PNNL-13490

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

Interim Report

Visual Sample Plan Version 1.0 User's Guide

J. R. Davidson, Jr.[a] J. E. Wilson[b]
N. L. Hassig R. O. Gilbert

March 2001

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

(a) Davidson and Davidson, Inc., Kennewick Washington
(b) AIMTECH, Grand Junction, Colorado

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

This document was printed on recycled paper.
(8/00)

PNNL-13490

Interim Report

Visual Sample Plan
Version 1.0
User's Guide

J. R. Davidson, Jr.[(a)]
N. L. Hassig
J. E. Wilson[(b)]
R. O. Gilbert

March 2001

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

Pacific Northwest National Laboratory
Richland, Washington 99352

(a) Davidson and Davidson, Inc., Kennewick, Washington
(b) AIMTECH, Grand Junction, Colorado

Abstract

This user's guide describes Visual Sample Plan (VSP) Version 1.0 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 1.0 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, Millenium Edition, 2000, and Windows NT). 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.

Contents

Figures

Tables

Abbreviations

DCGL	Derived Concentration Guideline Limit
DPC	Decision Performance Curve
DPGD	Decision Performance Goal Diagram
DQO	Data Quality Objectives
DXF	Drawing Exchange Format
GIS	geographical information system
GPS	global positioning system
LBGR	lower bound of the gray region
MARSSIM	Multi-Agency Radiation Survey and Site Investigation Manual
MQO	Measurement Quality Objectives
VSP	Visual Sample Plan
WRS	Wilcoxon Rank Sum

1.0 Introduction

1.1 What is Visual Sample Plan (VSP)?

VSP is a software tool being developed to select the right number and location of environmental samples so that the results of statistical tests performed to provide input to environmental decisions have the required confidence and performance. Version 1.0 of VSP provides sample-size equations or algorithms needed by specific statistical tests appropriate for the following three environmental sampling objectives:

1. Comparing representative measurements from a defined geographical area to a fixed threshold (upper limit) value.

2. Comparing representative measurements from one defined geographical area with representative measurements from another geographical area.

3. Looking for areas of elevated concentration ("hot spots") at a study site.

VSP is easy to use, highly visual, and graphic. It is a significant help in implementing the sixth and seventh steps of the Data Quality Objectives (DQO) planning process ("Specify Tolerable Limits on Decision Errors" and "Optimize the Design for Obtaining Data," respectively). In particular, the VSP user is asked to specify decision performance requirements by constructing the Decision Performance Goal Diagram (DPGD) discussed in EPA (2000a). 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.

1.2 System Requirements and Installation

VSP was designed originally for the Microsft® Windows® 95 operating system but has been used successfully on several closely related operating systems. *Visual Sample Plan (VSP) Models and Code Verification* (Gilbert et al. 2001) documents the successful installation of VSP on personal computers operating with Microsoft Windows 95, 98, Millennium Edition, and Windows 2000. VSP should run successfully also on Windows NT® systems.

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
- 32 MB RAM
- 6 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, VSP1.0.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".

1.3 Removing VSP and Expiration Date for Demonstration Versions

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**.

VSP has been released many times as a prototype version for testing. These demonstration (or beta) versions all 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 1.0 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 at the least cost. The VSP computer program will assist the environmental professional in developing an environmental-sampling design.

For example, VSP can be used to assist in the development of sampling plans as part of the DQO process. VSP can automate some of the mechanical details of calculating sample size, specifying random sampling locations, and comparing sample costs with decision error rates. These activities all can be accomplished in the context of your own site map displayed on screen 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, using simple techniques, a visual map of the study site to be sampled. The study site may be only a portion of a complex facility, the map of which has been imported into VSP (see Figure 1.1, upper left window).

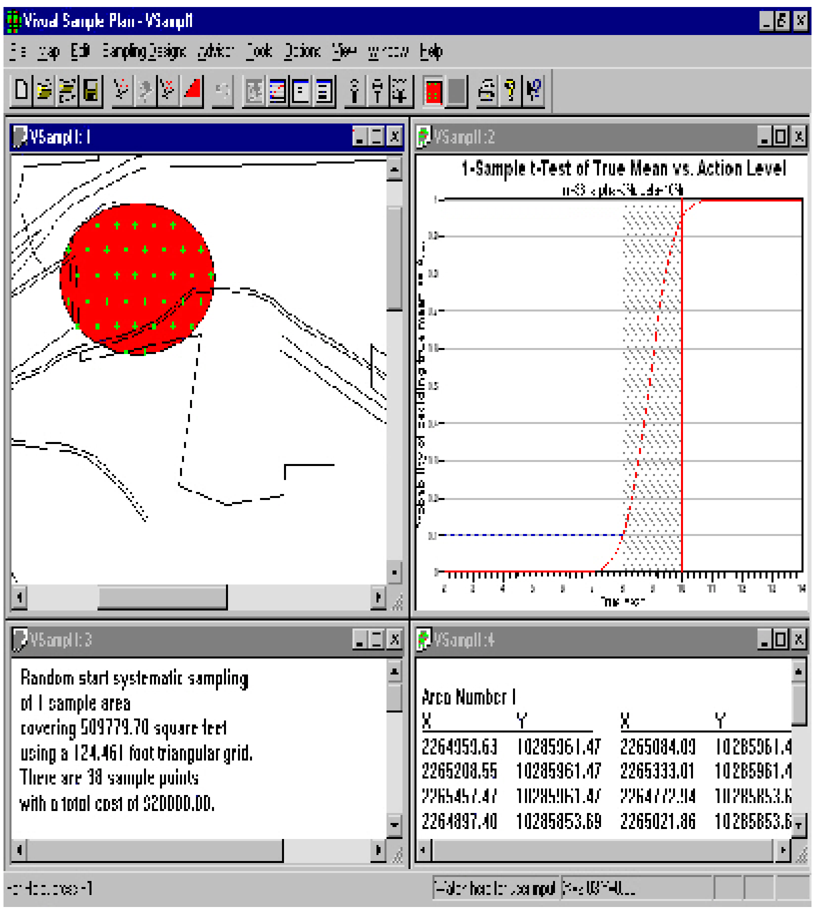

Figure 1.1. Screen Capture from VSP Using **Quad Window** Option

Then, for the sampling objective that you specify, 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). Moreover, VSP contains a Measurement Quality Objectives (MQO) module that enables you to assess the cost and benefits of making one, two, or three measurements per subsample (aliquot) randomly withdrawn from each field sample.

The locations of the samples over the study site are determined by the specific sampling design (pattern) that you select. Version 1.0 permits you to select simple random sampling, systematic sampling (using a square, rectangular or triangular grid), and judgment sampling. VSP provides two types of random number generators for randomly selecting a sampling location for the simple random sampling option: 1) a pseudo-random number generator for which each potential location has an equal and independent chance of being selected and 2) a quasi-random number generator (Press et al. 1993) for which locations are chosen to be somewhat more evenly spaced than would be obtained using the pseudo-random number generator. Version 1.0 of VSP also allows you to add sample points to a current design either manually (subjectively) or using an automatic algorithm called Adaptive-Fill. The Adaptive-Fill algorithm adds data points in such a way as to avoid existing sampling locations.

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) file for use in a geographical information system (GIS) or saved as a text file for use in global positioning system (GPS) software.

VSP also displays the user-specified DPGD for determining the number of samples (see Figure 1.1, upper right corner). The DPGD graphically shows the user-supplied acceptable probabilities of making decision errors based on the data obtained using the sampling design and number of samples. These probabilities are used by VSP to compute the number of samples needed. VSP also documents the selected sampling pattern (random, systematic or judgment), the minimum number of samples computed by VSP, and the total cost of sampling and measurements, as well as any relevant statements on the probability that the selected sampling design will detect “hot spots” (see Figure 1.1, lower left corner).

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

Understanding how to use VSP as one tool in the multifaceted DQO process is a key to providing a defensible sampling plan.

1.5.1 What Is the DQO Process?

The DQO process is a seven-step planning approach to develop sampling designs for data collection activities that support decision making. This process uses systematic planning and statistical hypothesis testing to differentiate between two or more clearly defined alternatives. See *Guidance for the Data Quality Objectives Process, EPA QA/G-4* (EPA 2000a) for an extended discussion of the DQO process. A summary of the seven steps is presented in Table 1.1.

Table 1.1. The Seven-Step Data Quality Objectives Process

Step1. State the Problem
Step 2. Identify the Decision
Step 3. Identify the Inputs to the Decision
Step 4. Define the Boundaries of the Study
Step 5. Develop a Decision Rule
Step 6. Specify Tolerable Limits on Decision Errors
Step 7. Optimize the Design for Obtaining Data

The final outcome of the DQO process is a design for collecting data (e.g., the number of samples to collect, and where, when, and how to collect the samples), together with the limits on the probabilities of making decision errors.

1.5.2 How Does VSP Fit into the DQO Process?

VSP is a sampling design tool that can be used to facilitate Steps 6 and 7 of the DQO process. To produce an appropriate sampling design with VSP, users must successfully complete Steps 1 through 6 of the DQO process. This is because some of the inputs for VSP (the limits on the probabilities of making two types of decision errors, the width of the gray region, estimates of sampling, and analytical variability) come from those first six steps. Step 7 of the DQO process, Optimize the Design for Obtaining Data, is where VSP can be used to try out different sampling designs and find the optimal design for the current problem.

1.6 What Are the Current Limitations of VSP?

VSP Version 1.0 does not provide

- sequential sampling designs
- simulation-based sampling designs
- sampling designs for statistical tolerance intervals

- sampling designs for estimating standard deviations or variances

- sampling designs using geostatistics-based methods.

We hope that as VSP is developed, you, the user, can provide us, the developers, with input on what new capabilities or sampling designs would be most useful in solving your sampling-design problems.

The experienced user can find innovative ways of expanding the capabilities in VSP to encompass more complex problems and more complex sampling designs than described in this user's guide.

2.0 Mechanics of Running VSP

2.1 Setting Up a Map

VSP is designed not only to provide a recommended minimum sample size but to display a set of random sampling locations directly on your selected Study Area. You may obtain a map (drawing) of the Study Area and the associated site in any of three ways:

1. Import the site map from a DXF file.

2. Import the site map from a previous VSP project saved in a VSP-format file.

3. Draw the Study Area using VSP's drawing tools.

We next illustrate these three methods.

2.1.1 Importing a Study Area or Site Map from a DXF File

You can draw a complex site map in an architectural drawing program such as Autodesk® AutoCAD® and save the drawing to a DXF file. The resulting DXF file can be imported into VSP, and a Study Area can be selected and a sampling design developed.

The Millsite.DXF file is an example DXF provided with VSP. The following steps illustrate how to use this file in VSP:

1. Start VSP using **Start ÷ Programs ÷ Visual Sample Plan ÷ Visual Sample Plan**. If VSP is already running, start a new project using Main Menu option **File ÷ New Project**.

2. Load the Millsite.DXF file using Main Menu option **Map ÷ Load DXF**. A quick alternative is to click on the **Load Map** button on the VSP toolbar.

3. The site map should appear on your screen as illustrated in Figure 2.1.

It is possible that your DXF file might contain information that the current version of VSP cannot read. If you have a file that cannot be imported successfully into VSP, please contact John Wilson at the e-mail address or phone number listed in the **Help ÷ About VSP** box.

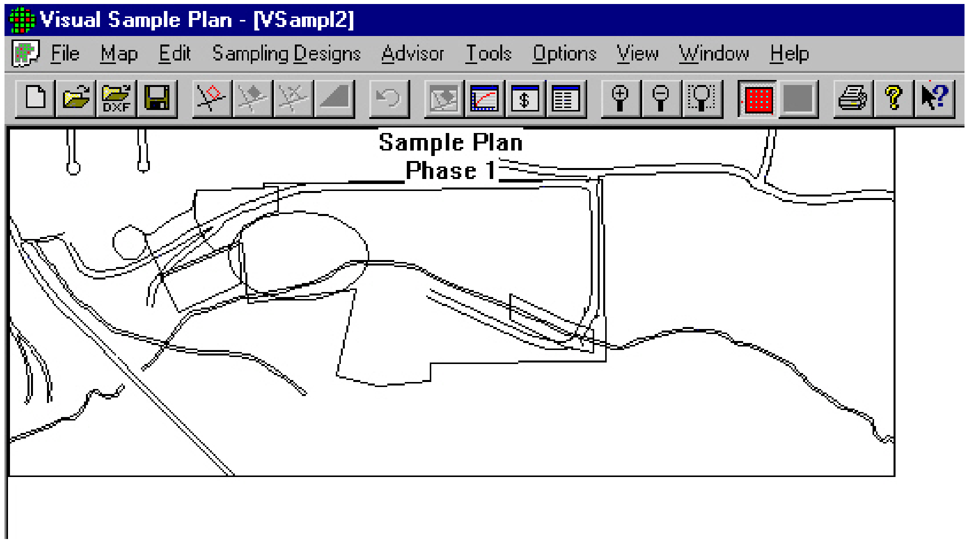

Figure 2.1. The Millsite.DXF File Opened in VSP

2.1.2 Exporting and Importing a Study Area or Site Map File in the VSP Format

No matter how you imported or created a site map or Study Area for VSP, you can always save the information in VSP's own file format. For example, to save the DXF file you imported into VSP (as described in Section 2.1.1),

1. Use Main Menu option **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.

2. To open a VSP-format file, use Main Menu option **File ÷ Open Project** or use the **Open** button on the VSP toolbar.

2.1.3 Drawing a Study Area or Site Map Using VSP's 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:

1. Create a new project by choosing **File ÷ New Project** on the Main Menu or by clicking the **New** button on the toolbar. Expand the project window by pressing the **Maximize** button on the upper right corner of the project window.

2. 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.

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.

Draw MARSSIM Room. Click the **Draw Room** button on the drawing toolbar or select Main Menu option **Map ÷ Draw MARSSIM Room**. A yellow tooltip displays the three ways to draw a room using this tool:

- Enter the room's dimensions from the keyboard.

- Enter the room's corner coordinates from the keyboard.

- Use the mouse to establish the corner points and the wall height.

More detail on using these methods is provided below.

1. Enter the room length, width, and height dimensions and press the Enter key. Separate the dimensions with the letter "x." For example, type **12x10x8** and press the Enter key. A dialog box will ask if you want to include the ceiling. Figure 2.2 is an example with the ceiling chosen. The status bar at the bottom of the VSP screen displays the dimensions as you enter them.

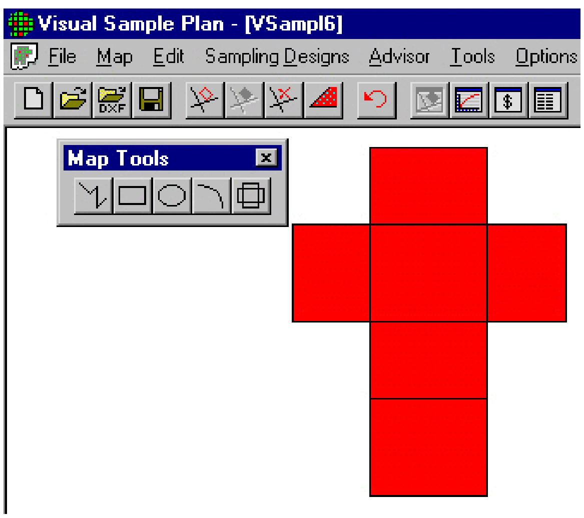

Figure 2.2. Example of MARSSIM Room with Ceiling Option Selected

2. Enter the corner coordinates of the room on the keyboard and the wall height. For example, start a new project using Main Menu option **File ÷ New Project.** Select Main Menu option **Map ÷ Draw MARSSIM Room** or use the **Draw Room** button on the drawing toolbar. Type **50,50** and press the Enter key. For the opposite corner coordinate type **90,90** and press Enter. Type 9 for the wall height and press Enter. Answer **No** to the "Include Ceiling?" question. Your screen should be similar to that shown in Figure 2.3.

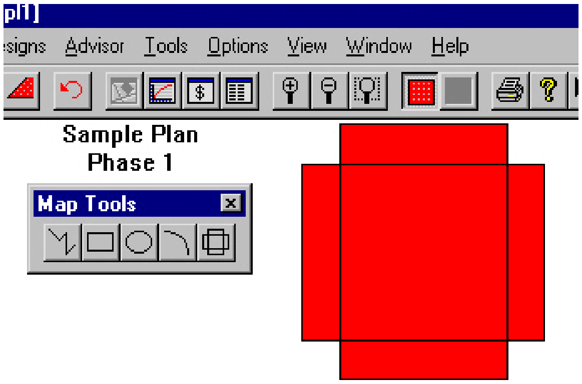

Figure 2.3. Example of MARSSIM Room Without Ceiling Option Selected

3. The third method is to click the mouse for the first corner, then the second corner, and finally click the third time after dragging the cursor out a distance to indicate the wall height.

2.2 Selecting a Study Area and Zooming In on a Study Area

2.2.1 Selecting a Study Area

A Study Area, i.e., an enclosed region in which to place sampling locations, must be defined in order to make sampling locations available. (You can use any of the sampling designs without a Study Area, but they will not create sampling locations.) Press the **New Area** button on the VSP toolbar (or choose Main Menu option **Edit ÷ Sample Areas ÷ Define New Sample Area**). A **Color** dialog box appears. Use this dialog to choose the color of the Study Area. After the color is selected, a tooltip box appears on the map to provide information on the selection method. There are two basic ways in which to select the Study Area:

1. **One-Step Method**. Position the cursor inside one of the enclosed areas on the map and right-click with the mouse. The Study Area is created, and a dialog box appears. This dialog box shows the size of the Study 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 Study 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 Study Area, click the **Finish Area** button on the VSP toolbar or select Main Menu option **Edit ÷ Sample Areas ÷ Finish New Sample Area**. The area dialog box appears, allowing you to change the map units.

Note: A Study Area cannot cross over itself. If this happens, an error message appears, stating "The area is invalid and will be removed."

2.2.2 Zooming In, Out, or Windowing a Study Area

The **Zoom In**, **Zoom Out**, and **Zoom Window** buttons in the middle of the VSP toolbar provide methods to focus in on a Study 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 a 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.

2.3 Importing, Exporting, and Removing Sampling Locations

2.3.1 Importing Sampling Locations

There are two ways to import sampling locations:

1. The first way to import sampling locations is to 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 Study 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 2.4.

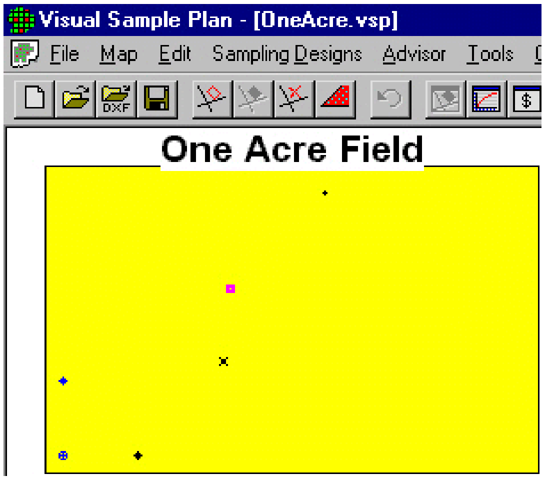

Figure 2.4. The OneAcre.VSP Project with Sampling Locations Added from Windows Clipboard

You can place the mouse cursor on one of the sample points and right-click to see information about that sampling location. See Figure 2.5 for an example.

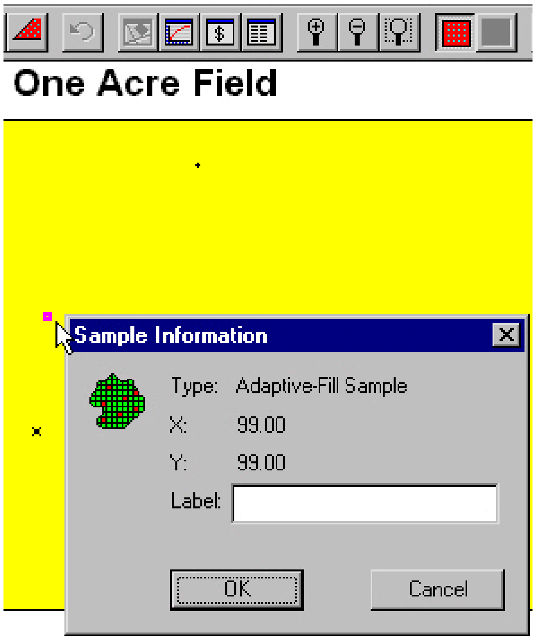

Figure 2.5. Example of Sample Information Box

2. A convenient way to import sampling locations is 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.

2.3.2 Exporting Sampling Locations

To export sampling locations to a text file (for example, to use the coordinates in a GPS),

1. Select the Study 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**.

2.3.3 Removing Sampling Locations

This option is best explained with an example:

1. Start VSP and or open a new project using Main Menu option **File ÷ New Project**.

2. Open the Millsite.DXF file using Main Menu option **Map ÷ Load DXF**.

3. Click on the **New Area** button and, after choosing a color, select the large ellipse by right-clicking inside the oval. If you accidently 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.

4. Choose the Main Menu option **Sampling Designs ÷ Nonparametric ÷ True Mean or Median vs. Action Level**. Click the **Apply** button to place samples in the Study Area. You should now have a Study Area similar to Figure 2.6 with 24 sampling locations.

5. Using 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 ten or twelve rows (sampling locations) from file Points.Txt. Save the remaining rows to a new file named Remove.Txt. These are the locations that will be removed from the Study Area.

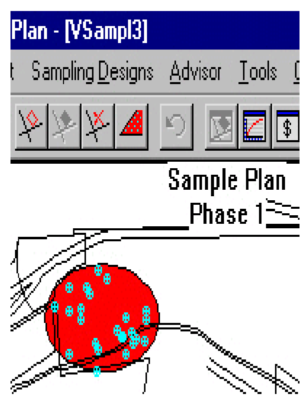

Figure 2.6. Example Study Area with Sampling Locations

7. Finally, to remove the sampling locations listed in Remove.Txt from the Study Area, choose Main Menu option **Map ÷ Sample Points ÷ Remove**. Select the file Remove.Txt to have its locations removed from the current Study Area. For example, in Figure 2.7, there are now only 11 sample points instead of the original 24 shown in Figure 2.6.

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

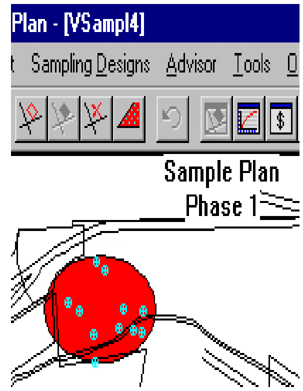

Figure 2.7. Example Study Area after Sampling Locations Have Been Removed

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 type of samples (surface soil, air, etc.), and how to take samples and analyze them. We identified the general areas of where to take samples in Section 2.2, **Set Up Sampling Areas**. In this section, we discuss where *within the Sampling Area* to locate the samples. We also discuss *how many* samples to take. The type of samples to take, i.e., soil vs. groundwater, wet vs. dry, in situ vs. send off to a lab, is determined during Step 3 of the DQO process (Define Inputs) and is not addressed directly in VSP. The MQO module in VSP (See Section 5.4) discusses analytical method type.

3.1.1 Defining the Purpose of Sampling

As stated in Section 1.3.2, 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 that the current data collection effort will be used to resolve. The following types of problems are addressed currently in VSP. Future versions will expand on this list:

1. Compare a population parameter (such as the mean, the median, or a proportion) to a threshold. This is called a one-sample problem in statistics terminology.

2. Compare the population parameters (such as the mean, the median, or a proportion) of two populations to each other. 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.

3. Estimate the mean of a population, providing a confidence interval.

4. Identify hot spots, i.e., small pockets of contamination.

5. Choose sampling locations based on expert judgment or determine the number of sampling locations by a method not currently available in VSP.

The major sampling designs currently provided by VSP are summarized in Table 3.1. The statistical test or method appropriate for each option is listed.

Table 3.1. Statistical Tests and Methods Available in VSP

Problem Type	Assumptions	How to Pick Locations	Option in VSP Sampling Designs Menu	Statistical Test or Method Used
One-Sample Hypothesis Test	Parametric	SRS[a]	True Mean vs. Action Level	One-Sample t-Test
		Grid	True Mean vs. Action Level	One-Sample t-Test
	Nonparametric	SRS	True Mean or Median vs. Action Level	Wilcoxon Signed Rank Test
		Grid	True Mean or Median vs. Action Level	Wilcoxon Signed Rank Test
		SRS	Proportion vs. Given Proportion	One-Sample Proportion Test
		Grid	Proportion vs. Given Proportion	One-Sample Proportion Test
		SRS	MARSSIM Sign Test	MARSSIM Sign Test
		Grid	MARSSIM Sign Test	MARSSIM Sign Test
Two-Sample Hypothesis Test	Parametric	SRS	True Mean vs. Reference Area True Mean	Two-Sample t-Test
		Grid	True Mean vs. Reference Area True Mean	Two-Sample t-Test
	Nonparametric	SRS	Comparison of Two Populations	Wilcoxon Rank Sum Test
		Grid	Comparison of Two Populations	Wilcoxon Rank Sum Test
		SRS	Comparison of Two Proportions	Two-Sample Proportion Test
		Grid	Comparison of Two Proportions	Two-Sample Proportion Test
		SRS	MARSSIM WRS Test	MARSSIM WRS Test
		Grid	MARSSIM WRS Test	MARSSIM WRS Test
Confidence Interval on a Mean	Parametric	SRS	Confidence Interval on True Mean	One-Sided or Two-Sided C.I.
		Grid	Confidence Interval on True Mean	One-Sided or Two-Sided C.I.
Hot-Spot Sampling	See Section 3.2.3	Grid	by Probability and Hot Spot Size	ELIPGRID-PC Algorithm
		Grid	by Probabability and Grid Size	ELIPGRID-PC Algorithm
		Grid	by Cost and Hot Spot Size	ELIPGRID-PC Algorithm
		Grid	by Predetermined Grid Size	ELIPGRID-PC Algorithm
Number of Samples Determined by User	User-Supplied	SRS	Predetermined	User-Supplied Test/Number of Samples
		Grid	Predetermined	User-Supplied Test/Number of Samples
		Manual	Judgment Sampling ÷ Manually Add Samples	Not Applicable

(a) SRS = simple random sampling.

The screens shown in Figures 3.1 through 3.3 are provided to help you relate Table 3.1 to the VSP menu options.

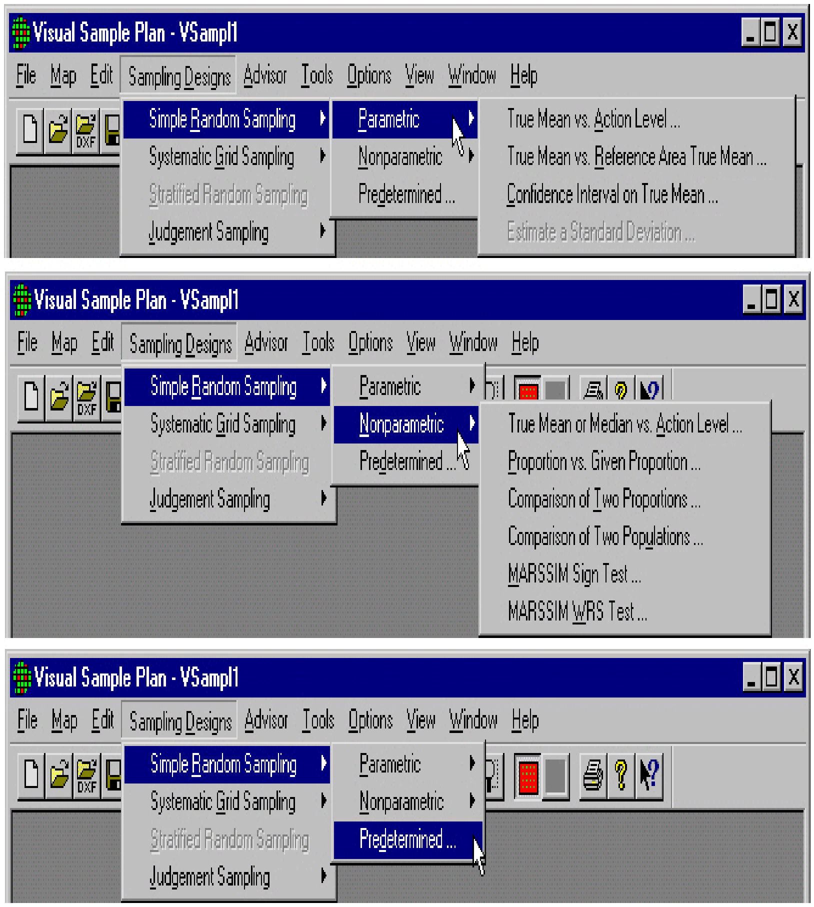

Figure 3.1. Simple Random Sampling Options in VSP

3.1.2 Selecting a Sampling Design

If you are a new VSP user, the multitude of available methods likely is confusing; you may be unsure about which test or method is appropriate for your situation. A key question can help narrow down the choices:

C What regulatory guidance will be used in the analysis of the collected data?

VSP is based primarily on two guidance documents. The majority of the sample-size equations used in VSP come from EPA's *Guidance for Data Quality Assessment–Practical Methods for Data Analysis–EPA QA/G-9* (EPA 2000b). Table 3.2 lists VSP's sampling design options that use the sample-size equations in QA/G-9.

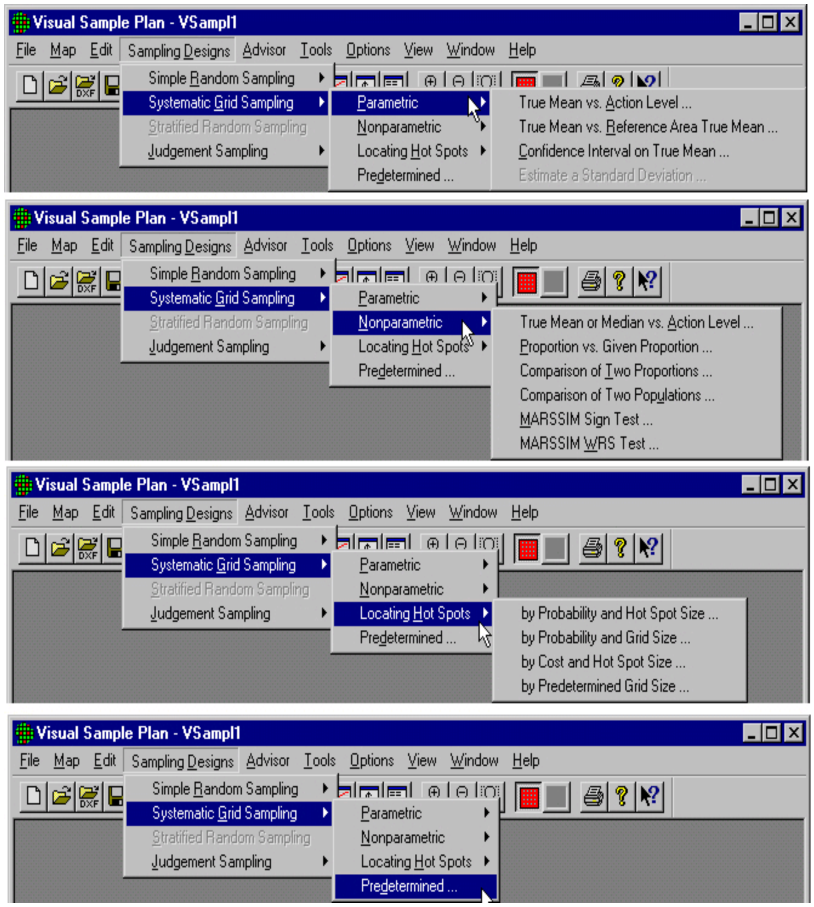

Figure 3.2. Options Based on Locating Sampling Points on Nodes of Grid

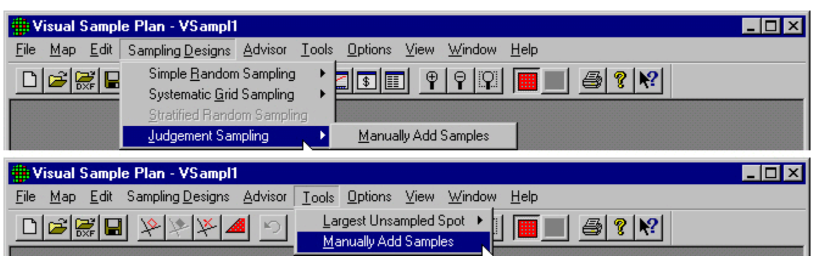

Figure 3.3. Options Allowing Selection of Sampling Locations Based on Judgment

The *Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM)* (EPA 1997) is the other regulatory guidance used to provide sample-size equations for VSP. Table 3.2 lists the sections in MARSSIM used by VSP for these equations.

Table 3.2. References in QA/G-9 Used by VSP

VSP Sampling Design Option	EPA QA/G-9 Reference
True Mean vs. Action Level	Section 3.2.1.1 The One-Sample t-Test
True Mean or Median vs. Action Level	Section 3.2.1.2 The Wilcoxon Signed Rank (One-Sample) Test for the Mean
Proportion vs. Given Proportion	Section 3.2.2.1 The One-Sample Proportion Test
True Mean vs. Reference Area True Mean	Section 3.3.1.1 Student's Two-Sample t-Test (Equal Variances)
Comparison of Two Populations	Section 3.3.3.1 The Wilcoxon Rank Sum Test
Comparison of Two Proportions	Section 3.3.2.1 Two-Sample Test for Proportions

Table 3.3. References in MARSSIM Used by VSP

VSP Sampling Design Option	MARSSIM Reference
MARSSIM Sign Test	Section 5.5.2.3 Contaminant Not Present in Background—Determining Numbers of Data Points for Statistical Tests
MARSSIM WRS Test	Section 5.5.2.2 Contaminant Present in Background—Determining Numbers of Data Points for Statistical Tests

The **Locating Hot Spots** problem is discussed in *Statistical Methods for Environmental Pollution Monitoring* (Gilbert 1987, pp. 119-131) and *ELIPGRID-PC: Upgraded Version* (Davidson 1995).

Where and when samples are taken is one of the most important components of a sampling design. There are many options for spatially and temporally selecting units to be included in a sample -- units can be selected randomly, on a regular pattern, they can be selected by strata, they can be selected sequentially—the options are many. More than one design is suitable for most problems; however, nuances of the problem or unique site conditions may suggest one design over another. For a complete discussion of sampling designs and the impact of different designs on selecting test statistics and estimation formulas, see EPA (2000c).

VSP assumes that you typically will use a probabilistic sampling design, i.e., a design that has known probabilities of inclusion of population units into the sample. VSP contains the most basic probability-based sampling designs. Most statistical tests require probability-based designs. The combination of probabilistic sampling design and statistical tests are at the core of being able to make confidence statements about conclusions drawn from sample data.

VSP currently deals only with two-dimensional spatial designs; however, with a little effort, 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 options:

- **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 and orientation randomly selected. The grid pattern is selected in the screen that comes up after you have selected the problem type. Figure 3.6 shows this screen.

- **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. Note that "formal" stratified sampling is not currently available in VSP Version 1.0. This option is shown as "greyed out" in Figures 3.1, 3.2, and 3.3. However, you can divide a site into a number of sample areas and allow VSP to proportionally allocate the number of samples to each sample area based on its physical area. This is done automatically for all selected sampling areas.

- **judgment sampling—**You simply point and click anywhere in a sampling area. These sampling locations are based on the judgment of the user.

When you select the **Systematic Grid Sampling** design, you have an option of what type of grid pattern to use. In Figure 3.4, we see the screen where grid pattern is selected. This screen appears when the **Grid** tab is selected. There are three options for the grid pattern: square, triangular (chosen in Figure 3.4), and rectangular. 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**. Once all selections have been made, hit **Apply**.

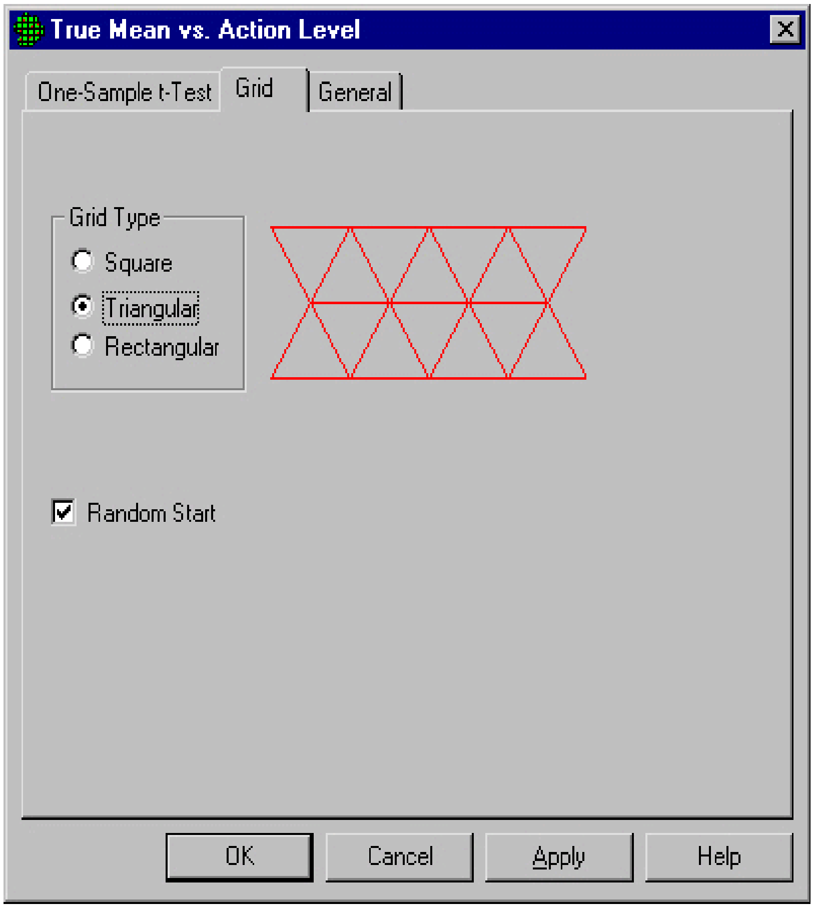

Figure 3.4. Dialog Box for Entering Type of Grid Design When **Systematic Grid Sampling** Is Selected

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 3.5 shows **Judgment Sampling** selected in VSP and six sampling locations selected manually.

If **Judgment Sampling** is selected, you are on your own to work out the details of the design and the statistical support for conclusions, similar to when the **Predetermined** option is select under Simple Random or Systematic Designs (see Figure 3.3).

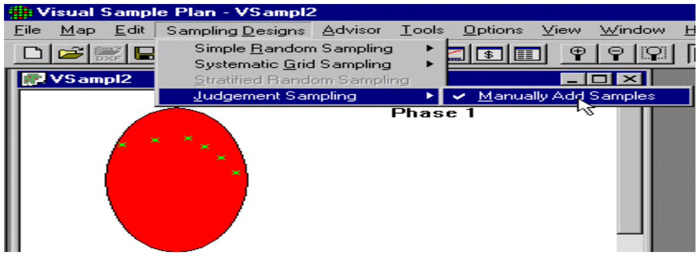

Figure 3.5. **Judgment Sampling** in VSP

3.2 DQO Inputs and Sample Size

The inputs needed for VSP's sample-size calculations are decided upon during the DQO process (see Section 1.3.1). 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 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 DQO inputs that affect sample size for the problem types that deal with parametric and nonparametric population parameters are as follows:

- *Null Hypothesis Formulation*—The null hypothesis is the working hypothesis. There must be convincing evidence in the data before the null hypothesis can be rejected. 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 latest version of its 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 the Action Level will be released as clean to the public, i.e., a "consumer's" risk. 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 latest version of the DQO guidance (EPA 2000a). 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 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. This can be thought of as a "producer's" risk, in which the responsible party is considered the producer.

- *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, this is the square root of the variance.

Note: The **Help** function in VSP provides a description of each of these units. You can put the curser in the input box for any of the DQO inputs, and a definition of what is being asked for will appear in a **Help** window.

We now provide a series of examples that illustrate both the mechanics of using VSP and some of the considerations involved in setting the DQO inputs.

3.2.1 Example 1: Comparing a Mean or Median to an Action Level

Problem Statement 1: A site has one Study Area. The Action Level for the contaminant of interest is 6 pCi/g in the top 6 inches 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 H_0: $\mu \geq$ AL.

We desire an alpha error rate of 1%. We also desire a beta error rate of 1% as recommended by EPA QA/G-4 guidance (EPA 2000a, p. 6-11). 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, the error rates, 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 1.1: 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.

VSP Solution 1.1: We start by choosing VSP option **Sampling Designs ÷ Systematic Grid Sampling ÷ Parametric ÷ True Mean vs. Action Level**. A grouping of the input dialogs is shown in Figure 3.6.

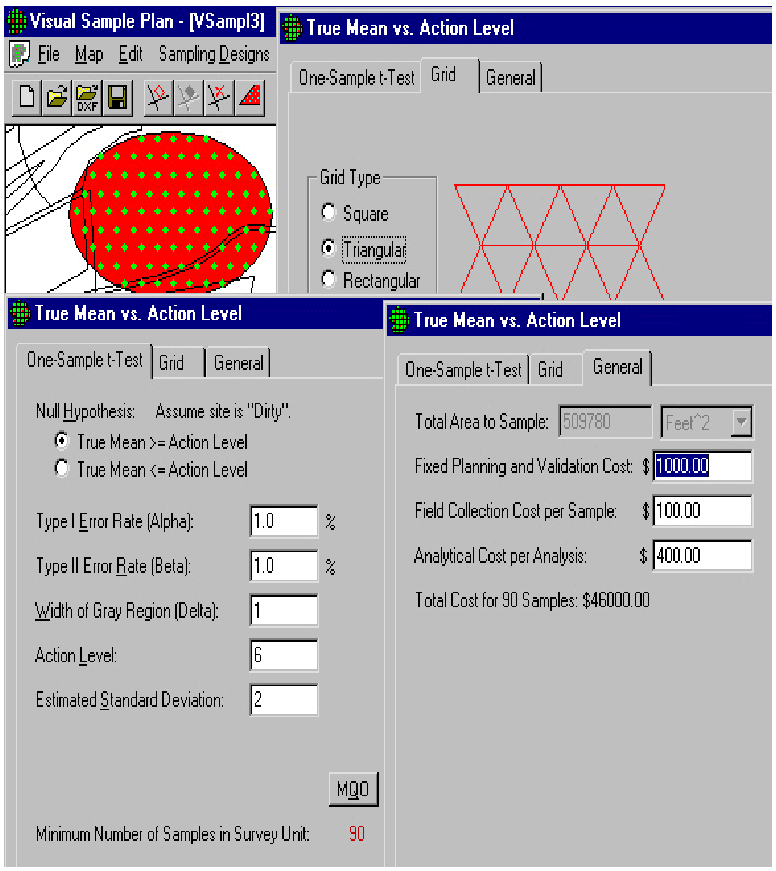

Figure 3.6. Input Boxes for Case 1.1 with Original Error Rates

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% is acceptable. Figure 3.7 reveals that those changes lead to a significant reduction in the sampling cost, now $19,000 for n = 36 samples.

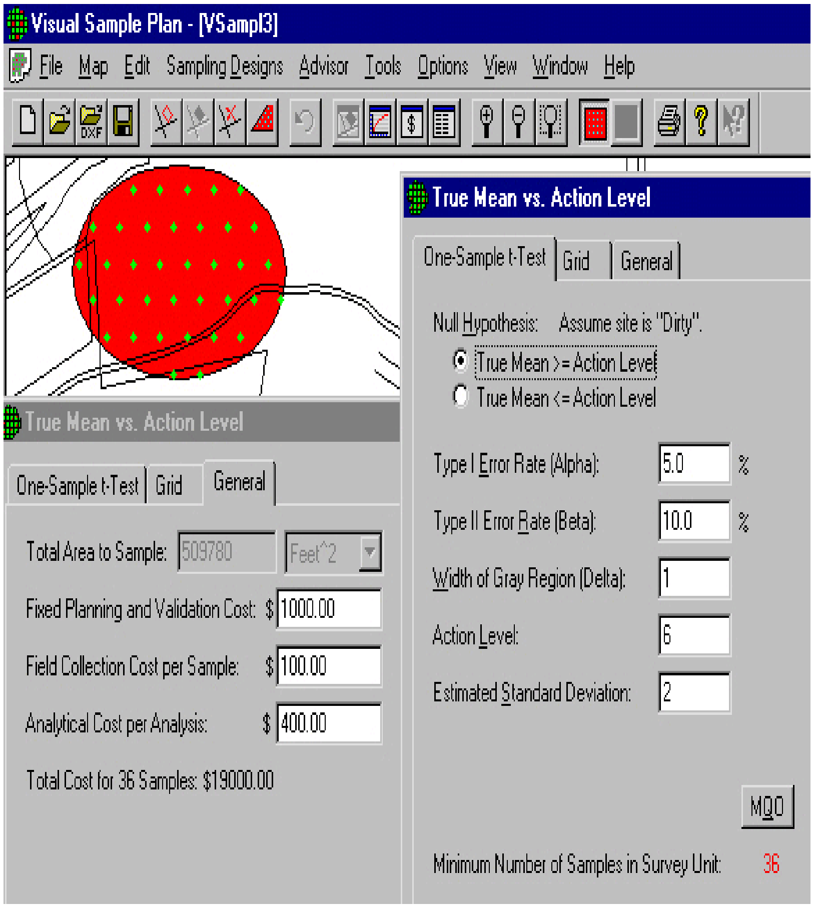

Figure 3.7. Input Boxes for Case 1.1 with Increased Error Rates

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 QA/G-9* (EPA 2000b, p. 3-5).

Case 1.2: 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 1.2: We start by choosing VSP option **Sampling Designs ÷ Systematic Grid Sampling ÷ Nonparametric ÷ True Mean or Median vs. Action Level**. A grouping of the input dialogs is shown in Figure 3.8.

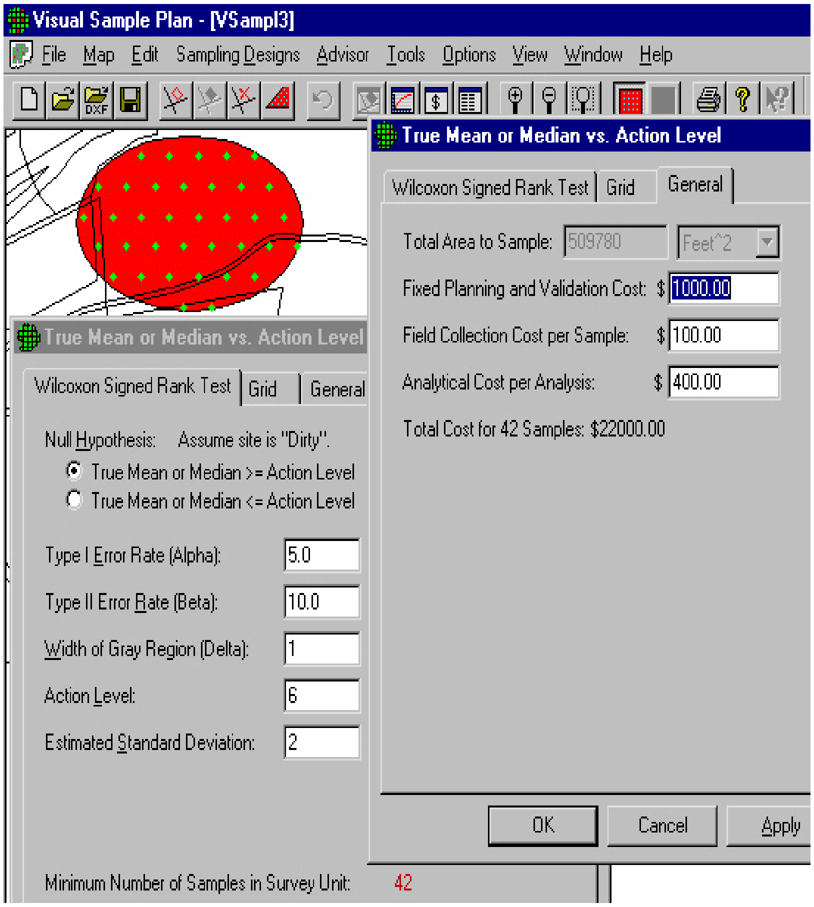

Figure 3.8. Input Boxes for Case 1.2 Using Nonparametric Option

For our inputs, and assuming that we will use a nonparametric Wilcoxon Signed Rank 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.

3.2.2 Example 2: Comparing Mean or Median to Reference Area Mean or Median

Problem Statement 2: A site has one Study Area. The Action Level for the contaminant of interest is 5 pCi/g above background in the top 6 inches 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 H_0: Site μ - Reference Area μ $ AL. 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% as recommended by EPA QA/G-4 guidance (EPA 2000a, p. 6-11). 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, the error rates, 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, since 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 2.1: 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 2.1: We start by choosing VSP option **Sampling Designs ÷ Systematic Grid Sampling ÷ Parametric ÷ True Mean vs. Reference Area True Mean**. A grouping of the input dialogs is shown in Figure 3.9.

We see that for our inputs, using a two-sample t-test will require taking175 field samples in the Study 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?

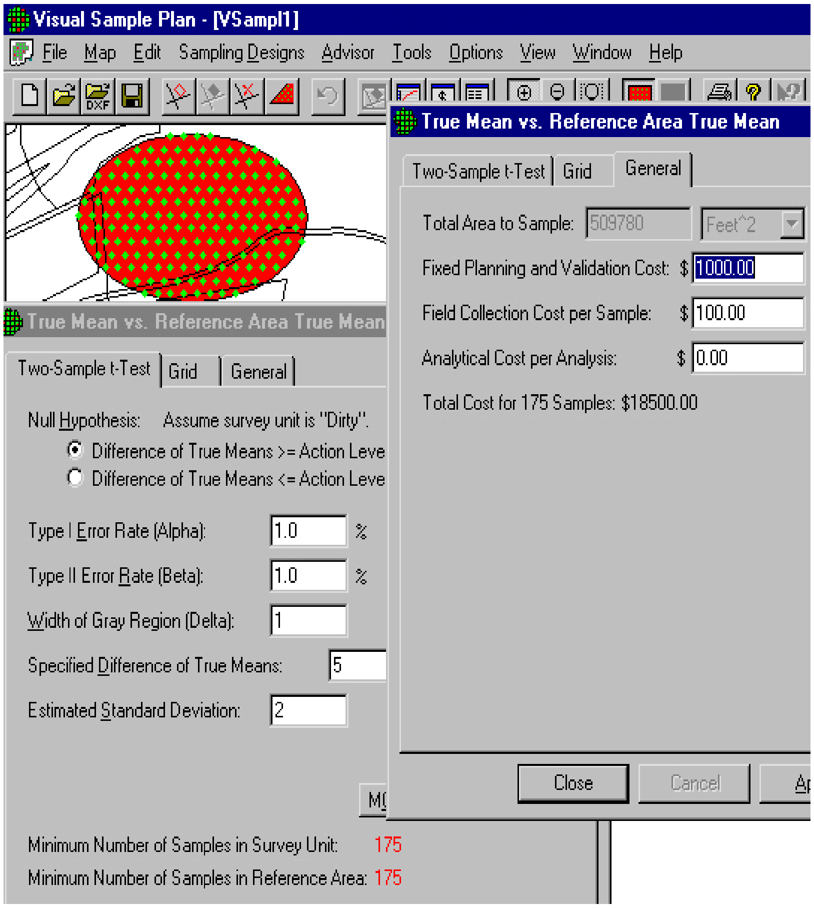

Figure 3.9. Input Boxes for Case 2.1 with Original Error Rates

In Figure 3.10, you can see that 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. This was done by increasing both the alpha error rate and the beta error rate to 5%.

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.

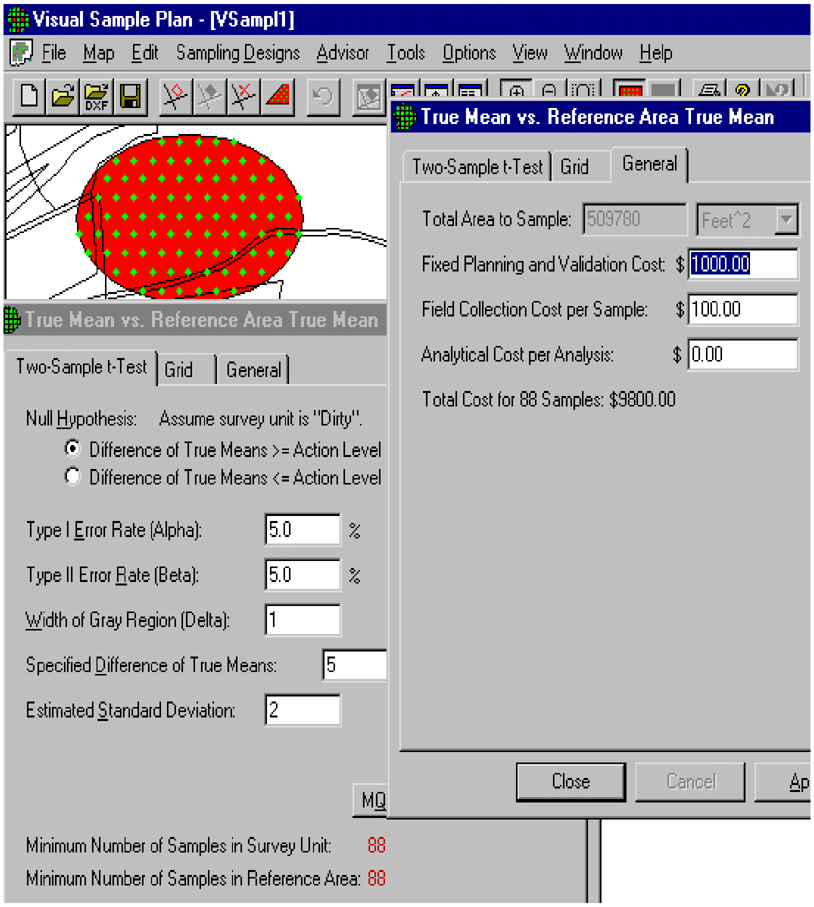

Figure 3.10. Input Boxes for Case 2.1 with Increased Error Rates

What about the assumption of approximately equal standard deviations for the measurements? When we collect the data, we will need to check that assumption. See *Guidance for Data Quality Assessment Practical Methods for Data Analysis EPA QA/G-9* (EPA 2000b, p. 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 2.2: We now look at the case in which the nonparametric Wilcoxon Rank Sum test is planned for the data analysis phase of the project.

VSP Solution 2.2: We start by choosing VSP option **Sampling Designs ÷ Systematic Grid Sampling ÷ Nonparametric ÷ Comparison of Two Populations**. A grouping of the input dialogs is shown in Figure 3.11.

In Figure 3.11, 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.

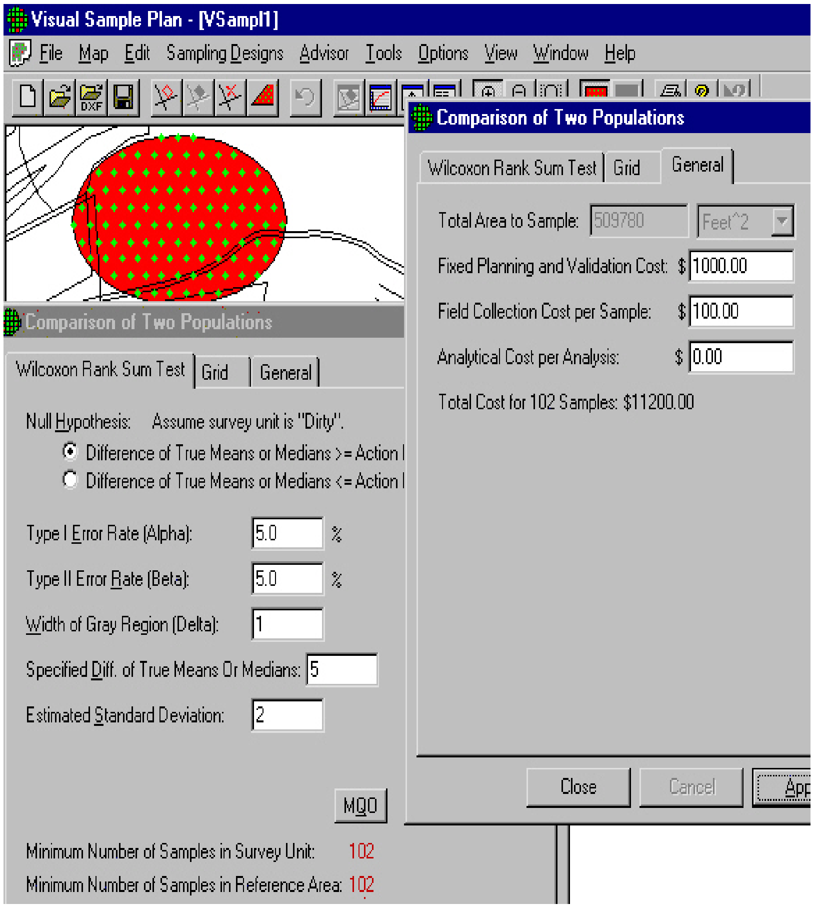

Figure 3.11. Input Boxes for Case 2.2 Using Nonparametric Wilcoxon Rank Sum Test

Case 2.3: 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 adust the various DQO input parameters and try to discover a sampling plan that will meet the new goals.

VSP Solution 2.3: Figure 3.12 shows that with an alpha of 5%, a beta of 20%, and a gray region width of 1.25, 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.

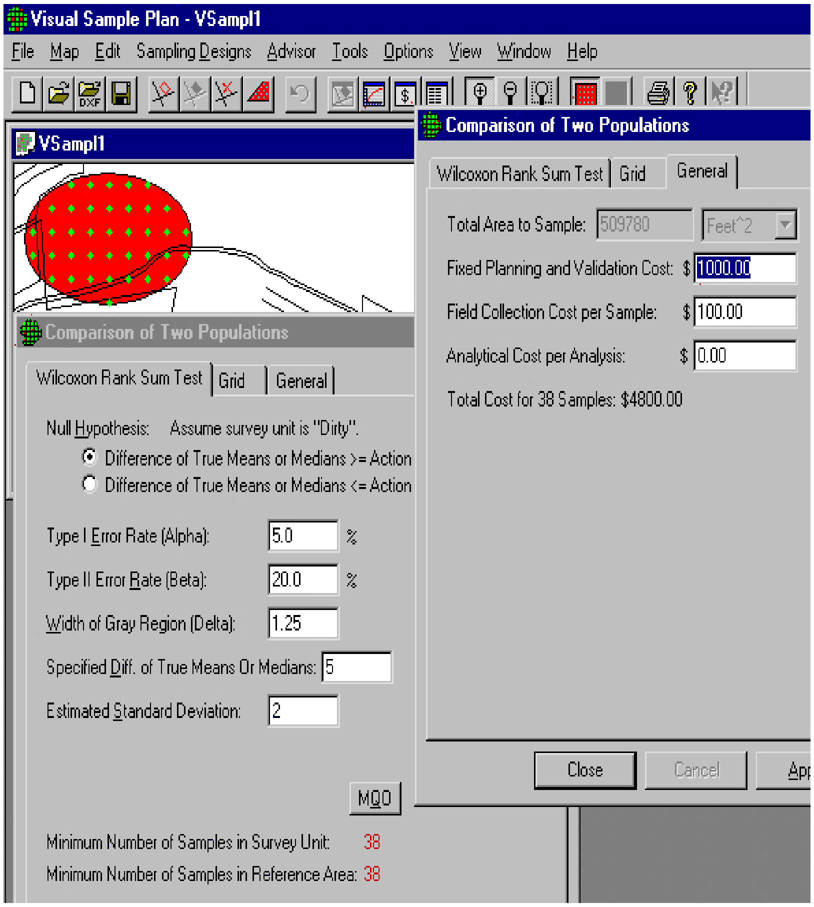

Figure 3.12. Input Boxes for Case 2.3 Using Nonparametric Wilcoxon Rank Sum Test

Case 2.4: Suppose our combined sampling budget is reduced to $5,000. Can VSP provide a sampling design that meets that goal?

VSP Solution 2.4: Figure 3.13 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.

There are definite consequences of reducing sampling requirements to fit a minimum budget. The consequences could include a greater chance of concluding that a dirty site is clean or a clean site is dirty.

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.

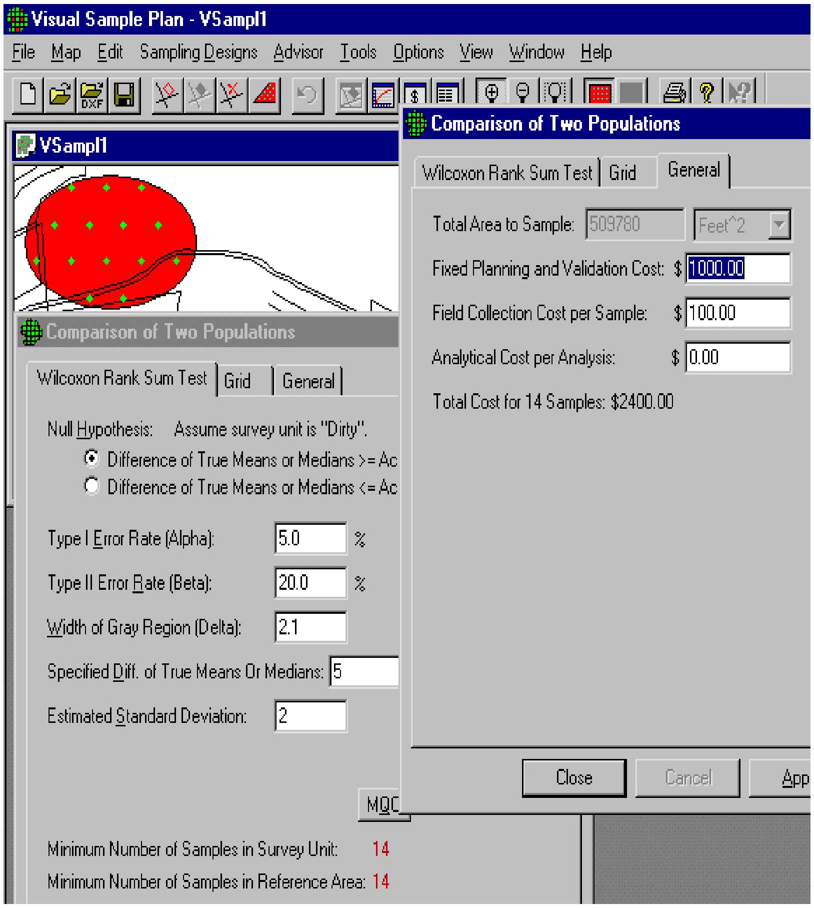

Figure 3.13. Input Boxes for Case 2.4 Using Nonparametric Wilcoxon Rank Sum Test

Case 2.5: To illustrate the GIGO problem, 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?

VSP Solution 2.5: Figure 3.14 shows the new sample size has jumped to 53, almost a four-fold increase over the 14 samples used in VSP Solution 2.4. 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.

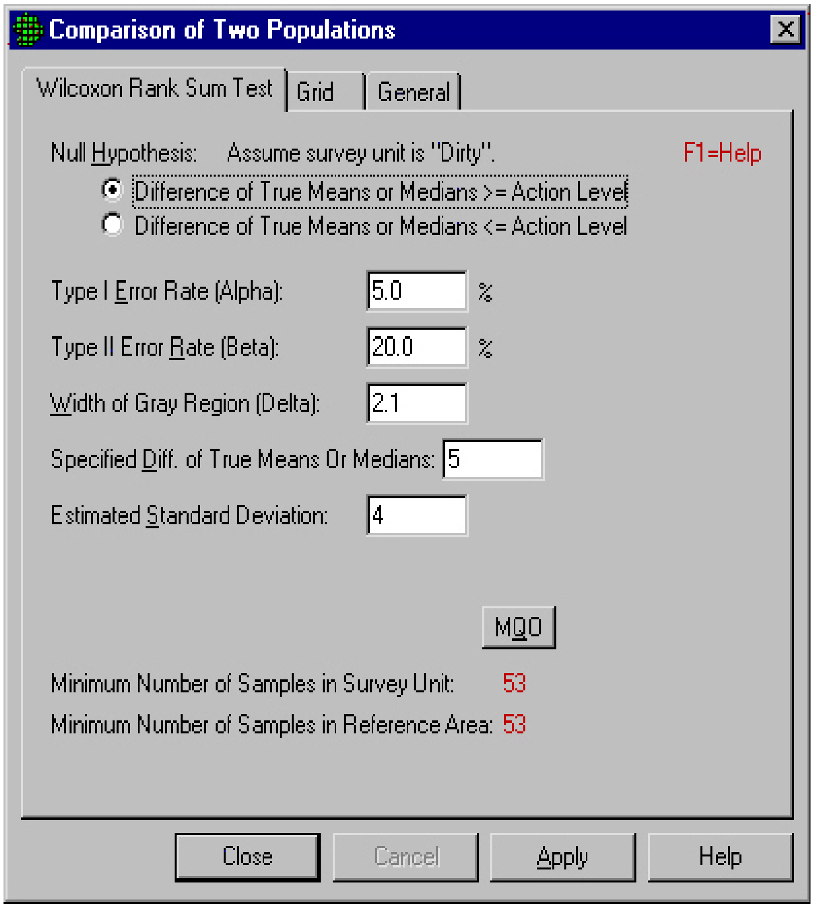

Figure 3.14. Input Boxes for Case 2.5 with Larger Standard Deviation

3.2.3 Example 3: Locating Hot Spots

Problem Statement 3: A site has one Study Area of one acre. 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%. More information on this general problem can be found in *Statistical Methods for Environmental Pollution Monitoring* (Gilbert 1997). 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 3.1: We assume that the assumptions listed in Gilbert (1987, p. 119) are valid for our problem. We specify a hit probability of 95% and a radius of 15 feet. We will let VSP calculate the length of the side of the equilateral triangular grid needed for these inputs.

VSP Solution 3.1: First, open the file OneAcre.Vsp using VSP Main Menu option **File ÷ Open Project**. Next, select VSP option **Sampling Designs ÷ Systematic Grid Sampling ÷ Locating Hot Spots ÷ by Probability and Hot Spot Size**. A grouping of the input dialogs is shown in Figure 3.15.

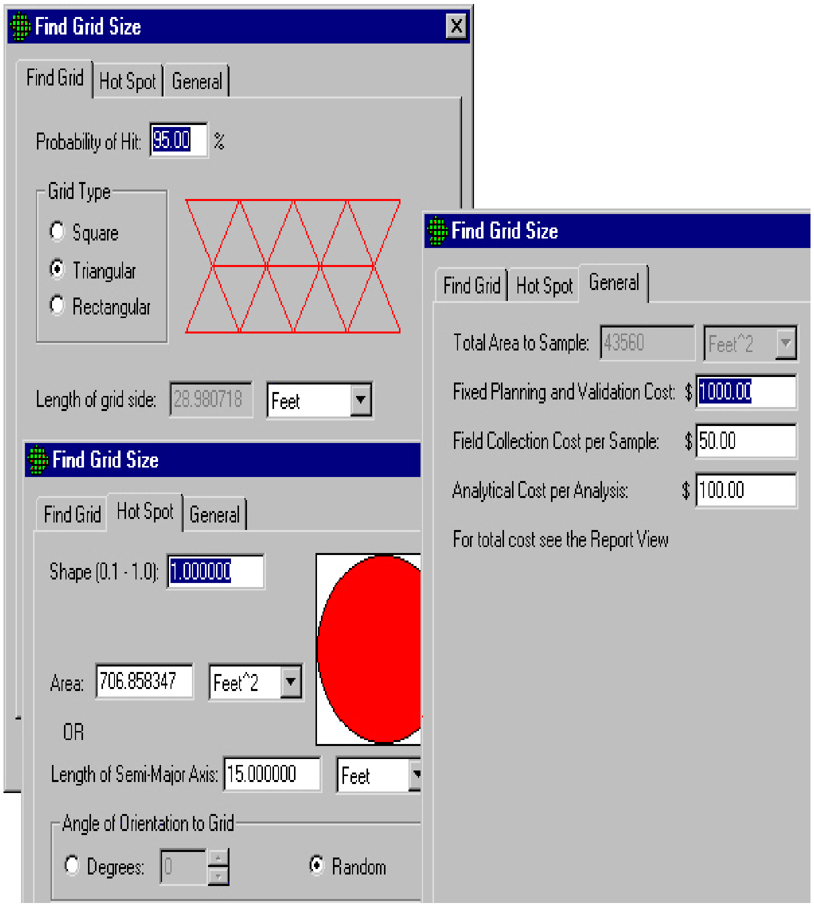

Figure 3.15. Input Boxes for Case 3.1

The recommended length of grid side is shown in the dialog box with the **Find Grid** tab. It is about 29.98 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. However, the required number of samples given in the **Report View** 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 Study Area's outside edges.

Figure 3.16 shows two examples of reported sample sizes for Case 3.1. In the top **Report View**, the sample size is given as 54; in the bottom **Report View**, the sample size is given as 63.

Also note that the curve in the graph is for the theoretical case of no edge effects. Therefore, values obtained from the graph will often differ from the **Report View** values. This also explains why the blue line on the graph will often change as you repeatedly press the **Apply** button.

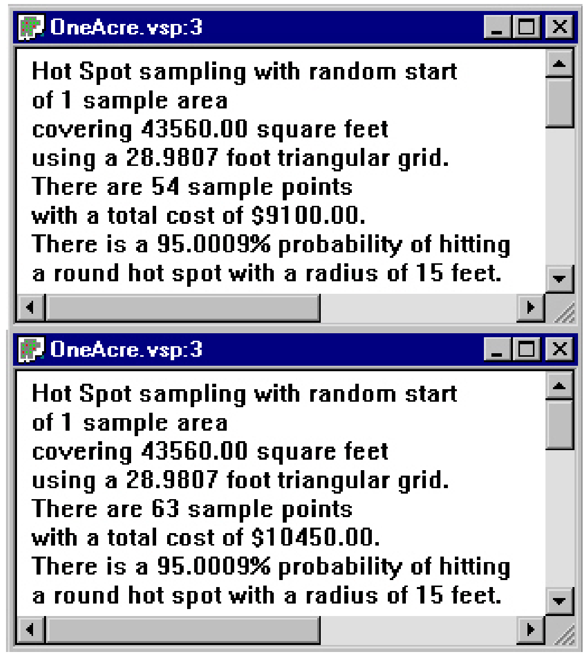

Figure 3.16. Two Different Sample Sizes for Same Inputs to Case 3.1 Hot-Spot Problem

3.3 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, click on the **General** tab to enter costs. A sample **General** screen is shown in Figure 3.17.

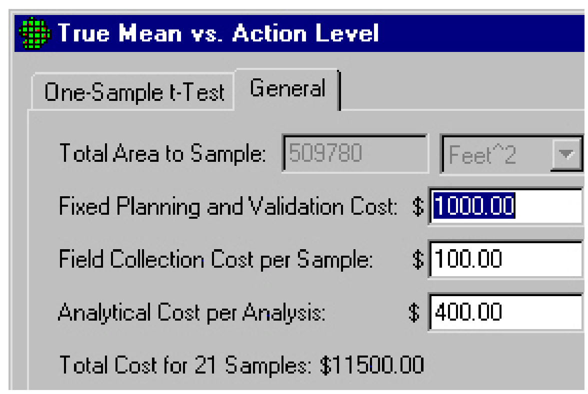

Figure 3.17. Screen for Entering Sampling Costs for Sampling Design

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 cost. Examples of per-unit field costs are the costs paid to 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.

4.0 Assessment of Sampling Plans

VSP provides multiple displays for allowing you to assess the sampling plan that has been designed/selected. 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), and 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 desired VSP sampling plan.

4.1 Display of Sampling Design on the Map

In Section 2.1, we described how to set up a **Map**. In Section 2.2, we described how to set up a **Study 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 Study Area. We then selected option **Sampling Designs ÷ Systematic Grid Sampling ÷ Locating Hot Spots ÷ by Probability and Hot Spot Size.** 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 be selecting a **Shape** of **1.0**. We press **Apply**, and when we return to the map, we see a display similar to that shown in Figure 4.1. 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 3.3, we described how to enter costs. We can see the result of applying those costs to the design we selected by selecting **Report** from the View menu. What is displayed is a report that shows, among other things, the total cost of the design we have chosen. An example of a report is shown in Figure 4.2. We see that for a Study Area of 2577.6 square feet, using a 7.17-foot square grid, there are 55 sample points or grid nodes, with a total cost of $27,500.

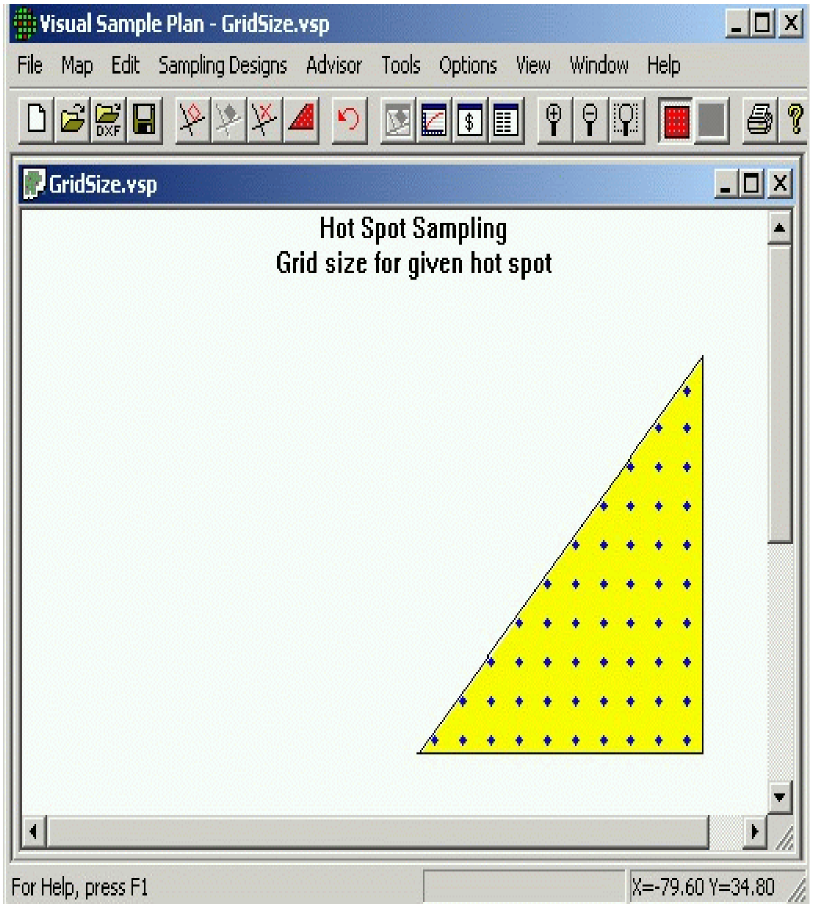

Figure 4.1. Display of Sampling Locations on Map

4.3 Display of Performance of Design (Graph)

VSP has a Performance of Design display that shows a graph of the performance of a sampling design that has just been created. Each Type of Problem has a performance display tailored to its situation. Once a **Sampling Design** has been selected, the **DQO inputs** entered on the dialog box input screen, and the **Apply** button pressed to apply the design to the Study Area, the display of the performance can be seen by selecting **View ÷ Graph** from VSP's main menu.

Some of the displays available for various types of problems are described in the following paragraphs.

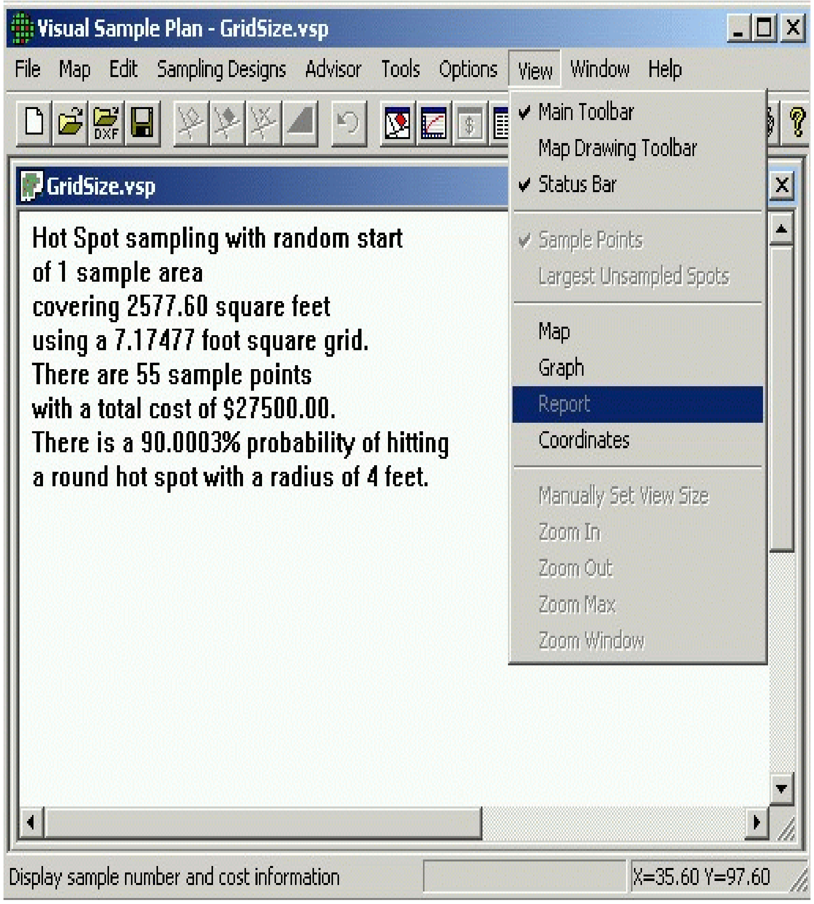

Figure 4.2. Display of **Report** Showing Total Costs of Sampling Design

4.3.1 Performance of Design for Hot Spot Problem Type

The first sampling design performance assessment display we show is a graph of the probability of hitting a hot spot of a specified size vs. the number of samples using whatever type of grid was specified (a **square** grid in our example). The more sampling points (i.e., the tighter the grid spacing and hence the higher cost), the higher the probability of hitting the hot spot with one of the nodes on the grid. A dashed blue line on the graph shows the desired input of a **90%** probability of hitting the hot spot of radius **4 feet** and the required **55** sample points to achieve this.

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 4.3, we have the solid line positioned at a 60% probability of hitting a hot spot of radius 4 feet when there are only 31 grid samples in an area of 2577.6 square feet.

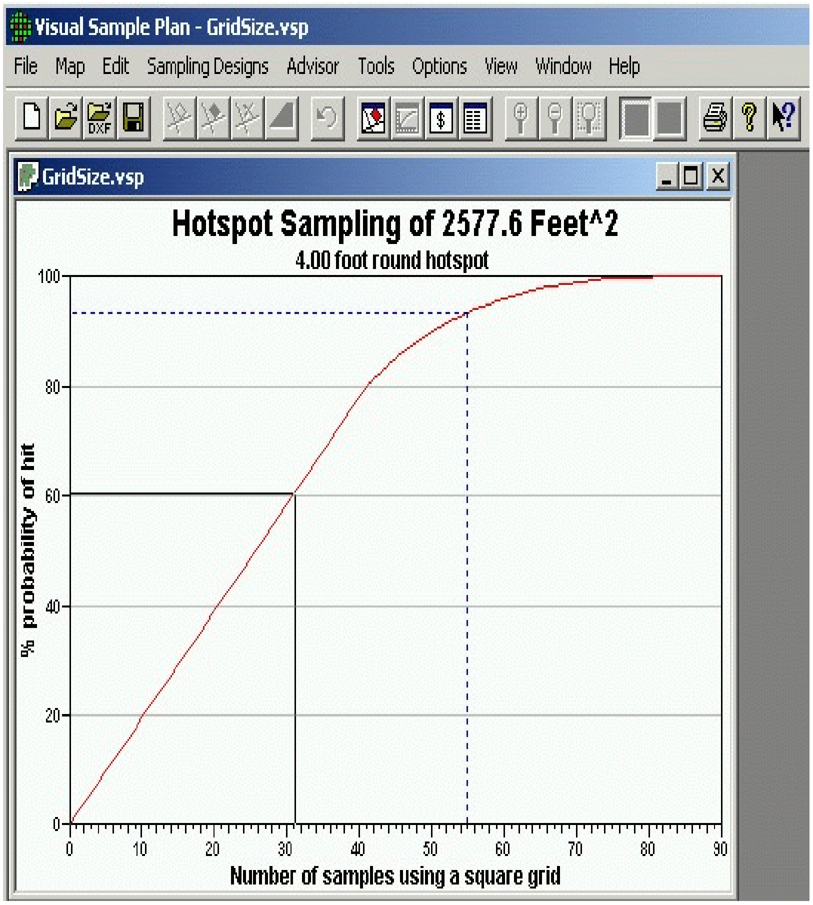

Figure 4.3. Display of Graph Showing Probability of Hitting Hot Spot vs. Number of Samples

4.3.2 Performance of Design for the Comparison of a Mean to an Action Level

The next performance assessment display we show is a graph of the probability of deciding the sample area is dirty on the vertical axis vs. a range of possible true mean values on the horizontal axis. Figure 4.4 is the Decision Performance Curve (DPC) described in EPA's QA/G-4 guidance (EPA 2000a, pp. 6-7 to 6-11). The DPC shows the range of possible true values of the parameter of interest on the x-axis and the range of probabilities (0 to 1) on the y-axis of deciding the parameter of interest exceeds the Action Level using sample data. (If the null hypothesis is "Site is Dirty," the y-axis reads greater than *or equal to* the Action Level).

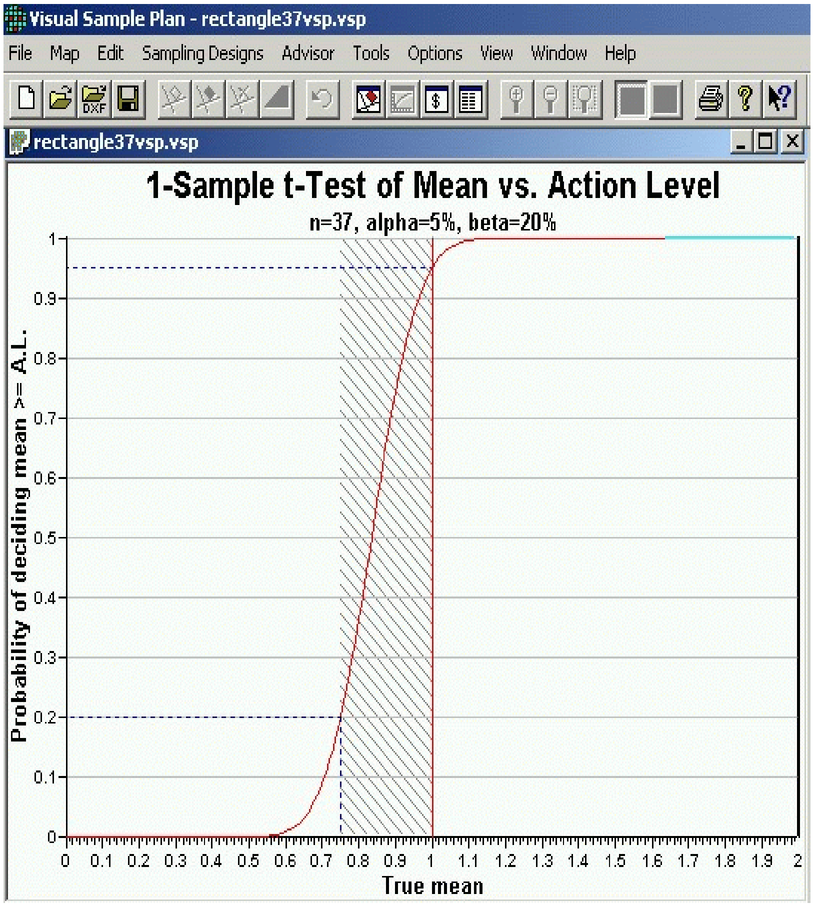

Figure 4.4. Decision Performance Curve for H_0: True Mean >= Action Level for Comparing Mean vs. Action Level

The parameter of interest for this example is the mean. The solid vertical red line is positioned at a true mean value of **1.0**, which corresponds to the Action Level. The area in gray hash marks is the gray region shown here from 0.75 to 1.0 and input as a delta (width) of **0.25**. 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 **20%**. 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).

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.

The DPC graph in Figure 4.4 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 4.4 also is telling us

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

- Dirty sites with true mean just *above* the Action Level will lead to mostly *correct* decisions.

However, when we reverse the null hypothesis and state it as **H_0: True Mean <= Action Level**, i.e., assume “Site is Clean,” we see in Figure 4.5 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.

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, the width of the gray region, total sampling costs, and costs of incorrect decisions.

4.3.3 Performance of Design for the Confidence Interval for a Mean Problem Type

The display for assessing a confidence interval for a mean is somewhat different from the previous two examples because this is not a testing problem but an estimation problem. As such, there is only one type of decision error rate, alpha. Shown in Figure 4.6 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 **95% one-sided** confidence interval on the mean was desired.

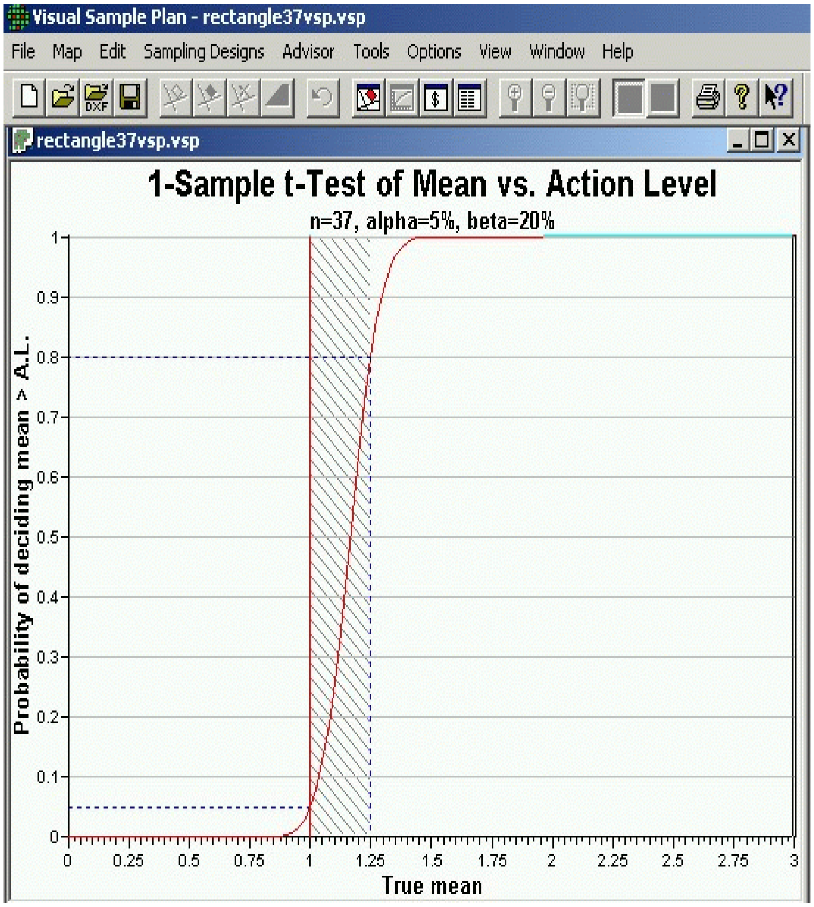

Figure 4.5. Decision Performance Curve for H_0: True Mean <= Action Level for Comparing Mean vs. Action Level

VSP calculated that a sample size of 26 was required. The 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 4.6, 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 99.54 and 0.50.

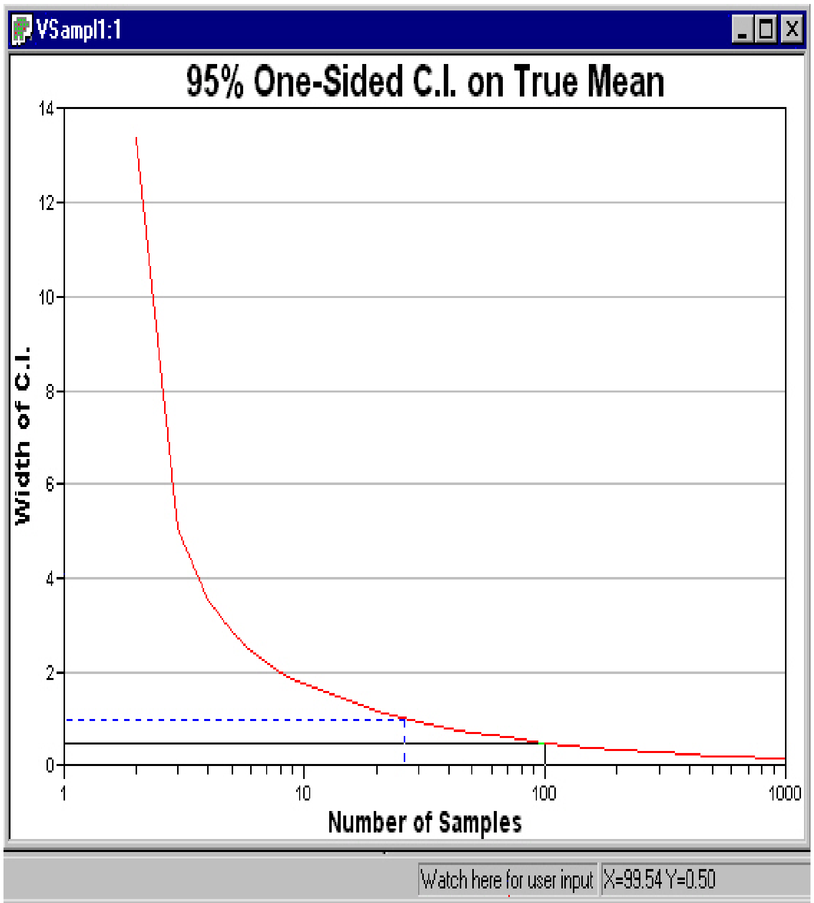

Figure 4.6. Decision Performance Graph for One-Sided 95% Confidence Interval

4.3.4 Performance of Design for Comparing a Proportion to a Given Proportion

The sampling design assessment display for comparing a proportion to a fixed proportion is a graph of the number of samples vs. beta. Remember, if the null hypothesis is "Site is Dirty," beta is the probability of deciding a specific clean site is dirty if the true proportion is equal to the lower bound of the gray region (LBGR).

For this problem, we set alpha to **1%** and beta to **5%,** and the width of the gray region to **0.15.** VSP calculated a sample size of 122. The dashed blue line terminates on the y-axis at 122 samples and on the x-axis at a beta of **5%**.

Note the solid black line in Figure 4.7 and the values in the status bar. They show that for a beta of 20%, the minimum number of samples is reduced to 82.84, i.e., 83 samples. Moving the black line is a quick way to play "what-if" games regarding sample sizes and beta error rates for a given alpha.

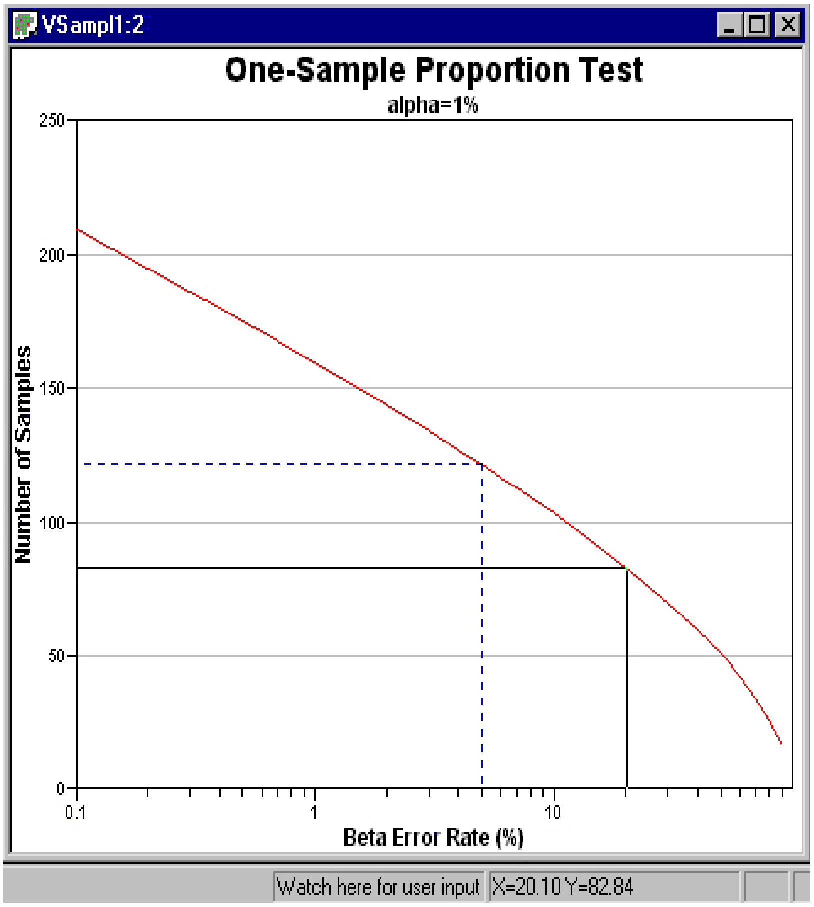

Figure 4.7. Decision Performance Graph for Comparing Proportion to Fixed Proportion

4.3.5 Performance of Design for Parametric Comparison of Two Populations

The sampling design assessment display for comparing two populations using the parametric two-sample t-test is a graph of the probability of deciding the difference of true means is greater than or equal to the Action Level vs. various differences of true means. This graph is similar to the Decision Performance Curve discussed in Section 4.3.2, but this time we are dealing with two populations and the x-axis is a range of possible *differences* between the two population means.

The graph shown in Figure 4.8 is for **H_0: Difference of True Means or Medians >= Action Level**, where the specified difference (Action Level) is **5**, delta is **2**, alpha = **5%**, beta = **10%**, and the estimated common standard deviation = **3**.

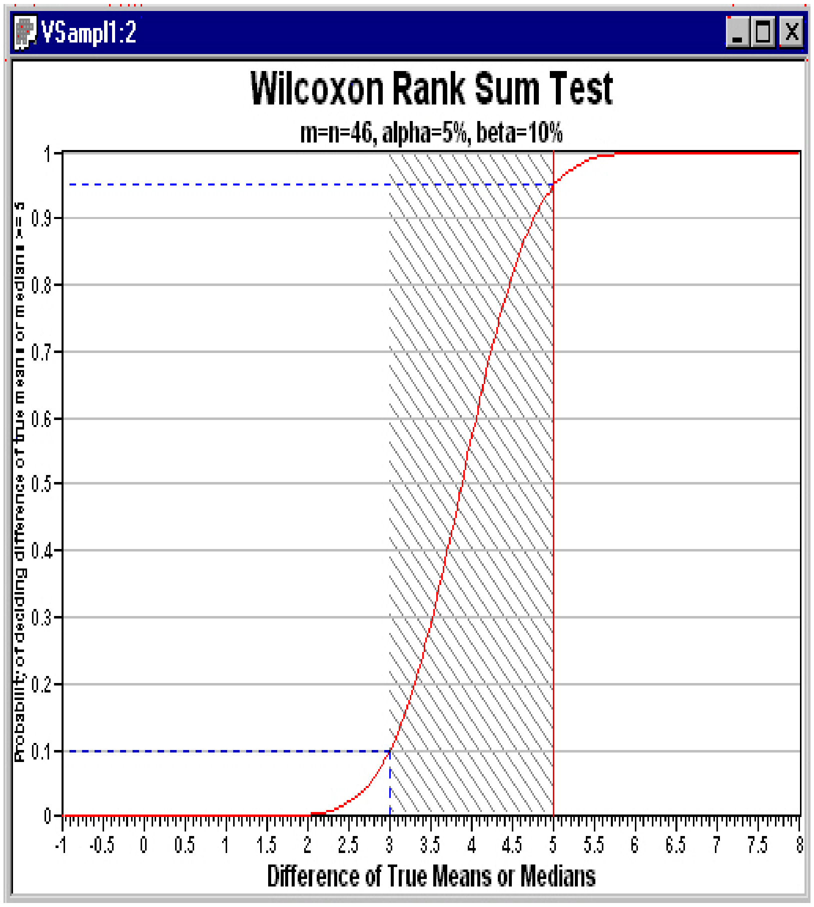

Figure 4.8. Decision Performance Graph for Comparing Means of Two Populations

Figure 4.8 shows us the specific Decision Performance Curve for the given inputs. The probabilities of deciding the Study Area is 5 or more units (pCi/g, ppm, etc.) above the Reference Area are plotted against potential true differences in means between the Study Area population and the Reference Area (background) population.

4.3.6 Performance of Design for Comparing the Mean or Median of a Population to an Action Level (MARSSIM Sign Test)

While the Sign Test is a standard statistical test, we have included it here with reference to MARSSIM, the multi-agency document *Multi-Agency Radiation Survey and Site Investigation Manual* (EPA 1997, pp. 5-31 to 5-35). MARSSIM suggests using the Sign Test to detect uniform failure of remedial action throughout a Survey Unit when the contaminant in not present in background or only at negligible levels.

The example we use to illustrate the MARSSIM Sign Test Decision Performance Curve is adapted from MARSSIM (EPA 1997, pp. 5-33 and 5-34).

A site has one Survey Unit. The Derived Concentration Guideline Limit (DCGL) for the contaminant of interest is 140 Bq/kg (3.9 pCi/g) in soil. The contaminant is not present in background; data from previous investigations indicate an estimated standard deviation of 3.7 Bq/kg (MARSSIM inadvertently refers to this number as an average.) For this problem, the alpha error rate is set to 5% and the beta error rate is set to 1%. The LBGR is set to125 Bq/kg, so delta is 140 Bq/kg - 125 Bq/kg = 15 Bq/kg.

Determine the number and placement of sampling locations if the Survey Unit is a 15-ft-long by 14.8-ft-wide by 9.8-ft-high room.

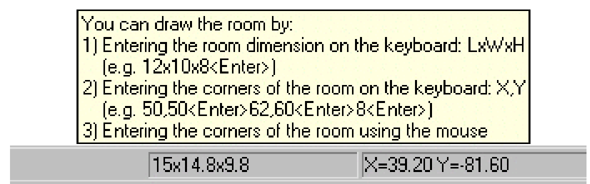

Figure 4.9. Status Bar and Tool Tip for **Draw MARSSIM** Room Option

We begin by using VSP Main Menu option **File ÷ New Project**. Next we create the room using option **Map ÷ Draw MARSSIM Room**. We enter the dimensions as **15 x 14.8 x 9.8** and press the **Enter** key (the letter "x" occurs between the dimensions). Note that Figure 4.9 shows the small yellow tool tip that comes up when this option is selected. We have moved it down just above the status bar at the bottom of the screen where the room dimensions are displayed.

Next, we click the **No** button to exclude the ceiling from our drawing. The result with inputs is shown in Figure 4.10.

VSP recommended a minimum number of samples in the Survey Unit of 20. The value of 20 comes from the calculated sample size of 16 plus a user-supplied percent overage, set here at 20% as discussed in MARSSIM, p. 5-33.

The graph in Figure 4.11 is the Decision Performance Curve or Decision Performance Goal Diagram. (EPA [2000a, p. 6-11] makes a distinction between the DPC and the DPGD. For historical reasons, we do not.) The dashed blue horizontal line at the top highlights the level for alpha, as described earlier, and the gray hash marks denote the gray region. The dashed blue line for beta at the bottom is mostly hidden by the curve.

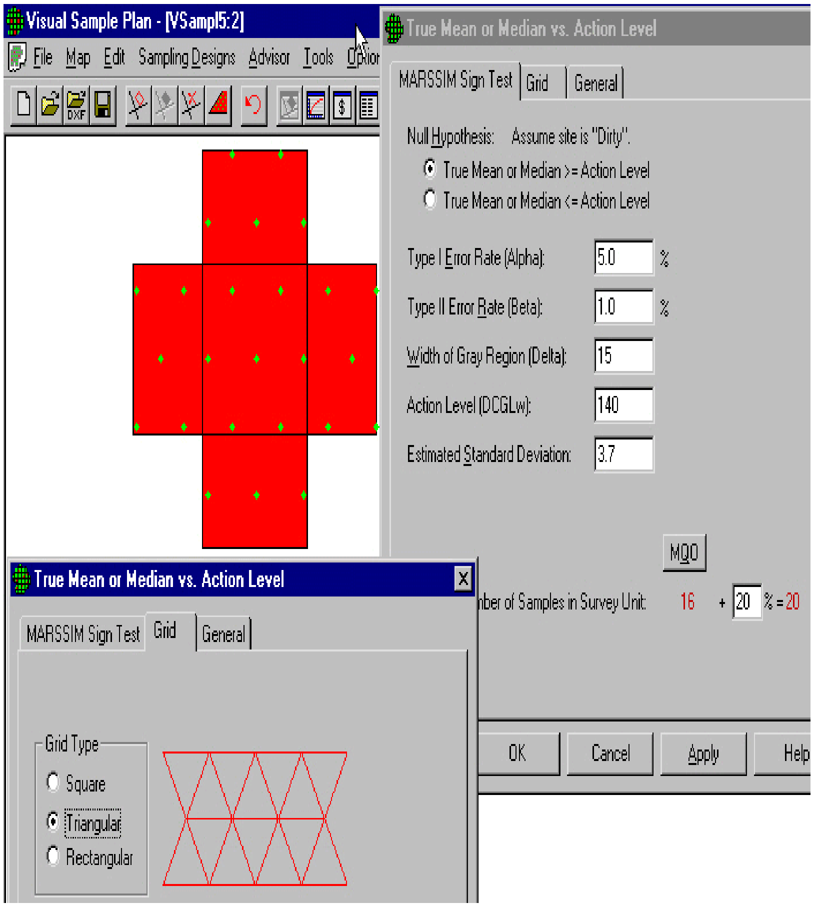

Figure 4.10. MARSSIM Sign Test Inputs and Room

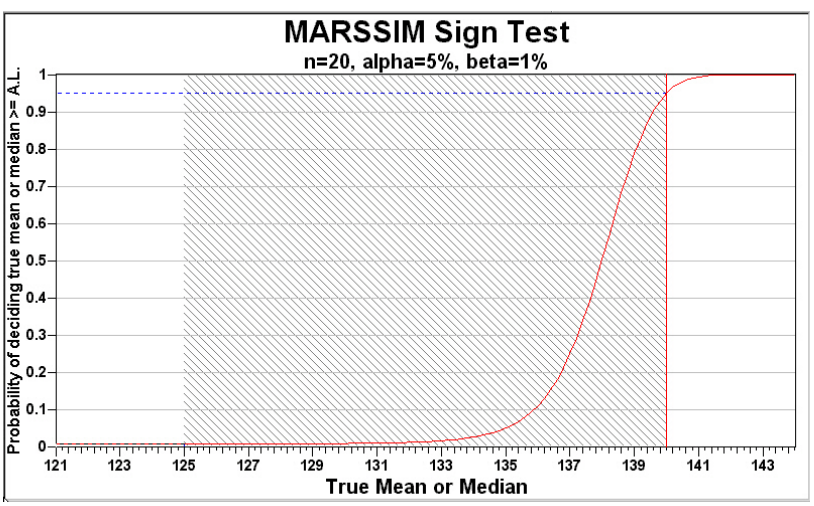

Figure 4.11. Decision Performance Graph for MARSSIM Sign Test

4.3.7 Performance of Design for Comparing the Mean or Median of a Population to a Background Level (MARSSIM WRS Test)

The WRS, or Wilcoxon Rank Sum test, is actually a two-sample statistical test. As discussed in MARSSIM (EPA 1997, pp. 5-25 to 5-31), it is used to test the null hypothesis that the mean or median concentration in the survey unit exceeds that in the reference area (i.e., background area) by more than the $DCGL_W$. In other words, the contaminant is present in background.

The example we use to illustrate the MARSSIM WRS Test Decision Performance Curve is adapted from MARSSIM (EPA 1997, pp. 5-29 to 5-31).

A site has 14 Survey Units and 1 Reference Area. The $DCGL_w$ for the contaminant of interest is 160 cpm. The contaminant is present in background at a level of 45 cpm. Data from previous investigations indicate an estimated standard deviation of 20 cpm. For this problem, the alpha error rate is set to 5% and the beta error rate is set to 5%. The LBGR is initially set at 80 cpm, resulting in delta equal to 80 cpm, i.e., 160 cpm - 80 cpm.

Our task is to determine the number and placement of sampling locations for a rectangular Survey Unit 45 ft long by 20.8 ft wide.

We begin by using VSP Main Menu option **File ÷ New Project**. Next, we create the rectangular Survey Unit using option **Map ÷ Draw Rectangle**. We enter the corner coordinates of the Survey Unit as 0,0 and 55,20.8 and press the **Enter** key (here a comma occurs between the coordinates). The result with inputs and the **Report View** is shown in Figure 4.12.

VSP recommended a minimum number of 10 samples in the Survey Unit. The value of 10 comes from the calculated sample size of 8 plus a user-supplied percent overage, set here at 20% as discussed in MARSSIM (EPA 1997, p. 5-29). Notice that there are actually 12 sampling locations shown in the Survey Unit and listed in the **Report View**. This occurs because of the way this particular random grid fit into the Survey Unit. If one presses the **Apply** button repeatedly, this number will change depending how each grid fits into the Survey Unit.

We recommend either obtaining a grid with the exact number recommended in the VSP dialog box or sampling at all the locations on a grid with a larger number of nodes than recommended by the sample size calculations. Arbitrarily dropping two locations, as this example could lead you to do, would result in uneven coverage of the Survey Unit.

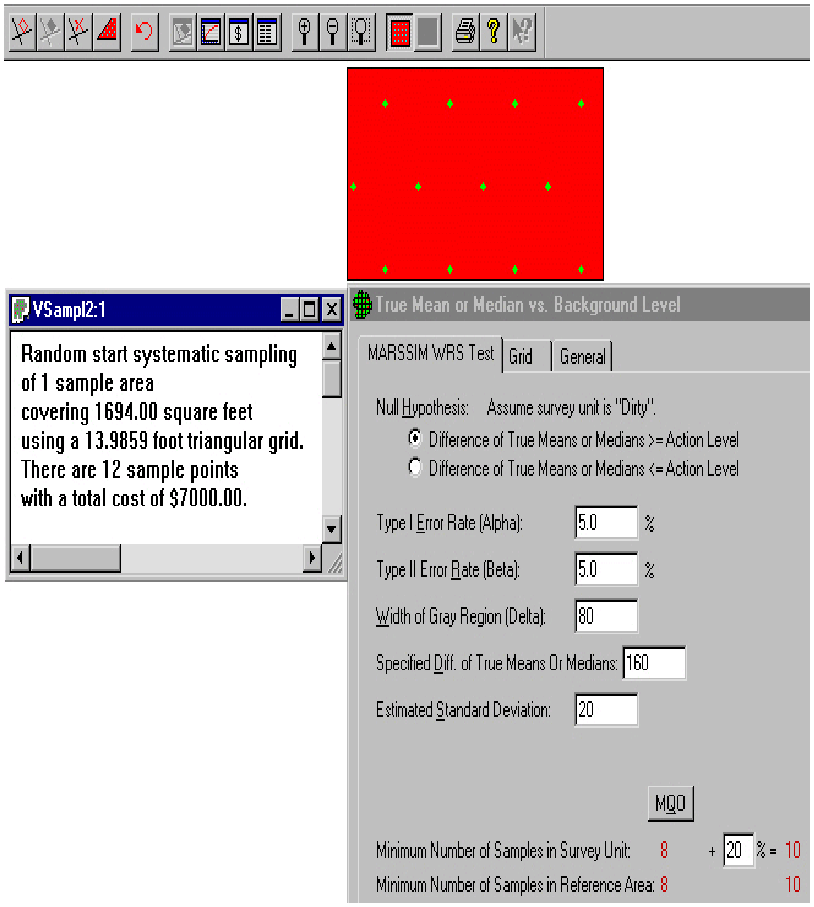

Figure 4.12. MARSSIM WRS Test Survey Unit, Input Dialog Box, and Report View

Another point of potential confusion relates to the MARSSIM recommendation of 9 sampling locations for each Survey Unit instead of the 10 locations given by VSP. This occurs because of the way VSP adds the percent overage and then rounds the number of locations for each area.

The graph in Figure 4.13 is the Decision Performance Curve. The dashed blue horizontal line at the top highlights the level for alpha, as described earlier, and the gray hash marks indicate the gray region. The dashed blue line for beta at the bottom is mostly hidden by the curve.

The black line in the middle of the graph allows you to move the cursor and see different combinations of x-axis values with the associated y-axis values in the status bar at the bottom of the screen. For example, Figure 4.14 shows that when the difference of true means or medians is 140 cpm, the associated probability of deciding the site is dirty is 47%. This large probability of error is what distinguishes the gray region from the very clean region on the far left and the very dirty region on the far right.

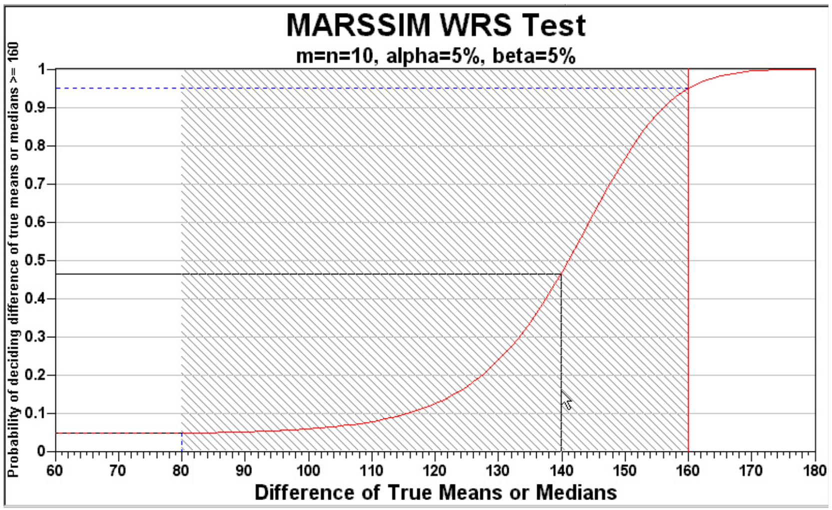

Figure 4.13. Decision Performance Graph for MARSSIM WRS Test

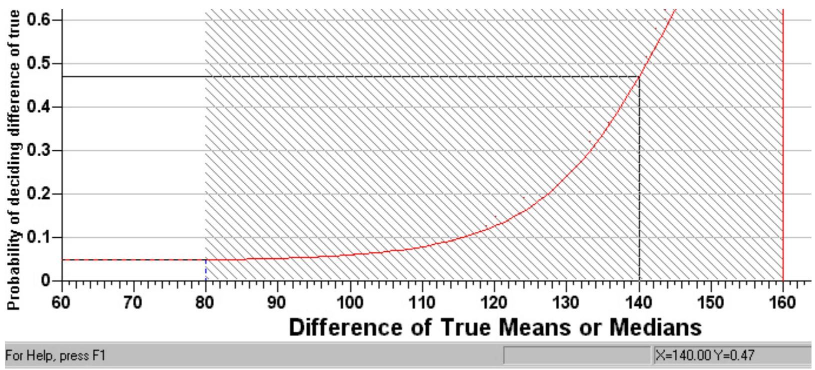

Figure 4.14. Cursor Location Line for MARSSIM WRS Test

4.4 Display of the Report

VSP provides a report for each sampling design calculation. We can see this display by selecting Main Menu option **View** ÷ **Report**. The report lists the type of sampling--random, systematic, or hot spot--used in the current sampling design. The area of the decision unit, sample sizes, and total sampling costs are also given. An example report is shown in Figure 4.15. Each time VSP calculates a new sample size, or adds points to an existing design, a new report will be generated. In future VSP releases, this report section will be improved significantly and will output a document that can be placed within a Quality Assurance Project Plan section.

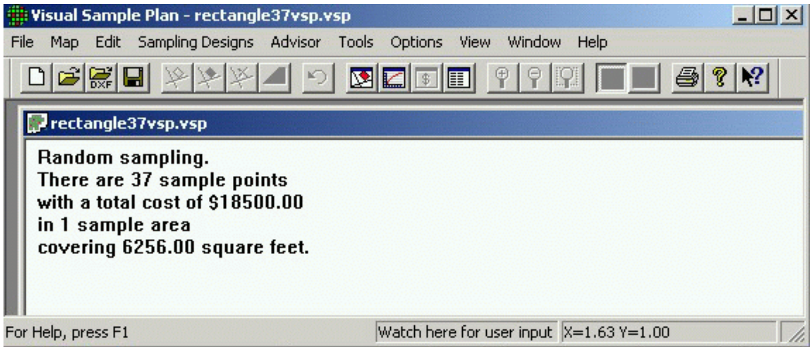

Figure 4.15. Report for Comparing Mean vs. Action Level

4.5 Display of Coordinates

The final type of display in VSP is the list of coordinates for each sample point on the map. We can see this display by selecting Main Menu option **View ÷ Coordinates**. The x and y coordinates are displayed for each sample point. 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 4.16 is an example of the Coordinates view.

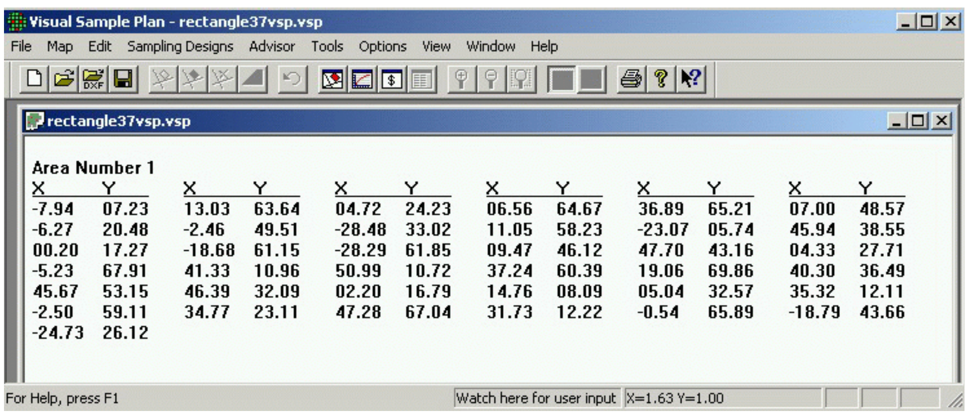

Figure 4.16. Coordinates Display of Sampling Locations

4.6 Multiple Displays

Multiple displays can be brought up on the same screen. Under **Window** from the Main Menu, select **Double Window** to see the Map and the Graph together on the same screen; select **Triple Window** to see the Map, Graph, and Report together on the same screen; and select **Quad Window** to see the Map, Graph, Report, and Coordinates together on the same screen. Figure 4.17 shows the results of the **Quad Window** option.

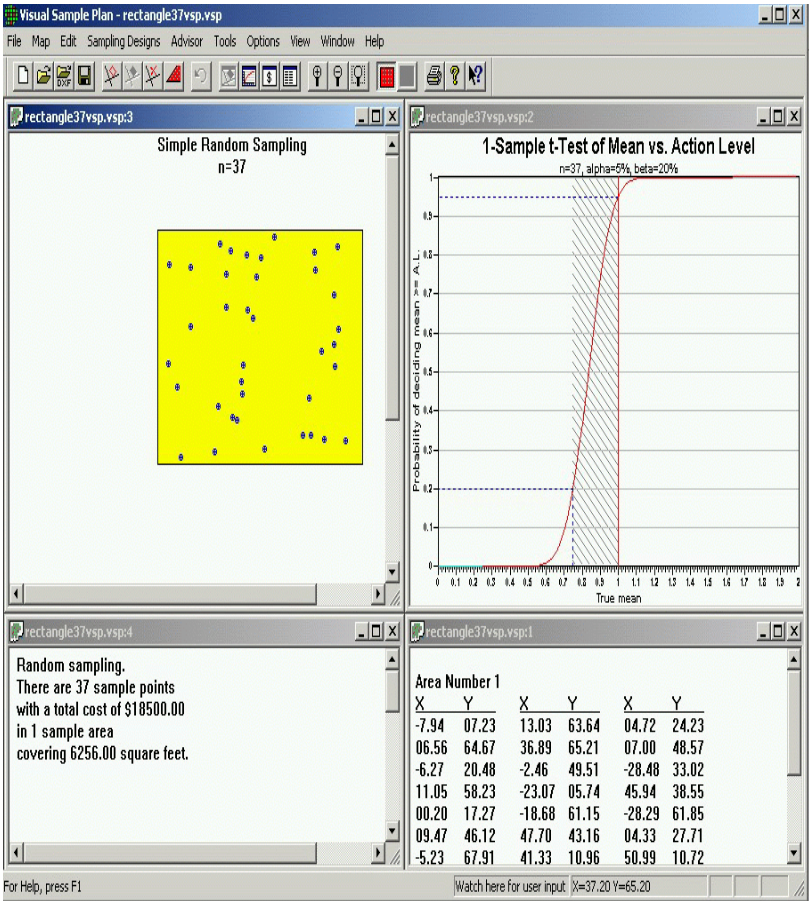

Figure 4.17. Quad Display of Map, Graph, Report, and Coordinates on Same Screen

To summarize, in Figure 4.18 we show the selection of a **Sampling Design** by specifying the Problem Type for our example problem, we have input the **DQO inputs** into the dialog box for that Problem Type, we have **Applied** the design to our **Study Area**, and displayed the Map, Graph, Report, and Coordinates simultaneously using the **Quad Window** from the Windows menu.

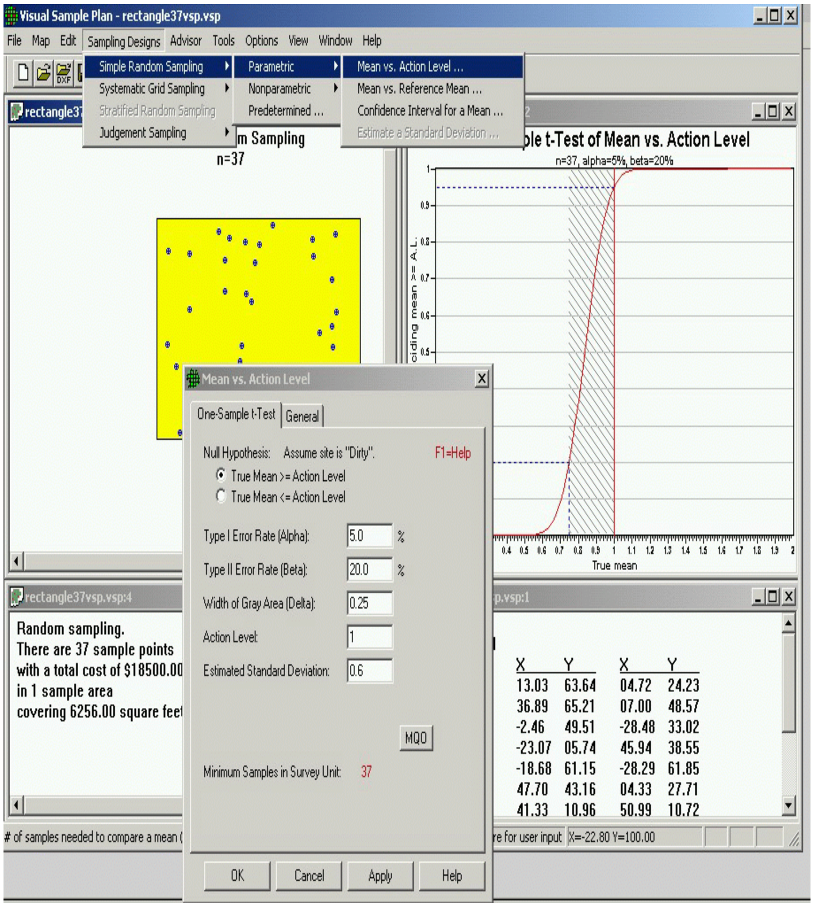

Figure 4.18. Combined Display of VSP Inputs and Outputs

5.0 Extended Features of VSP

VSP has many extended features that have not been described so far in this user's guide. In Section 5.0, 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 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.

5.1 Multiple Areas To Be Sampled

If you want to sample multiple areas, VSP allows you to select multiple areas as sampling areas. 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 5.1. We drew two sample areas, the rectangle and the ellipse. We next assumed that a sampling-design algorithm not currently in VSP called for $n = 25$ samples. Using option **Sampling Designs ÷ Simple Random Sampling ÷ Predetermined**, VSP allocated 7 of the 25 requested samples to the rectangle and 18 to the ellipse. This is because the ellipse covers an area approximately 2.5 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.

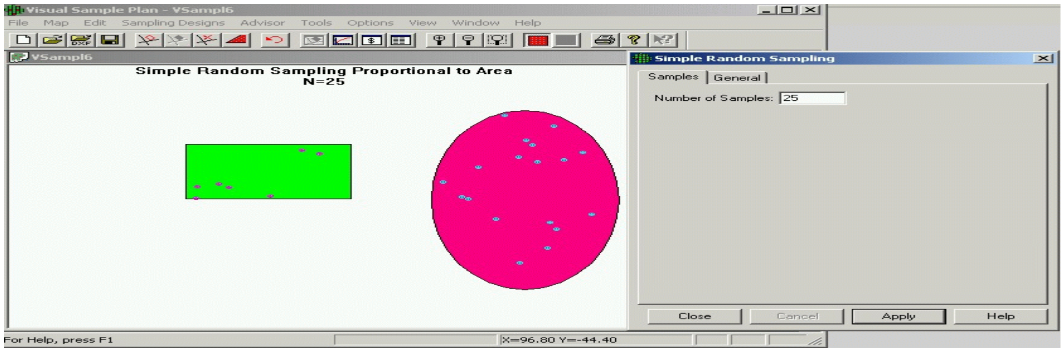

Figure 5.1. Proportional Allocation of Samples to Multiple Sample Areas

5.2 Largest Unsampled Spot

If VSP has generated a sampling design for a Study 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 Study Area without overlapping a sample point. To find this area, use Main Menu option **Tools ÷ Largest Unsampled Spot ÷ Find**. This is shown in Figure 5.2.

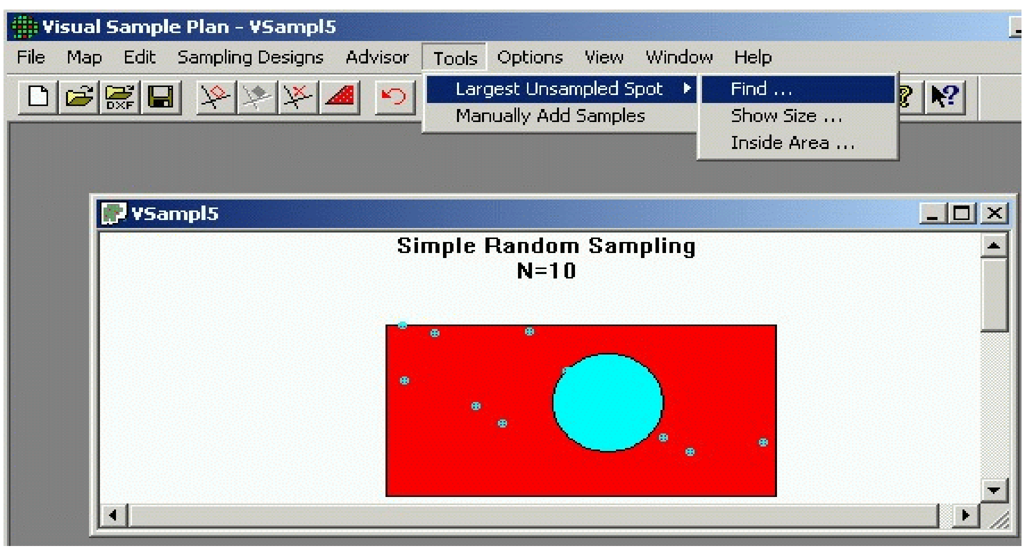

Figure 5.2. **Largest Unsampled Spot** Displayed on Rectangular Study Area

A dialog box, displayed in Figure 5.3, gives you the option of specifying the accuracy of the circle's radius, whether you want to consider area corners as additional sample points, and whether to allow the spot to overlap the Study Area.

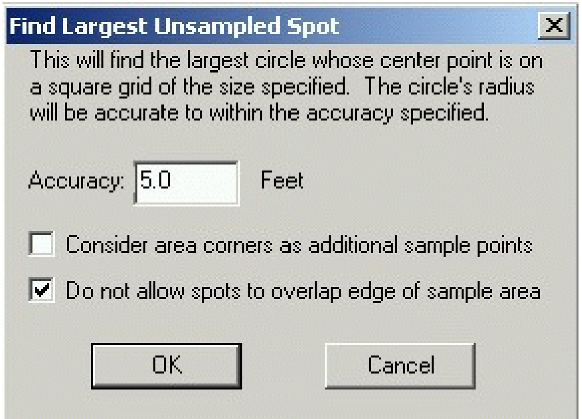

Figure 5.3. Dialog Box for **Largest Unsampled Spot** Allowing Accuracy and Other Options To Be Specified

Two other displays are available: **Show Size...** and **Inside Area....** The **Show Size...** dialog box is shown in Figure 5.4. It indicates that the radius of the circle is 12.51 feet. The total area of the circle is 492.02 square feet.

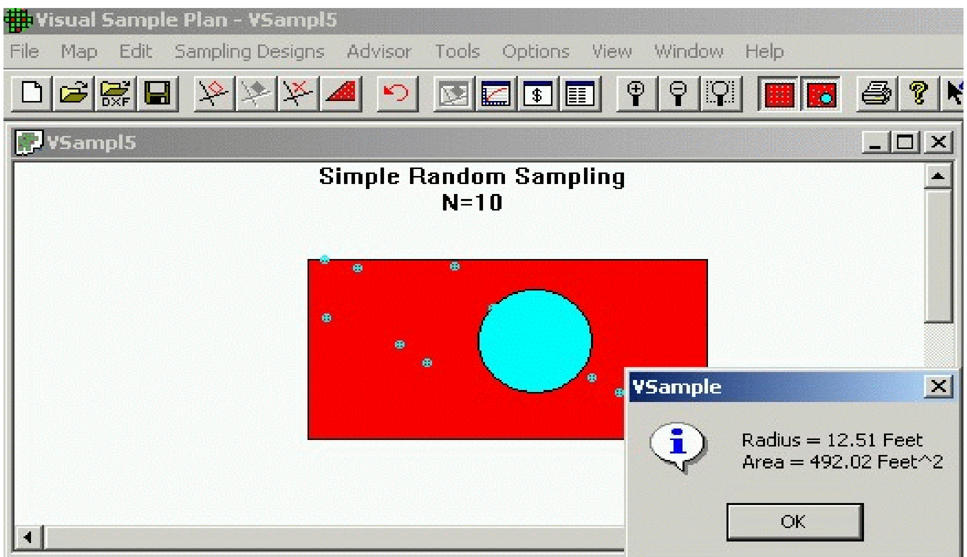

Figure 5.4. Dialog Box for **Largest Unsampled Spot** Showing Size of Circle That Would Fit into Largest Unsampled Area

The other display is the **Inside Area...** dialog box, shown in Figure 5.5. It indicates that 100% of the circle is within the Study Area. If you choose to allow the largest unsampled spot to overlap the Study Area edges when using the **Tools ÷ Largest Unsampled Spot ÷ Find** option, there will be situations in which the circle extends beyond the boundary of the Study Area.

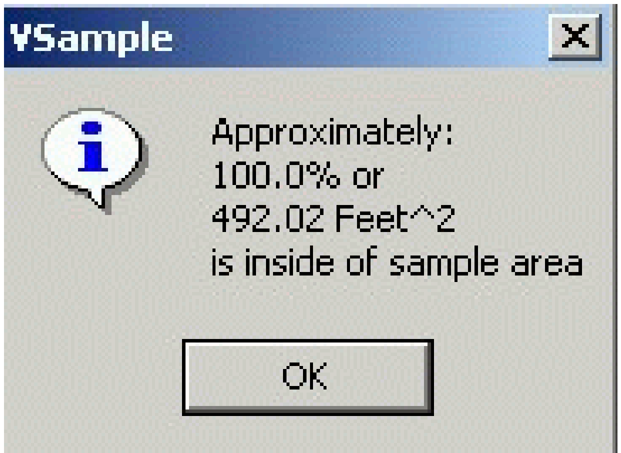

Figure 5.5. Dialog Box for Largest Unsampled Spot Showing Percentage of Circle Within Study Area

5.3 Pseudo-Random and Quasi-Random Sampling

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.

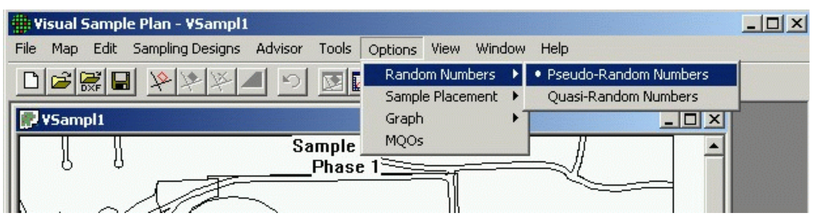

Figure 5.6. Menu for Selecting Type of Random Number Generator

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.

Sampling locations chosen with a quasi-random number generator are restricted as to proximity. The sampling locations are chosen to avoid previously selected locations in the *current* sampling design. Conceptually, one can think of electrons in a box. At any moment, each electron is in a random location but they all are avoiding other electrons because of like negative charges.

Quasi-random numbers are generated in pairs. The sequence of paired numbers is generated in such a way that sample points are 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 *Visual Sample Plan (VSP) Models and Code Verification* (Gilbert et al. 2001).

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 5.6 to handle the problem of avoiding existing sampling locations.

5.4 Measurement Quality Objectives Module

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.

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 *Visual Sample Plan (VSP) Models and Code Verification* (Gilbert et al. 2001).

The MQO option is selected from the screen that pops up after a Sampling Design has been selected. In Figure 5.7, we see the dialog box that appears on the screen after pressing the MQO button. 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.

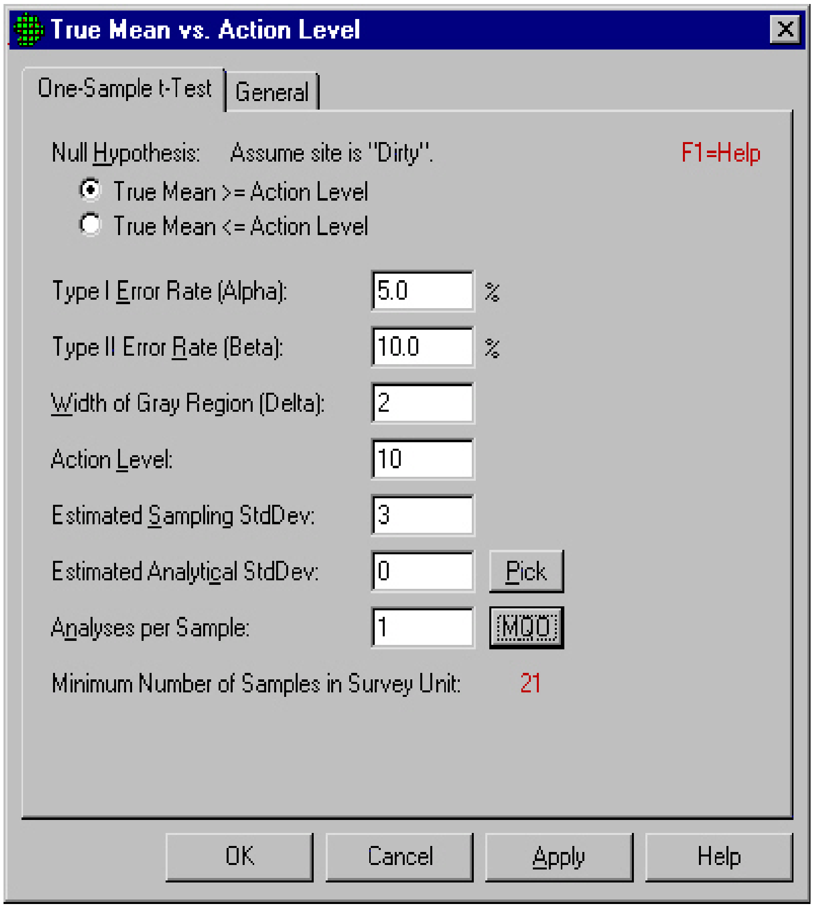

Figure 5.7. **MQO Input Dialog Box** with Default Values Displayed

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 5.7 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 5.8

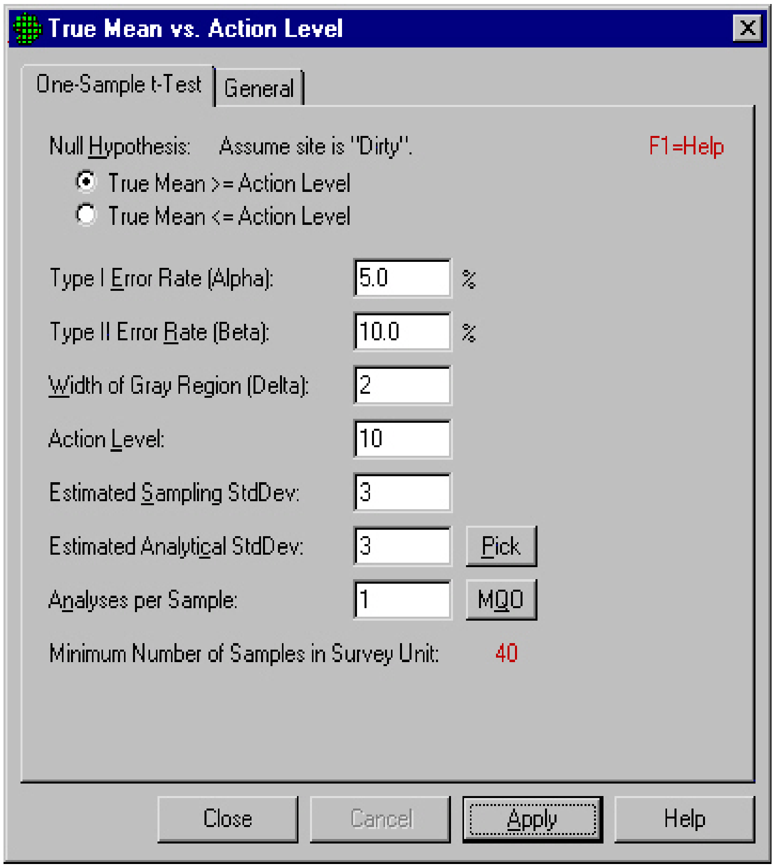

Figure 5.8. MQO Input Dialog Box Showing Positive Value for **Estimated Analytical Standard Deviation** with **1 Analysis per Sample**

If we take two repeated measurements of each sample (**Analyses per Sample** set to **2**), we see in Figure 5.9 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.

When you select the **General** tab at the top of the screen, a new display and set of inputs is shown. This is shown in Figure 5.10. 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

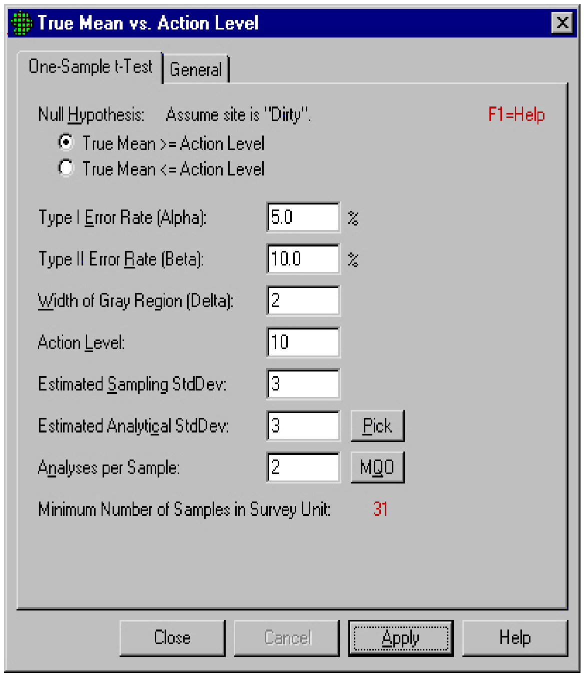

Figure 5.9. MQO Input Dialog Showing Positive Value for **Estimated Analytical Standard Deviation** with **Multiple Analyses per Sample**

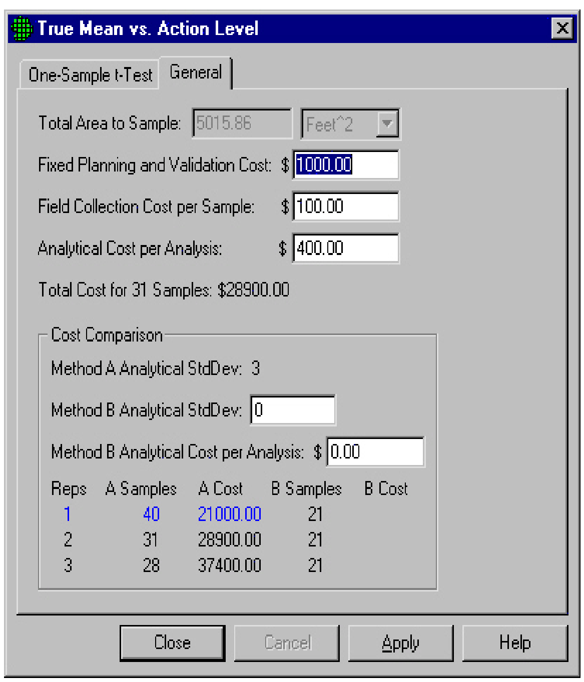

Figure 5.10. Cost Input Dialog Box for MQO Option

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 5.10. 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.

Next we show input values for Method B. Here, we enter a **Method B Analytical Standard Deviation** of **4** (somewhat higher than Method A), but with a lower **Cost per Sample** (shown here as $100). In Figure 5.11 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.

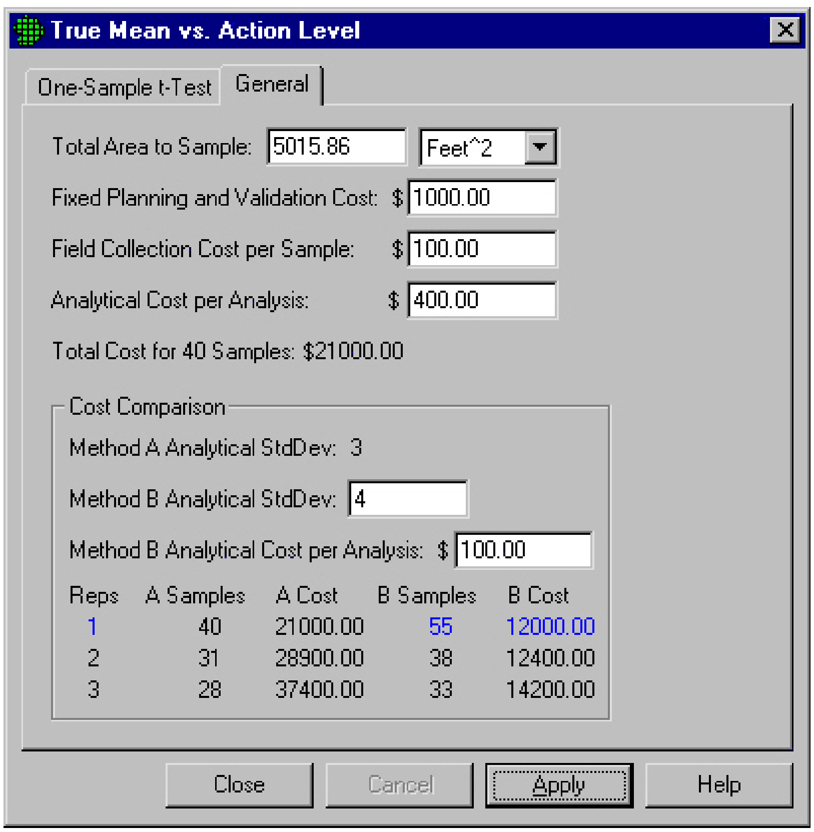

Figure 5.11. Display of **Cost Comparison** for Method A and Method B from MQO Module

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. Figure 5.12 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.

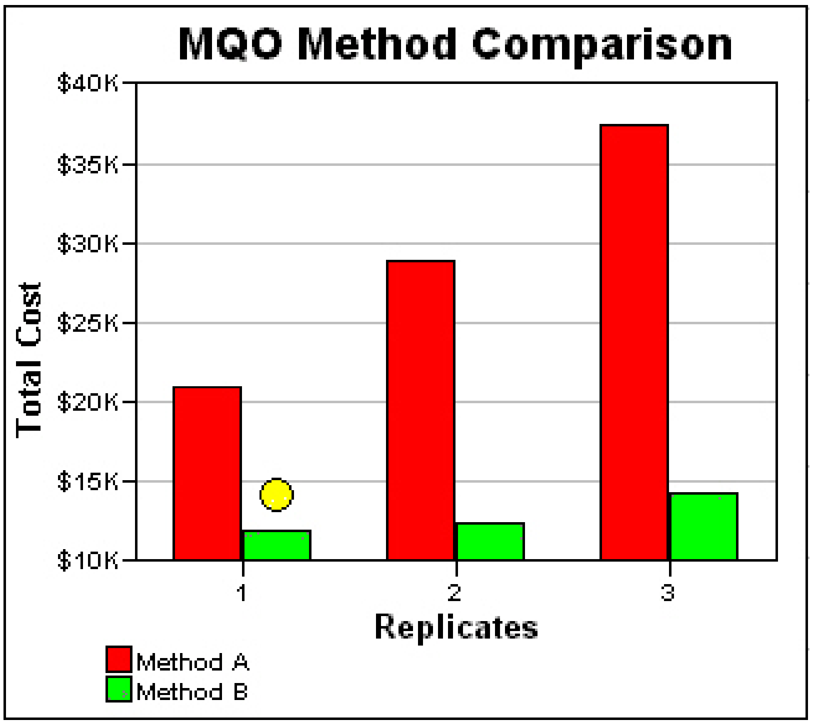

Figure 5.12. MQO Method Comparison Chart

5.5 Judgment Sampling/Manually Adding Samples

VSP allows you to manually add sampling locations to the Study Area. This is not a recommended way to sample because it injects the bias of the sampler into the project, with the result being the probability structure that underlies probability sampling is lost. However, if you have justification for the intervention, VSP allows the option.

You may manually add sampling locations in either of two equivalent ways: the **Sampling Designs ÷ Judgment Sampling ÷ Manually Add Samples** option, or the **Tools ÷ Manually Add Samples** option. In order to do either one, a Study Area must first be selected. In Figure 5.13, we selected as our Study Area the large elliptical area in the Millsite.dxf file. After selecting **Tools ÷ Manually Add Samples**, we move the mouse over the Study Area; the mouse cursor becomes a cross-hair. If we left-click the mouse, we see a small "x" appear on the map where we verify that a sample should be taken. We can add as many samples as we choose in this way. In Figure 5.13, we added 3 samples manually.

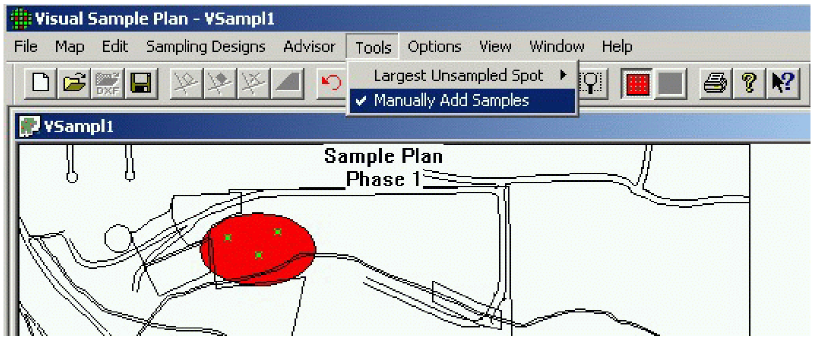

Figure 5.13. **Judgment Sampling** with Three Sampling Locations Added Manually

Note that when **Judgment Sampling** is used, no graph of the performance can be displayed. Displaying the graph will show a "**No Graph**" title. The report also tells us that no performance goal diagram is available for this user-supplied design. This is because VSP has no way of assessing the performance (or decision error limits) of a design where the number and placement of samples is arbitrary and not grounded in statistical sampling theory.

5.6 Adaptive-Fill Option for Placement of Sample Points

The **Adaptive-Fille** 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 Study 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 5.3, are placed in the Study 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 Study Area using an algorithm to maximally avoid pre-existing sampling locations. The Adaptive-Fill algorithm can be used with either random number generator. The **Adaptive-Fill** option is shown in Figure 5.14.

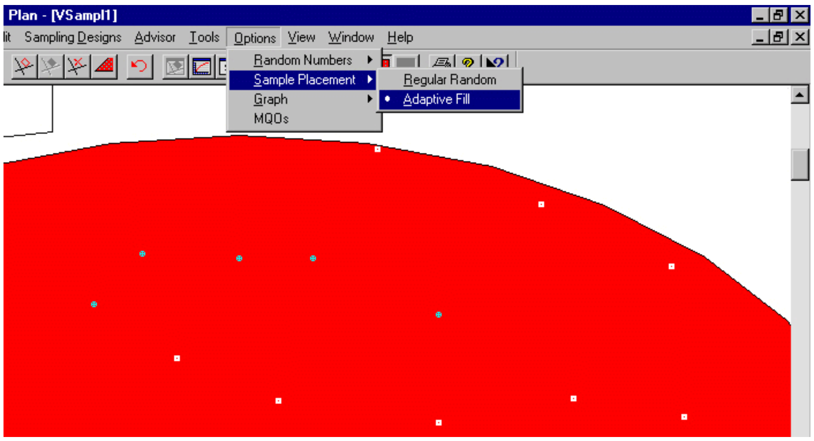

Figure 5.14. **Adaptive-Fill** Option for Sample Placement
(Shown Here with Study Area from Millsite Map)

Note that in Figure 5.14 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** message 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 only displayed in the **Sample Information** window and the exported text file of sampling locations (see Figure 5.16). See Figure 5.15 for an example of a message for an Adaptive-Fill sampling location.

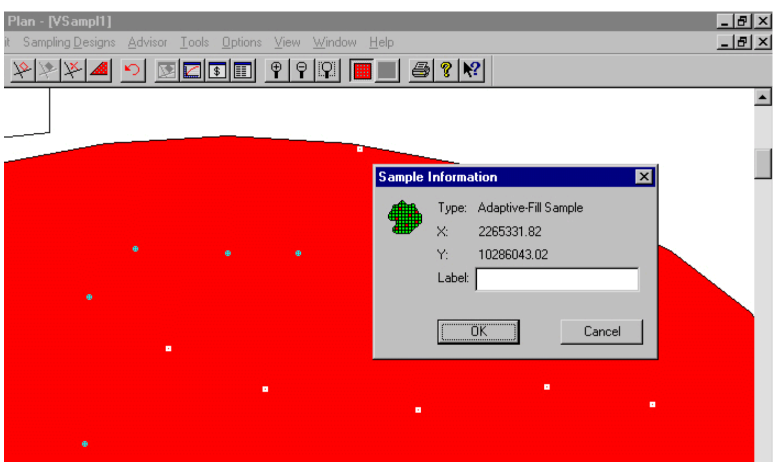

Figure 5.15. **Sample Information** Window Displayed When the User Right-Clicks on Selected Sample Points on Map

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 5.16.

As VSP is developed further, additional expanded features are planned.

```
2265358.48  10285484.16 Random
2264994.53  10285772.97 Random
2265495.23  10285746.35 Random
2264959.10  10285877.01 Random
2264998.96  10285465.40 Adaptive-Fill
2265557.41  10285973.36 Adaptive-Fill
2265262.83  10285752.67 Adaptive-Fill  Example Label
2265715.07  10285790.78 Adaptive-Fill
2264985.31  10285625.38 Adaptive-Fill
2265166.99  10285424.58 Adaptive-Fill
2264824.44  10285914.72 Adaptive-Fill
```

Figure 5.16. Sample Exported Text File of Sampling Locations

6.0 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. R. Davidson, J. E. Wilson, and B. A. Pulsipher. 2001. *Visual Sample Plan (VSP) Models and Code Verification*. PNNL-13450, 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. 2000c. *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.