possible confidence interval widths vs. number of samples for the problem specified. The dashed blue line terminates at the y-axis at a confidence interval width of 1.0, as specified by the user, and at the x-axis at the recommended minimum sample size of 26.

The solid black line is a locating aid you can slide up and down the graph to easily read the trade-offs between increased width of the confidence interval and increased number of samples. In Figure 4.4, the x-axis value (number of samples) and the y-axis value (width of confidence interval) for the current solid black line can be seen in the status bar as X = 2.72 and Y = 5.96.

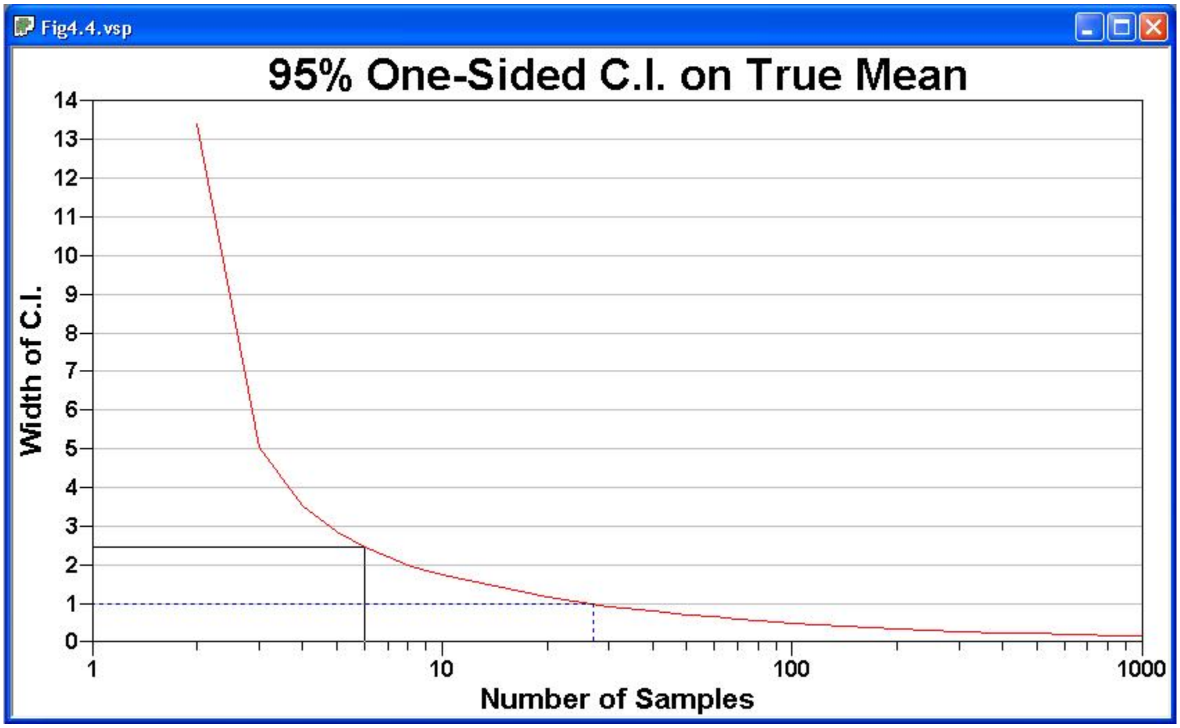


Figure 4.5. Decision Performance Graph for One-Sided 95% Confidence Interval

4.3.3 Performance of Design for Sampling Goal: Comparing a Proportion to a Fixed Threshold

The sampling design assessment display for comparing a proportion to a fixed threshold is a graph of the number of samples vs. the decision error, beta. These parameters were selected for the performance graph because they can be directly calculated and the graph provides a visual display of how increasing the number of samples decreases one of the error rates (beta).

Note: If the appropriate statistical test is used, the test is designed to achieve the level of significance, or alpha. It is beta and the power of the test (1-beta) that are affected by sample size.

For this sampling goal, there is no clear distinction between “Site Dirty” and “Site Clean,” depending on how the null hypothesis is formulated. If the proportion we are talking about is the proportion of 1-acre lots in a building development that have trees, then exceeding a threshold would be a “good thing.” However, if the proportion is the proportion of acres that have contamination greater than 10 pCi, then exceeding the threshold would be a “bad thing.” Alpha and beta are still defined as false acceptance and false rejection rates, but the user must formulate the hypotheses and select limits on the error rates consistent with the goals of the project and which type of error is most important to control.

In the example in Figure 4.5, the null hypothesis was set to **True Proportion >= Given Proportion.** As such, beta is the probability of deciding the proportion exceeds the threshold when the true proportion is equal to or less than the lower bound of the gray region. For this problem, we set alpha to **1%** and beta to

5%, and the lower bound of the gray region to **0.35 (i.e.,** width of gray region = 0.15). The proportion we want to test against (Action Level) is 0.5. This Action Level is the default for VSP because it is the most conservative. That is, the most number of samples are needed to differentiate a proportion from 0.5 (vs. differentiate a proportion from any other percentage). VSP calculated a sample size of 169. The dashed blue line terminates on the y-axis at 169 samples and on the x-axis at a beta of **5%.** The heading for this graph reminds the user that the one-sample proportion test is the assumed test that will be used for making the decision because the sample size formula in VSP is based on the one-sample proportion test.

Note the solid black line in Figure 4.5 and the values in the status bar. The black line shows that for a beta of 20%, the minimum number of samples is reduced to 109. Moving the black line is a quick way to play "what-if" games regarding sample sizes and beta error rates for a given alpha.

4.3.4 Performance of Design for Sampling Goal: Compare Average to Reference Average

The sampling design performance display for comparing the true means of two populations when the **assumption of normality can be made** is a graph of the probability of deciding if the difference of true means is greater than or equal to the specified difference (Action Level) vs. various differences of true means. This graph is similar to the Decision Performance Curve discussed in Section 4.3.1, but this time we are dealing with two populations, and the x-axis is a range of possible *differences* between the two population means.

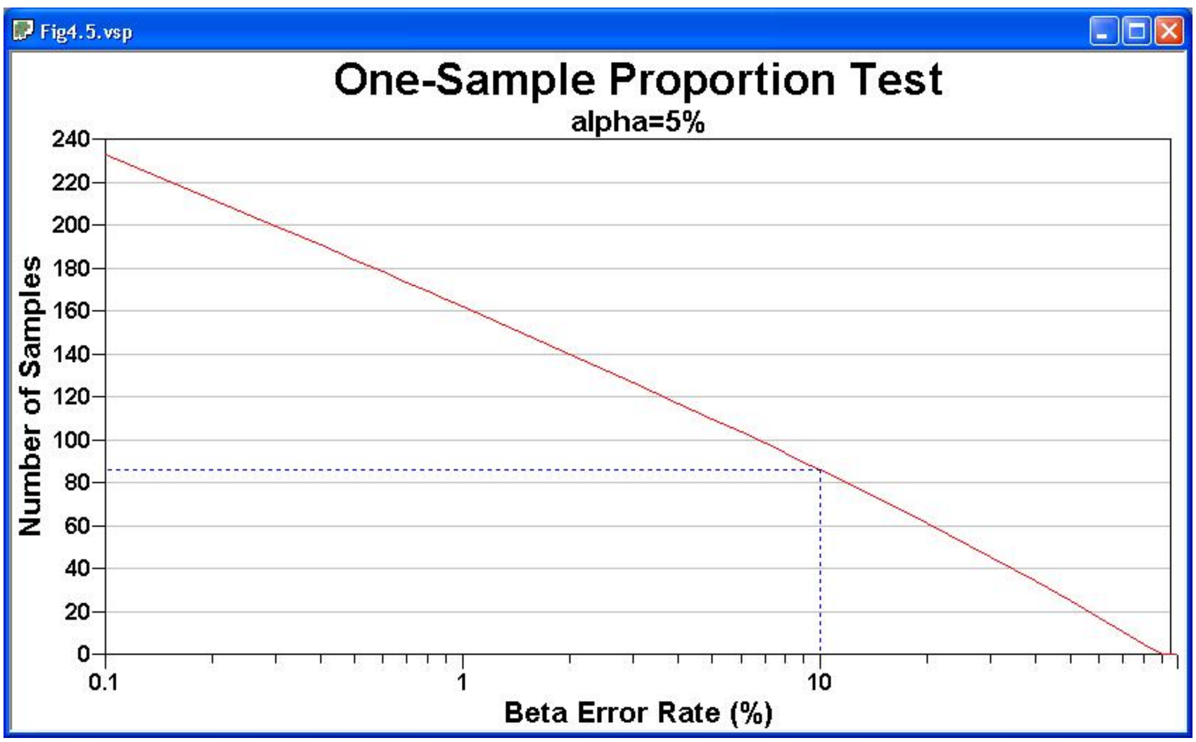


Figure 4.6. Decision Performance Goal Diagram for Comparing a Proportion to a Fixed Threshold

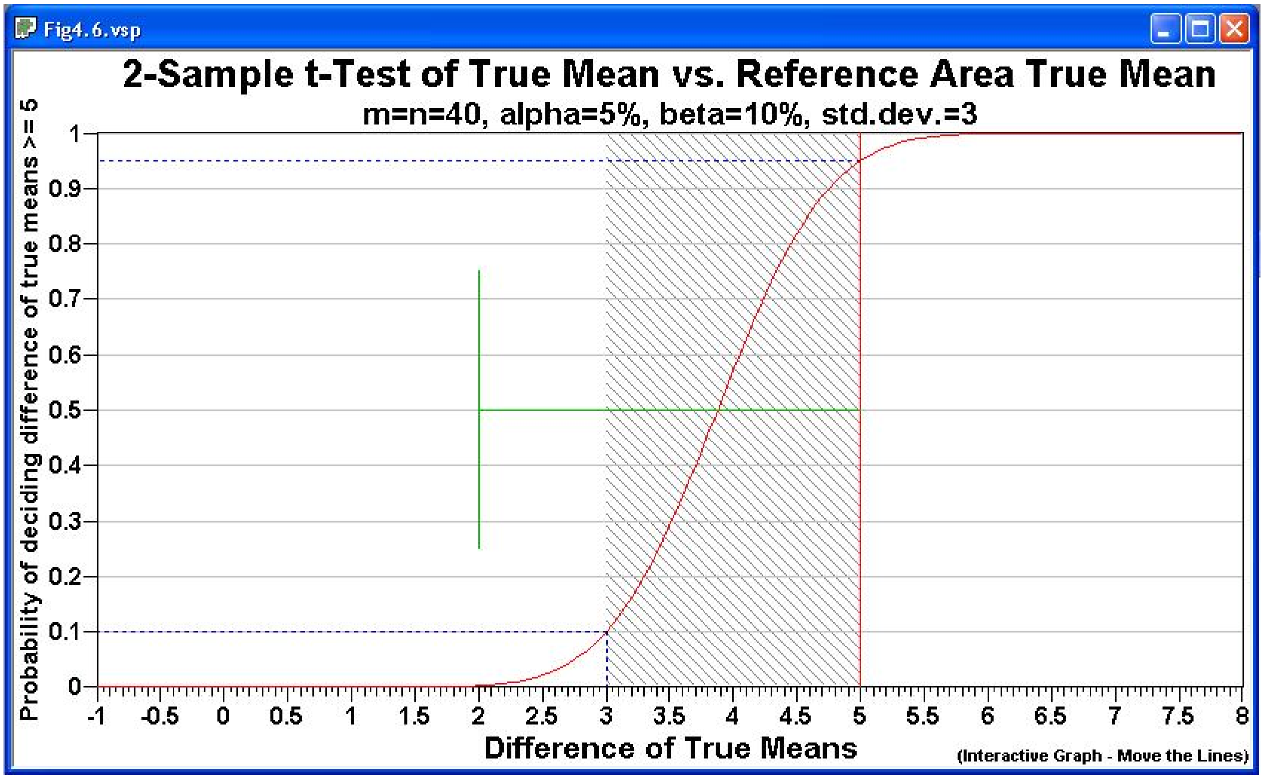


Figure 4.7. Decision Performance Goal Diagram for Comparing a Sample Area Mean to a Reference Area Mean

The graph shown in Figure 4.6 is for **H_0: Difference of True Means >= Action Level.** We revert back to the notion that this null hypothesis implies a " Dirty Site" condition. If the action level is a positive number, we would classify the site as greater than background or "Dirty." For this problem, the specified difference of the two means (Action Level) is **5,** the width of the gray region is **2,** alpha = **5%,** beta = **10%,** and the estimated common standard deviation = **3.**

Note: The standard deviation is the average expected difference between the individual units in a population and the overall mean for that population. It is assumed that both populations (Sample Area and Reference Area) have the same standard deviation. The graph is labeled "2-Sample t-Test" because it is assumed that the two-sample t-test will be used as the statistical test.

Figure 4.6 shows that we need to take 40 samples both in the Sample Area and 40 samples in the Reference Area. The probabilities of deciding the Sample Area is 5 or more units (pCi/g, ppm, etc.) above the Reference area are plotted against the true differences in means. The standard deviation is shown as the green line at a distance of 3 from the Action Level.

When the assumptions of **data are not required to be normally distributed** is made, we can see from the pull-down menu lists under Sampling Goals that two non-parametric statistical tests are proposed – the Wilcoxon rank sum test and the MARSSIM WRS (Wilcoxon rank sum) test.

The Decision Performance Goal Diagrams for the two nonparametric tests are similar to the DPGD for the parametric two-sample t-test. In Figure 4.7, we see the DPGD for the MARSSIM WRS test using the same inputs as the problem in Figure 4.6. The difference of true means or medians is plotted on the x-axis, and the probability of deciding the difference is equal to or greater than the action level of 5 is shown on the y-axis. For the MARSSIM formulation of the WRS test, the action level is the DCGLw. The lower bound of the gray region is the difference in means or medians where we want to limit the beta error. It can be shown that using the same parameters as in the two-sample t-test, the Wilcoxon rank sum test requires 46 samples in both

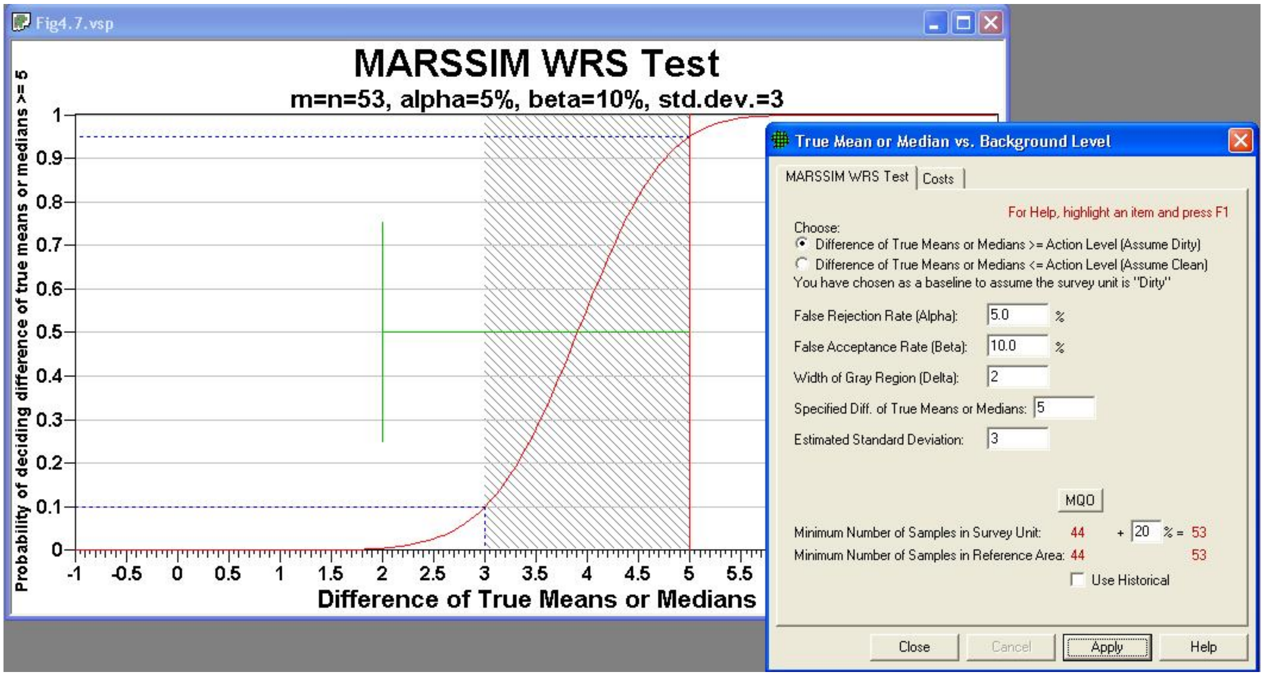


Figure 4.8. Decision Performance Graph for Comparing a Sample Area Mean to a Reference Area Mean (Nonparametric Version, MARSSIM WRS

the Study and Reference Areas while the MARSSIM WRS test requires only 44 samples. The larger sample sizes required for the nonparametric tests reflects the premium that must be paid for not making any assumptions about the underlying distribution of the populations.

Shown in Figure 4.7, the input dialog for the MARSSIM WRS test allows the user to supply a percent overage to apply to the sample size calculation. MARSSIM suggests that the number of samples should be increased by at least 20% to account for missing or unusable data and for uncertainty in the calculated values of Sample Size, n (MARSSIM, p. 5-29). With the extra 20%, the sample size now becomes 53 samples required in both the Sample Area (i.e., Survey Unit or Study Area) and Reference Area.

4.3.5 Performance of Design for Sampling Goal for Hot Spot Problem

The Decision Performance Goal Diagram for the hot spot problem is a graph of number of samples on the x-axis and the probability of hitting a hot spot of a specified size on the y-axis. The heading of the performance graph lists the size of the hot spot and the size of the sample area. The trade-off displayed is that by increasing the number of samples (i.e., a tighter grid spacing and hence the higher cost), and/or changing the grid type (say from square to triangular), there is a higher probability of hitting the hot spot with one of the nodes on the grid. This is almost a straight-line relationship until we get into larger sample sizes, and then the efficiency is diminished.

Returning to the problem we laid out in Section 4.1, for the sampling goal of **Sampling Goal > Locating a Hot Spot > Systematic Sampling > Minimize the number of samples,** the graph is shown in Figure 4.8. This graph is for the 2577.6 ft^2 Sample Area shown in Figure 4.1 and for finding a **4-ft** round hot spot. The graph shows the desired input of **90%** probability of hitting the circular hot spot of radius 4 ft. and the **51** samples required to achieve this. The dashed blue may show slightly different values, see discussion below.

This example shows how VSP may place a slightly different number of sampling points (nodes) on a map than the exact number calculated. Shown on the performance curve, and displayed in the Report, are anywhere from 45 to 55 samples placed on the map. The difference between the calculated number of sample and the number of samples placed on the map is 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas. Repeatedly pressing the **Apply** button from the dialog box will select a different random starting point for the grid and may change the number of samples that will fit in the sample area.

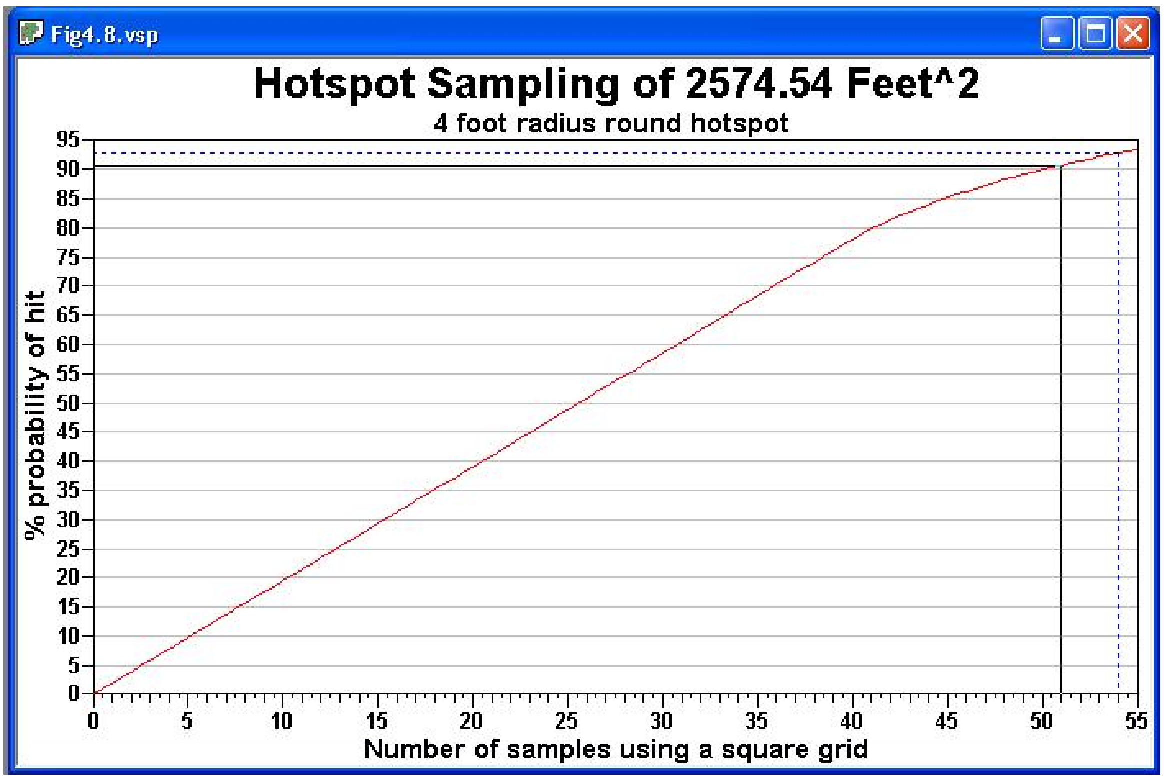


Figure 4.9. Probability of Hitting a Hot Spot vs. Number of Samples

The probability of hit is a geometric relationship between the grid spacing and the hot spot size and shape. The probability of hit is *not* a function of number of samples. On the graph, however, grid spacing is translated to the number of samples on a *theoretical* sampling area. The number of theoretical samples is shown on the graph because it is a more meaningful metric for the user than grid spacing. The dashed blue line on the performance curve shows the number of samples that fit on the *actual* sample area given the starting point. The report also lists the actual number of samples placed on the map.

Important note: Regardless of where the dashed blue line occurs on the graph, the probability of hit for *your* sampling design is the one you specified and is shown on the sampling goal dialog. This is true because the probability of hit is a geometric relationship between the grid spacing and the hot spot size and shape.

Deselecting the Random Start on the dialog box removes the random assignment of the grid and keeps the grid fixed with each repeated hit of the Apply button, keeping the same sample size.

Note that you can use the mouse to move the solid black line up and down the graph. You can use this solid line to easily read off the probability vs. sample size trade-off options from the horizontal and vertical axes. In Figure 4.8, we have the solid black line positioned at a 60% probability of hitting a hot spot of radius 4 ft when there are only 31 grid samples in an area of 2577.6 ft^2.

4.3.6 Performance of Design for Sampling Goal of Compare Proportion to a Reference Proportion

The graph for displaying the performance of the design for comparing a proportion to a reference proportion is similar to the comparison of two population means (see Figure 4.6). As such, the difference between the two true proportions is shown on the x-axis, and the probability of deciding that the difference between the two true proportions is greater than a specified difference (i.e., the Action Level) is shown on the y-axis. The two proportions being compared could be, say, the proportion of children with elevated blood lead in one area compared to the proportion in another area, or it could be the percentage of 1-m squares within an acre that have contamination greater than 1 ppm of dioxin. The comparison might be to compare the amount of contamination (stated as a percentage remaining at a site after it has been remediated) to a background or reference area. Using the naming convention in EPA (2000b, pp. 3-27 – 3-31), the site (also called the survey unit, Sample Area) is Area 1, and the reference or background area is Area 2. The document can be downloaded from the EPA site: http://www.epa.gov/quality.qa_docs.html.

In Figure 4.9, we see the inputs from the dialog box along with the Decision Performance Goal Diagram. The example has as the null hypothesis "no difference between site and background,"or Ho: P1 – P2 <= 0. The two estimated proportions are required to calculate the standard deviation for the pooled proportion used in the sample size formula. With this formulation, the specified difference (Action Level) is 0, and the false acceptance error rate (beta = 5%) is set at the difference of P1 – P2 = 0.10. Thus, 0.10 is the upper bound of the gray region, which VSP requires to be greater than the Action Level. When the null hypothesis is changed to Difference of Proportions >= Specified Difference, the lower bound of the gray region is less than the action level.

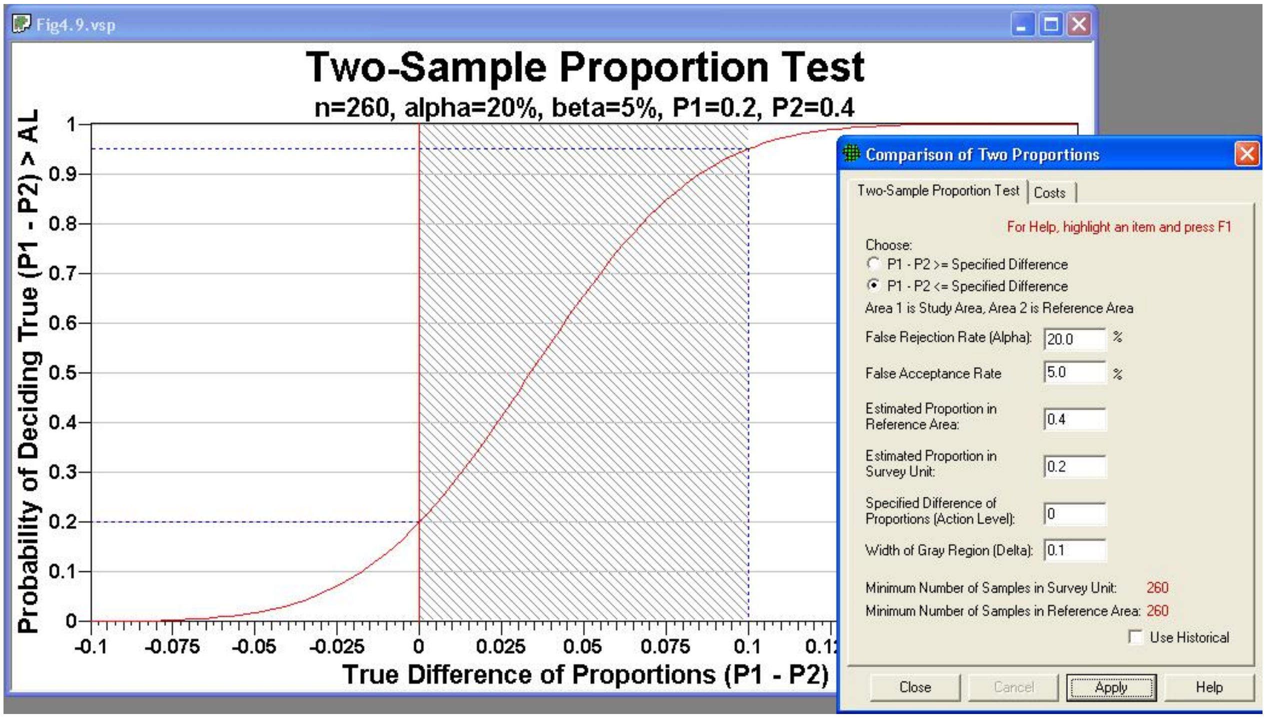


Figure 4.10. Decision Performance Goal Diagram for Comparing a Sample Area Proportion to a Reference Area Proportion. Input dialog box for design shown as insert.

The graph in Figure 4.9, labeled the Two-Sample Proportion Test, lists the inputs of alpha, beta, and the two estimated proportions in the heading line. The S-shaped curve shows that for larger differences in the true proportions, the probability of correctly deciding the difference exceeds the Action Level increases. This is intuitive because the greater the difference between two populations, the easier it is to correctly distinguish that difference from a fixed threshold (Action Level).

4.3.7 Performance of Design for Sampling Goal of Establish Boundary of Contamination

In Figure 3.9a we see the performance of the sampling design defined by the inputs shown in Figure 3.40 applied to the Sample Area of the large “oval” in the Millsite.dxf file (see Map View in Figure 3.42). This problem is for the **Enclosed Boundary**. In order to have a 95% confidence of finding a hot spot of diameter of **45 ft**, we need 12 segments. This dictates we need **60** Primary Sampling Locations (12 x 5 = 60). The relationship between diameter of hot spot and number of primary sampling locations is shown in the dashed blue line positioned on the performance curve shown in Fig. 3.9a. The dashed line shows the current number of Primary Sampling Locations for this design (n = 60), which may differ from the optimum number because of rounding and bump-out effects. We can see from the cross hairs positioned on the performance curve (right-click on graph, and toggle the **Display Cross Hairs** to “on”), that if we expand our Primary Sampling Locations to 100, we could detect a hotspot of diameter 27 ft. with the sample level of 95% confidence.

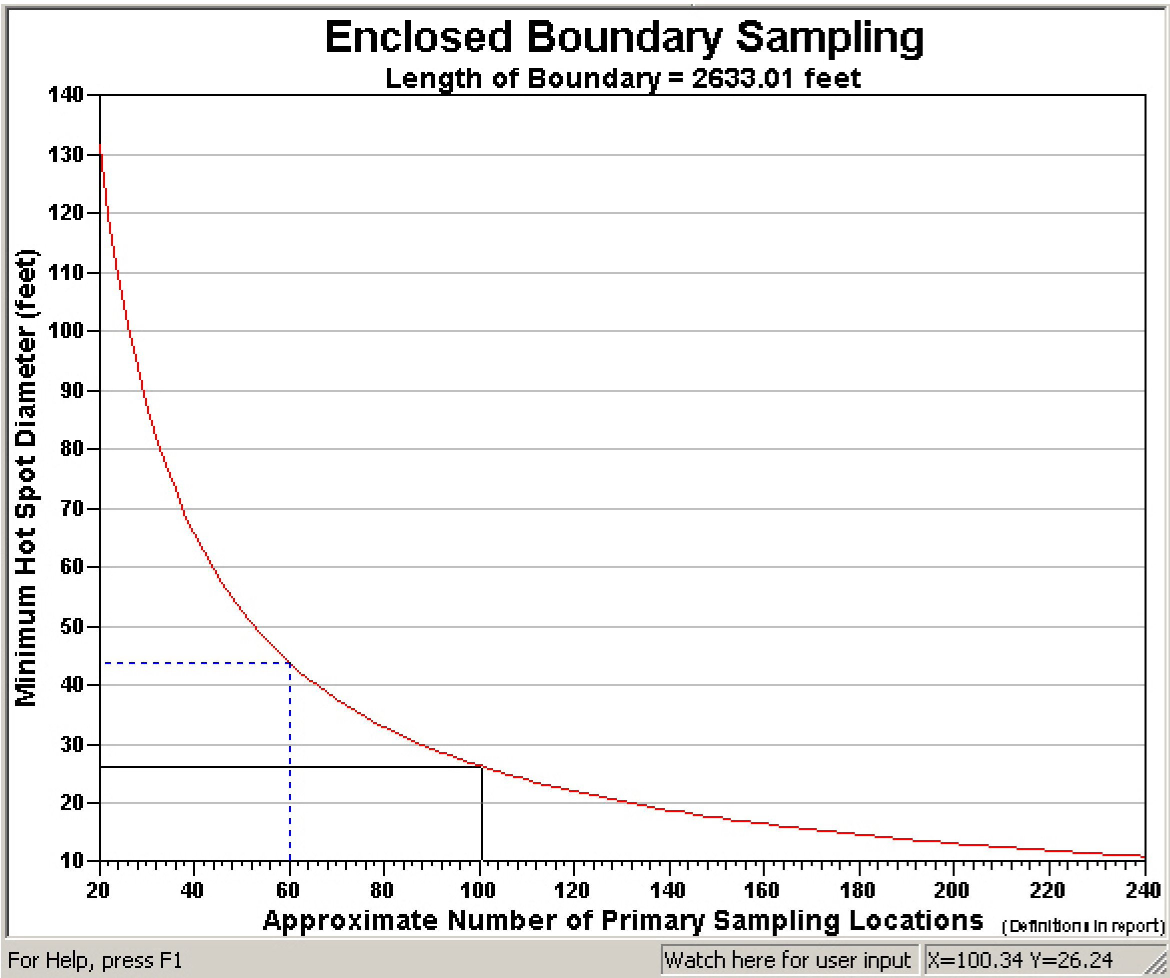


Figure 4.11. Curve of trade-off Between Primary Sampling Locations and Size of Hot Spot that can be Detected

The sample type of graph is produced for the **Open Boundary** Sampling Problem.

4.4 Display of the Report

One of the most valuable outputs from VSP is the Report that is generated for each application of a Sampling Design to a Sample Area. The Report View for a sampling design is available by either selecting the **REPORT VIEW** button on the toolbar, or by selecting **View > Report** from the main menu. The only Sampling Goals that do not produce Reports are **Non-statistical sampling approaches** and **Comparisons of two proportions.**

The Report provides the VSP user with a complete documentation of the sampling design selected. The report includes:

- statement of sampling objective,

- the assumptions of the design,
- sample size formula,
- inputs provided by the user,
- summary of VSP outputs including sample size and costs,
- list of samples with their coordinates and labels,
- map with sample locations identified,
- Performance Goal Diagram,
- Peer-reviewed technical references for designs and formulas,
- technical discussion of the statistical theory supporting the sampling design and sample size formula.

The reports are suitable for incorporation into a quality assurance project plan or a sampling and analysis plan. The report for some of the sampling designs include:

- recommended data analysis activities for how data should be used in the appropriate statistical test to make a decision,
- insight into options presented in the Input Dialog Box,
- sensitivity tables showing how sample number changes as input parameters change, and
- extended statistical discussions and support equations.

Some of the output from VSP, for some designs, is viewable only within the report. VSP users can use the information in the **Report** as an additional source of **Help**.

A few selections from the report for the selection **Sampling Goal > Compare Average to a fixed threshold > Assume data normally distributed > Simple Random Sampling** are shown in Figure 4.10. Each time VSP calculates a new sample size, changes VSP input, or adds points to an existing design, the report is updated automatically. The complete report can be copied to the clipboard for pasting into a word processing application like Microsoft Word selecting **Edit > Copy** from the main menu when the report view is active. The text and graphics are copied using rich text format (RTF) to preserve formatting. The user opens Microsoft Word, selects **Paste,** and the entire report is copied into a Word document.

The sensitivity table in the **Report View** allows the user to do "what-if" scenarios with VSP input and output. For one sampling goal, the sensitivity table shows how sample size changes with changes in the standard deviation and the two decision error rates, alpha and beta. Different sampling goals and sets of assumptions have different variables and parameters in their sensitivity table. The user can change the variables and range of values shown in the sensitivity table by right-clicking anywhere in the report. A dialog box, as shown in Figure 4.11, is displayed to allow the user can choose which of up to four variables that will be displayed, along with each variable's starting and ending value, and the step-size shown in the sensitivity table. Displayed in the table can be the number or samples, cost, or both. Certain sampling designs have the option to show parameters other than cost.

Random sampling locations for comparing a mean with a fixed threshold (parametric)

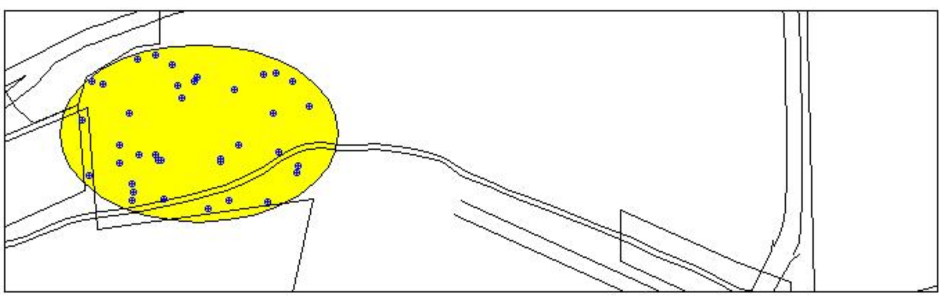
Summary

This report summarizes the sampling design used, associated statistical assumptions, as well as general guidelines for conducting post-sampling data analysis. Sampling plan components presented here include how many sampling locations to choose and where within the sampling area to collect those samples. The type of medium to sample (i.e., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

The following table summarizes the sampling design developed. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	37
Number of samples on map [a]	37
Number of selected sample areas [b]	1
Specified sampling area [c]	509779.78 ft^2
Total cost of sampling [d]	$19500.00

[a] This number may differ from the calculated number because of 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas.
[b] The number of selected sample areas is the number of colored areas on the map of the site. These sample areas contain the locations where samples are collected.
[c] The sampling area is the total surface area of the selected colored sample areas on the map of the site.
[d] Including measurement analyses and fixed overhead costs. See the Cost of Sampling section for an explanation of the costs presented here.



Area: Area 1					
X Coord	**Y Coord**	**Label**	**Value**	**Type**	**Historical**
2265128.4318	10285981.0061		0	Random	
2265305.4734	10285630.1118		0	Random	
2265161.3618	10285860.6159		0	Random	
2265563.9628	10285920.1073		0	Random	

Figure 4.12. Report View of the Sampling Goal: Compare Average to a Fixed Threshold, Normality Assumed, Simple Random Sampling

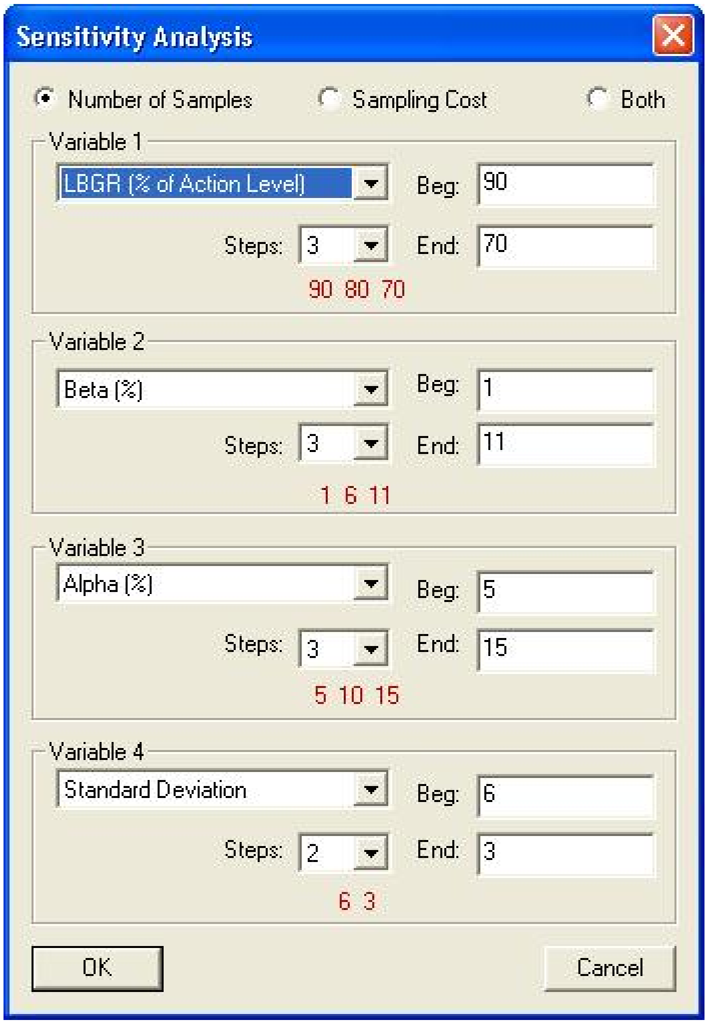


Figure 4.13. Dialog Box for Changing Variables Displayed, and Range for Variables Shown, in Sensitivity Table in Report View. Shown here is input dialog for sampling goal of compare average to threshold, normality assumed (parametric), simple random sampling.

4.5 Display of Coordinates

The fourth type of display in VSP is the list of coordinates for each sample point on the map. We can see this display by **COORDINATE VIEW** button on the toolbar, or by selecting Main Menu option **View** > **Coordinates.** The x and y coordinates are displayed for each sample point. Also displayed are the sample points label, a value (if entered by the user), the type (e.g., random, systematic, RSS), and a "true/false" indicator of whether or not this sample point is an historical sample (previously taken sample). Coordinates are segregated by Sample Area. These coordinates can be copied and pasted into a spreadsheet or word processing file using Main Menu option **Edit > Copy.** Figure 4.12 is an example of the Coordinates view.

Area: Area 1

X Coord	Y Coord	Label	Value	Type	Historical
2265128.4318	10285981.0061	VSP-1	0	Random	
2265305.4734	10285630.1118	VSP-2	0	Random	
2265161.3618	10285860.6159	VSP-3	0	Random	
2265563.9628	10285920.1073	VSP-4	0	Random	
2265578.9547	10285591.6967	VSP-5	0	Random	
2264986.3670	10285519.3750	VSP-6	0	Random	
2265259.0015	10285457.0271	VSP-7	0	Random	
2265149.0585	10285908.0085	VSP-8	0	Random	
2264934.5577	10285689.4339	VSP-9	0	Random	
2264825.5277	10285581.4322	VSP-10	0	Random	
2265206.7974	10285920.6213	VSP-11	0	Random	
2264979.1298	10285547.6926	VSP-12	0	Random	
2265004.4190	10285655.0953	VSP-13	0	Random	
2265492.7605	10285807.0948	VSP-14	0	Random	
2265301.9146	10285639.4444	VSP-15	0	Random	
2264833.7042	10285919.5649	VSP-16	0	Random	
2265476.4405	10285483.6637	VSP-17	0	Random	
2265625.4932	10285831.6183	VSP-18	0	Random	
2265068.7330	10285656.2455	VSP-19	0	Random	
2264936.0599	10285623.3110	VSP-20	0	Random	
2265514.4665	10285663.3566	VSP-21	0	Random	
2265458.2032	10285945.5990	VSP-22	0	Random	
2265066.2328	10286016.7997	VSP-23	0	Random	
2265085.2622	10285636.6912	VSP-24	0	Random	
2264799.7844	10285778.3154	VSP-25	0	Random	
2265505.2629	10285953.7490	VSP-26	0	Random	
2265218.2016	10285936.4287	VSP-27	0	Random	
2265369.7326	10285690.3660	VSP-28	0	Random	
2265075.2469	10285634.8206	VSP-29	0	Random	
2265003.4862	10286002.0137	VSP-30	0	Random	
2265352.9186	10285889.5687	VSP-31	0	Random	
2264877.7457	10285910.0798	VSP-32	0	Random	
2264973.4826	10285804.1923	VSP-33	0	Random	
2265584.5676	10285612.5624	VSP-34	0	Random	
2265334.9627	10285491.0034	VSP-35	0	Random	
2265096.8930	10285495.7522	VSP-36	0	Random	
2264983.5288	10285491.0735	VSP-37	0	Random	

Figure 4.14. Coordinates Display of Sampling Locations

4.6 Multiple Displays

Multiple displays can be brought up on the same screen. Table 4.3 lists the options available under main menu item **Window**.

Table 4.3. Window Menu Commands

New Window	Creates a new window that views the same project
Cascade	Arranges windows in an overlapped fashion
Tile	Arranges windows in non-overlapped tiles.
Arrange Icons	Arranges icons of closed windows
Double Window	Shows map view and graph view
Triple Window	Shows map, graph, and report views
Quad Window	Shows map, graph, report, and coordinate views

The user can select the **QUAD WINDOW** button from the toolbar for a quick way to display the Quad Window. Figure 4.13 shows the results of the **Quad Window** option.

To summarize, in Figure 4.14 we show the selection of a **Sampling Goal** and sample type **(Simple Random Sampling),** we have entered the **DQO inputs** into the dialog box, **Applied** the design to our **Sample Area,** and displayed the Map, Graph, Report, and Coordinates simultaneously using the **Quad Window** from the Windows menu.

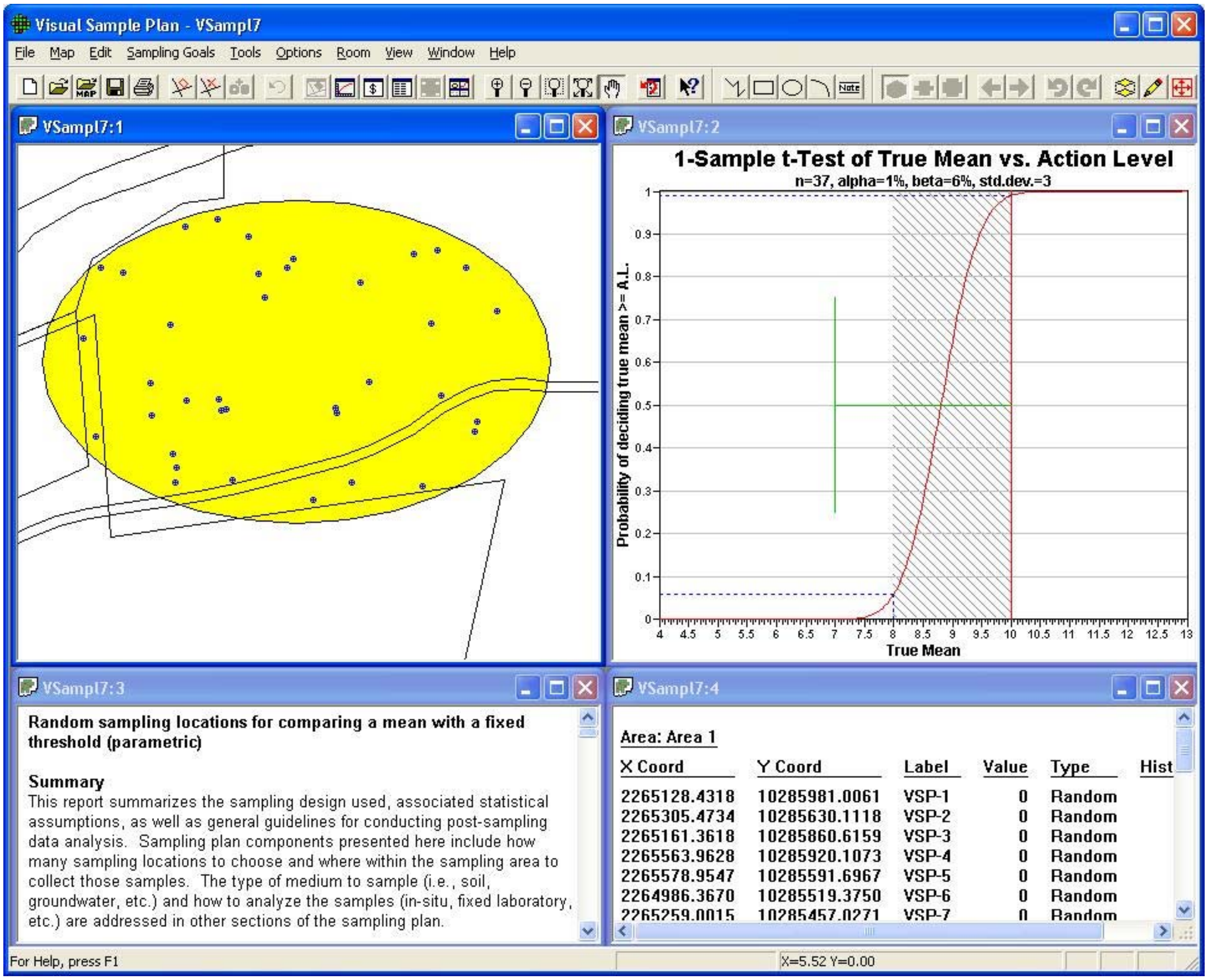


Figure 4.15. Quad Display of Map, Graph, Report, and Coordinates on Same Screen

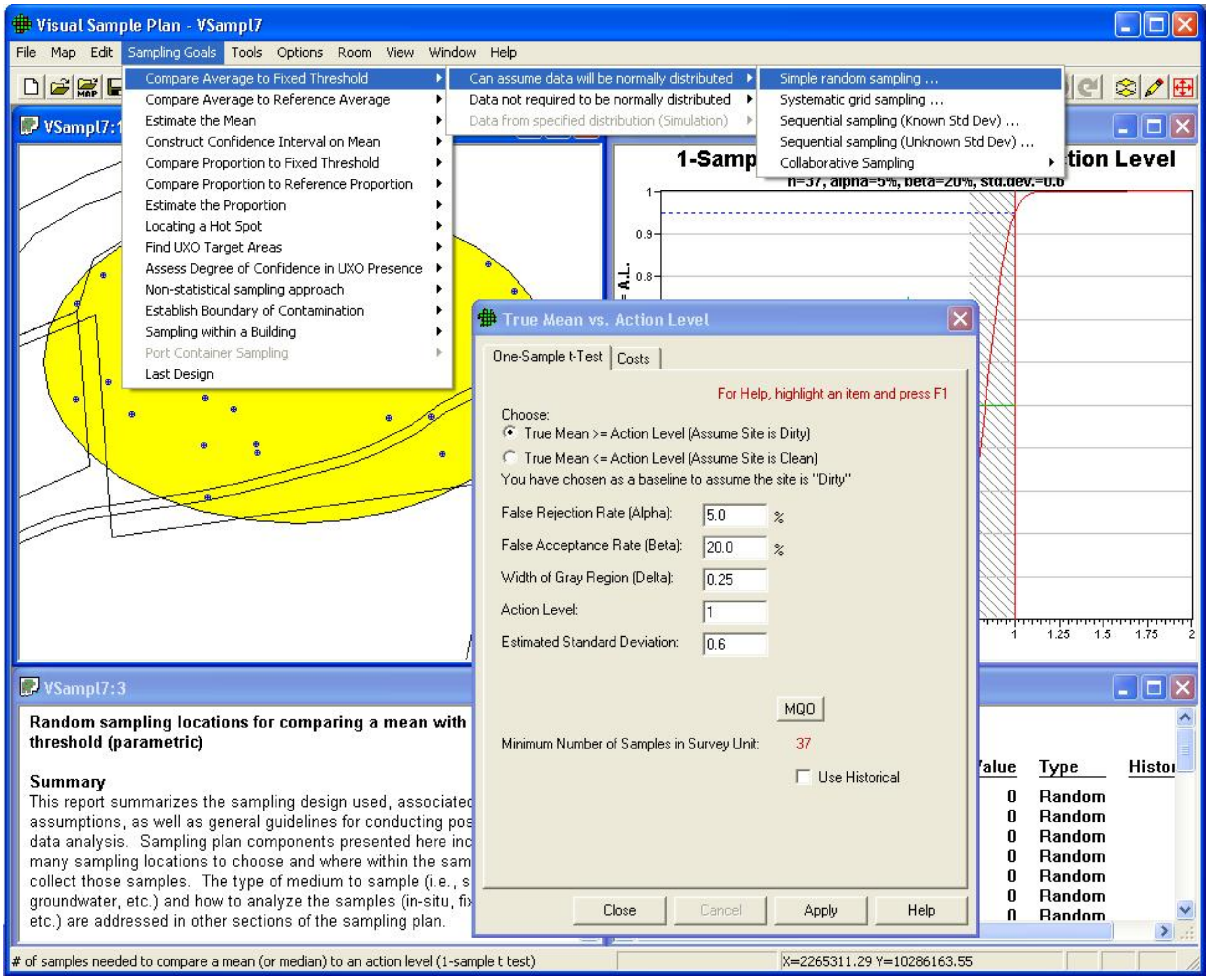


Figure 4.16. Combined Display of VSP Inputs and Outputs

5.0 Extended Features of VSP

VSP 3.0 has many extended features that have not been described so far in this User's Guide. In this section, we discuss some key extended features. The beginning user may not need these features, but a more experienced user will find them invaluable. These features expand on VSP's core capabilities. They are useful once a user has identified a basic sampling design and now wants to explore variations of the design, explore features of the design that are not part of the initial selection parameters, and add more capability to VSP.

The extended features fall into three categories:

- Features found in Main Menu items: **Tools, Options, and View**, and
- Features found in the Dialog Box for individual sampling designs, e.g., the **Cost Tab,** and the **Data Analysis Tab**
- Multiple Areas to be Sampled

5.1 Tools

5.1.1 Largest Unsampled Spot

If VSP has generated a sampling design for a Sample Area and you want to know the largest unsampled area, VSP can display this information. The largest unsampled spot is defined as the largest circle that will fit inside a Sample Area without overlapping a sample point.

In Figure 5.1, we opened the VSP Project File Example1.vsp included with the standard VSP installation. From the Main Menu we select **Tools > Largest Unsampled Spot > Find**. A dialog box tells you that VSP will search the Sample Area to find the largest circle that would fit into the unsampled area. The user is given the option of specifying the accuracy of the circle's radius, whether to consider area corners as additional sample points, and whether to allow the spot to overlap the Sample Area. After hitting the **OK** button, VSP searches the Sample Area, and places the spot on the Map, and displays an Information Box that says the radius of this circle is **205.22 ft**. (see Figure 5.2).

Two other displays are available: **Show Size**... and **Inside Area**.... The Show Size... displays the same Information Box shown in Figure 5.2. **Inside Area**… brings up the Information Box shown in Figure 5.3. It says that 100% of the circle is within the Sample Area. If the option to allow the largest unsampled spot to overlap the Sample Area edges had been selected, there may be situations where the circle extends beyond the boundary of the Sample Area resulting in a percentage less than 100%.

5.1.2 Reset Sampling Design

This command clears the current sampling design and removes all samples from the map (including unselected sample areas).

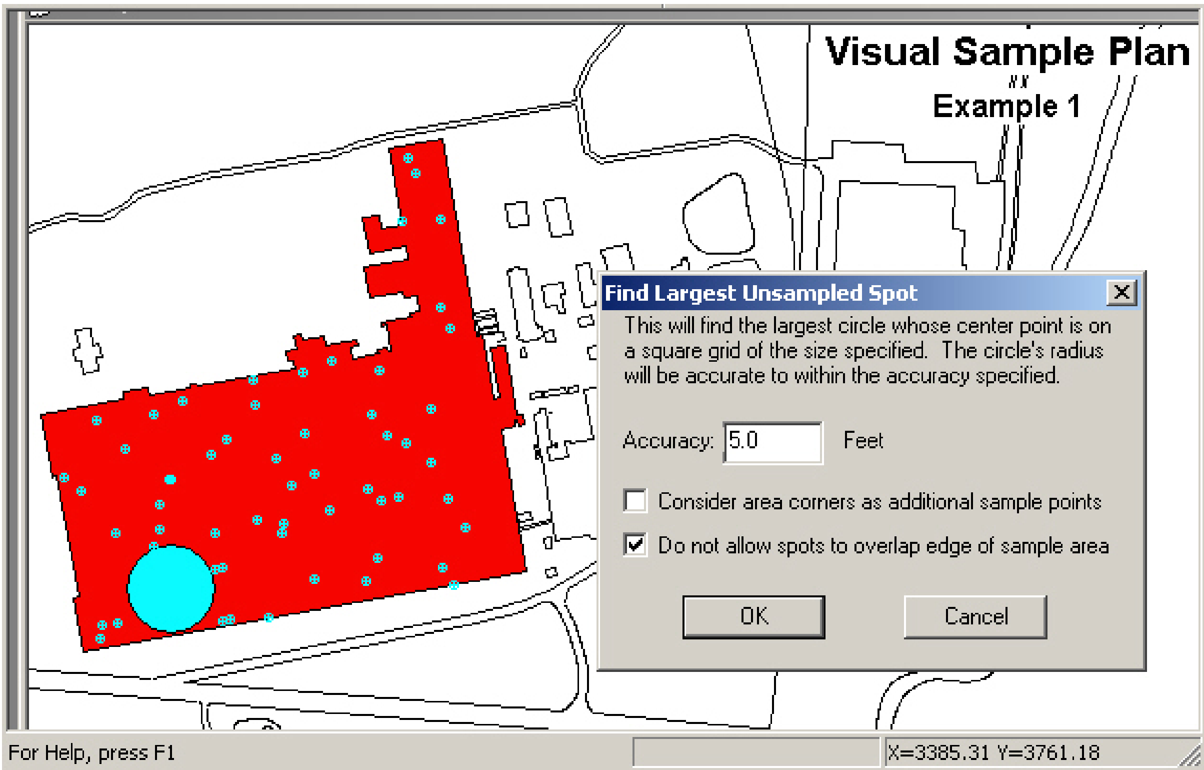


Figure 5.1. Largest Unsampled Spot Displayed on Map

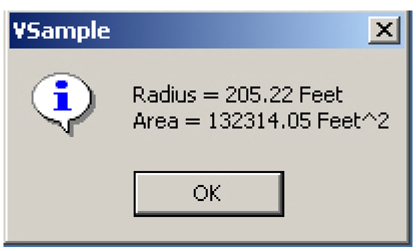


Figure 5.2. Information Box for Largest Circle that Would Fit Within an Unsampled Area

5.1.3 Measure Distance

Use this tool to measure distances on the map. After selecting this command, the cursor will become a ruler. Click on the map or enter a location (x, y) on the keyboard to anchor the first point. A line will be drawn from the anchor point to the cursor as you move the mouse. The status bar will also indicate the distance from the anchor point to the cursor. After clicking on a second point or entering a second point on the keyboard, a dialog will appear displaying the distance. In Figure 5.4, we see that the distance from the sampling point to the building edge is **547.33 ft**.

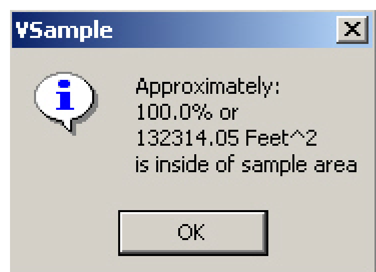


Figure 5.3. Information Box Showing Percentage of Circle Within the Sample Area

Hold the Shift key down to attach either point to an existing point on the map.

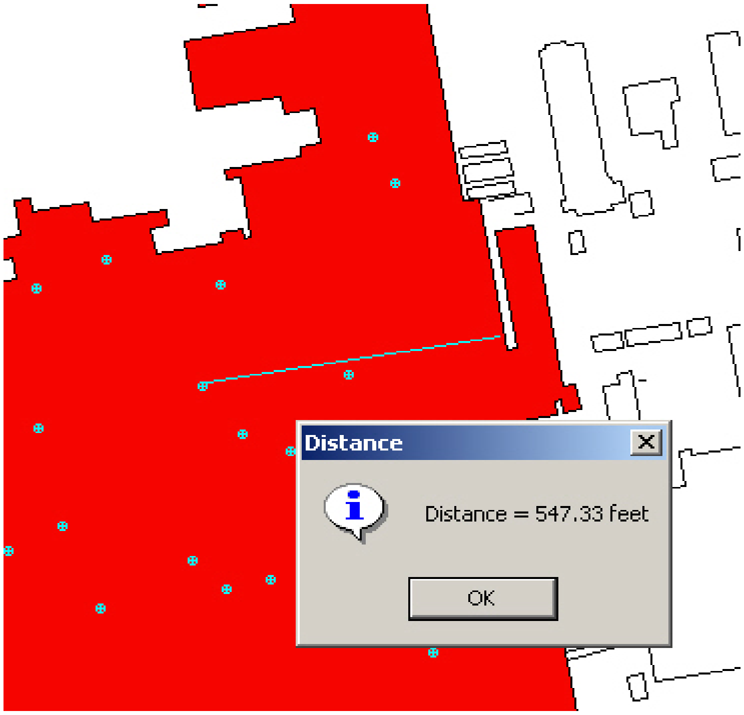


Figure 5.4. Measuring Tool in VSP

5.1.4 Make Sample Labels

Individual samples can have labels and values associated with them. This tool lets the user design the sample label. Selecting **Tools > Make Sample Label** brings up the Dialog Box shown in Figure 5.5. VSP assumes the user will want to assign a unique number to each sample within a Map, so all labels start with "VSP-<NNN>". Other information can be added to the label, such as the Local X Coordinate and Local Y Coordinate as shown in Figure 5.5, by selecting the information variable names on the list and hitting the **Add** button. (The information can also be added by double-clicking on the list item or by typing in the label format edit box.). Once the **OK** button is pushed, the user sees the current set of Sample Labels in Map View.

Sample labels are discussed in Section 2.4

5.2 Options

5.2.1 Random Numbers

VSP allows the user two options when selecting how random numbers are generated. The random numbers are used to pick coordinates for sampling locations when the design calls for either a random-start grid or random placement of all points. The user selects the desired random number generator using **Options > Random Numbers** from the Main Menu. The two options are **Pseudo-Random Numbers and Quasi-Random Numbers**. The user "toggles" between these two options. This is shown in Figure 5.6. Note that once an option is selected, it remains active until changed. VSP is initialized with the **Pseudo-Random Numbers** option active.

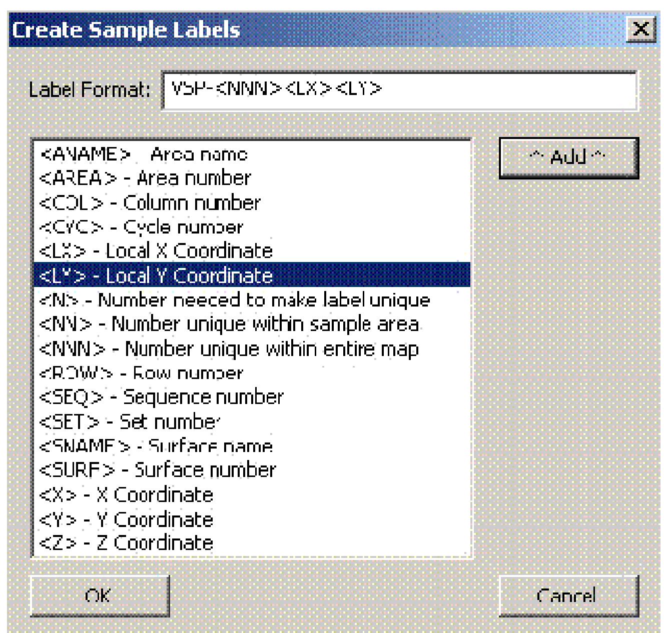


Figure 5.5. Dialog Box for Creating Sample

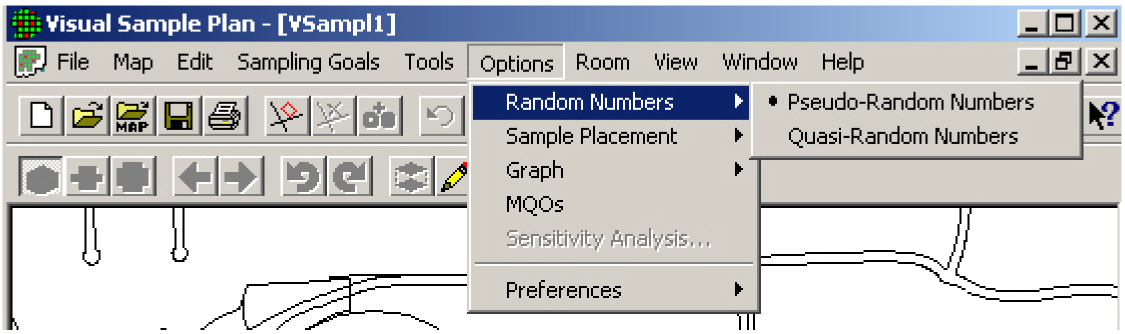


Figure 5.6. Menu for Selecting Type of Random Number Generator

Sampling locations (i.e., the x and y coordinates of the location) chosen with a pseudo-random number generator are not restricted in any way. The first location chosen and the tenth location chosen can be right next to each other or far apart, like throwing darts at a dart board. The locations where the darts hit can be clumped together or spread out, depending on chance.

Quasi-random numbers are generated in pairs. One member of the pair is used for the X coordinate; the other member is used for the Y coordinate. The sequence of paired numbers is generated in such a way that sample points tend to be spread evenly over a sample area. VSP's quasi-random-number generator uses Halton's Sequence. For a discussion of the algorithms used for both the pseudo- and the quasi-random number generator, see Version 2.0 Visual Sample Plan (VSP) Models and Code Verification (Gilbert et al. 2002).

If the current sampling design is being added to a study area with existing sampling locations, the quasi--random number generator will have no knowledge of those locations and might by chance put a new sampling location right next to an existing location. See the **Adaptive-Fill** option in Section 5.2.2 to handle the problem of avoiding existing sampling locations.

5.2.2 Sample Placement

The Adaptive-Fill option allows the addition of "random" sampling locations in such a way as to avoid existing sampling locations. Adaptive Fill has to do with the placement of the sampling locations, not the number of samples. The basic idea is to place new sampling locations so as to avoid existing locations and still randomly fill the Sample Area. The current Sampling Design option determines the number of locations.

VSP usually places new sampling locations using the default option, **Options > Sample Placement > Regular Random**. When Regular Random is selected, the sampling locations produced by either of the two random number generators discussed in Section 5.2.1 are placed in the Sample Area without regard to pre-existing samples. In fact, VSP removes all previous sampling locations prior to placing the new set of sampling locations.

When the **Options > Sample Placement > Adaptive-Fill** option is selected, all pre-existing sampling locations are left in place, and new sampling locations are placed in the Sample Area using an algorithm to maximally avoid preexisting sampling locations. The Adaptive-Fill algorithm can be used with either random number generator. The Adaptive-Fill option is shown in Figure 5.7.

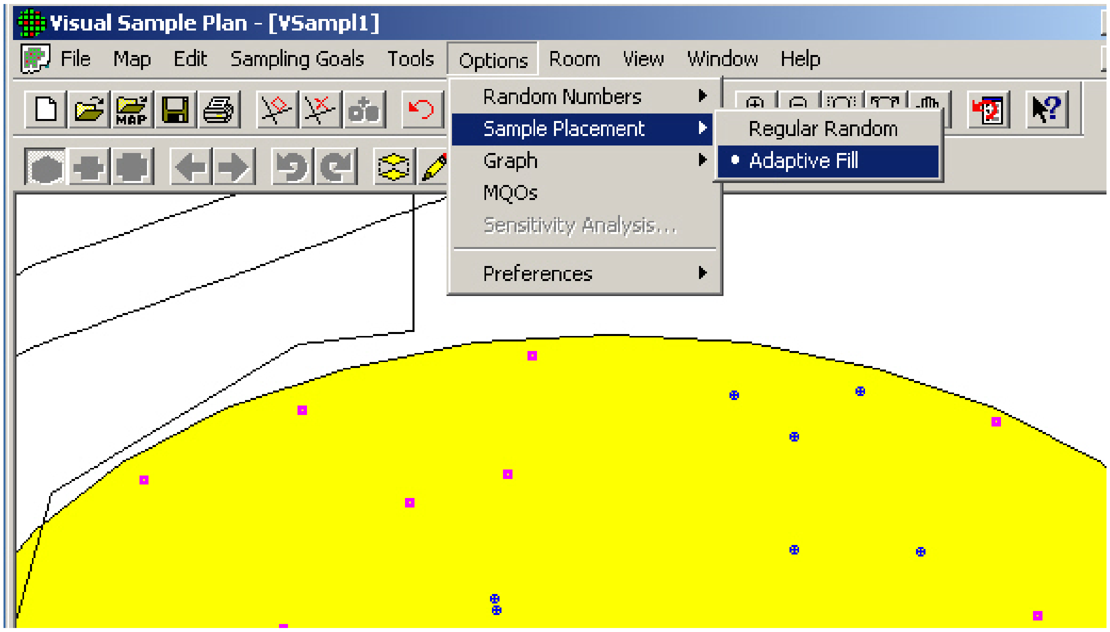


Figure 5.7. Adaptive-Fill Option for Sample Placement (Shown Here with Sample Area from Millsite Map)

Note that in Figure 5.7 the original sampling locations are marked with a circular symbol. In contrast, the Adaptive-Fill sampling locations are marked with a square symbol. If you right-click on a sampling-location symbol, a Sample Information dialog will display the type of sample, the coordinates, and a label input field. The label input field allows a specific sampling location to be given an ID number or remark. The label information is displayed in the Sample Information dialog, the report view, and the coordinate view. The label is also exported along with other sample information when exported to a text file (see Figure 5.9). See Figure 5.8 for an example of right-clicking on an Adaptive-Fill sampling location.

If the sampling locations are exported to a text file using **Map > Sample Points > Export**, an Adaptive-Fill location will be noted and any label the user might have added will be saved. An example text file is shown in Figure 5.9.

5.2.3 Graph

Graphs can be displayed with many different options. Figure 5.10 shows the options that can be selected using **Options > Graph**. Options are selected by clicking the option on or off. Once selected, that option will be in place for all Graphs. Note that we saw these same options in Chapter 4, Figure 4.2a, by right-clicking on a Graph. Table 4.2 describes these options in more detail.

5.2.4 Measurement Quality Objectives (MQOs)

The Measurement Quality Objectives (MQO) module in VSP provides a way to extend the sampling design to consider not only the number and placement of samples in the field but also what happens in the measurement or analysis process. After all, it is the final result of the "measured sample value" that gets reported back to the project manager and used in statistical tests to make a decision.

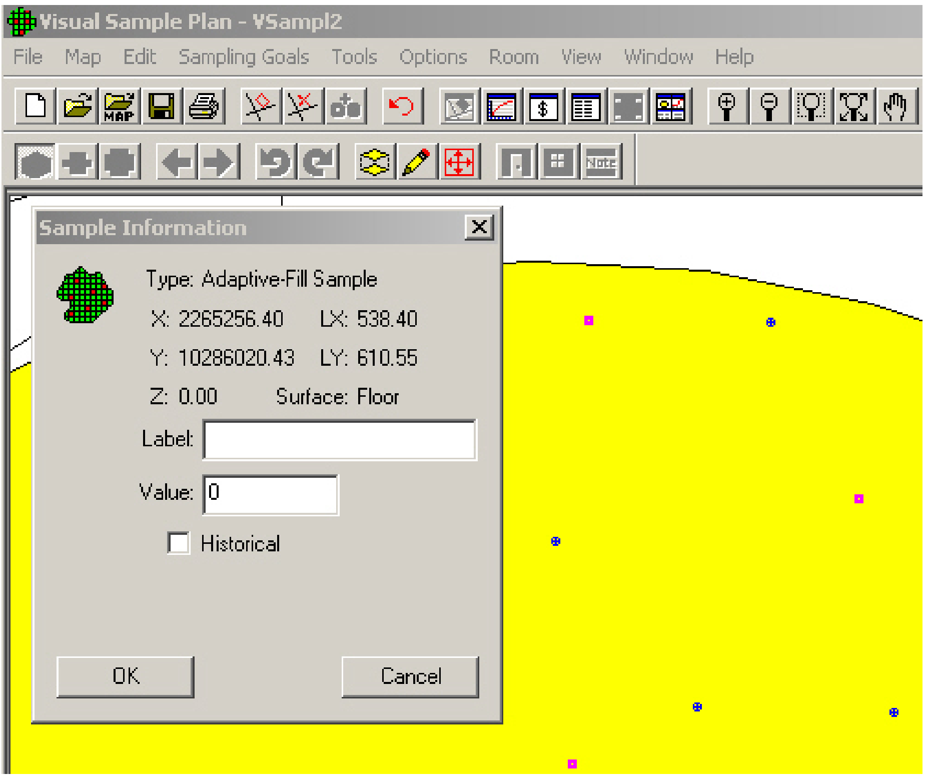
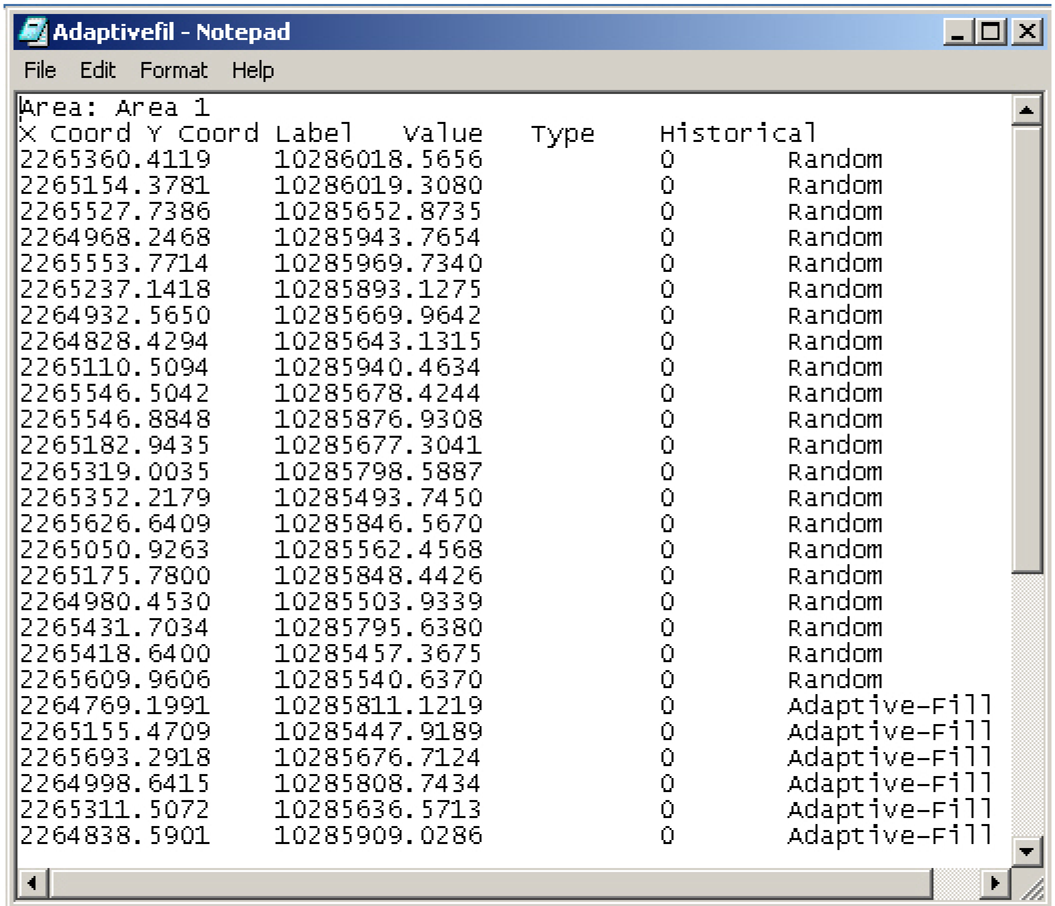


Figure 5.8. Sample Information Window Displayed When the User Right-Clicks on Selected Sample Points on Map

There is a trade-off between taking more samples using a crude (i.e., less precise) measuring device vs. taking fewer samples using a precise measuring device and/or method. This is because total decision error is affected by the total standard deviation of the samples. The total standard deviation includes both sampling variability and analytical measurement variability.

There is also a trade-off between taking more measurements (i.e., replicate measurements) when using these less precise analytical measuring devices and/or methods vs. taking few measurements and using more precise analytical measuring devices and/or methods. The MQO module in VSP lets the user play "what-if" games with various combinations of sampling standard deviation, analytical (i.e., measurement) standard deviation, number of analyses (i.e., replicates) per sample, and number of samples to take. More discussion of this topic and the sample size equations behind the VSP calculations can be found in Version 2.0 Visual Sample Plan (VSP) Models and Code Verification (Gilbert et al. 2002).



Area: Area 1

X Coord	Y Coord	Label	Value	Type	Historical
2265360.4119	10286018.5656		0	Random	
2265154.3781	10286019.3080		0	Random	
2265527.7386	10285652.8735		0	Random	
2264968.2468	10285943.7654		0	Random	
2265553.7714	10285969.7340		0	Random	
2265237.1418	10285893.1275		0	Random	
2264932.5650	10285669.9642		0	Random	
2264828.4294	10285643.1315		0	Random	
2265110.5094	10285940.4634		0	Random	
2265546.5042	10285678.4244		0	Random	
2265546.8848	10285876.9308		0	Random	
2265182.9435	10285677.3041		0	Random	
2265319.0035	10285798.5887		0	Random	
2265352.2179	10285493.7450		0	Random	
2265626.6409	10285846.5670		0	Random	
2265050.9263	10285562.4568		0	Random	
2265175.7800	10285848.4426		0	Random	
2264980.4530	10285503.9339		0	Random	
2265431.7034	10285795.6380		0	Random	
2265418.6400	10285457.3675		0	Random	
2265609.9606	10285540.6370		0	Random	
2264769.1991	10285811.1219		0	Adaptive-Fill	
2265155.4709	10285447.9189		0	Adaptive-Fill	
2265693.2918	10285676.7124		0	Adaptive-Fill	
2264998.6415	10285808.7434		0	Adaptive-Fill	
2265311.5072	10285636.5713		0	Adaptive-Fill	
2264838.5901	10285909.0286		0	Adaptive-Fill	

Figure 5.9. Sample Exported Text File of Sampling Locations

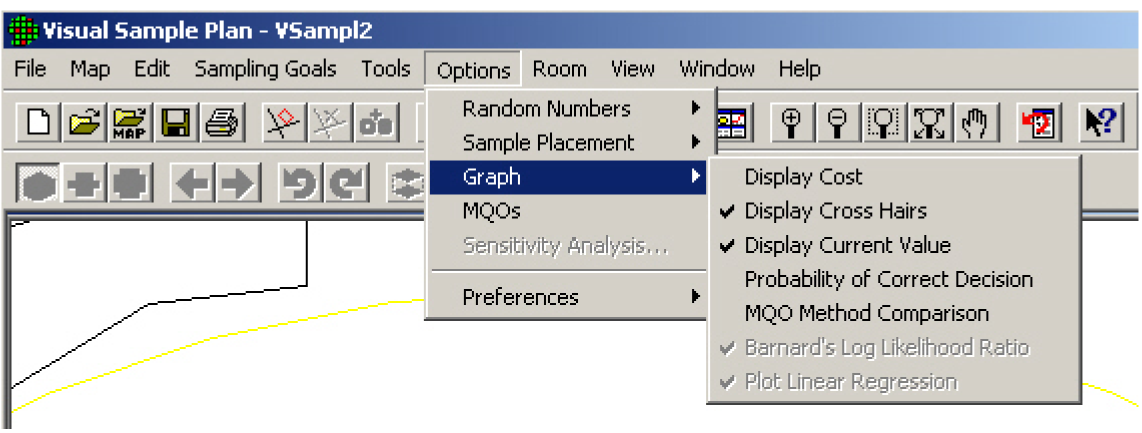


Figure 5.10. Graph Options

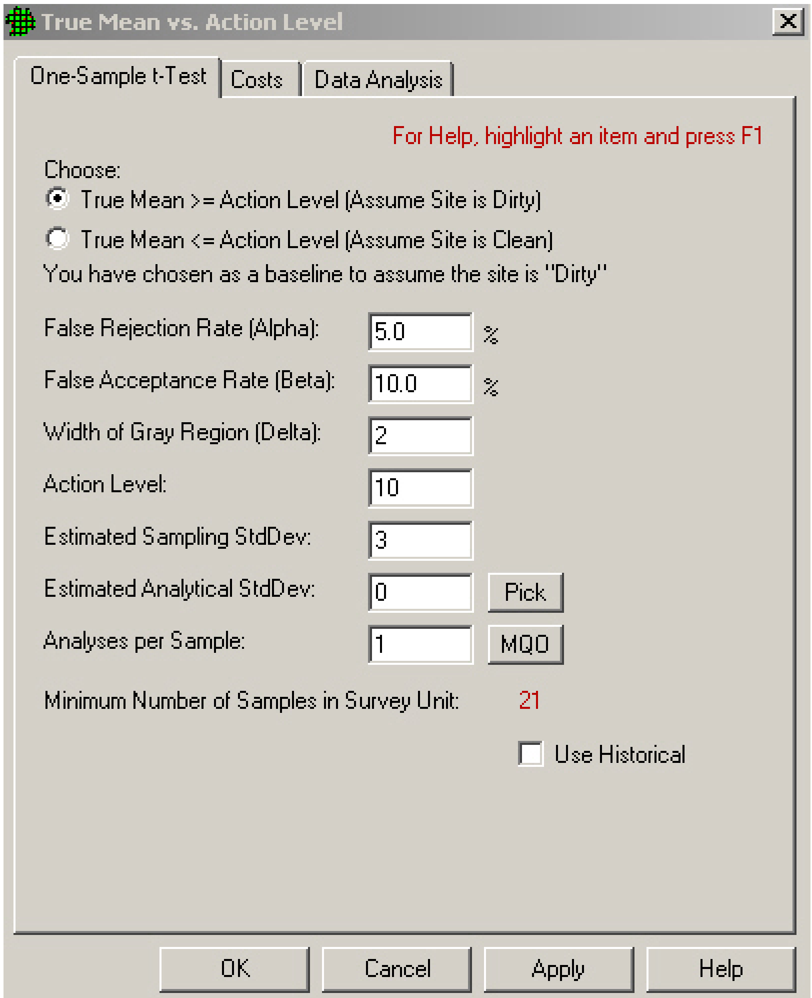


Figure 5.11. MQO Input Dialog Box with Default Values Displayed

The MQO option is selected from the dialog that pops up after a Sampling Design has been selected. The MQO option can also be toggled using **Options > MQOs** from the main menu. In Figure 5.11, we see a dialog box that contains the MQO button (**Sampling Goals > Compare Average to Fixed Threshold**). This dialog box allows you to provide additional inputs, such as the analytical standard deviation and number of analyses per sample. There is also a **Pick** button (not active at this time but planned in future versions of VSP) to provide access to a library of standard analytical methods with their reported analytical standard deviations.

Note that the default values are **0** for the **Estimated Analytical Standard Deviation** and **1** for the **Analyses per Sample**. This means that the user-selected analytical or measurement method does not add a significant component of variability to the total standard deviation; i.e., the method provides essentially the same numeric value when repeated measurements are made on a sample. Using the input parameter values shown in Figure 5.11 and with these default MQO values, we get n = 21 samples.

Now let's start changing the MQO input values. First, we change the **Estimated Analytical Standard Deviation** to **3**. We still take only one analysis per sample. We see VSP now tells us we need to take **40** field samples to obtain the desired error rates we specified. This is shown in Figure 5.12.

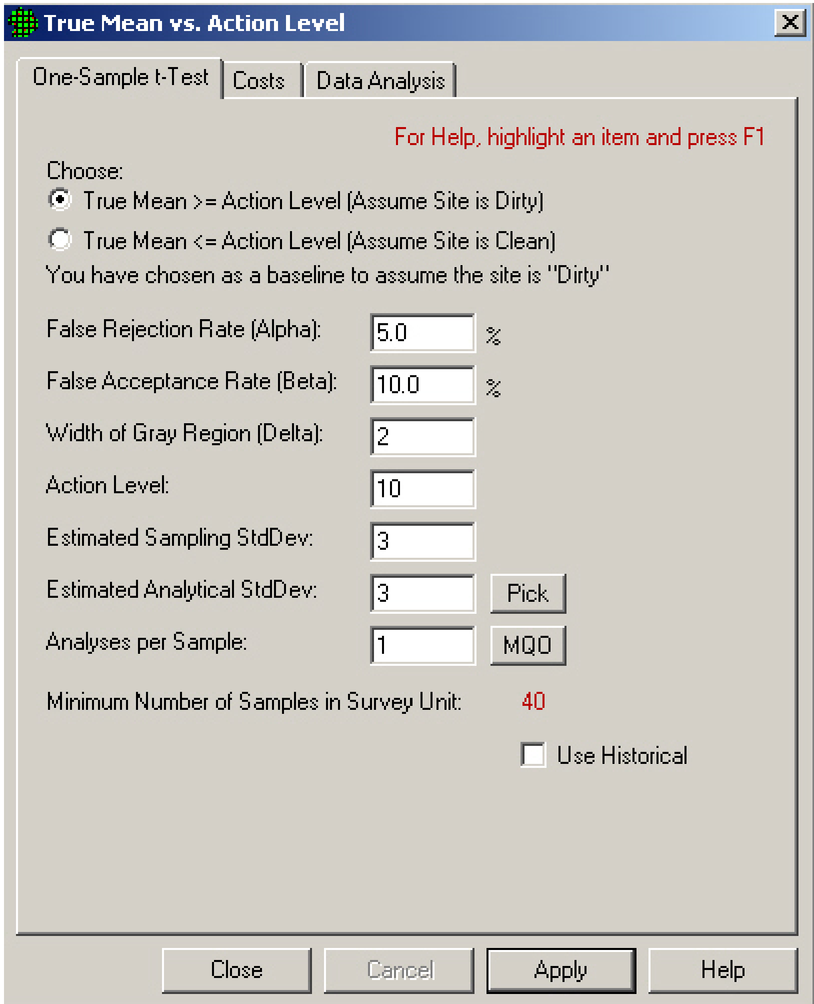


Figure 5.12. MQO Input Dialog Box Showing Positive Value for **Estimated Analytical Standard Deviation** with **1 Analysis per Sample**

If we take two repeated measurements of each sample (**Analyses per Sample** set to **2**), we see in Figure 5.13 that the number of field samples is now only **31**.

You can try different values in the MQO input boxes and see the effect on the resulting number of field samples.

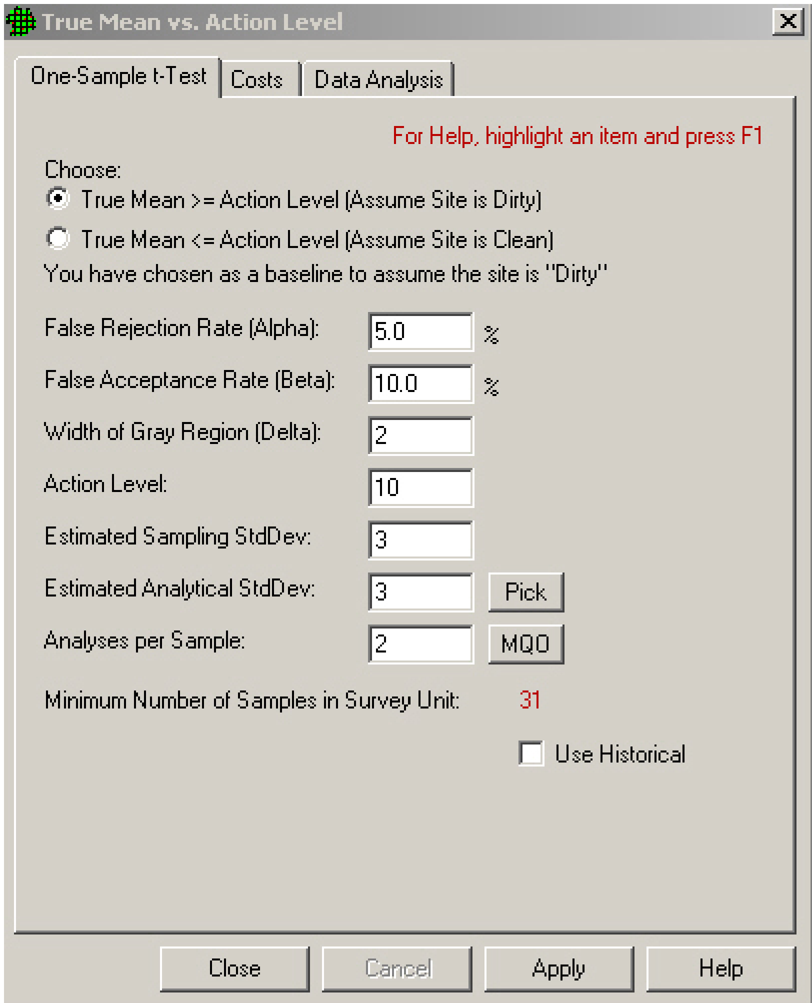


Figure 5.13. MQO Input Dialog Showing Positive Value for **Estimated Analytical Standard Deviation** with **Multiple Analyses per Sample**

When you select the **COSTS** tab at the top of the screen, a new display and set of inputs is shown. This is shown in Figure 5.14. In this dialog box, we can enter costs for **Field Collection** (shown here as **$100** per sample) and **Analytical Cost per Analysis** (shown here as **$400** per analysis). This screen also provides a **Cost Comparison** between two possible options, Analytical Methods A and B. We see the Method A Analytical Standard Deviation of **3** that we entered on the previous screen. We can also enter an Analytical Standard Deviation for Method B. Initially, VSP displays the default values of **0** for Method B as shown in Figure 5.14. VSP displays the comparison for one, two, or three replicate analyses for only Method A because Method B has an analysis cost of $0.00.

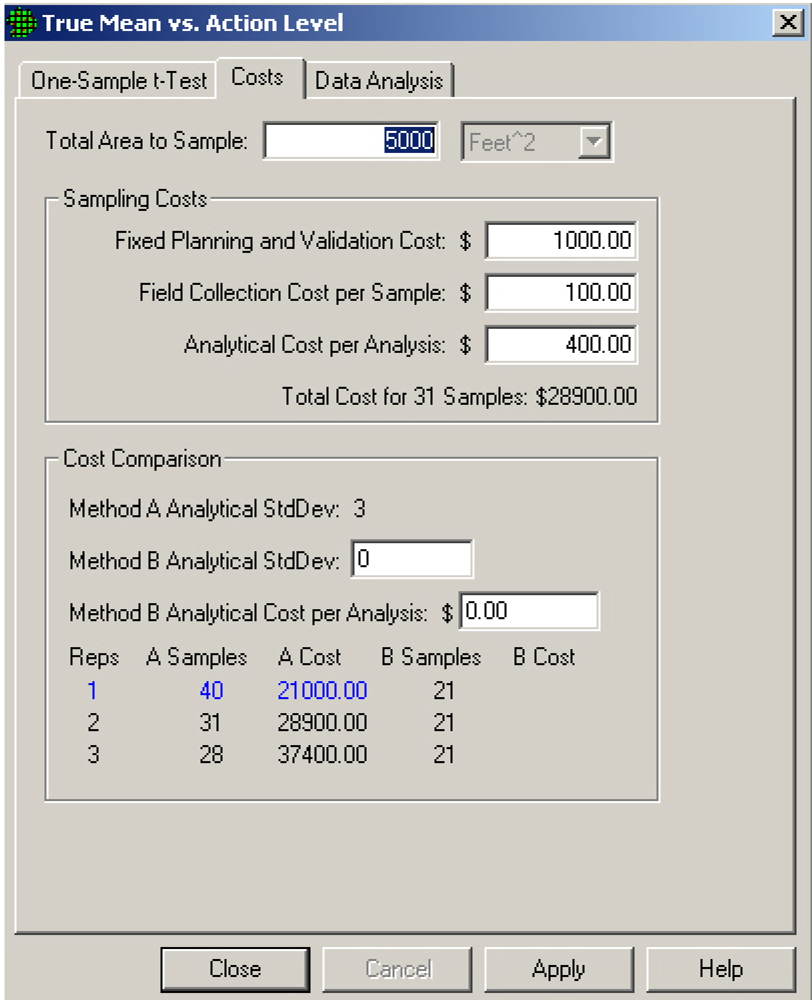


Figure 5.14. Cost Input Dialog Box for MQO Option

Next we show input values for Method B. Here, we enter a M**ethod B Analytical Standard Deviation** of **4** (somewhat higher than Method A), but with a lower **Cost per Sample** (shown here as **$100**). In Figure 5.15 we see that the Method Comparison is now filled in with the new values. The lowest cost option (Method B with 1 Analysis per Sample) is highlighted in blue.

Notice that the lowest cost sampling design for this problem has the most field samples, n = 55. This is because Method B has a very low analysis cost of only $100 vs. the much higher cost for Method A of $400. Therefore, Method B can reduce the uncertainty in the final decision by allowing many more field samples to be analyzed compared with Method A.

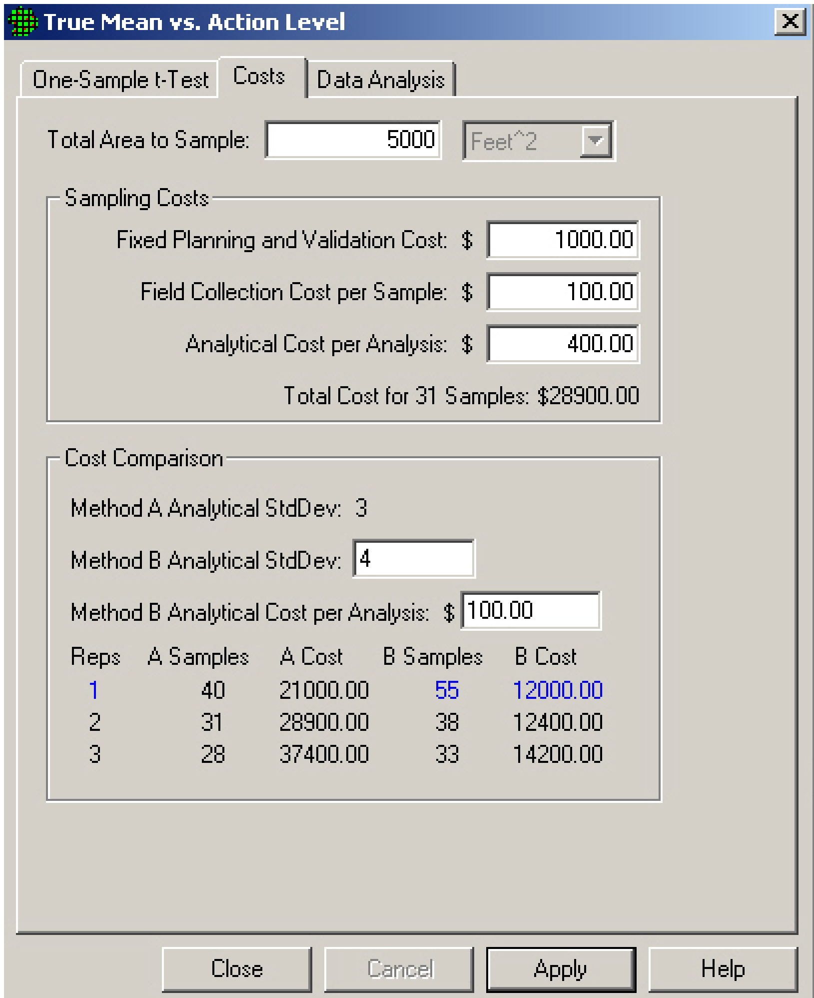


Figure 5.15. Display of Cost Comparison for Method A and Method B from MQO Module

Note also that the sampling design will not automatically change to the Method B case highlighted in blue. If you want a sampling design based on Method B, you must update the **Analytical Cost per Analysis** for Method A to match the Method B cost. Then return to the **One-Sample t-Test** tab, change the **Estimated Analytical Standard Deviation** value to match the Method B value, and press the **Apply** button to get the Method B-based sampling design.

A graphical comparison of the analytical methods is shown on the Decision Performance Curve when **Options > Graph > MQO Method Comparison** is checked. You must go to **View > Graph** to see the chart. Figure 5.16 shows an example.

The yellow circle is placed above the lowest-cost sampling design that meets the objectives. In this case, the circle is above a green bar representing the cost of using sampling design Method B with one analysis per sample.

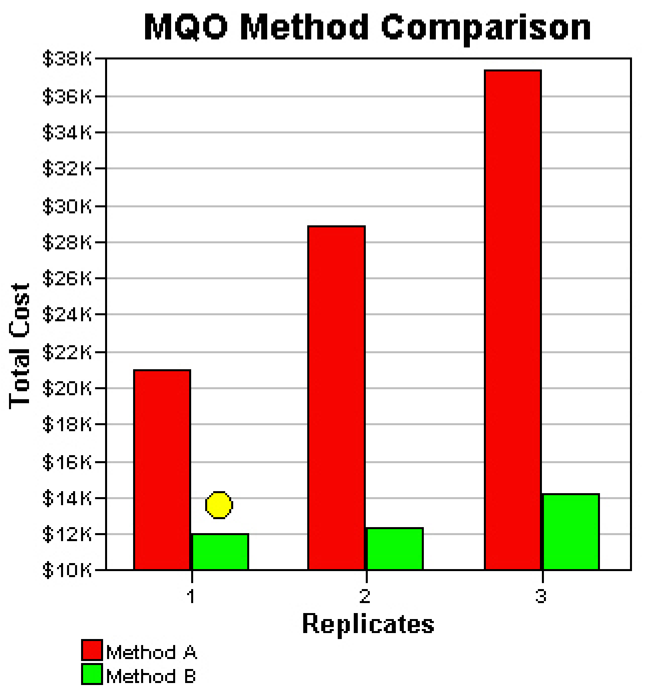


Figure 5.16. MQO Method Comparison Chart

5.2.5 Sensitivity Analysis

This option accesses the sensitivity analysis parameters on the Report View. The sensitivity analysis parameters may also be accessed by right-clicking on the Report View itself. The example shown in Figure 5.17 is from the VSP Project File Example1.vps. With **View** > **Report** selected, scroll down to the section on Sensitivity Analysis. Now select **Options** > **Sensitivity Analysis**, and the Dialog Box shown in Figure 5.17 is displayed. The user can do sensitivity analysis on up to 4 variables (only 3 are shown here), can select a starting and ending value for each variable, and can specify the number of steps for incrementing the variable. VSP displays the values for each step in Red below the Step window. For this example, in Figure 5.17, we say we want to see the number of samples required at values of the Standard Deviation s= 1.2, and 0.6. For each of these levels, we want to look at three levels of Beta, and three levels of Alpha. VSP calculates sample size for each of these 2 x 3 x 3 = 18 options and displays the values in the table.

This option is a very powerful tool for looking at “what if” scenarios and determining trade-offs for risk and cost. Tables for Number of Samples, Sampling Cost, or both can be displayed. Figure 5.18 shows 4 of the DQO Parameters being changed, and shows the results of the sensitivity analysis for both number of samples and cost.

5.2.6 Preferences

Figure 5.19 shows the Preferences available in VSP. Table 5.1 provides a brief description of each Preference. Consult VSP **Help** > **Help Topics** > **Menus** > **Options menu** > **Preferences** for a more detail discussion of each menu item.

5.3 View Menu

The View Menu offers the user options for how VSP displays information. Table 5.2 gives a brief description of each option. Consult VSP **Help** > **Help Topics** > **Menus** > **View menu** for a more detail discussion of each menu item. Many of these items have been discussed previously in this manual and will be mentioned only briefly here.

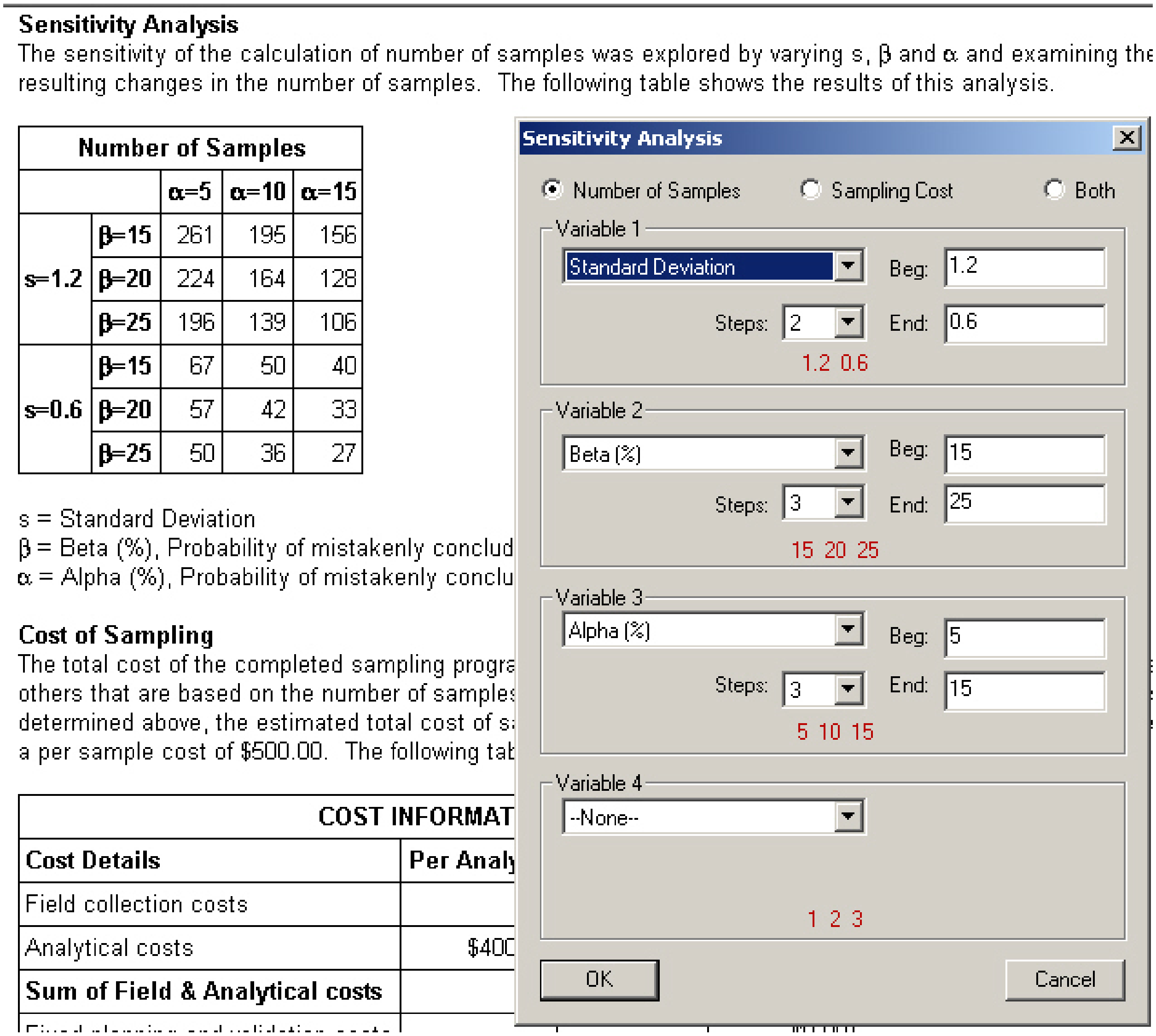
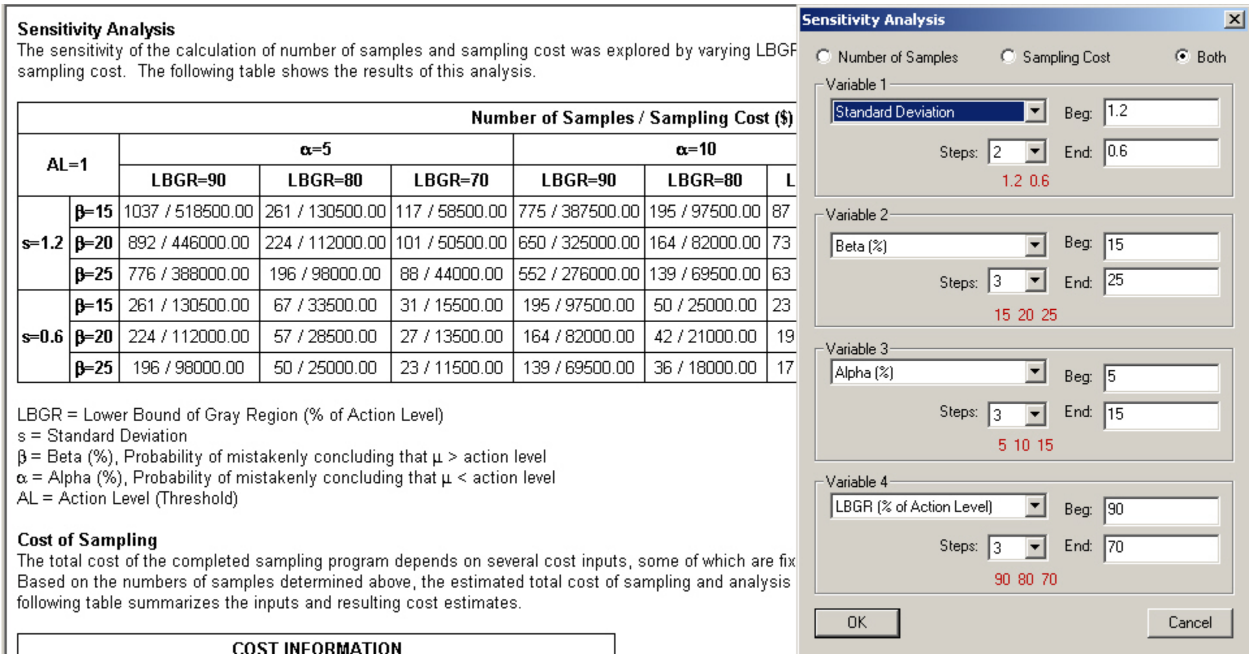


Figure 5.17. Sensitivity Analysis for 3 DQO Input Parameters. Results are shown for Number of Samples, as displayed in a table in Report View.



Sensitivity Analysis
The sensitivity of the calculation of number of samples and sampling cost was explored by varying LBGF
sampling cost. The following table shows the results of this analysis.

AL=1		Number of Samples / Sampling Cost ($)					
		α=5			α=10		
		LBGR=90	LBGR=80	LBGR=70	LBGR=90	LBGR=80	L
s=1.2	β=15	1037 / 518500.00	261 / 130500.00	117 / 58500.00	775 / 387500.00	195 / 97500.00	87
	β=20	892 / 446000.00	224 / 112000.00	101 / 50500.00	650 / 325000.00	164 / 82000.00	73
	β=25	776 / 388000.00	196 / 98000.00	88 / 44000.00	552 / 276000.00	139 / 69500.00	63
s=0.6	β=15	261 / 130500.00	67 / 33500.00	31 / 15500.00	195 / 97500.00	50 / 25000.00	23
	β=20	224 / 112000.00	57 / 28500.00	27 / 13500.00	164 / 82000.00	42 / 21000.00	19
	β=25	196 / 98000.00	50 / 25000.00	23 / 11500.00	139 / 69500.00	36 / 18000.00	17

LBGR = Lower Bound of Gray Region (% of Action Level)
s = Standard Deviation
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level
AL = Action Level (Threshold)

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fix
Based on the numbers of samples determined above, the estimated total cost of sampling and analysis
following table summarizes the inputs and resulting cost estimates.

COST INFORMATION

Figure 5.18. Sensitivity Analysis for 4 DQO Input Parameters. Results are shown for Number of Samples and Cost, as Display in a Table in Report View

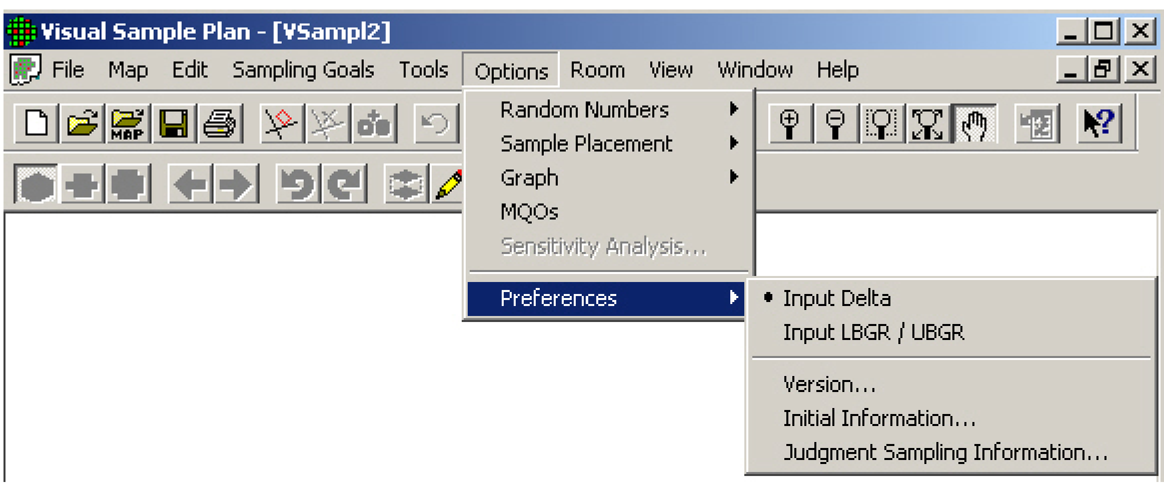


Table 5.1. Preferences Menu Items

Input Delta	Allows the input of Delta in design dialog boxes
Input LBGR / UBGR	Allows the input of LBGR / UBGR in design dialog boxes.
Version	Changes the sub-version of VSP (which eliminates some of the sampling design choices)
Initial Information	Displays the initial information associated with the chosen sub-version of VSP.
Judgment Sampling Information	Displays informational dialog when judgment sampling is used.

Table 5.2. View Menu Items

Main Toolbar	Shows or hides the main toolbar
Map Drawing Toolbar	Shows or hides the toolbar used for drawing on maps.
Ranked Set Toolbar	Shows or hides the toolbar used for ranked set sampling
Room Toolbar	Shows or hides the toolbar used for room manipulation
Status Bar	Shows or hides the status bar.
Labels	Shows or hides the sample labels on the map. Any combination of Labels, Coordinates, Local Coordinates and Values can be displayed.
Background Picture	Shows or hides the background picture.
Transparent Sample Area	Allow background picture to be seen behind sample areas.
Largest Unsampled Spots	Shows or hides the largest unsampled spots.
All Grid Cells	Shows or hides all grid cells used for adaptive cluster sampling
Leading Edge	Shows only the leading edge of an open-type sample area
Map Scale	Shows or hides the coordinate scale on the map.
Map Legend	Shows or hides the map size legend.
Swath Corners	Shows the coordinates at the corners of swaths
Room North Arrow	Shows or hides the north arrow in the room view.
Room Perspective Ceiling	Shows or hides the ceiling in the perspective room view.
Map	Change current project view to map view.
Graph	Change current project view to graph view.
Report	Change current project view to report view.
Coordinates	Change current project view to coordinate view.
Room	Change current project view to room view.
Zoom In	Increase the map view size.
Zoom Out	Decrease the map view size.
Zoom Max	Change map view size to fit window
Zoom Window	Increase map view size to selected area.
Pan	Move the visible portion of the map.

5.4 The Cost Tab: Setting Up Sampling Costs – Inputs for the General Screen

VSP allows users to enter sampling costs so that the total cost of a sampling program is available. Once a sampling design is selected and the DQO inputs have been entered into one of the dialog boxes, click on the Costs tab to enter costs. A sample **Costs** screen is shown in Figure 5.20. The inputs for this example were entered in the Dialog Box shown in Figure 5.21.

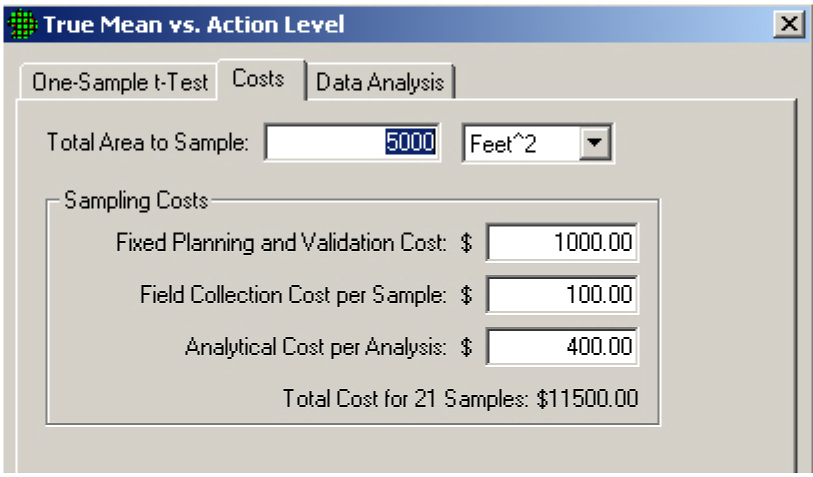


Figure 5.20. Screen for Entering Sampling Costs for a Sampling Design – Accessed through the Cost Tab

VSP enables you to break down costs into the following categories:

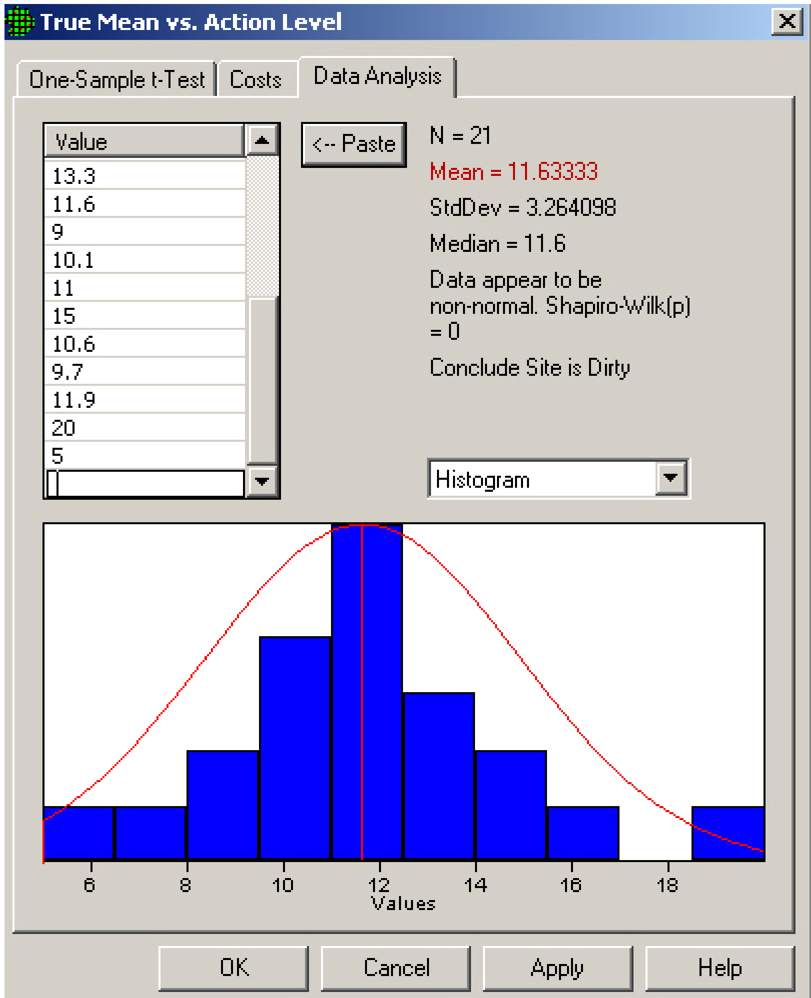
- *fixed planning and validation costs* - This is the fixed cost that is incurred, regardless of how many samples are taken. Examples of fixed costs are the cost to mobilize a sampling crew and get the equipment into the field.
- *field collection cost per sample* - This is the per-sample Figure 5.22. Screen for Entering cost. Examples of per-unit field costs are the costs paid to Sampling Costs for Sampling Design technicians to collect the sample and package and transport it.
- analytical cost per analysis - This is the cost to analyze a specimen or a sample. As discussed in Section 5.4, you can specify how many repeated analyses you want taken per sample or specimen.

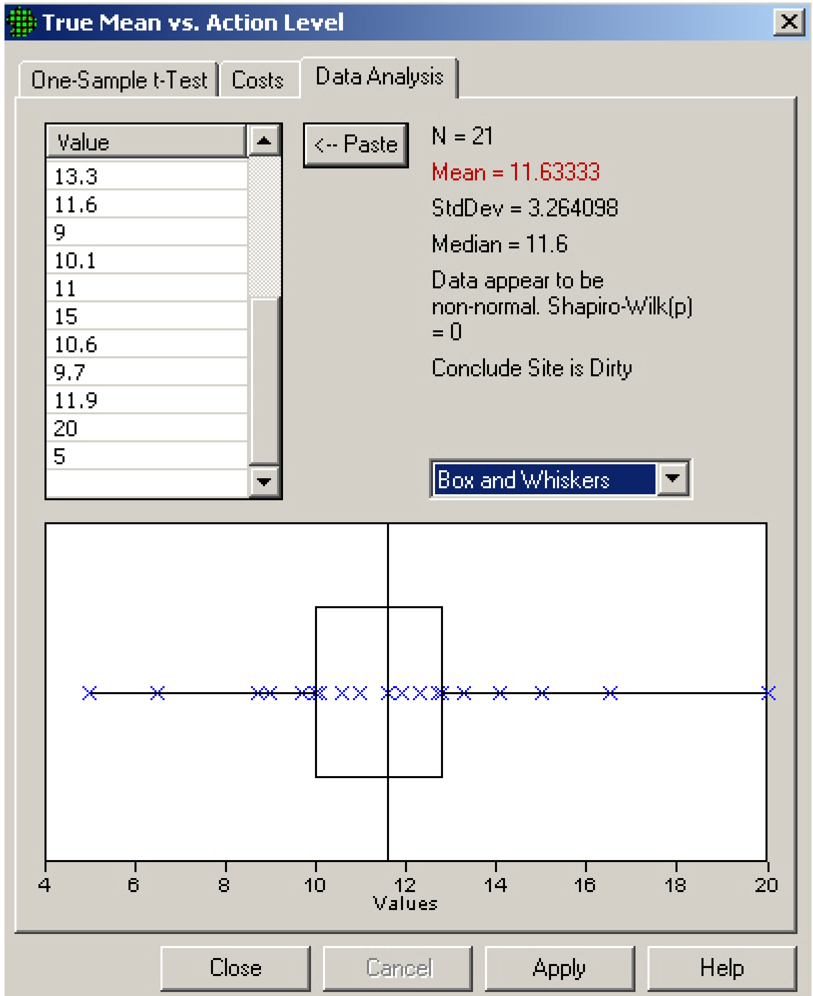
VSP calculates a total cost for the design specified, shown here as $11,500. Total cost is the sum of the fixed cost, shown here as $1,000, plus per-sample field collection cost of $100, plus analytical cost per analysis of $400, multiplied by the number of samples, 21. No duplicate analyses were specified, so the total per-unit cost is $500. Thus, the total sampling cost is $1000 + 21 x $500 = $11,500.

The hot spot sampling goal has some unique cost features. First, costs are displayed in one of the tables in the **Report View** and not on the **Cost** tab of the dialog box. Second, this is the only Sampling Goal for which you can specify a cost as a design criteria and VSP will calculate the number of samples to meet that goal (see Section 3.2.8). This is done by selecting from the main menu **Sampling Goal > Locating a Hot Spot > Systematic Grid Sampling > Predetermined fixed cost**.

5.5 Multiple Areas To Be Sampled

VSP allows the user to select multiple areas as sampling areas. All the examples shown so far involved a single Sample Area. When multiple areas are selected, VSP allocates the samples to the areas in proportion to the area of the respective individual sample areas. For example, if one area is twice as large as the other sample area, it will receive twice as many sample points. This is shown in Figure 5.23.





In Figure 5.23, we show two sample areas -- a rectangle and a circle. We next assume that a sampling-design algorithm not currently in VSP called for n = 25 samples. Using option Sampling Goals > Non-statistical sampling approach > Predetermined number of samples > Simple random sampling, VSP allocated 7 of the 25 requested samples to the rectangle and 18 to the circle. This is because the circle covers an area approximately three times larger than the rectangle.

Note that when multiple sample areas are drawn on a Map, you can select or deselect sample areas using Main Menu option Edit > Sample Areas > Select/Deselect Sample Areas. Alternatively, you can select or deselect a sample area by clicking on it with the mouse.

The Change Color option can be used to change a sample area's color. First, select those sample areas to be given a new color. Then use the Edit > Sample Areas > Change Color sequence and choose the new color for the currently selected sample areas.

Note that when multiple sample areas are selected, VSP-derived sampling requirements assume that the decision criteria and summary statistic of interest (mean, median) apply to the combined sample areas and not to the individual areas.

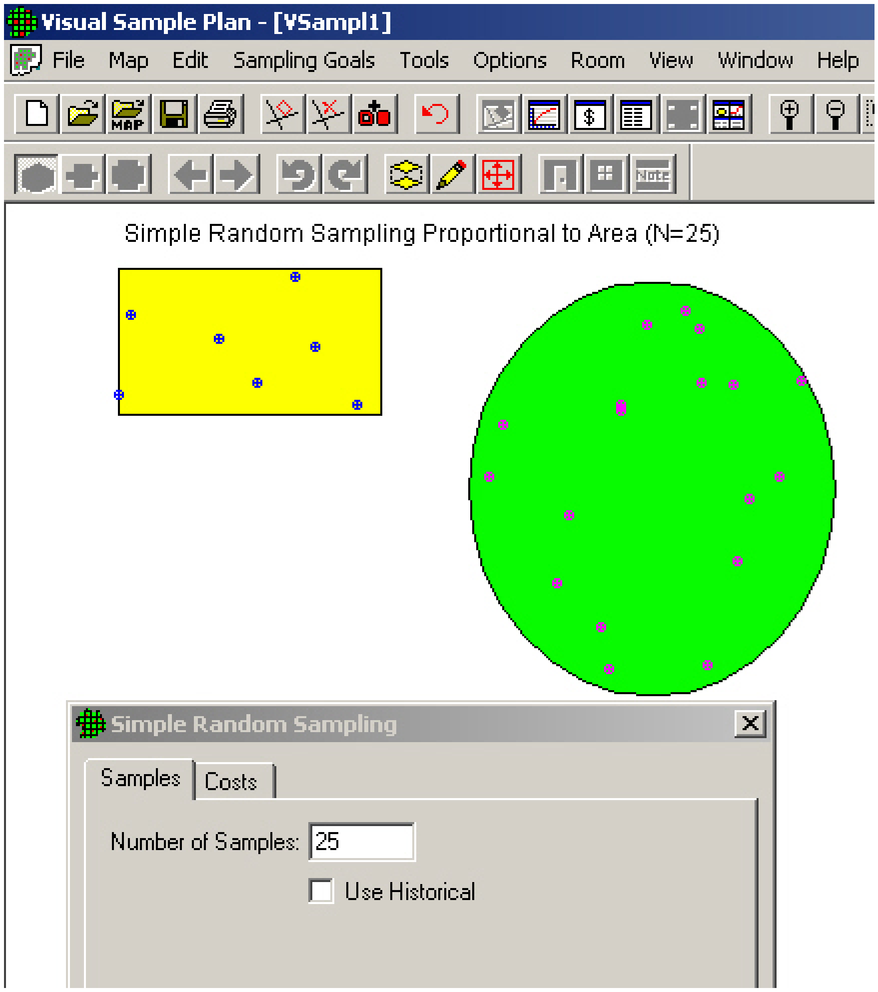


Figure 5.23. Proportional Allocation of Samples to Multiple Sample Areas

6.0 Room Features in VSP

In VSP versions prior to 2.5, support for rooms was limited to a **Draw Room** tool on the Map menu. This provided a way to easily draw a floor, four walls and an optional ceiling. It didn't provide a way to easily correlate sample locations on the map to sampling locations in the real world.

Starting with VSP version 2.5, the old Draw Room tool and has been replaced by a complete Room Supporting Infrastructure, including sampling locations in 3 dimensions and a Local Coordinate System.

Under this new system, a room in VSP is any Sample Area that has a height greater than zero. The floor, walls and ceiling can be viewed in the **Room View**. Objects like doors and windows can be placed in a room. Samples can be placed on the walls and ceiling of a room as well as on the floor. To display the Room View, select **View > Room** from the main menu or click on the **Room View** tool button on the main toolbar.

Most of the menu commands discussed in this chapter also have tool button shortcuts available on the **Room Toolbar**. To active the Room Toolbar, select **View > Room Toolbar** on the main menu.

Currently, all the VSP sampling designs (except Swath sampling, which doesn't make sense for rooms) allow placement of samples on surfaces within rooms. Future versions of VSP will have additional Sampling Goals that deal with making decisions about rooms and buildings. This functionality will support decisions on when contaminated buildings can be re-occupied following a release event(s) within the building.

6.1 Room Creation and Manipulation

6.1.1 Creating a Room from a Sample Area

Remember that a room in VSP is simply a sample area that has had its height set greater than zero. So the simplest way to create a room is to create a sample area and set its height to something like 8 feet. To set the height of a room, right-click on the sample area on the map to display the Sample Area Information dialog box (see Figure 6.1). The room height is one of the parameters that can be modified in this information dialog box. The **Room > Set Room Height** menu command (or **Set Height** tool button) displays a dialog that sets the height of all selected sample areas when the Map View is active, or sets the height of only the current room when the Room View is active. (See Section 6.2.1 for a definition of current room.)

The Sample Area Information dialog also allows you to include or exclude the floor and the ceiling from a room. The floor and ceiling *are* included by default. If these surfaces are excluded they will not receive sample points from any sampling design and they will not be displayed in the Room view.

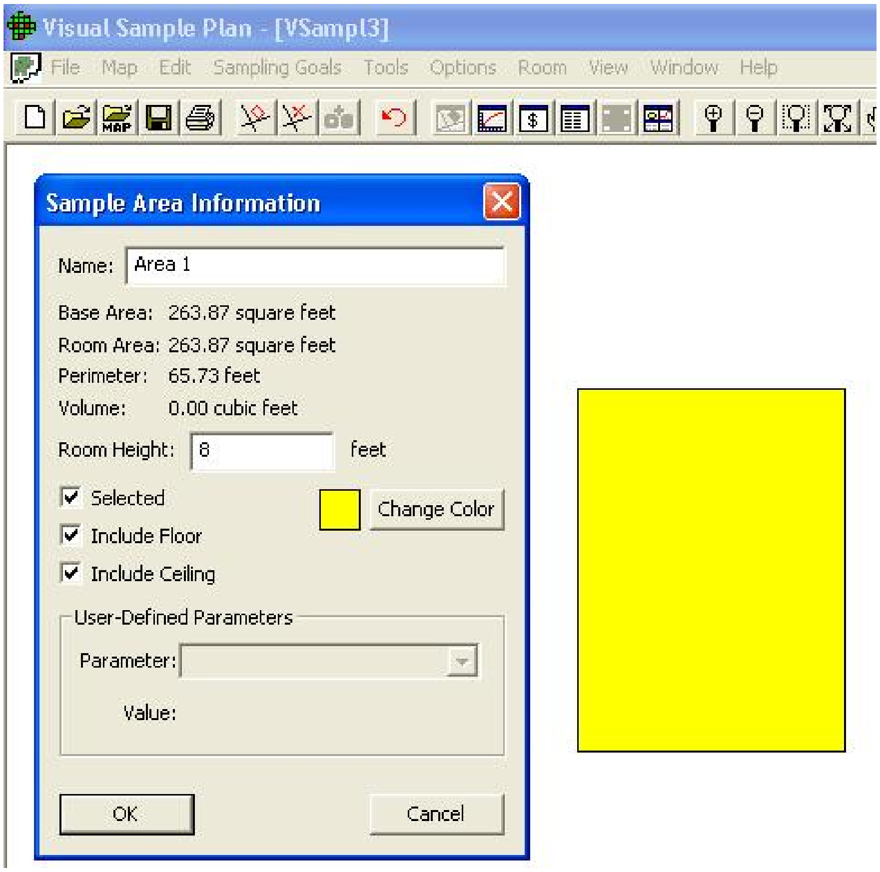


Figure 6.1. Sample Area Information Dialog

6.1.2 Drawing a Room

The **Room > Draw Room** menu command (or the Draw Room tool button) allows you to draw a room on the map. Drawing a room is similar to drawing a rectangle, except that it automatically sets the wall height for you. Define one corner of the floor rectangle by clicking on the map with the left mouse button or by entering the coordinates on the keyboard (e.g., 100,200<**Enter**>). Then enter the opposite corner with the mouse or keyboard. Holding the Ctrl key while clicking on the map limits the floor rectangle to a square.

An alternate method of defining a room is to enter the length, width, and height on the keyboard (Length x Width x Height <**Enter**>). If the room is defined this way, then one corner is placed at the origin (0,0).

6.1.3 Room Delineation Method

If you have a building floor-plan map, the **room delineation** method is usually the fastest and easiest way to create rooms in VSP. Choosing **Room > Delineate Rooms** from the menu (or clicking the Delineate Rooms tool button) toggles the room delineation mode on or off. The room delineation mode allows you to create rooms with right angles inside existing map shapes. This mode is for use in the map view. If a room contains some walls that are not at right angles to other walls, those rooms will need to be created by the first method outlined in Section 6.1.1.

While in this mode, left-clicking with the mouse creates a red rectangle. The red rectangle is framed by moving up, down, left and right from the point clicked and stopping at the first line encountered in that direction. Several red rectangles may be needed to fill up the space inside an irregular shaped room on the map.

After the room is filled with adjoining red rectangles, right-click with the mouse to combine all the rectangles into a room. If there are disjoint red rectangles, each group of red rectangles becomes a separate room. Figure 6.2 illustrates the delineation method in action.

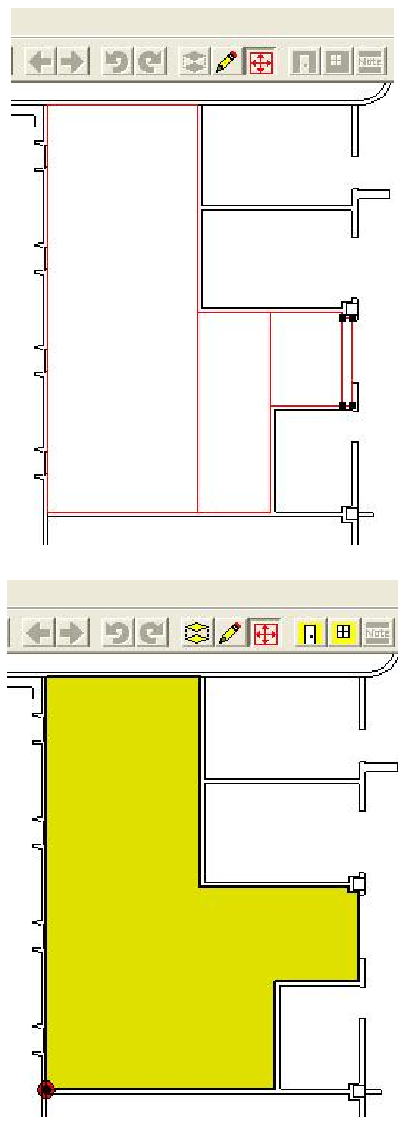


Figure 6.2.   Room Delineation Mode

If the desired room is a simple rectangle, right-click with the mouse to create the rectangle and convert it to a room in a single step.

6.1.4 Room Manipulation

The room can be modified in the map view by inserting a point into a wall and then moving the wall section. To insert a new point into a wall, use the **Map** > **Insert Point** menu command. To move a wall section, select the wall section by holding down the Shift key while clicking on the center of a wall segment. Once selected, the wall segment can be dragged with the mouse. Hold the Ctrl key down while dragging to the wall aligned at right angles. Figure 6.3 illustrates a new point inserted on the right side of a room, the segment selected, and the segment being dragged outward.

You can also set the exact length of a selected wall segment by right-clicking on it. This displays a dialog that allows you to change the length of the segment. Figure 6.4 illustrates a segment length being changed with the dialog.

The On-line Help functions provide more detailed instructions on room manipulation.

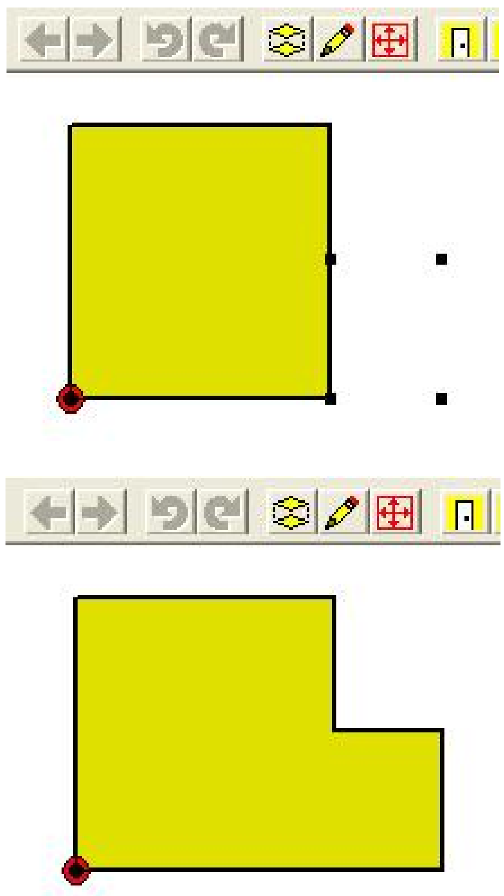


Figure 6.3. Room Manipulation

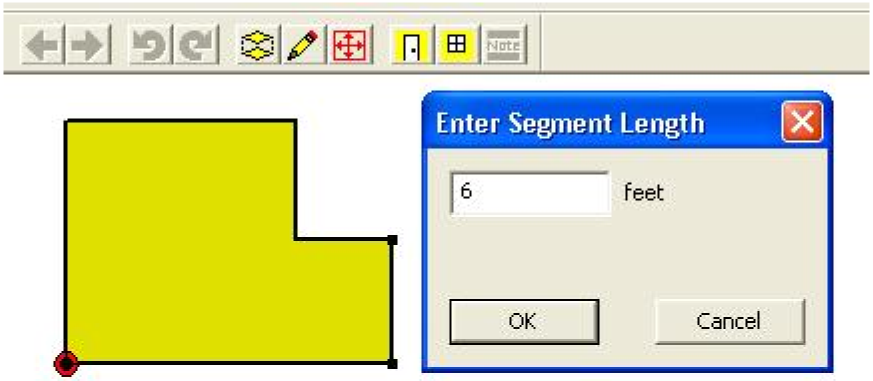


Figure 6.4. Changing Segment Length

6.2 Room Display Options

6.2.1 Current Room

When one or more rooms are selected on a map, exactly one of the rooms is the current room. The current room is the one that is displayed in the Room View. The current room is also indicated on the map by a thick black border and a darker shade. The last room to be created, selected, or viewed with the Sample Information dialog becomes the current room. Using the **Room > Next Room or Room > Previous Room** menu commands scrolls through the list of selected rooms. Each of these commands also has a tool button. Figure 6.5 shows the current room in the Map View.

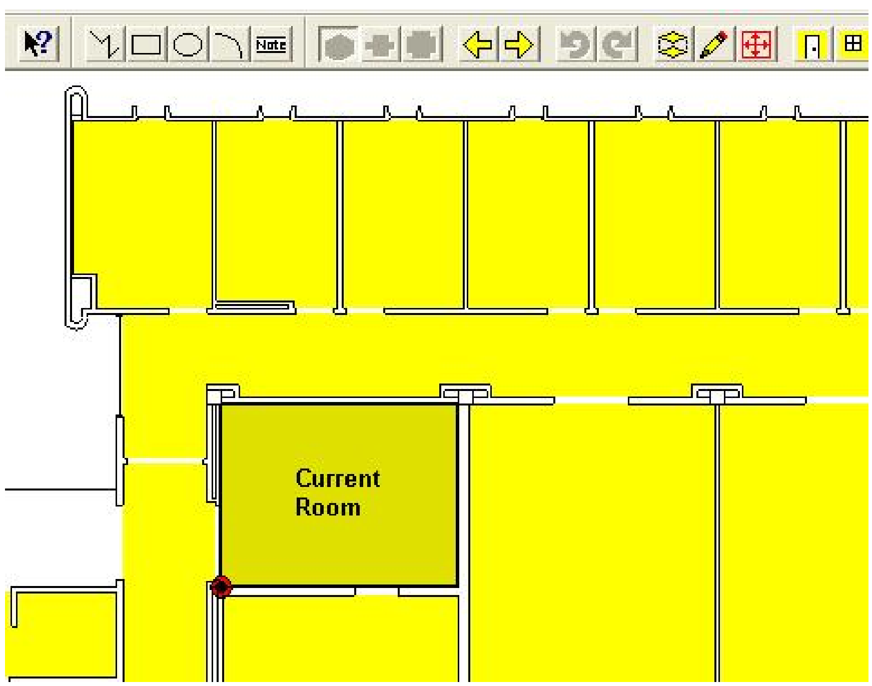


Figure 6.5. Current Room

6.2.2 Room View Types

VSP provides 3 different room view types: perspective, wall-strip and splayed. These View Types can be selected using the menu commands:

Room > **Perspective Room** (or **Perspective Room** tool button)

Room > **Room with Wall Strip** (or **Wall Strip Room** tool button)

Room > **Splayed Room** (or **Splayed Room** tool button)

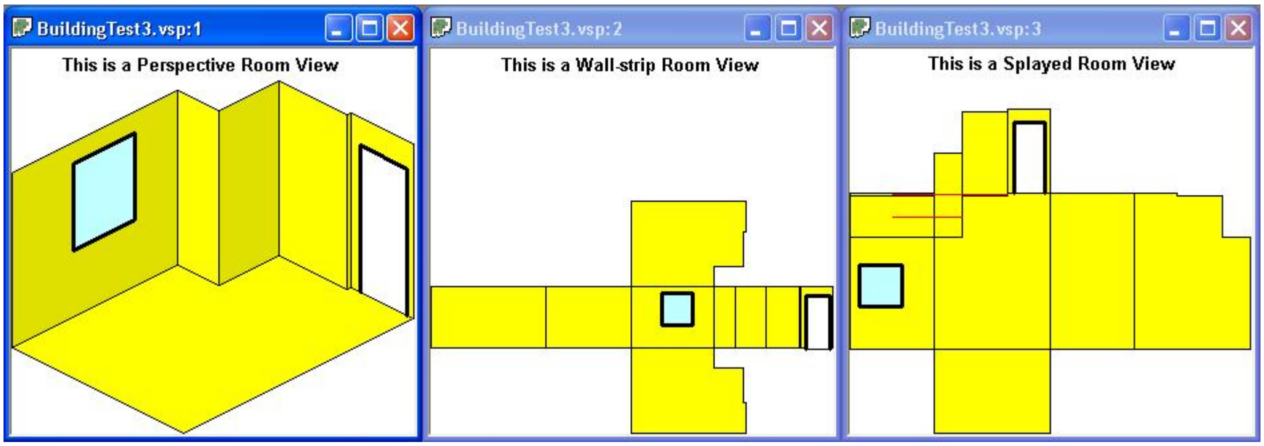


Figure 6.6. Room View Types

6.2.2.1 Perspective Room

The perspective room view shows the room as a 3-dimension semi-perspective view. Inside wall surfaces that are facing you are displayed; inside wall surfaces that are facing away are not displayed. The floor is show tilted down and the ceiling (if included) is shown tilted up. The wall sections are shown attached to the floor surface. (If the floor is excluded and the ceiling is included, then the wall sections are shown attached to the ceiling surface.)

The **Room** > **Rotate Room Left** and **Room** > **Rotate Room Right** menu commands (or the Rotate Left and Rotate Right tool buttons) rotate the room by 90 degrees, to allow viewing from a different perspective. This does not change the coordinates of the room or any object associated with the room, such as sample locations.

6.2.2.2 Wall-Strip Room

The wall strip room view shows a floor (if included), a mirror-image ceiling (if included), and a strip of wall sections lain edge to edge. The floor and ceiling surfaces are shown attached to the proper wall section.

The **Room > Rotate Room Left** and **Room > Rotate Room Right** menu commands (or the Rotate Left and Rotate Right tool buttons) rotate the room by 90 degrees, to allow viewing the floor and ceiling from different angles. This does not change the coordinates of the room or any object associated with the room, such as sample locations.

6.2.2.3 Splayed Room

The splayed room view shows a floor, a mirror-image ceiling, and individual wall sections lain outward from the floor. The wall sections are originally laid down so that they do not overlap. If the wall section is not shown attached to the floor, a red line shows where it should be attached.

To move a wall section to a new location on the display, simply click and drag with the mouse. Right-click with the mouse on a wall section to move it back to its attached location. Moving a wall section on the display does not change the coordinates of the room or any object associated with the room, such as sample locations.

6.2.3 Room North Arrow

By default, VSP places a north arrow on the room floor and ceiling surface. Figure 6.7 shows the north arrows on the floor and ceiling. Use the **View > Room North Arrow** command to show or hide the north arrow. The location of the north arrow can be changed by grabbing the center of the north arrow with the mouse and dragging it. (The center to grab is between the N and the arrow.)

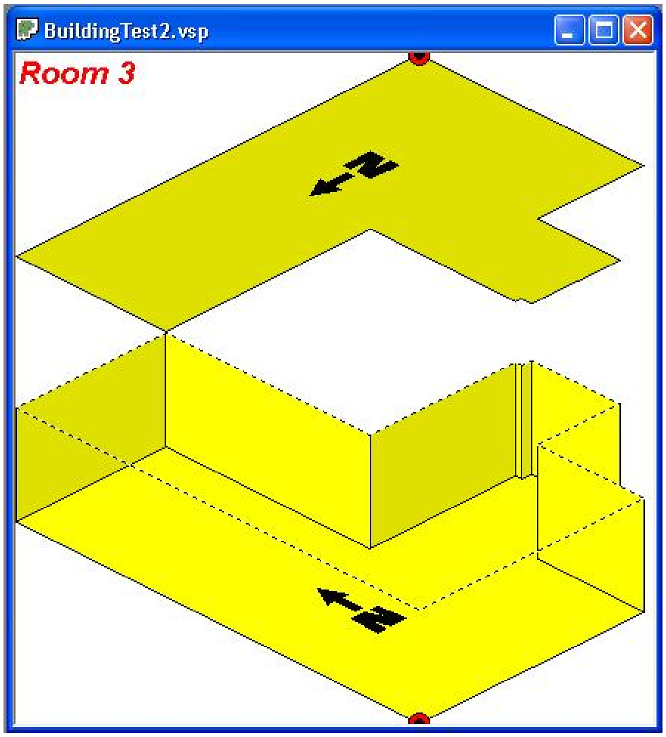


Figure 6.7. Room North Arrows

VSP assumes by default that north is the top of your map. If this is not the case, select the **Map** > **Set Map Extents** menu command. Change the North Offset in the Map Extents dialog box to correct the direction of north on your map.

6.2.4 Perspective Ceiling

If you prefer not to see the ceiling in the perspective view, use the **View > Room Perspective Ceiling** command to hide the ceiling.

6.3 Room Objects

Three different types of objects can be inserted into a room: Doors, Windows and Annotations. Figure 6.8 shows a room with all three objects.

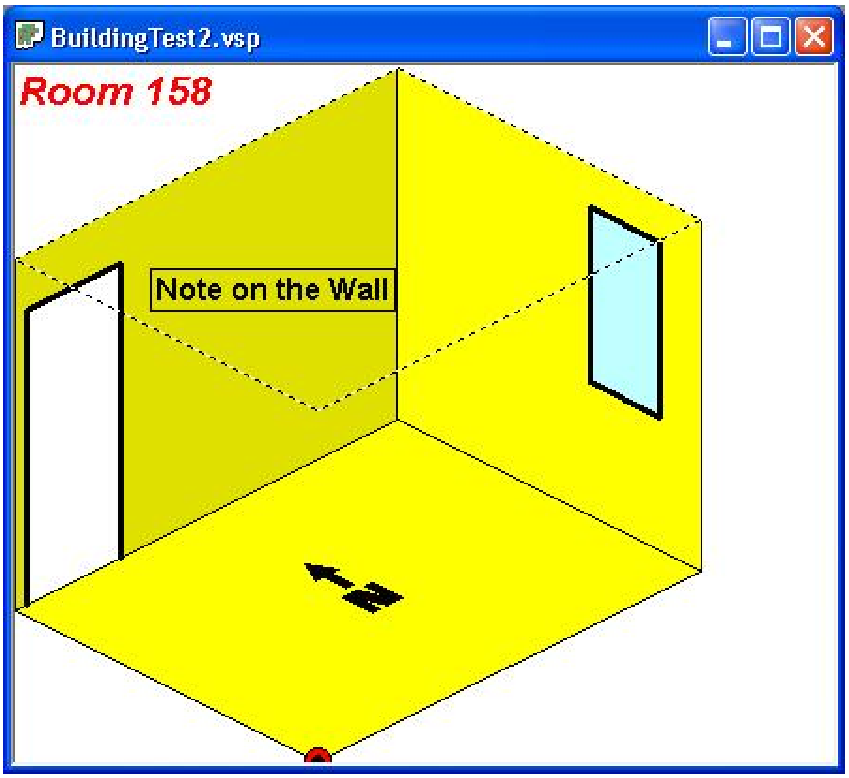


Figure 6.8. Room Objects

6.3.1 Doors

To insert a door object into the current room, use the **Room > Insert Door** menu command (or **Insert Door** tool button). This command can be used in the Map View or the Room View.

After selecting this command, click on one edge of the door and then on the other edge of the door. On the Map View, the coordinates can also be entered on the keyboard. Hold the shift key while clicking to attach the door to the nearest point on the map. Both edges of the window must be on the same wall section. Note that doors are inserted into the current room (see Section 6.2.1).

Doors can be moved, resized or deleted by using the **Object Information** dialog which is accessed by right clicking on the door. This dialog also allows you to change the door sub-type: Open Door (sample locations will be excluded from the door surface); or Fixed Door (sample locations can be included on the door surface). Figure 6.9 shows the Object Information dialog for a door.

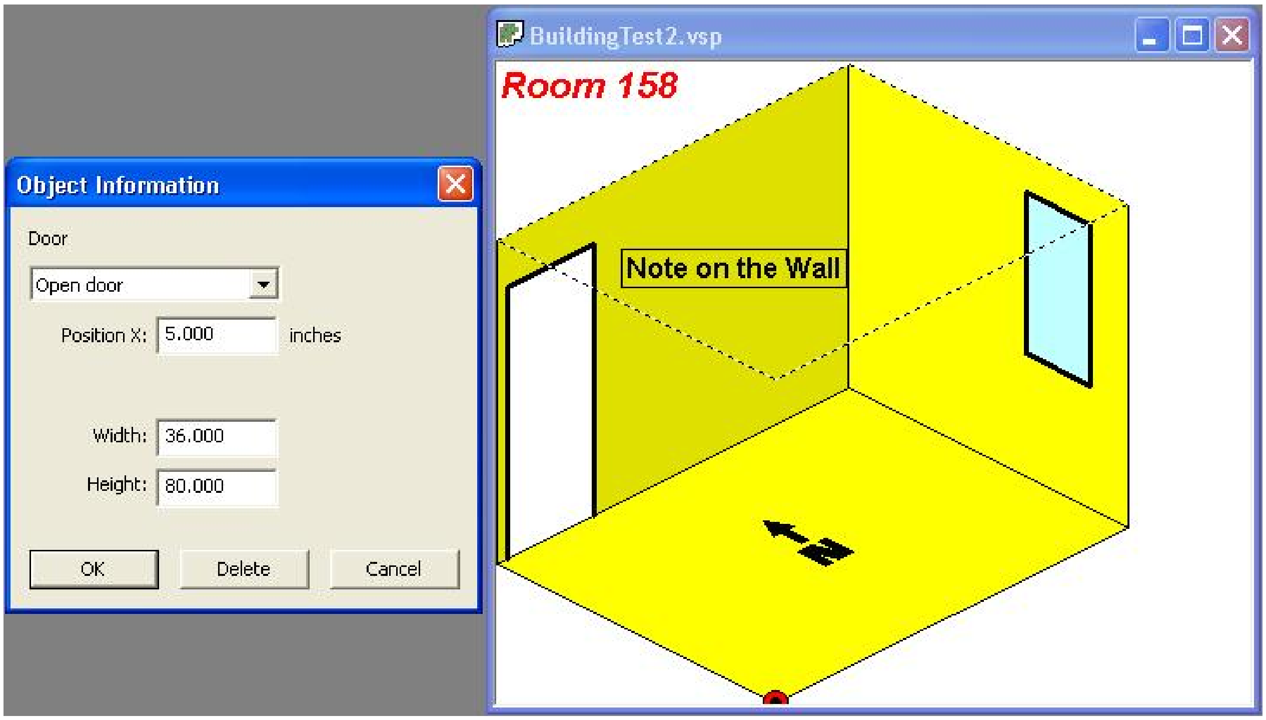


Figure 6.9.   Object Information Dialog Box

6.3.2 Windows

To insert a window object into the current room, use the **Room > Insert Window** menu command (or **Insert Window** tool button). This command can be used in the Map View or the Room View.

To use the command in the Map View, click on one edge of the window and then click on the other edge of the window. The window will be placed and sized with default heights. The edge coordinates can also be entered on the keyboard. Hold the shift key while clicking to attach the window to the nearest point on the map. Both edges of the window must be on the same wall section. Note that doors are inserted into the current room (see Section 6.2.1).

To use this command in the Room View, click on one corner of the window and then click on the opposite corner of the window.

Windows can be moved, resized, or deleted by using the **Room Object Information** dialog which is accessed by right clicking on the door. This dialog also allows you to change the door sub-type: Open Window (sample locations will be excluded from the window surface); or Fixed Window (sample locations can be included on the window surface).

6.3.3 Notes

To insert a note object into the current room, use the **Room > Insert Annotation** menu command (or **Insert Note** tool button). This command can be used only in the Room View.

Click on the location where you want to place a Note. The text, font, color and other attributes can be modified from the **Annotation Object Dialog Box** which is access by right clicking on the Note. Notes can be moved by selecting them (by clicking on them) and then dragging them with the mouse.

6.4 Other Room Features

6.4.1 Surface Labels

To automatically create a label (Annotation or Note Object) on each surface of a room, use the **Room > Label Surfaces** command. This command creates a label for each surface in the current room when used in the Room View, or for all selected rooms when used in the Map View. The default labels are:

Floor

Ceiling

Wall 1

Wall 2

Wall…

This command allows you to choose the Font Type, Size, Style and Color as well as other attributes associated with labels. Figure 6.10 shows room surface labels.

After being created, the surface labels may be edited by right-clicking on them. If this command is run again, it will delete any existing surface labels and replace them with the default surface labels. The surface label is also reported on the Sample Information dialog (accessed by right-clicking on a sample location). Figure 6.11 shows a Sample Information dialog with a custom surface label.

6.4.2 Local Coordinates and Room Origin

VSP uses local coordinates to facilitate the location of samples inside of rooms (and also regular sample areas). Local (LX, LY) are relative coordinates. Local coordinates can be viewed on the Sample Information dialog in Figure 6.11.

In regular Sample Areas, the local coordinates are relative to the minimum X and minimum Y coordinates of the sample area.

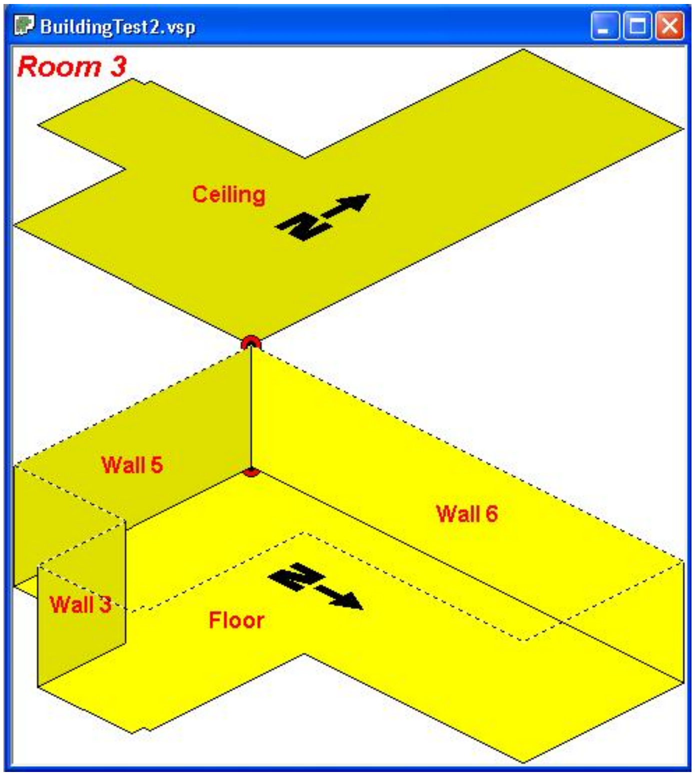


Figure 6.10. Room Surface Labels

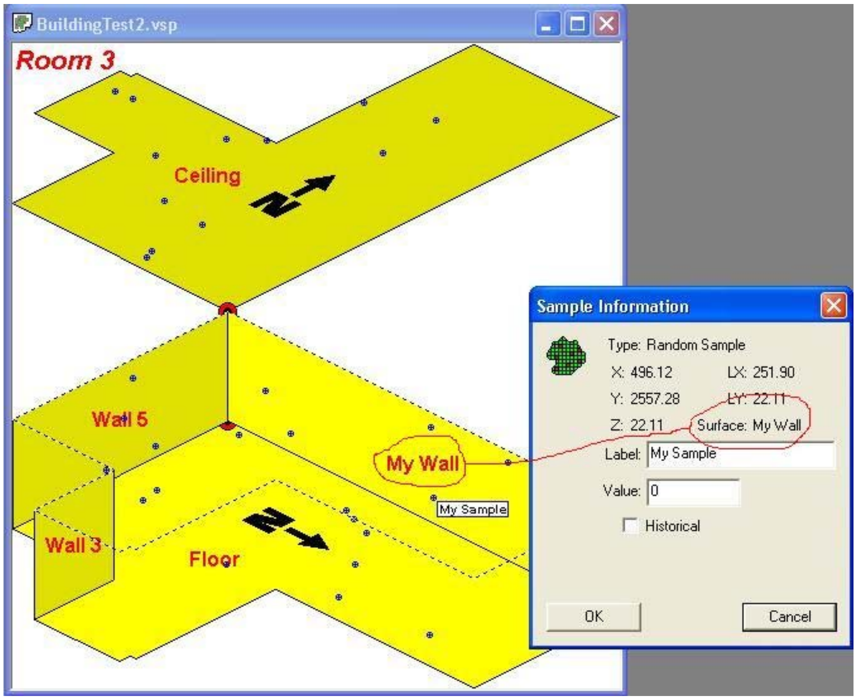


Figure 6.11. Room Surface Labels

In Rooms, the local coordinates are relative to the surface they are on. For walls, the local coordinates are relative to the lower left hand corner of the wall section as you are facing it. For the floor and ceiling, the local coordinates are relative to the **Room Origin**. The location of Room Origin is shown on the Room View by a red and black bull's-eye.

By default the Room Origin is the minimum X and minimum Y, but can be changed with the **Room > Set Room Origin** menu command. Select the new location for the Room Origin by clicking on the Map. Hold down the Shift key while clicking to attach the Room Origin to an existing point on the map. The Room Origin can also be set by entering the coordinates on the keyboard.

6.4.3 Room Label

VSP provides a fixed room label at the upper left corner of each Room View. Although this label cannot be hidden, the font, color and other label attributes can be modified by left clicking on the label. The same color and styles are used for each Room View.

7.0 References

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

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

Gilbert, R. O, J. E. Wilson, R. F. O'Brien, D. K. Carlson, D. J. Bates, B. A. Pulsipher, and C. A. McKinstry. 2002. *Version 2.0 Visual Sample Plan (VSP) Models and Code Verification.* PNNL-13991, Pacific Northwest National Laboratory, Richland, Washington.

Gilbert, R. O, J. E. Wilson, R. F. O'Brien, D. K. Carlson, B. A. Pulsipher, and

D. J. Bates. 2003. *Version 2.0 Visual Sample Plan (VSP): UXO Module Code Description and Verification.* . PNNL-14267, Pacific Northwest National Laboratory, Richland, Washington.

Gilbert, R. O. 2003. *Results of QA Testing of the Visual Sample Plan (VSP) Collaborative Sampling (CS) Module* PNNL Milestone Report to EPA Quality Staff, Pacific Northwest National Laboratory, Richland, Washington

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* - *QA/G-9, QA00 Update.* EPA/600/R/96/084, Office of Environmental Information, U.S. Environmental Protection Agency, Washington, D.C.

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

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1993. *Numerical Recipes in C: The Art of Scientific Computing.* 2nd ed. Cambridge University Press, United Kingdom.

7.0 References

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

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

Gilbert, R. O, J. E. Wilson, R. F. O'Brien, D. K. Carlson, D. J. Bates, B. A. Pulsipher, and C. A. McKinstry. 2002. *Version 2.0 Visual Sample Plan (VSP) Models and Code Verification.* PNNL-13991, Pacific Northwest National Laboratory, Richland, Washington.

Gilbert, R. O, J. E. Wilson, R. F. O'Brien, D. K. Carlson, B. A. Pulsipher, and

D. J. Bates. 2003. *Version 2.0 Visual Sample Plan (VSP): UXO Module Code Description and Verification.* . PNNL-14267, Pacific Northwest National Laboratory, Richland, Washington.

Gilbert, R. O. 2003. *Results of QA Testing of the Visual Sample Plan (VSP) Collaborative Sampling (CS) Module* PNNL Milestone Report to EPA Quality Staff, Pacific Northwest National Laboratory, Richland, Washington

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* - *QA/G-9, QA00 Update.* EPA/600/R/96/084, Office of Environmental Information, U.S. Environmental Protection Agency, Washington, D.C.

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

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1993. *Numerical Recipes in C: The Art of Scientific Computing.* 2nd ed. Cambridge University Press, United Kingdom.

Distribution

No. of
Copies

ONSITE

4	**Pacific Northwest National Laboratory**	
	R. O. Gilbert	
	N. L. Hassig	K5-12
	B. A. Pulsipher	K5-12
	J. E. Wilson	K5-12