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

Visual Sample Plan
Version 2.0
User's Guide

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

September 2002

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights**. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Visual Sample Plan
Version 2.0
User's Guide

N. L. Hassig   R. F. O'Brien

J. E. Wilson   B. A. Pulsipher

R. O. Gilbert   C. A. McKinstry

D. K. Carlson   D. J. Bates

September 2002

Acknowledgments

Abstract

This user's guide describes Visual Sample Plan (VSP) Version 2.0 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 2.0 provides sample-size equations or algorithms needed by specific statistical tests appropriate for specific environmental sampling objectives. The easy-to-use program is highly visual and graphic. VSP runs on personal computers with Microsoft Windows operating systems (95, 98, NT, 2000, Millennium Edition, and XP). Designed primarily for project managers and users without expertise in statistics, VSP is applicable to any two-dimensional geographical population to be sampled (e.g., surface soil, a defined layer of subsurface soil, building surfaces, water bodies, and other similar applications) for studies of environmental quality.

Contents

Figures

xi

Tables

1.0 Introduction

1.1 What is Visual Sample Plan?

Visual Sample Plan (VSP) is a software tool for selecting the right number and location of environmental samples so that the results of statistical tests performed on the data collected via the sampling plan have the required confidence for decision making. More than 1000 copies of VSP 1.0 are in distribution today; sponsors of this public domain software include the U.S. Environmental Protection Agency (EPA), Department of Energy, and Department of the Navy. VSP 2.0 is a major new release of the software and incorporates many new features.

VSP provides sample designs and sample-size equations needed by specific statistical tests appropriate for several types of environmental problems. Table 1.1 is a list of the sampling goals that can be addressed in VSP 2.0.

Table 1.1. List of Sampling Goals

Sampling Goal	User's Guide Section
Compare Average to Fixed Threshold	3.2.1
Compare Average to Reference Average	3.22
Estimate the Mean	3.23
Construct Confidence Interval on Mean	3.24
Compare Proportion to Fixed Threshold	3.25
Compare Proportion to Reference Proportion	3.2.6
Estimate the Proportion	3.2.7
Locating a Hot Spot	3.2.8

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

VSP is designed primarily for project managers and users who are not statistical experts, although those individuals with statistical expertise also will find the code very useful. VSP is applicable to any two-dimensional geographical population to be sampled, including surface soil, a defined layer of subsurface soil, building surfaces, water bodies, or other similar applications.

1.2 What's New in VSP 2.0?

VSP 2.0 offers expanded sampling designs—sequential, adaptive, ranked-set, swath sampling. It includes an **Advisor** menu selection to guide the user through the steps in designing a sampling plan. It has an expanded set of **Sampling Goals**, along with the appropriate statistical test options for use in the sample size formulas. It includes a **Report View** that describes in words the sampling plan designed by VSP, along with the assumptions and formulas used in sample size calculations. VSP 2.0 allows the user to format graphs, reports, and maps. It allows labeling of samples by attribute, input and tracking of historical samples, and expanded graphics on maps. Many operations are streamlined, like a "one-click" creation of **Sample Area** and an initial screen for selecting which version of VSP to use (different sponsors have requested stylized versions of VSP).

The experienced user will notice many other changes and enhancements. The expanded **Help** function provides an exhaustive list of formulas, definitions, and references used within VSP. Users are encouraged to liberally use the **Help** function as they navigate the expanded functionality within VSP 2.0. Help is available on a topic basis when accessed from the **Help** menu item on the main screen, and on a tutorial basis when accessed from the **Help** button at the bottom of the input screen for each of the sampling designs.

The next beta test reload of VSP will include a new Sampling Design goal: Find an Unexploded Ordnance (UXO) Target Area. This will add sampling along a parallel or grid swath and will provide post-survey target detection evaluation.

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

1.3 Installation and System Requirements

VSP 2.0 runs on Microsoft®Windows® 95 and later… *Visual Sample Plan (VSP) Models and Code Verification* (Gilbert et al. 2001) documents the successful installation of VSP on personal computers operating with Windows® 95, 98, Millennium Edition, and Windows 2000.

VSP 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".

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

New versions of VSP are often released as prototypes for testing. These demonstration (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 2.0 is not a demonstration version and does not have an expiration date.

1.4 Overview of VSP

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

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

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

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

The locations of the samples over the Sample Area are determined by the specific sampling design (pattern) that you select. For some Sampling Goals, and for some assumptions about the population, only certain designs are allowed from a statistical theory perspective. For example, sequential sampling is
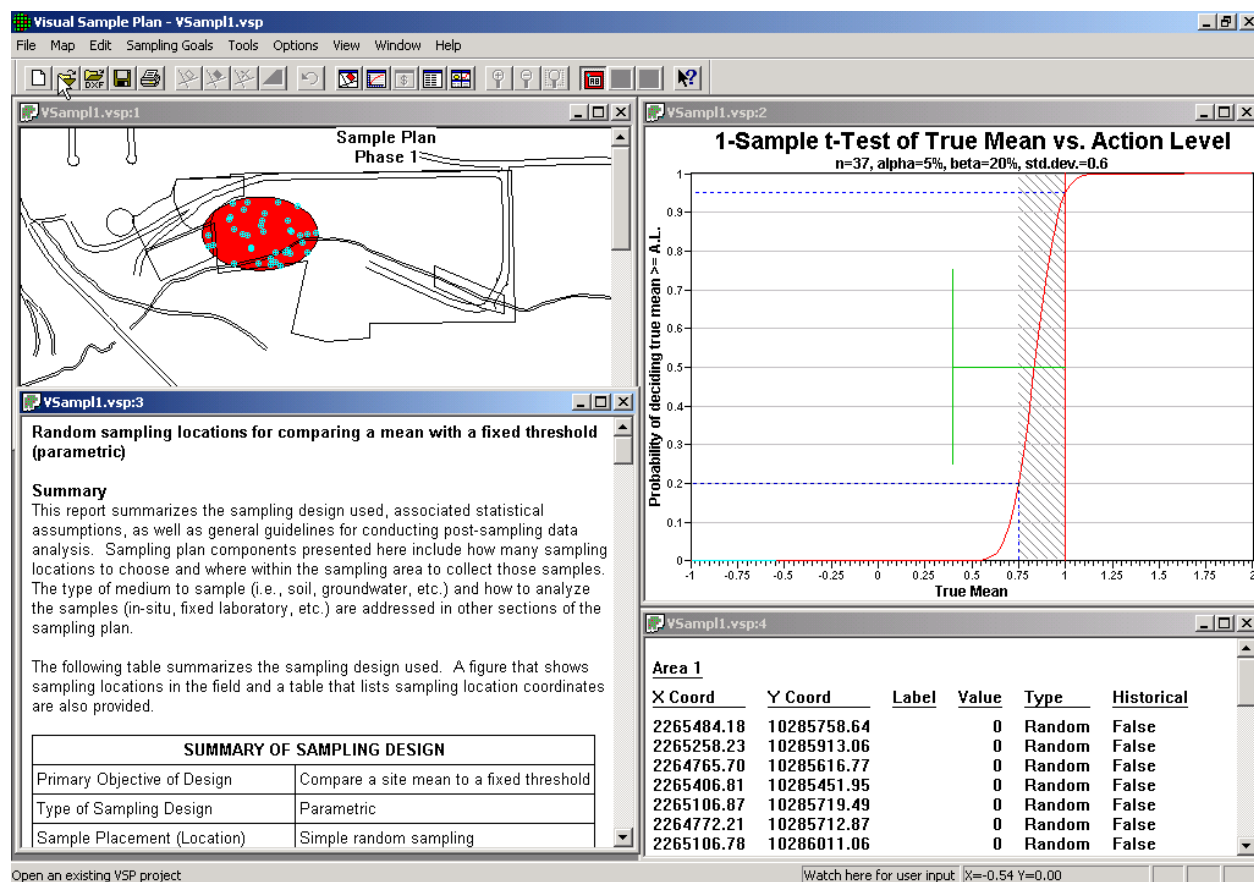
Figure 1.1. Screen Capture from VSP Using **Quad Window** Option

appropriate only for the sampling goal of **Compare Average to a Fixed Threshold** when the population units can be assumed to be distributed normally. When there is a choice of designs, VSP displays a drop-down menu of the allowable designs. VSP 2.0 has an expanded set of sampling designs from which to select, including adaptive sampling, sequential sampling, and ranked-set sampling.

VSP 2.0 allows the user to consider previously sampled locations (called historical samples) when deciding where to take new samples. Historical samples can be imported and located on the map. Samples can be given text labels and displayed on the map using a variety of symbol shapes. Samples can be grouped and visually identified by using different symbols for each sample group (e.g., samples taken at different times, or using different collection methods). Designed as a planning tool, VSP does not keep track of sample values except in the case of sequential sampling (see Section XX).

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.

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

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

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

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

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

2.0 Mechanics of Running VSP

2.1 Getting Started and Navigational Aids

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

The choice of VSP versions is offered because different versions of VSP have been developed for different sponsors. Versions were designed to simplify the options presented to the single-purpose user as VSP became more complex. For example, users interested only in MARSSIM applications can select the

Figure 2.1. VSP Welcome Screen with Version Selection Menu

MARSSIM version. That version contains menu items relating to only rooms and surfaces, and its statistical tests and sampling design options are limited to only those that are MARSSIM-approved. Currently, all versions implement the full range of VSP functionality except **EPA G-5S VSP Implementation Version**, which implements only those sampling designs discussed in *Guidance for Choosing a Sampling Design for Environmental Data Collection* (EPA 2001). Future VSP releases will implement this version control function.

Once a version has been selected, the second navigational menu pops up: **Need Help?** This Help screen appears when VSP is first launched and again whenever **Advisor > Show Advisor** is selected from the main menu (Figure 2.2). Clicking any of the items under **Need Help?** will launch Microsoft Word, where a document will load that has the relevant information on that topic from the User's Manual. Once in Word, you must click the Close button ("X") in the upper-most right-hand corner of the Word toolbar to return to VSP. You can close the **Need Help?** screen by either clicking in the "X" button on the top bar or selecting **Close** at the bottom of the screen.
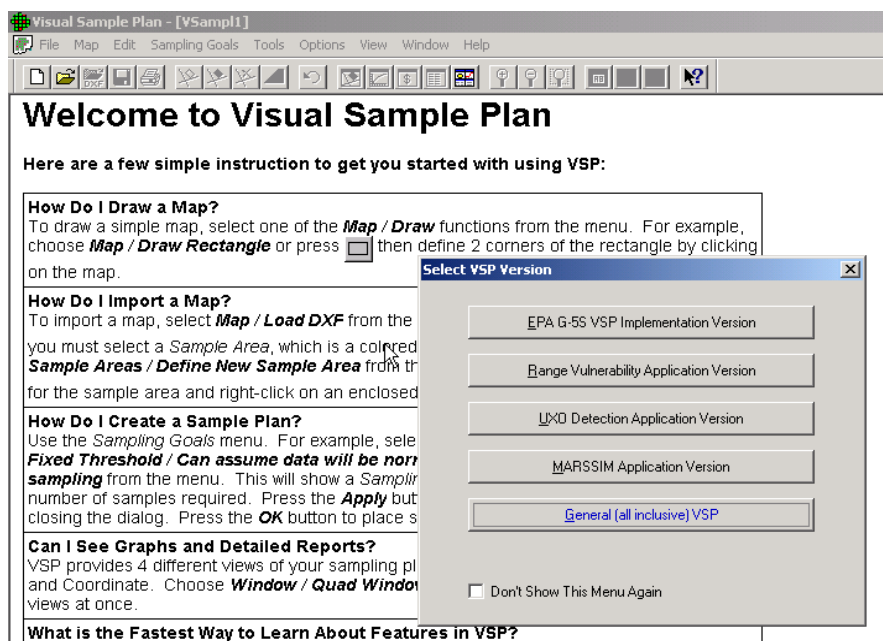
Figure 2.2. VSP Welcome Screen with **VSP Advisor**

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

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

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

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

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

2.2 Setting Up a Map

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

Figure 2.4. Pull-Down Menu Items Under **File**

1. Import the site map from a drawing interchange format (DXF) file.

2. Import the site map from a previous VSP project that was saved in VSP format (i.e., a .VSP file).

3. Draw the map or Sample Area using VSP's drawing tools.

These three methods are illustrated below. Because VSP has a graphical user interface, the user may specify the Map Extents to make the coordinates on the VSP map correspond to the actual coordinates and distances at the physical site being replicated. This is done by selecting **Map > Set Map Extents** from the Main Menu.

2.2.1 Importing a Site Map from a Drawing Interchange Format File

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

1. From the main menu, select **Map > Load DXF**. A quick alternative is to click on the **Load Map** button on the VSP toolbar.

2. A list of available DXF files in the Visual Sample Plan folder is displayed. Select Millsite.dxf. You may change folder names to search for the desired file using standard Microsoft file-searching if the file is not in the Visual Sample Plan folder. Double-click on Millsite.dxf.

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

Figure 2.5. The Millsite.dxf File Opened in VSP

It is possible that your DXF file contains 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.

2.2.2 Importing a Site Map File in the VSP Format

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

2.2.3 Importing Site Maps of Other Formats

Currently VSP imports only files in the DXF and VSP formats. Any other file types (such as .DWG) must be converted to a DXF file, which can be done by software such as AutoCAD. In the future, we hope to allow import of SHP formatted shape files, the format used by many geographical information system (GIS) software packages.

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

- Create a new project by choosing **File > New Project** on the Main Menu or by clicking the **New** button on the main toolbar. To dismiss the "Welcome to Visual Sample Plan" displayed upon opening a new project, set the Map Extents (**Map > Set Map Extents**). Click the **OK** button to use the VSP default values. You can now start drawing in the project window. Expand the project window by pressing the **Maximize** button on the upper right corner of the project window.

- Choose **View > Map Drawing Toolbar** from the Main Menu. This displays a toolbar used specifically for drawing a map. This toolbar also may be docked if you prefer to remove it from the project window. To dock the drawing toolbar, place the mouse cursor on the blue title bar and drag the drawing toolbar onto the VSP toolbar.

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

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

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 tooltip box 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 2.6 is an example with the ceiling chosen. The status bar at the bottom of the VSP screen displays the dimensions as you enter them.

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 2.7.

Figure 2.6. Example of MARSSIM Room with 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.
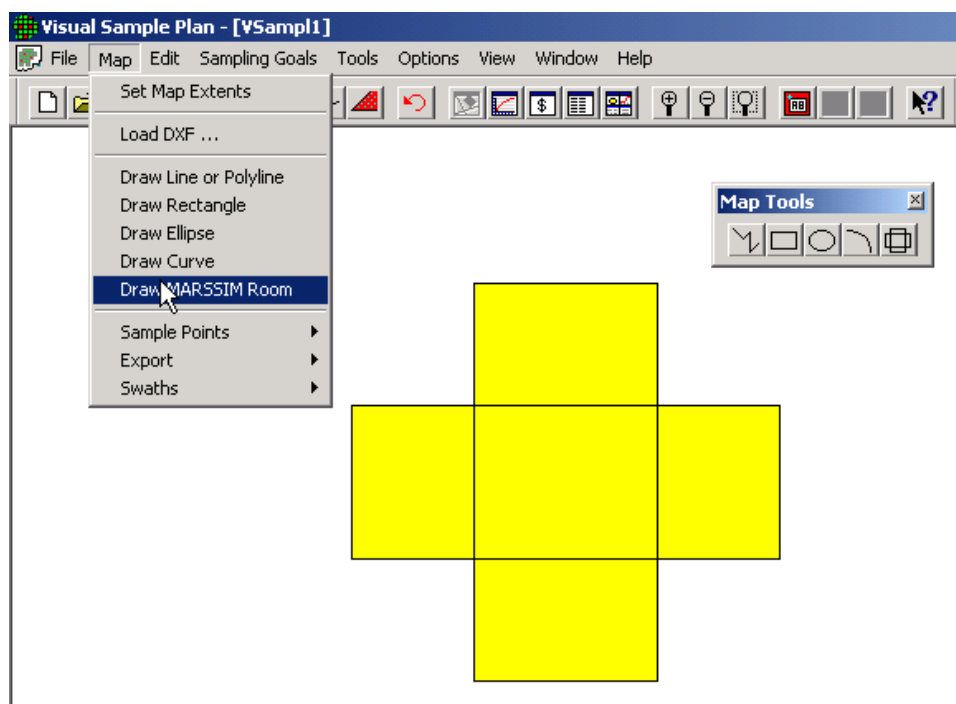
Figure 2.7. Example of MARSSIM Room Without Ceiling Option Selected

2.2.4 Working with Maps

Deleting Segments of a Map. VSP imports DXF files and turns the objects into polylines or a series of connected points. As such, if you want to remove a segment from either an imported map or a user-drawn map, you may click on a segment and hit the Delete key on your keyboard. Right-clicking on any segment in a map displays the vertices of the polyline in a outline of bold squares. With the outline in bold squares displayed, hit the Delete key on your keyboard and that segment is removed.

Zooming In and Out. The **Zoom In**, **Zoom Out**, and **Zoom Window** buttons in the middle of the VSP toolbar provide methods to focus in on a Sample Area or other region of a site map. Press once on the **Zoom In** button and then click on the site map to make it grow larger. Turn off this mode by pressing the **Zoom In** button again. The **Zoom Out** button works the same way except that it makes the site map shrink. The location on the site map where you click determines the area of the new focus.

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

2.3 Sample Areas in VSP

2.3.1 Creating a Sample Area

Once a map is created, a Sample Area must be created. A Sample Area is an enclosed region in which to locate samples. The user must identify the area to VSP in order to make sampling locations available. (Note: You can use any of the sampling designs except Judgement Sampling without a Sample Area defined, but they will not create sampling locations.)

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

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

1. **One-Step Method**. Position the cursor inside one of the enclosed areas on the map and right-click with the mouse. The Sample Area is created, and a dialog box appears. This dialog box shows the size of the Sample Area and allows you to change the units of the map. Click the **OK** button on the dialog when done.
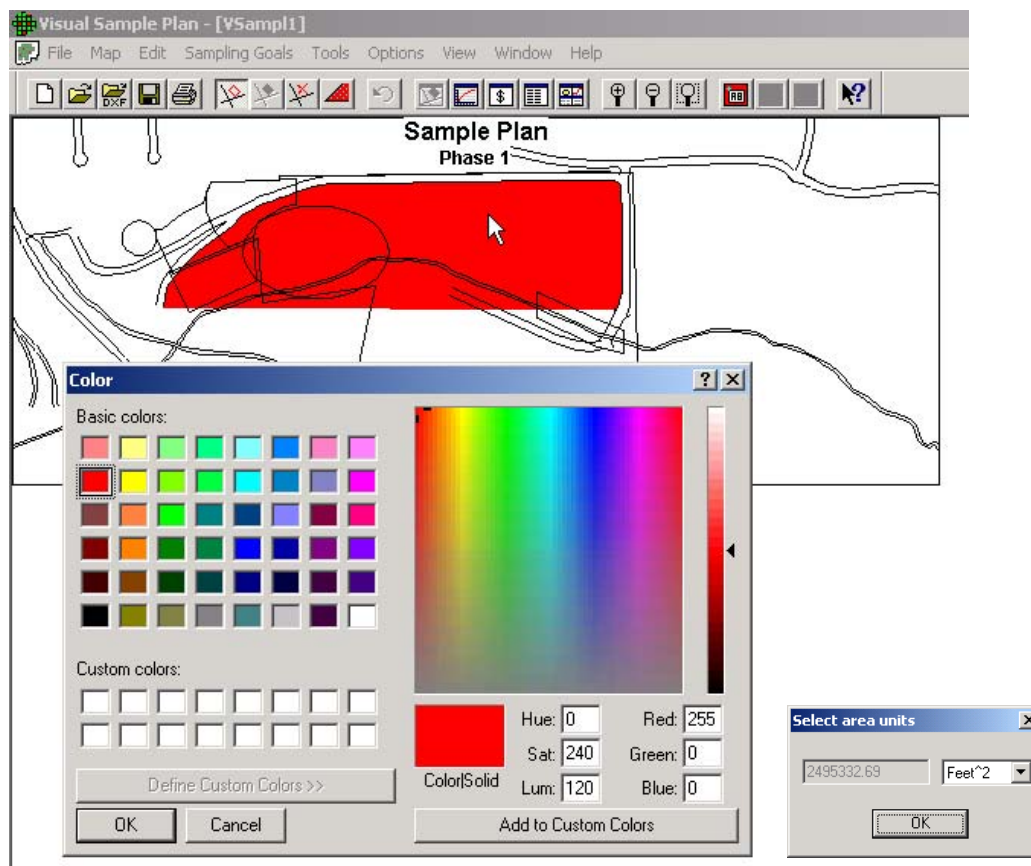
Figure 2.8. Map with a Single Sample Area

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

A map may contain a single Sample Area or multiple Sample Areas. For example, OneAcre.VSP (an example of a VSP file included with the program) is a single Sample Area, while Example1.VSP could have multiple Sample Areas because the map consists of several enclosed areas that could be selected as Sample Areas. When multiple Sample Areas are selected, samples located on the map by VSP are distributed across all the areas.

2.3.2 Selecting or Deselecting Sample Areas

VSP allows the user to control which Sample Areas are available for locating samples. Creating a Sample Area automatically "selects it" for locating samples. You know it is "selected" because it appears in a solid color on the map. "Deselected" Sample Areas appear with only the outline of the Sample Area in color and the interior blanked out. You may Select or Deselect a Sample Area in two ways: left click within the Sample Areas, or from the Main Menu select **Edit> Sample Areas > Select/Deselect Sample Areas**. The latter method brings up a dialog box that allows you to choose which areas to select or deselect. Figure 2.9 shows a VSP map with two of the five Sample Areas selected.
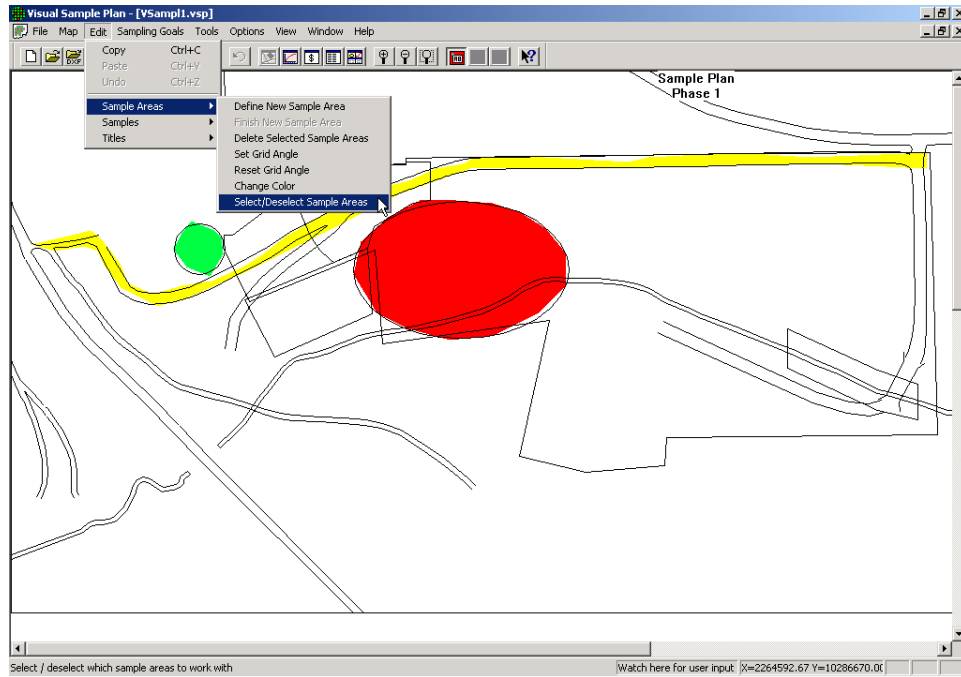
Figure 2.9. Map with Multiple Sample Areas Selected

2.3.3 Deleting Sample Areas

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

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

Individual samples have several attributes within VSP:

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

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

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

2.4.1 Importing Samples

There are two ways to import sampling locations:

1. Copy them from the Windows Clipboard. Edit the coordinates in a text editor, a word processor, or a spreadsheet. Each line (or row) represents a different sampling location. The first column is the x coordinate; the second column is the y coordinate. The third column is the sample Type and is optional. Valid sample Types are Random, Systematic, Hotspot, Manual, Adaptive-Fill, or Unknown. The fourth column is the sample label and is optional. Spaces or tabs should separate columns. (Tabs are preferable.) The coordinates must lie inside a selected Sample Area.

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

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

Now press Ctrl-A to select all the text and Ctrl-C to copy the text to the Windows Clipboard. Run VSP and load OneAcre.Vsp. Select the Main Menu option **View > Coordinates**. Paste the coordinates into VSP using either Ctrl-V or Main Menu option **Edit > Paste**. View the new sampling locations using Main Menu option **View > Map** or **Window > Quad Window**. Your map view should now look like Figure 2.10.
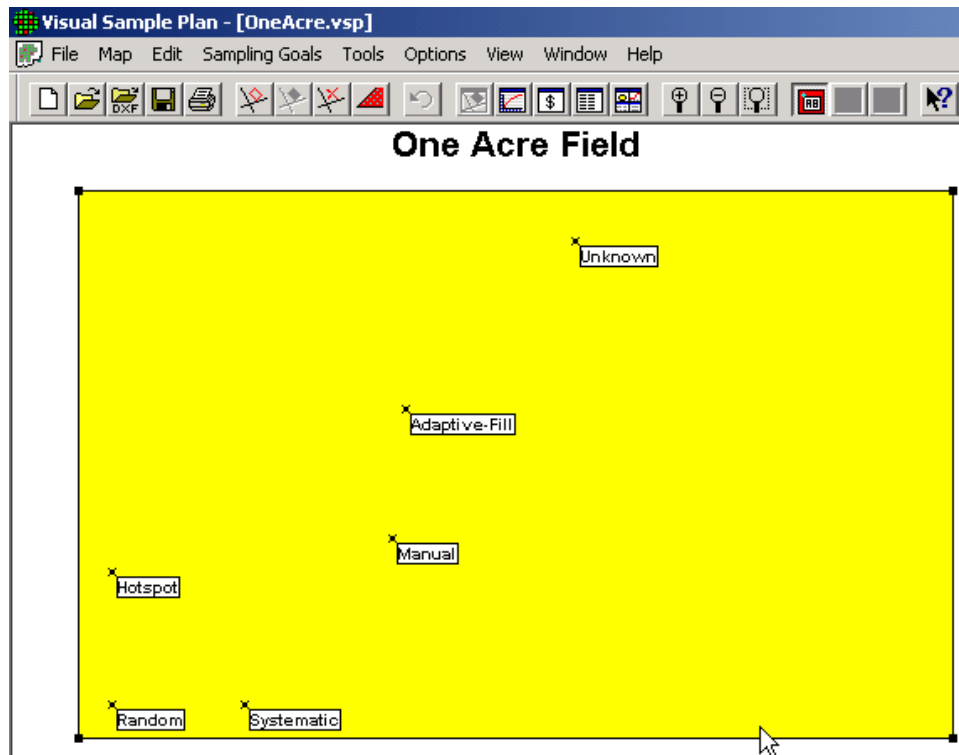
Figure 2.10. The OneAcre.VSP Project with Sampling Locations Added from Windows Clipboard

You can place the mouse on any sample point and right-click to see the attributes of the sample at that sample Location. Figure 2.11 shows the Sample Information VSP has for the sample near the arrow.

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

    Samples that are imported are assigned Shapes depending on the Type attribute assigned. Sample Type can be edited by selecting **Edit > Samples > Shapes** from the Main Menu. The Dialog box that appears shows both the shapes assigned to valid Types (use the pull-down menu to select among valid Types), and gives a picture of the Shape. Figure 2.12 shows that when a sample was collected according to a **Random** design, it will be displayed with a **Small Cross** within a **Circle** Shape. Shapes can be edited using this Dialog box for imported and manually located samples only.
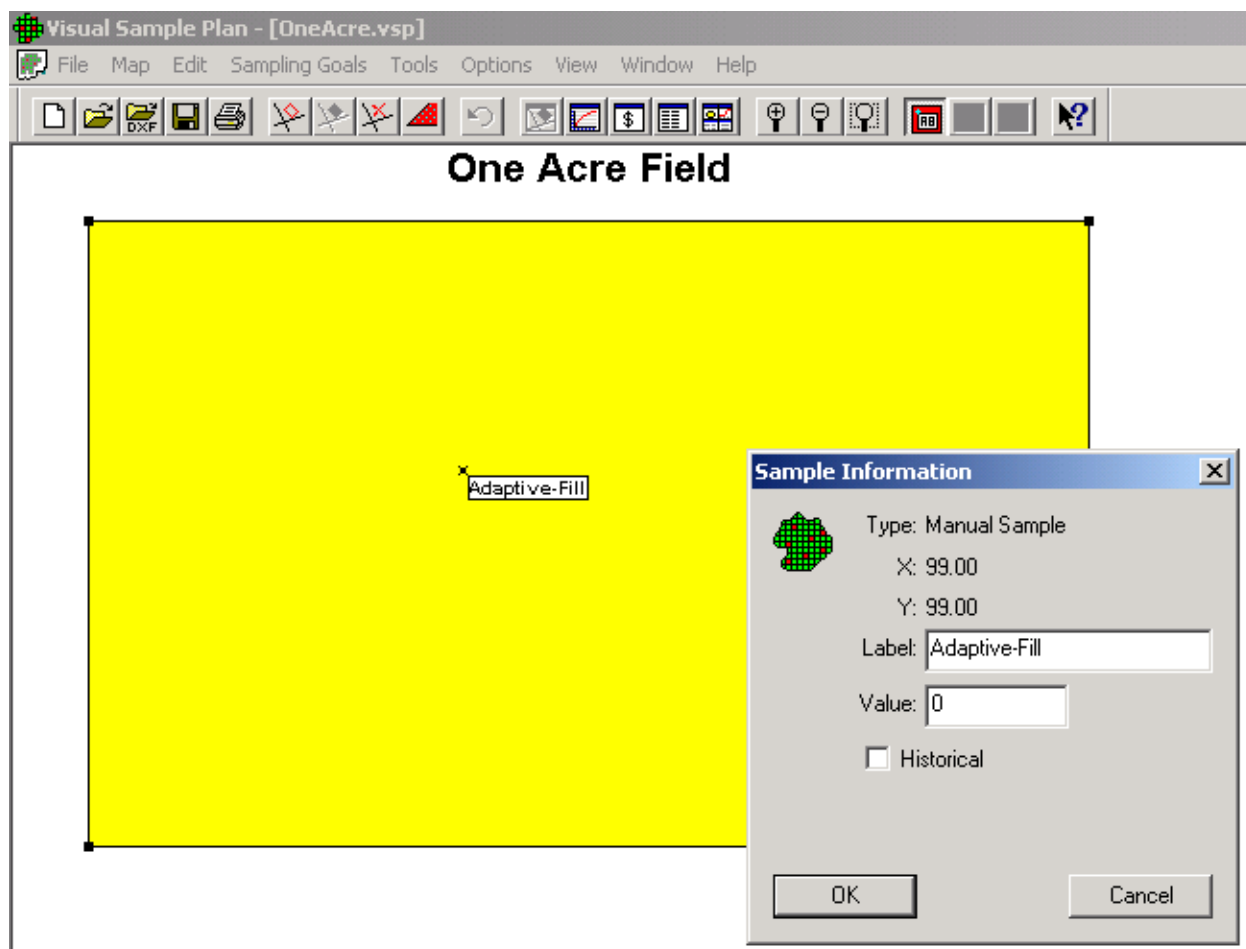
Figure 2.11. Example of Sample Information Box

2.4.2 Historical Samples

Sample locations with the Historical box checked (see Figure 2.11) have a unique role in VSP. VSP gives you "credit" for them in accounting for the total number of new samples needed. This is explained in Section 3.2. The important point to remember here is that if you import samples, manually add samples, or have a sampling design previously created within VSP, you can give specific samples a "Historical" status by placing your mouse over the sample location while in Map View and, in the dialog that comes up, checking the Historical box.

2.4.3 Exporting Sampling Locations

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

1. Select the Sample Area as described above and develop the sampling design as described in Section 3.

2. Choose Main Menu option **Map > Sample Points > Export**. Provide a name for the text file and click **Save**.

2.4.4 Removing Sampling Locations

This option is best explained with an example:

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

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

3. Click the **New Area** button on the toolbar and, after choosing a color, select the large ellipse by right-clicking inside the oval. If you accidentally get some other area, click the **Remove Areas** button and start over. Place the cursor as far from other objects as possible but still inside the ellipse.

4. Choose the Main Menu; select **Sampling Goals > Compare Average to Fixed Threshold > Data not required to be normally distributed > Simple Random Sampling (Wilcoxon signed ranks test).** Click the **Apply** button to place samples in the Sample Area. You should now have a Sample Area with 24 sampling locations similar to that shown in Figure 2.12.

5. Using the Main Menu option **Map > Sample Points > Export**, save all the sampling locations to a text file named Points.txt.

6. Now we are ready to remove some of the sampling locations. First, delete the first 16 rows (sampling locations) from file Points.txt using a text editor like Notepad. Save the remaining 5 rows to a new file named Remove.txt. These are the locations that will be *removed* from the Sample Area.
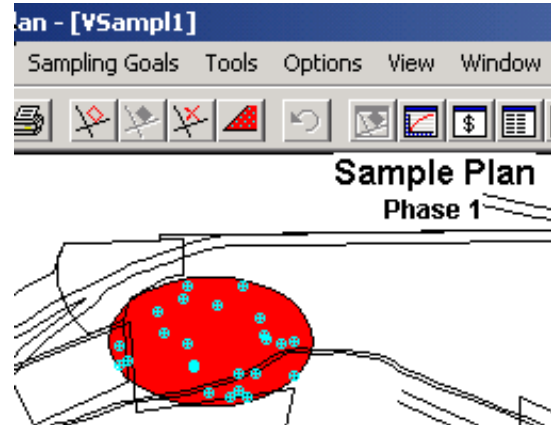
Figure 2.12. Example Sample Area with Sampling Locations

7. Finally, to remove the sampling locations listed in Remove.txt from the Sample Area, choose Main Menu option **Map > Sample Points > Remove**. Select the file Remove.txt and hit OPEN. You will see in Figure 2.13 there now are only 16 sample points instead of the original 24 shown in Figure 2.12.

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

2.5 Saving a VSP File

No matter how you imported or created a site map or Sample Area for VSP, you can always save the information in VSP's own file format. From the Main Menu, select **File > Save Project As** and provide a name for the project. VSP will add the VSP file extension automatically. Alternatively, you can use the **Save** button with the disk icon on the VSP toolbar. After you have created a sampling design as discussed later in this guide, saving your project as a VSP file also will save the input data, cost data, and recommended sample sizes.
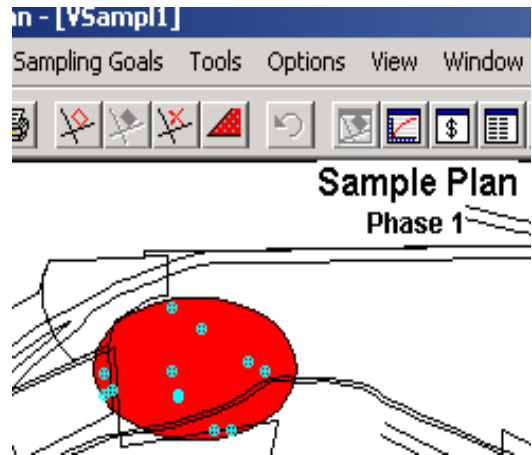
Figure 2.13. Example Study Area after Sampling

3.0 Sampling Plan Development Within VSP

3.1 Sampling Plan Type Selection

Sampling plan components consist of where to take samples, how many samples to take, what kind of samples (e.g., surface soil, air), and how to take samples and analyze them. We identified the general areas of where to take samples in Section 2.3, **Select a Sampling Area**. In this section, we discuss where *within the Sampling Area* to locate the samples. We also discuss *how many* samples to take. The kind of samples to take—i.e., soil vs. groundwater, wet vs. dry, 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 Measurement Quality Objectives module in VSP (Section 5.4) deals with how the method selected for analytically measuring the sample relates to other components of the sampling plan.

3.1.1 Defining the Purpose/Goal of Sampling

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

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 a population parameter (such as the mean or a proportion) and calculate a confidence interval.

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

VSP lists "Non-statistical sampling approach" under Sampling Goals, but this is not really a goal. Under this category, VSP allows the user to specify a predetermined sample size and locate the samples judgmentally. Because VSP has no way of knowing how the sample size and sample locations were chosen, the sampling approach is considered to be "non-statistical."

To give you an idea of how VSP threads from Sampling Goal to selection of a sampling design, Figure 3.1 shows the sequence of pull-down menus for one of the goals, **Sampling Goal of Compare Average to a Fixed Threshold**. All endpoints from the Sampling Goal main menu result in a dialog box where the user provides inputs for the specific design selected. VSP allows only certain options and
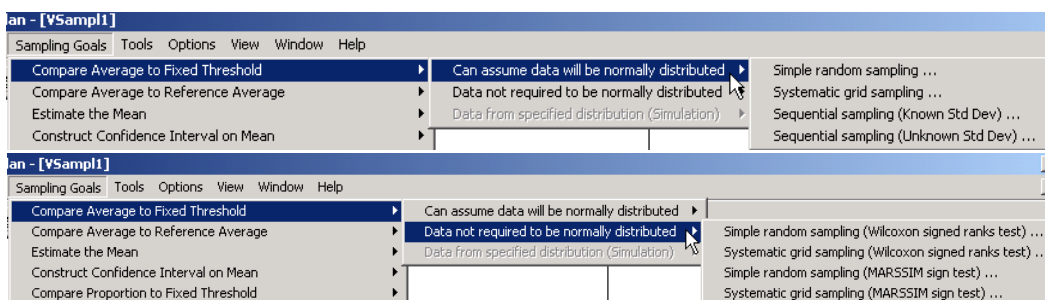
Figure 3.1. Options in VSP for Comparing an Average to a Fixed Threshold

designs (e.g., simple random, systematic) under each goal. This is because VSP contains the algorithms for calculating sample number and locating samples for only certain *goal-assumptions-statistical test or method* sequences. Future versions of VSP will expand on the number and type of algorithms offered.

3.1.2 Selecting a Sampling Design

A good discussion of how to select a sampling design is in EPA's *Guide for Choosing a Sampling Design for Environmental Data Collection* (EPA 2001). See Table 3-1 on pages 23-24 in that source for examples of problem types that one may encounter and suggestions for sampling designs that are relevant for these problem types in particular situations. Another guidance document, *Multi-Agency Radiation Survey and Site Investigation Manual* (MARSSIM) (EPA 1997), also provides insight into how to select a sample design. We highly suggest the user refer to these two documents before making a final selection on a sampling design.

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

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

We cannot discuss all the implications here, but the technical documentation for VSP (Gilbert et al. 2002) gives sample size formulas used in VSP and provides references. The online help in VSP also provides technical help and references. Finally, the reports that are available within VSP are a good source for definitions, assumptions, sample size formulas, and technical justification for the design selected.

VSP allows both probability-based designs and judgmental sampling:

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

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

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

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

  You can see an example of the grid pattern selected in the right-hand side of the dialog box in red. You may specify **Random Start** or a fixed start for the initial grid point using the check box next to **Random Start**. Choosing **Random Start** will generate a new random starting location for the first grid location each time the **Apply** button is pushed. Once all selections have been made, press **Apply**.
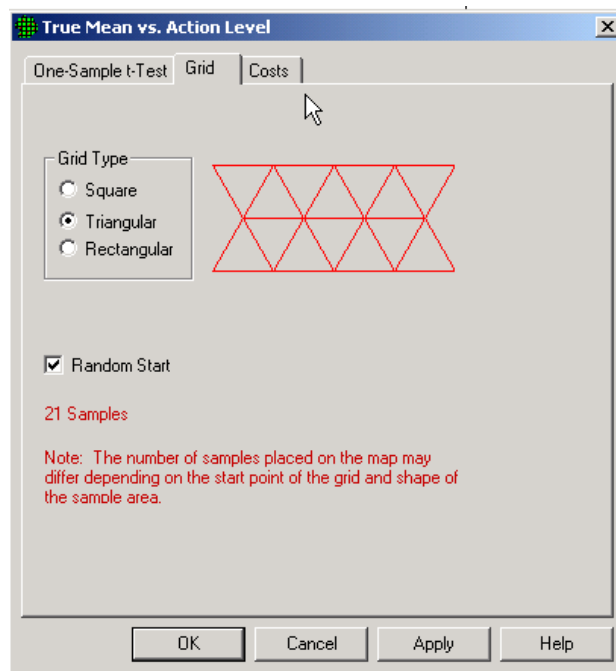
Figure 3.2. Dialog Box for Entering Type of Grid Design

- **stratified sampling** – Strata or partitions of an area are made based on a set of criteria, such as homogeneity of contamination. Samples are drawn from each stratum according to a formula that accords more samples to more heterogeneous strata.

- **adaptive cluster sampling** – An initial *n* samples are selected randomly. Additional samples are taken at locations surrounding the initial samples where the measurements exceed some threshold value. Several rounds of sampling may be required. Selection probabilities are used to calculate unbiased estimates to compensate for oversampling in some areas.

- **ranked set sampling** – In this two-phased approach, sets of population units are selected and ranked according to some characteristic or feature of the units that is a good indicator of the relative amount of the variable or contaminant of interest that is present. Only the *m*th ranked unit is chosen from this set and measured. Another set is chosen, and the *m*-1th ranked unit is chosen and measured. This is repeated until the set with the unit ranked first is chosen and measured. The entire process is repeated for *r* cycles. Only the *m* X *r* samples are used to estimate an overall mean.

- **sampling along a swath or transect** – Continuous sampling is done along straight lines (swaths) of a certain width using geophysical sensors capable of continuous detection. The goal is to find circular or elliptical targets. This design contains the two elements of traversing the target *and* detecting the target. VSP application is for unexploded ordnance (UXO).

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

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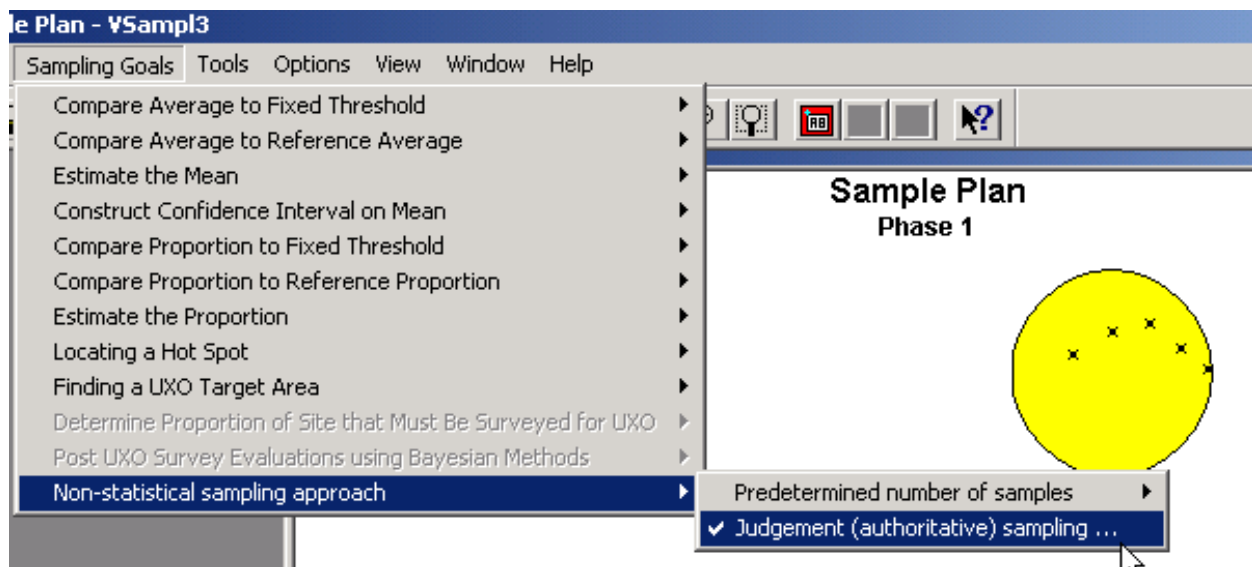
Figure 3.3. Judgement Sampling in VSP

3.2 DQO Inputs and Sample Size

The inputs needed for VSP's sample-size calculations are decided upon during the DQO process. 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 core set of DQO inputs that affect sample size for most of the designs are as follows:

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

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

- *Type II Error Rate (Beta)* – This is called the *false acceptance* rate in EPA's DQO guidance. This is the probability of not rejecting (accepting) a false null hypothesis. For the typical hypothesis test in which we assume the survey unit is dirty, beta is the chance a specific clean site will be condemned as dirty. Specifically, beta is the chance that a clean site with a true mean equal to or less than the lower bound of the gray region will be condemned as dirty. In general, beta is the maximum chance, outside the gray region, that a clean site will be condemned as dirty.

- *Width of Gray Region (Delta)* – This is the distance from the Action Level to the outer bound of the gray region. For the typical hypothesis test in which we assume the survey unit is dirty, the gray region can be thought of as a range of true means where we are willing to decide that clean sites are dirty with high probability. Typically, these probabilities are 20% to 95%, i.e., from beta to 1 - alpha. If this region is reduced to a very small range, the sample size grows to be extremely large. Determining a reasonable value for the size of the gray region calls for professional judgment and cost/benefit evaluation.

- *Estimated Sampling Standard Deviation* – This is an estimate of the standard deviation expected between the multiple samples. This estimate could be obtained from previous studies, previous experience with similar sites and contaminants, or expert opinion. Note that this is the square root of the variance.

Other inputs are required by some of the designs, and other inputs are required for design parameters other than sample size. For example, the stratified designs require the user to specify the desired number of strata and estimates of proportions or standard deviations for each of the stratum. These other inputs are described below and in the **Help** function.

Note: The **Help** function in VSP provides a description of each of the design inputs. 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. In addition, pressing the **Help** button at the bottom of each screen will bring up a file that contains a complete explanation of the design.

The next section contains a discussion of the inputs required by most of the designs available in VSP 2.0. The designs are organized by the Sampling Goal under which they fall. Not all options for all designs are discussed. The **Help** function describes all the options and is a good supplement to this User's Manual.

3.2.1 Compare Average to a Fixed Threshold

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

Continue where we left off in Section 2.3.3 with the Millsite.dxf map loaded. We selected a single Sample Area from the site. The Action Level for the contaminant of interest is 6 pCi/g in the top 6 in. of soil. Previous investigations indicate an estimated standard deviation of 2 pCi/g for the contaminant of interest. The null hypothesis for this problem is "Assume Site is Dirty" or H_0: True mean >=AL.

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

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

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

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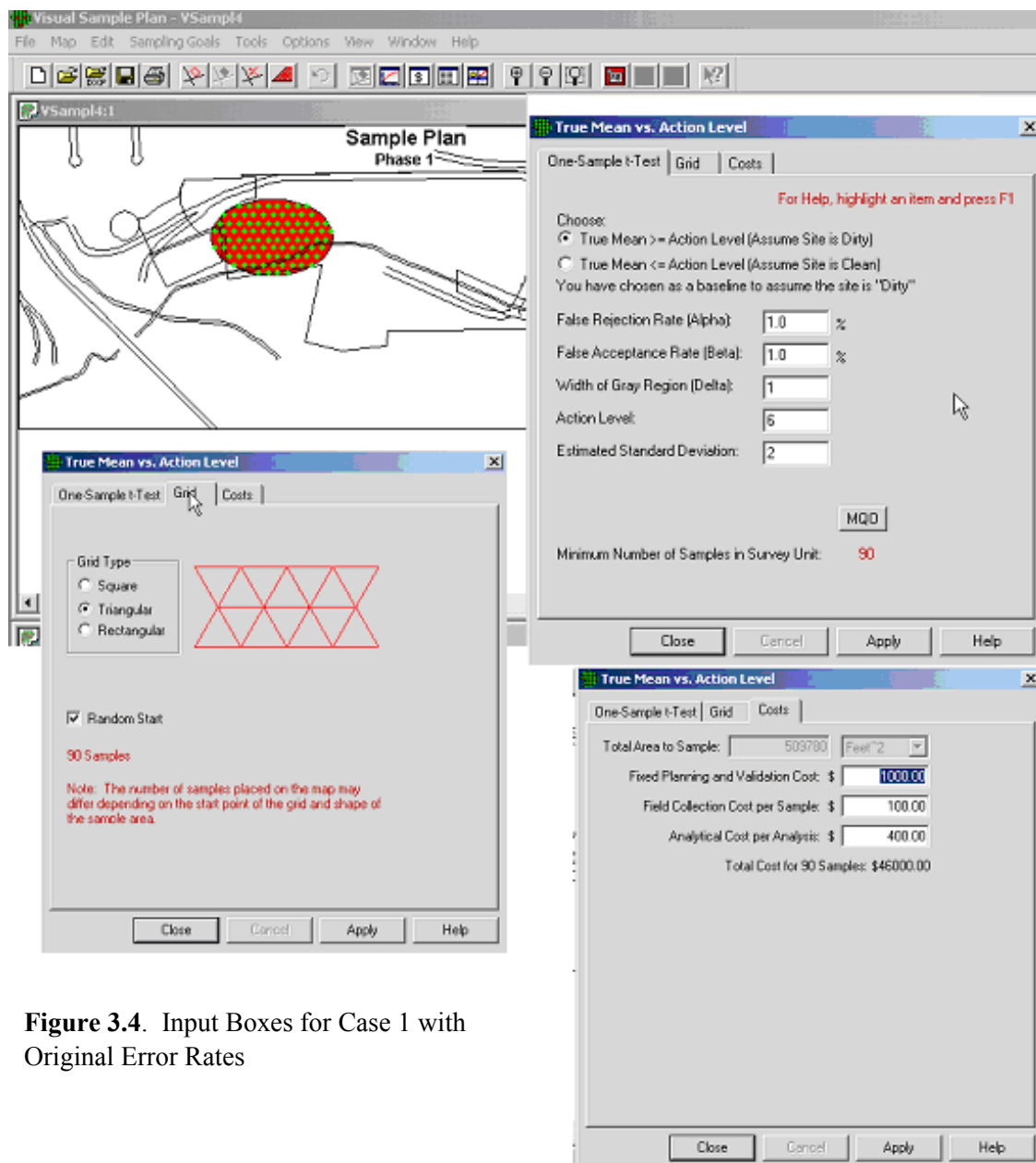
Figure 3.4. Input Boxes for Case 1 with Original Error Rates

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% are acceptable. Figure 3.5 reveals that those changes lead to a significant reduction in the sampling cost, now $19,000 for $n = 36$ samples.
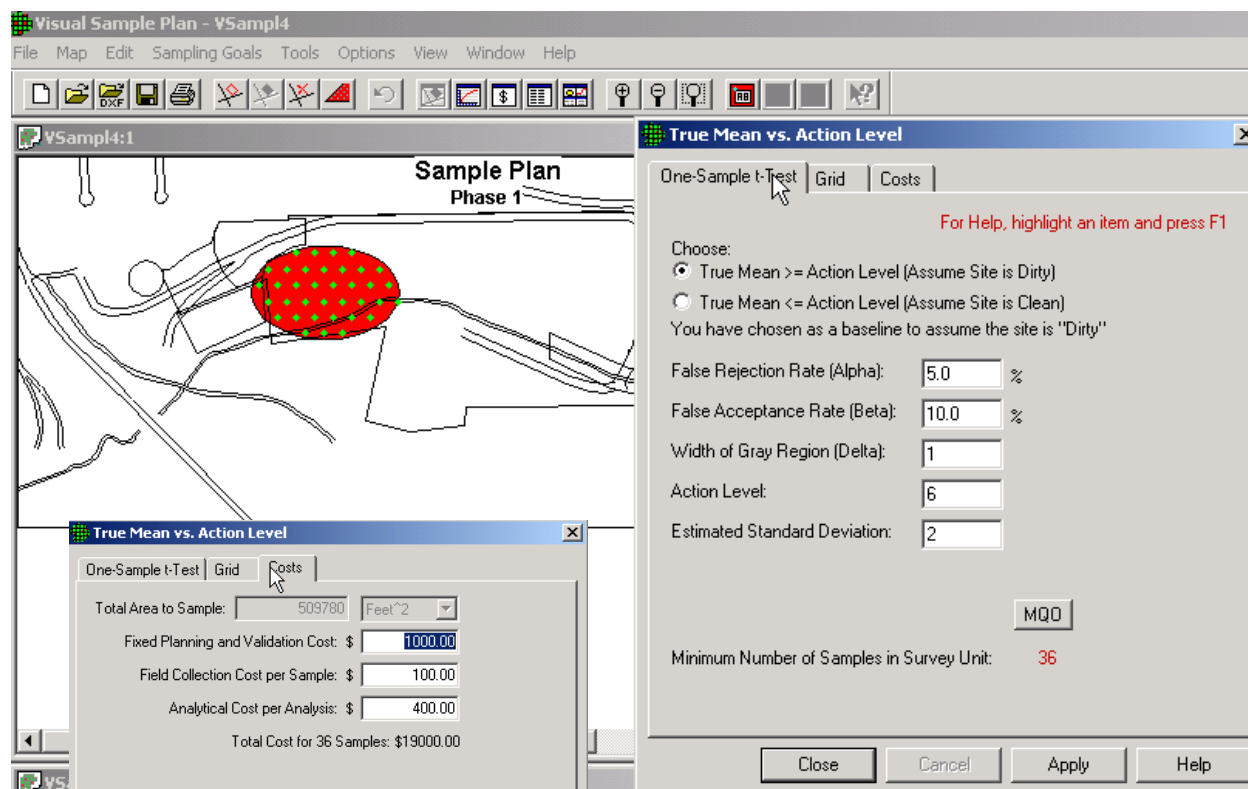
Figure 3.5. Input Boxes for Case 1 with Increased Error Rates

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* (EPA 2000b, pp. 3-5).

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

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

VSP Solution 2a: From VSP's main menu, select **Sampling Goal of Compare Average to a Fixed Threshold > Can assume data will be normally distributed > Sequential Sampling (Known Standard Deviation)**. The dialog box in Figure 3.6 appears. We begin by entering the DQO parameters for Alpha, Beta, Action Level, etc. Next, enter the **Number of Samples Per Round**, shown here as **3**. This parameter indicates how many samples you want to take each time you mobilize into the field. Each time you press the **Apply** button, VSP places a pattern of this many sampling locations on the map. When you close this design dialog, this pattern of sampling locations is locked or "frozen." In Figure 3.6, we see the results of pressing **Apply**, and three locations are placed on the Map labeled "Seq-1, Seq-2, Seq 3".
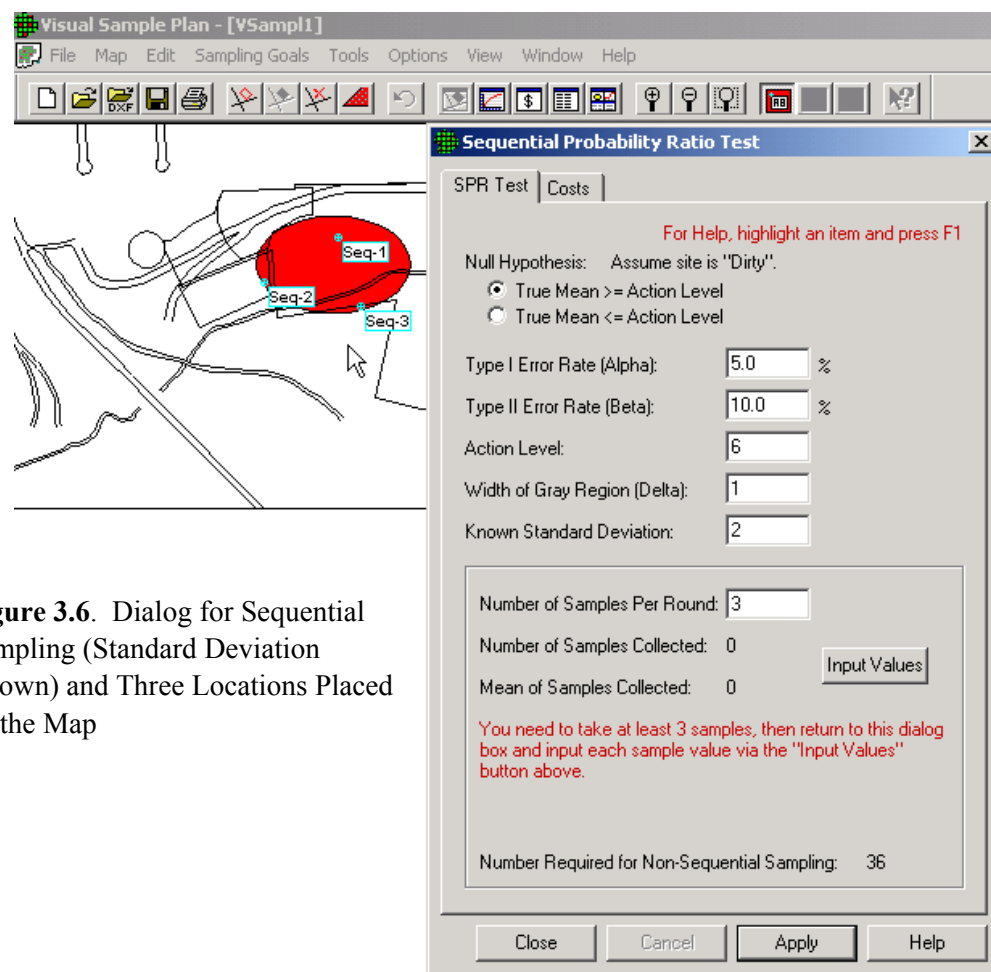
Figure 3.6. Dialog for Sequential Sampling (Standard Deviation Known) and Three Locations Placed on the Map

You must now exit this dialog (close the display by clicking the X in the upper right-hand corner of the display), go out and take the samples, and analyze them. Once the sample results are available, re-open the Sequential Probability Ratio Test (SPRT) design dialog box. You now see the **Number of Samples Collected** as **3**. Press the **Input Values** button and enter the measurement values for those three samples into the grid on the data input dialog. We enter these values as **6.5, 8, and 5**. Press the **OK** button and VSP returns to the original SPRT dialog box. We now see that VSP calculated a mean of 6.5 and a standard deviation of 1.5 for the values we entered. VSP cannot accept or reject the null hypothesis within the error limits we specified based on these values and suggests that **9** additional samples are needed to make a decision.
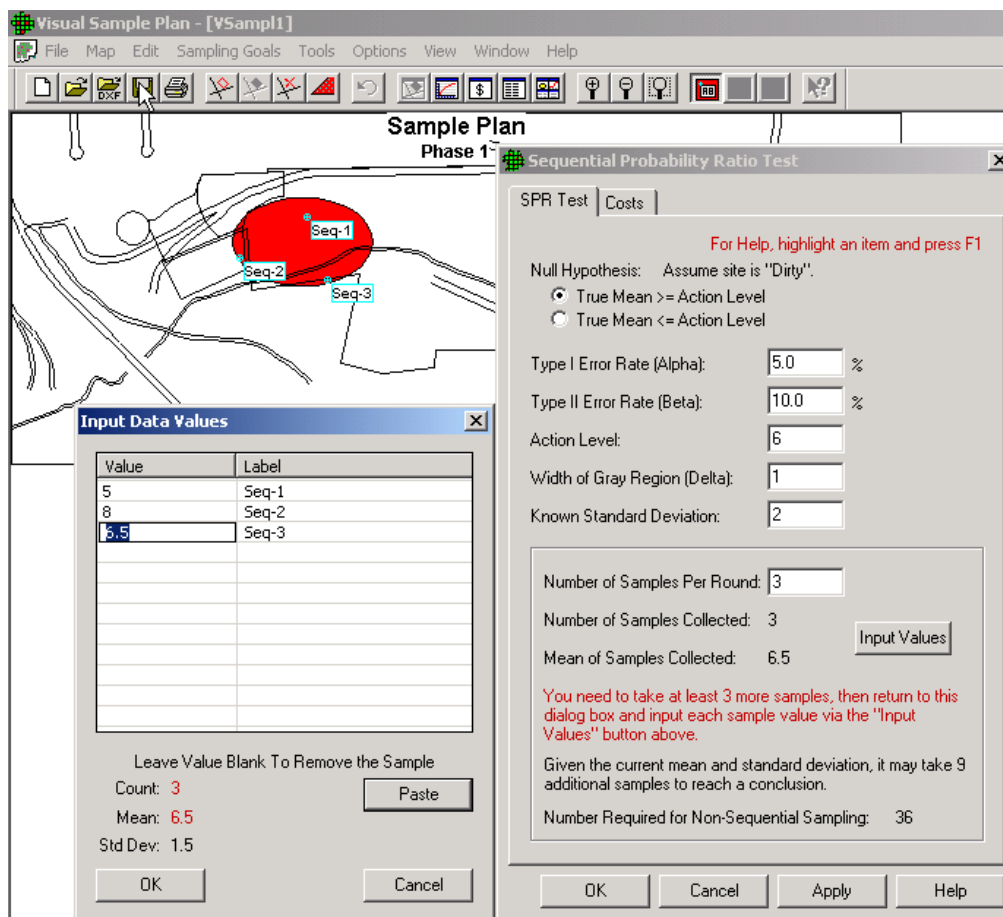
Figure 3.7. Data Input Dialog for Sequential Probability Ratio Test and Results from First Round of Sampling. Map View is shown in insert.

Switching over to the Graph View in Figure 3.8, we can see that in order to accept the null hypothesis that the site is dirty by taking just three samples, we need a sample mean of approximately 9 (i.e., cross-hair set at 3 on the x-axis, intersects the lower boundary of the red area at 9 on the y-axis). If we move the cursor to the green area, the cross-hair intersects the upper boundary of the green area at approximately 2.5. In other words, we need a sample mean of 2.5 in order to reject the null hypothesis and declare the
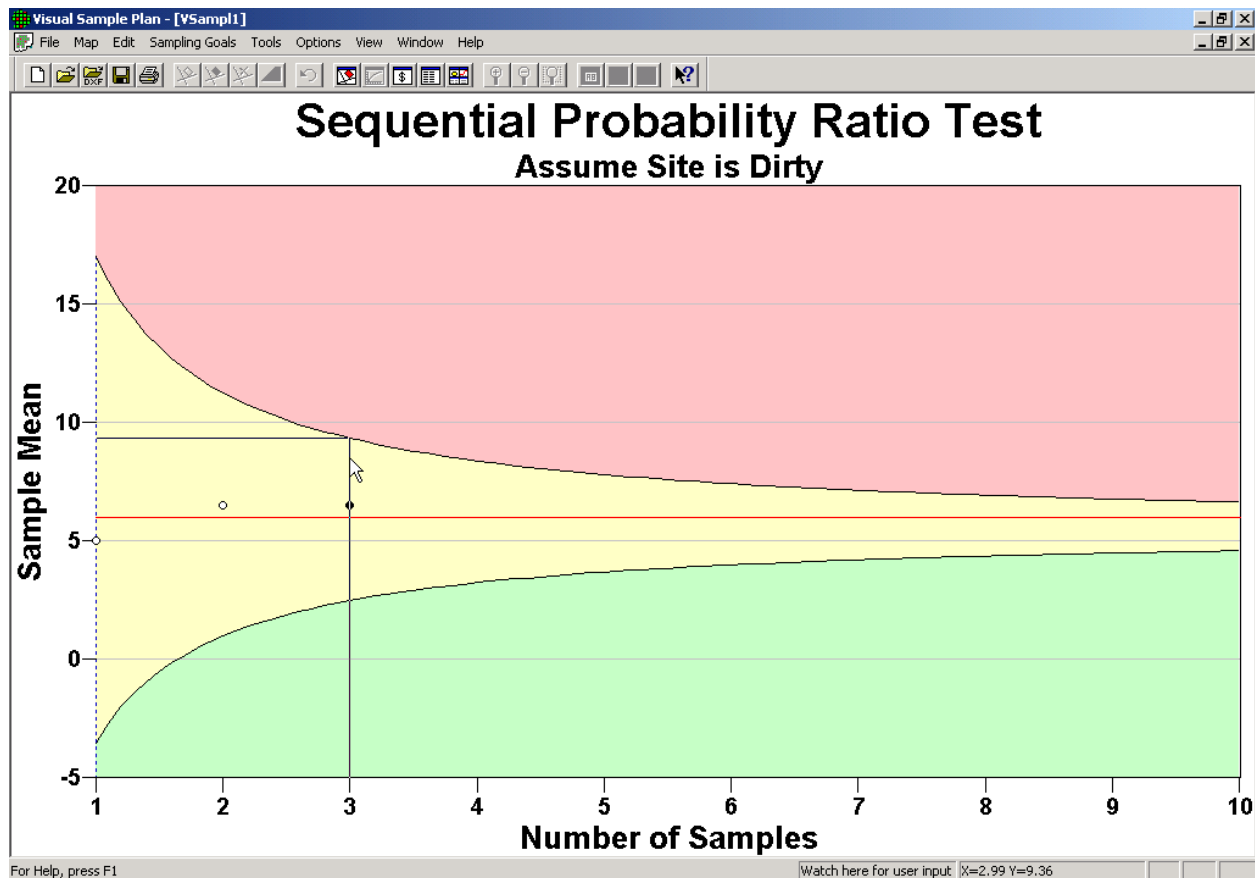
Figure 3.8. Graph View of Sequential Sampling

site clean with only three values. Remember, we told VSP that we *knew* the standard deviation to be 2, so that is what the program is using to make the projection of additional samples needed. If we are wrong, the VSP output will be misleading.

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

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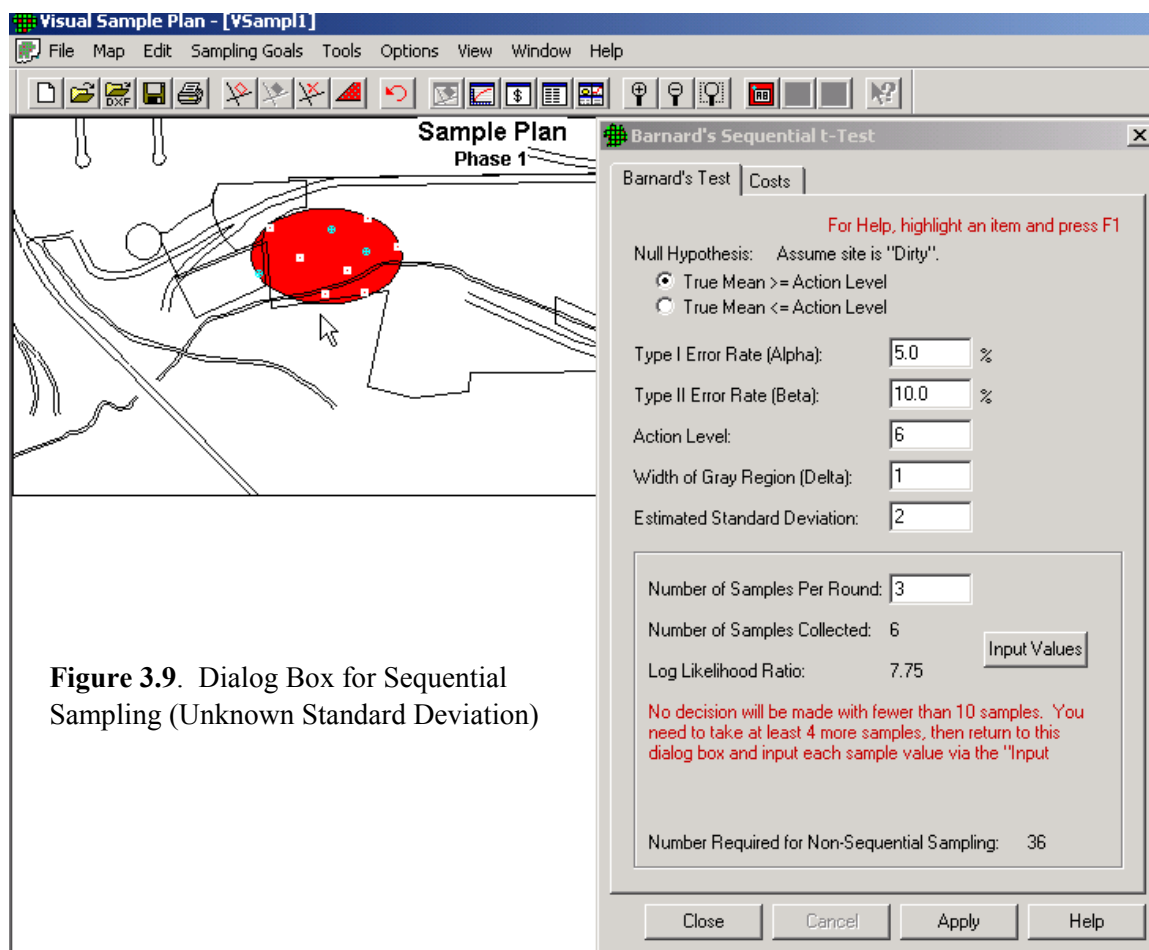
Figure 3.9. Dialog Box for Sequential Sampling (Unknown Standard Deviation)

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

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

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

Figure 3.10. Input Boxes for Case 3 Without the Normality Assumption

3.2.2 Compare Average to Reference Average

We again start with the Millsite.dxf map from Section 2.3.3 with a single Sample Area defined. The Action Level for the contaminant of interest is 5 pCi/g above background in the top 6 in. of soil. Background is found by sampling an appropriate Reference Area. Previous investigations indicate an estimated standard deviation of 2 pCi/g for the contaminant of interest. The null hypothesis for this problem is "Assume Site is Dirty" or H_0: Difference of True Means >= Action Level. In other words, the parameter of interest for this test is the *difference* of means, not an individual mean as was the case in the one-sample t-test.

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

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

Case 4: We assume that the populations we are sampling are approximately normal or that they are well-behaved enough so that the Central Limit Theorem of statistics applies. In other words, the distributions

of sample means drawn from the two populations are approximately normally distributed. If that is the case, the distribution of the differences also will be approximately normally distributed. We also assume the standard deviations of both populations are approximately equal. In addition, we determine that a systematic grid sampling scheme is preferable.

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

We see that for our inputs, using a two-sample t-test will require taking175 field samples in the Sample Area at a cost of $18,500. The sampling cost for the Reference Area also will be $18,500. The combined sampling cost of $37,000 is significantly beyond our budget of $20,000. What will be the result if we relax the error rates somewhat?

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

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

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

What about our planned use of a parametric test, the two-sample t-test? A sample size of 88 is large enough that we can probably safely assume the two-sample t-test will meet the assumption of normality for the differences of sample means. We should test this assumption after the data are collected.
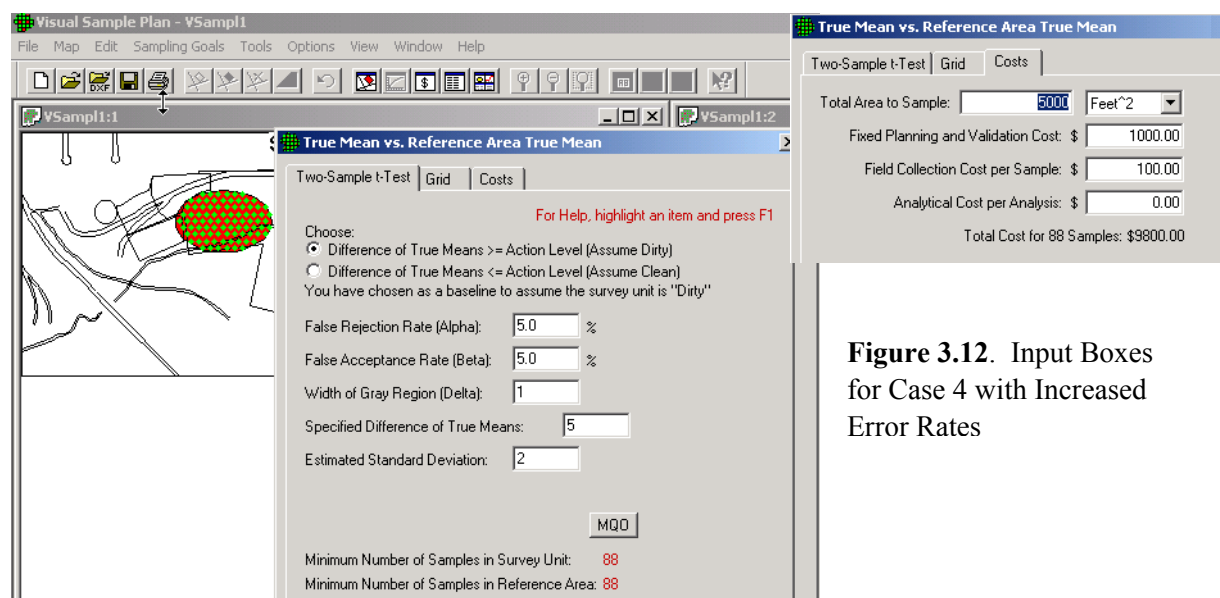
Figure 3.12. Input Boxes for Case 4 with Increased Error Rates

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

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

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

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

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

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

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

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

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

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

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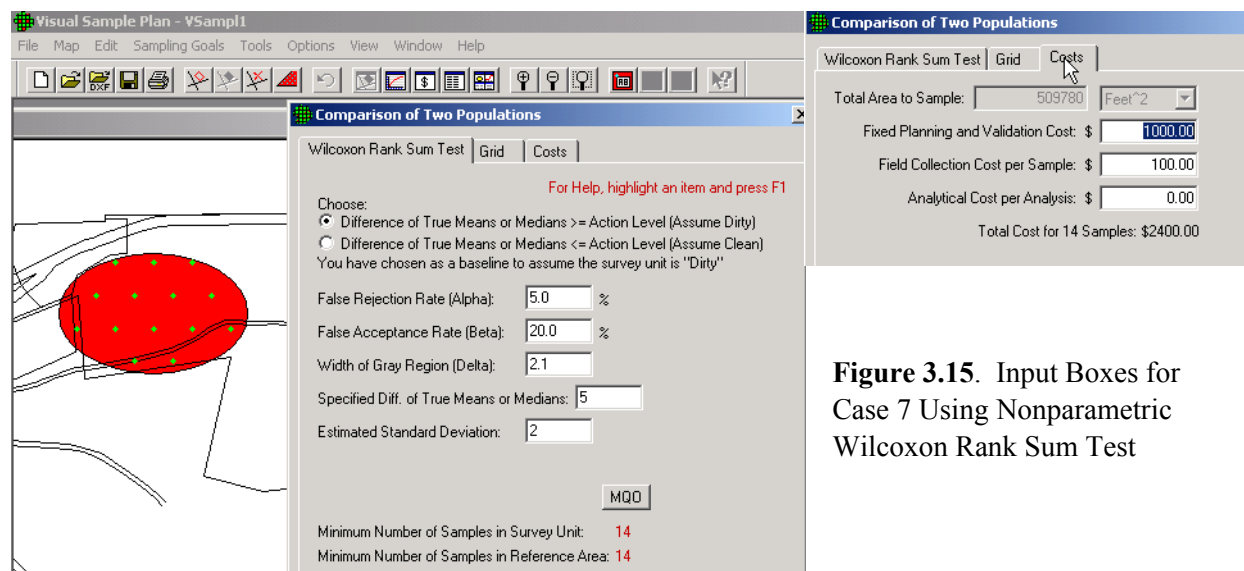
Figure 3.15. Input Boxes for Case 7 Using Nonparametric Wilcoxon Rank Sum Test

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. There is also a larger area of the gray region where you say you will not control (i.e., limit) the false acceptance error rate.

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

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

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

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

Figure 3.16. Input Boxes for Case 8 with Larger Standard Deviation

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

3.2.3   Estimate the Mean

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

- stratified sampling
- ranked set sampling
- adaptive cluster sampling.

3.2.3.1   Stratified Sampling

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

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

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

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

3.2.3.2   Ranked Set Sampling

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

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

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

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

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

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

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

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

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

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

Figure 3.18. Dialog Boxes for Ranked Set Sampling Design

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

Ranked set field sampling locations are generated with a label having the following format:

RSS-c-s-i

where   c = the cycle number

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

i = a unique identifier within the set.

Use **View > Labels** on the main menu or the **AB** button on the main toolbar (button also on the RSS toolbar) to show or hide the labels for the field sample locations. Figure 3.20 shows the labels on the map for field sample locations associated with **Cycle 3, All Sets**.
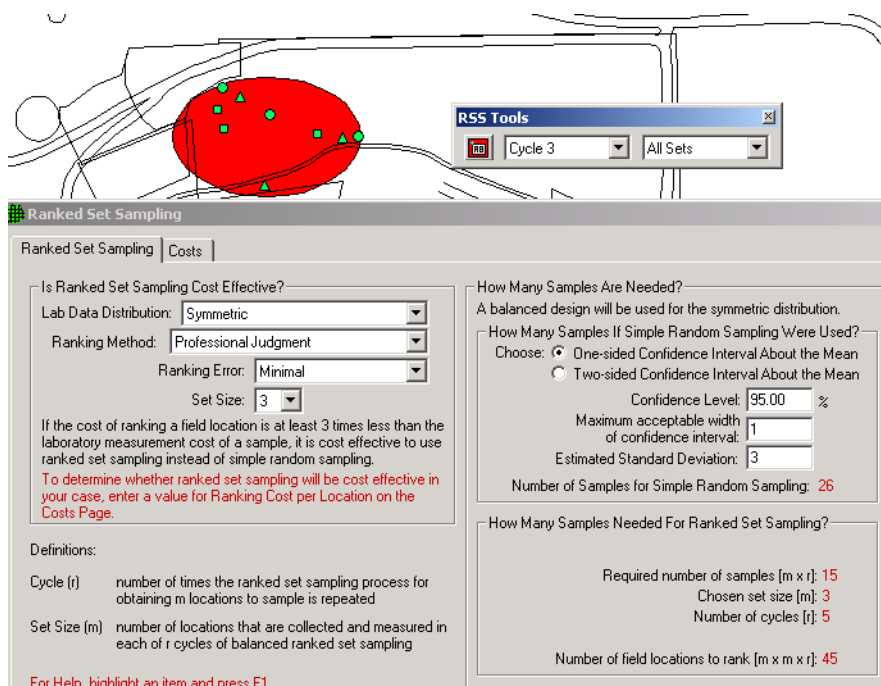
Figure 3.19. Map of RSS Field Sample Locations for All Sets in Cycle 3, Along with RSS Toolbar
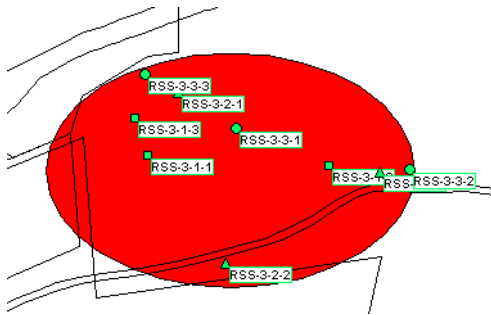
Figure 3.20. Map of RSS Field Sample Locations Along with Their Labels

3.2.3.3 Adaptive Cluster Sampling

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

The first step in this design is to divide the Sample Area into a grid of sampling units. VSP automatically does this for the user, but as we will see later, the initial grid size and grid angle can be changed. We can see this grid by selecting the **Map View** button on the toolbar or selecting **View > Map** from the main menu. One or more Sample Areas have previously been selected, as evidenced by the colored sample areas on the map. (Hint: select a color for the Sample Area OTHER THAN yellow, green, or red.) We suggest using the **Zoom In** button on the toolbar to make the Sample Area fill the screen so the grids can clearly be seen. The grid is overlain on the Sample Area by selecting the **View All Grid Cells** button on the toolbar. The grid is shown in Figure 3.21.
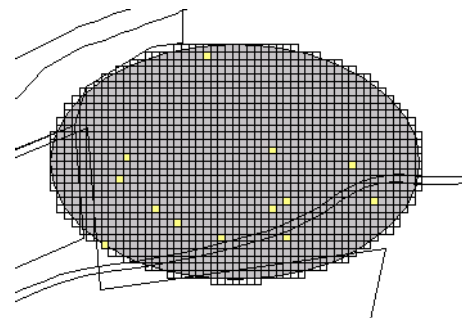
Figure 3.21. Map of the Sample Area with Grid Cells Displayed

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

In Figure 3.22, we can see the 13 initial samples as yellow squares on the map.
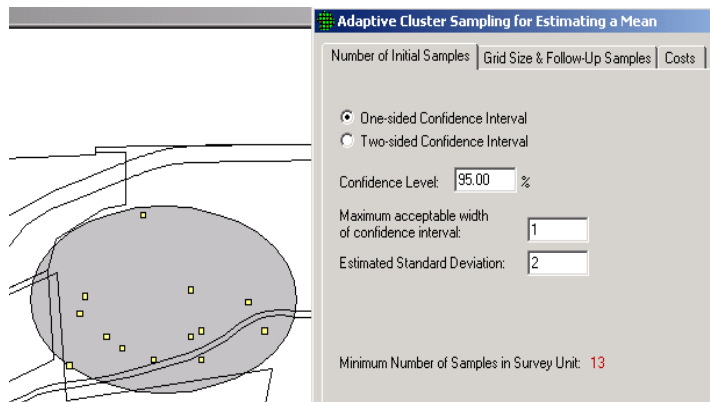
Figure 3.22. Map of Sample Area with Initial Samples for Adaptive Cluster Sampling Shown as Yellow Squares, Along with Dialog Box
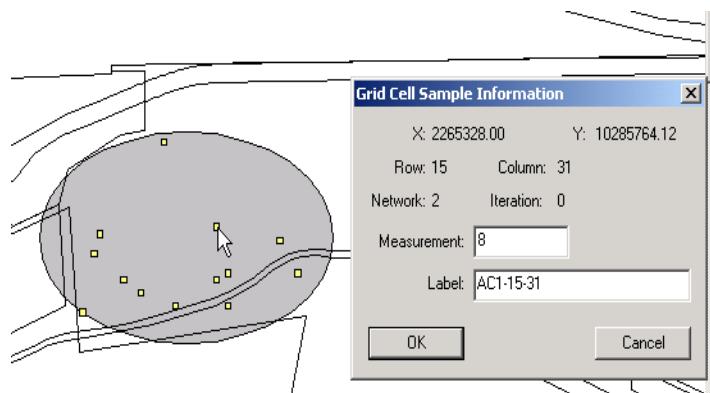
Figure 3.23. Dialog Input Box for Entering Sample Measurement Values and Labels for Initial Samples in Adaptive Cluster Sampling

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

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

Once **Measurement** values have been entered, the yellow squares turn to either green, indicating the sample did not exceed the threshold, or red, indicating the sample exceeded the threshold. The red samples are surrounded with additional yellow squares that now must be sampled. This process continues until there are no more yellow grid cells. In Figure 3.24, we see examples of green, single yellow, red surrounded by yellow, and red surrounded by green. Sampling and measurement continues until all the initial samples are green or red and all the added samples are green or red.
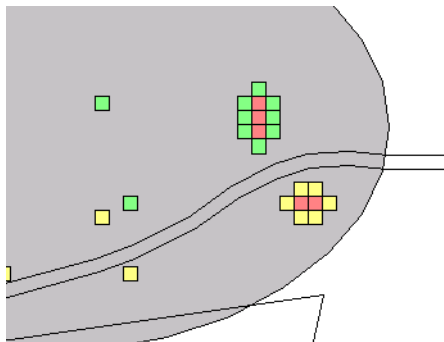
Figure 3.24. Examples of Combinations of Initial and Follow-Up Samples from Adaptive Cluster Sampling, with Dialog Box for Grid Size and Follow-Up Samples

Costs are entered using the **Cost** tab on the dialog box. The total cost for the initial 13 samples is displayed. The **Report** for adaptive cluster sampling shows the total cost for all the initial samples plus follow-up samples and provides an (unbiased) estimate of the mean and its standard error. Refer to VSP's **Help** for a complete discussion of adaptive cluster sampling.

3.2.4 Construct Confidence Interval on Mean

Only three DQO inputs are required for the confidence interval sampling goal: the confidence you want to have that the interval does indeed contain the true value of the mean, how large that interval should be (width of confidence interval), and an estimate of the standard deviation between individual units of the population. You must also decide whether you want the confidence interval to bound values on just one side of the mean (one-sided confidence interval) or on both sides of the mean (two-sided confidence interval). The two-sided confidence interval, smaller width sizes, and larger variation generally require more samples. In Figure 3.25, we see an example of the design dialog for the Confidence Interval on the Mean sampling goal, along with the sample size of **4** that VSP calculated.

3.2.5 Compare Proportion to Fixed Proportion

For comparing a proportion to a threshold, the designs available in VSP do not require the normality assumption. A one-sample proportion test is the basis for calculating sample size. The inputs required to calculate sample size are shown in the design dialog in Figure 3.26. The DQO inputs are the same ones we have seen before, but since the variable of interest is a proportion (percentage of values that meet a certain criterion or fall into a certain class) rather a measurement, the action level is stated as a value from 0.01 to 0.99. Based on the example shown in Figure 3.26, VSP calculates that a sample size of 23 is required.
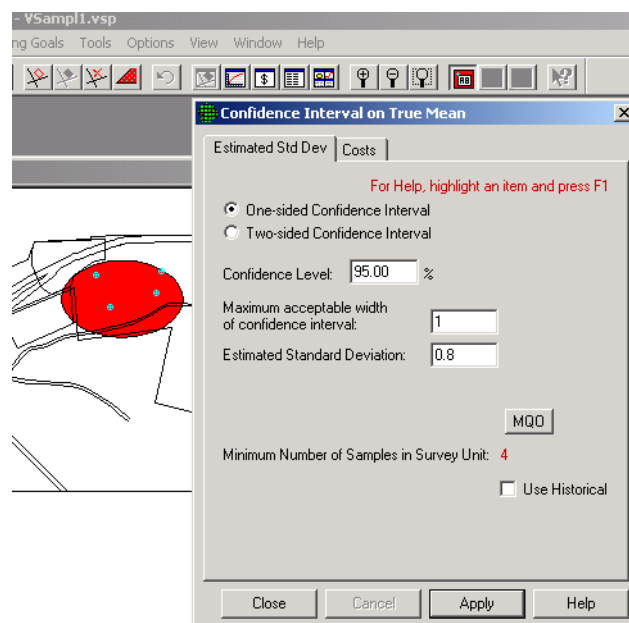
Figure 3.25. Design Dialog for Confidence Interval on the Mean

3.2.6 Compare Proportion to Reference Proportion

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

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

Note that if the proportion of interest is the proportion of positive units in the environment, say the proportion of one-acre lots within a development area that have trees, then we need to select the null hypothesis that affords us the greatest protection against a false acceptance. In Figure 3.27, we see an example of the design dialog for this sampling goal. VSP calculates that we need 38 samples in the survey unit and 38 samples in the reference area for this set of inputs.

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

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

3.2.7 Estimate the Proportion

Similar to the designs available for estimating the mean, we see that VSP offers stratified sampling for the sampling goal of estimate the proportion because a stratified design may be more efficient than either simple random sampling or systematic sampling. In Figure 3.28 we see the dialog box for inputting parameters for Stratified sampling. Prior to running VSP to calculate sample sizes for the strata, the user must have pre-existing information to use as the basis for dividing the site into non-overlapping strata. The strata should be more homogeneous internally than for the entire site (i.e., all strata). They must be homogeneous in the proportion of units that fall into one classification or another. The strata are the individual Sample Areas that the user selected and can be seen using **Map View**.

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

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

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

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

3.2.8 Locating a Hot Spot

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

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

- For a predetermined grid spacing, VSP will calculate the probability of finding a hot spot of a certain size.

- For a predetermined fixed cost, you can divide the area to be sampled into grids based on the budget, and VSP calculates the probability of finding a hot spot of a certain size.

- For a given probability and a given hot spot size, VSP calculates the minimum number of samples (minimum grid spacing) required to hit the hot spot.

For a given probability and a given grid spacing, VSP calculates the smallest hot spot that can be detected.

The basic structure for all these problems is that there are four variables (grid spacing, size of hot spot, probability of hitting a hot spot, and cost). You can fix any three of them and solve for the remaining variable. Figure 3.29 shows the screens for selecting which formulation of the problem you want and the dialogs that pop up in the active window after one formulation is selected.

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

3.2.9 Hot Spot Screens and Dialogs

Other unique features of the hot spot problem, such as how the reports look, how the resulting design depends on the size of the area to be sampled, how the number of samples calculated might not be equal to the number of samples fitted onto the map, and how orientation of the grid plays a role, are discussed in this section.

Problem Statement: A site has one Sample Area of one acre (43,560 square feet). We wish to determine the triangular grid spacing necessary to locate a potential circular pocket of contamination with a radius of 15 feet. We desire the probability of detecting such a hot spot, if it exists, to be at least 95%. 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 9: We assume that the assumptions listed in Gilbert (1987, p. 119) are valid for our problem. We specify a hit probability of 95%, a shape of 1.0 (circular), and a radius (Length of Semi-Major Axis) of 15 feet. We will let VSP calculate the length of the side of the equilateral triangular grid needed for these inputs.

VSP Solution 9: First, open the file *OneAcre.vsp* using VSP Main Menu option **File > Open Project**. (Note: This is a VSP-formatted project file and it contains a previously defined Sample Area of the entire acre.) Next, from the VSP Main Menu select **Sampling Goals > Locating a HotSpot > Systematic Grid Sampling > Maximum Number of Samples**. A grouping of the input dialogs is shown in Figure 3.29.
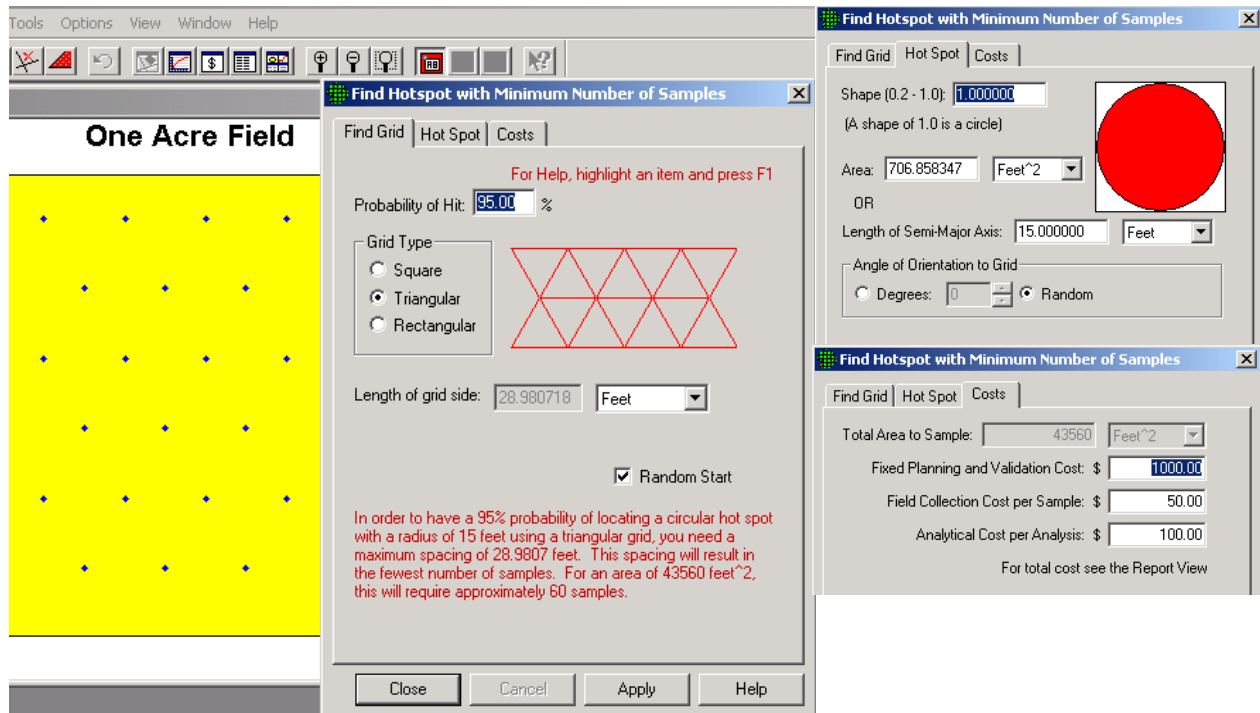
Figure 3.29. Input Boxes for Case 9

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

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

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

- Placing the cursor in the **Length of Semi-Major Axis** on the **Hot Spot** tab and right-clicking displays a black line on the picture of the circle for the radius.
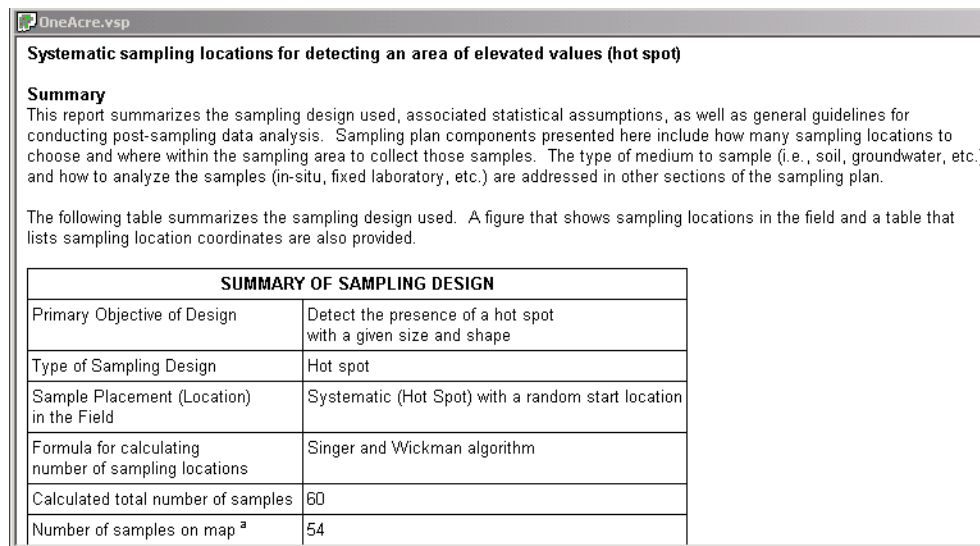
- **Shape** controls how "circular" the hot spot is. Smaller values (0.2) result in a more elliptical shape; 1.0 is a perfect circle.

- The user can specify the **Area** of the hot spot or the **Length of the Semi-Major Axis**. Both fields have pull-down menus for selecting the unit of measurement.

- The Report provides additional information on the design such as the number of samples (both "on the map" and "calculated") and grid area.

Note: the Hot Spot Sampling Goal is the only Sampling goal in VSP that takes into account the **Total Area to Sample** (see this field on the **Cost** tab) when calculating total number of samples. This is because of the nature of the algorithms used to calculate sample size for the various designs. All the other designs use the standard deviation to control sample size.

Figure 3.30 shows two examples of reported sample sizes for Case 9. In the top **Report View**, the *Number of samples on map* is 54; in the bottom **Report View**, the *Number of samples on map* is 63. A different number occurs each time the **Apply** button is pressed.

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.

The discussion above provides an insight into how to use VSP and how VSP user input supplied on the dialog boxes is used to design a sampling plan.

OneAcre.vsp

Systematic sampling locations for detecting an area of elevated values (hot spot)

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

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

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Detect the presence of a hot spot with a given size and shape
Type of Sampling Design	Hot spot
Sample Placement (Location) in the Field	Systematic (Hot Spot) with a random start location
Formula for calculating number of sampling locations	Singer and Wickman algorithm
Calculated total number of samples	60
Number of samples on map [a]	54

(a)

Figure 3.30. Two Different Sample Sizes for Same Inputs to Case 9 Hot Spot Problem

OneAcre.vsp

Systematic sampling locations for detecting an area of elevated values (hot spot)

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

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

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Detect the presence of a hot spot with a given size and shape
Type of Sampling Design	Hot spot
Sample Placement (Location) in the Field	Systematic (Hot Spot) with a random start location
Formula for calculating number of sampling locations	Singer and Wickman algorithm
Calculated total number of samples	60
Number of samples on map [a]	63

(b)

Figure 3.30. (contd)

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 into one of the dialog boxes, click on the **Costs** tab to enter costs. A sample **Costs** screen is shown in Figure 3.31.

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.

Figure 3.31. Screen for Entering Sampling Costs for Sampling Design

- *analytical cost per analysis* - This is the cost to analyze a specimen or a sample. As discussed in Section 5.4, you can specify how many repeated analyses you want taken per sample or specimen.

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

The hot spot sampling goal has some unique cost features. First, costs are displayed in one of the tables in the **Report View** and not on the **Cost** tab of the dialog box. Second, this is the only Sampling Goal for which you can specify a cost as a design criteria and VSP will calculate the number of samples to meet that goal (see Section 3.2.3). This is done by selecting from the main menu **Sampling Goal > Locating a Hot Spot > Systematic Grid Sampling > Predetermined fixed cost**.

4.0 Assessment of Sampling Plans

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

There are two ways to select/change views:

- Press one of the display buttons in the middle of the tool bar (**MAP VIEW, GRAPH VIEW, REPORT VIEW, COORDINATE VIEW**)

- From the main menu select **View** > **Map** (or **Graph**, **Report**, **Coordinate**)

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

In Section 2.1, we described how to set up a **Map**. In Section 2.2, we described how to set up a **Sample Area**. In Section 3.1, we described how to select a **Type of Sampling Plan**. In this section, we find out how to view the results of the sampling design we have just developed, displayed on the map.

In Figure 4.1, we see the display of a simple triangle we drew as our map and selected the entire triangle as our Sample Area. This is available, in file GridSize.vsp, included with the VSP program. We then selected from the main menu **Sampling goals > Locating a Hot Spot > Systematic grid sampling >**

Minimize number of samples. We selected the Probability of Hit to be **90%** and selected a **Square** grid. We entered **4.0** feet for the **Length of Semi-Major Axis** and indicated that we wanted to detect a circular hot spot by selecting a **Shape** of **1.0**. We press **Apply**, and when we return to the map, 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. For most sampling designs, Total Cost (per unit plus fixed costs) is tallied and displayed on the same screen where we enter the per-unit costs—under the **Costs** tab on the dialog box
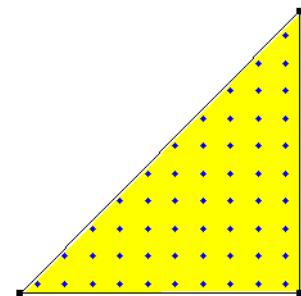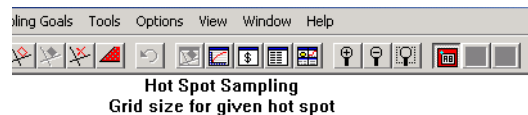
Figure 4.1. Display of Sampling Locations on Map

for entering design parameters. However, for the hot spot Sampling Goal, Total Cost is displayed only in the Report View (from the main menu, select **View > Report**). Reports are discussed in detail in Section 4.4.

As an example of how VSP displays the costs for a design, in Figure 3.31 we see that the Total Cost for 21 samples is $11,500.

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

VSP provides a display of the Performance of the Design for all of the sampling plans that result from sampling goals where a quantitative decision criterion is supplied. When the sampling goal is to "Estimate a Mean" or "Estimate a Proportion" and the only criterion the plans must meet is to minimize the variance of the estimate or minimize cost, there is little to graph in terms of the performance of the design, so no graph is available in VSP.

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

The following sections describe the major displays available for various types of problems. Displays not described are variants of those presented.

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

The display for goal of comparing an average to a fixed threshold (i.e., the Action Level) is a graph of the probability of deciding the true mean of the sample area is greater than or equal to the Action Level on the vertical (y) axis as opposed to a range of possible true mean values on the horizontal (x) axis. Figure 4.2 is the Decision Performance Curve (DPC) described in EPA's QA/G-4 guidance (EPA 2000a, pp. 6-7 – 6-11). Notice how the graph changes as we alternate the null hypothesis between "Assume Site Dirty" to "Assume Site Clean."
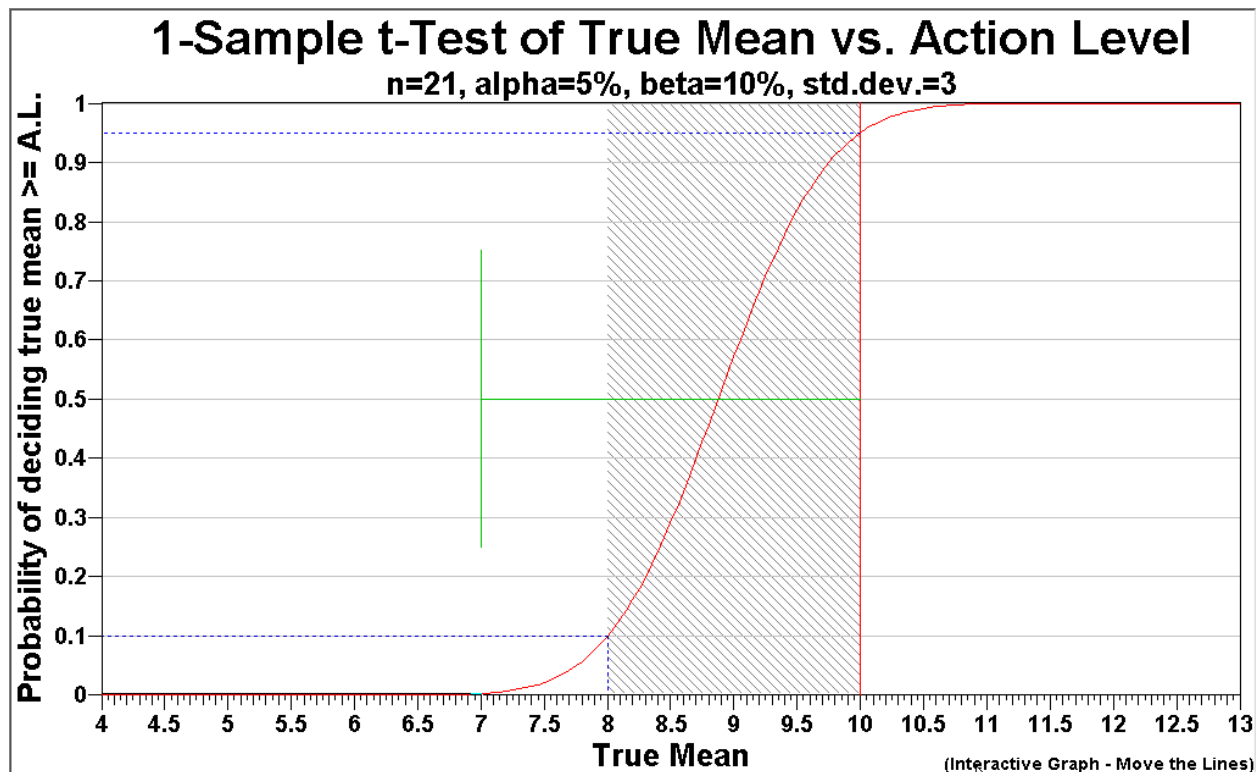
Figure 4.2. Decision Performance Curve for H_0: True Mean >= Action Level for Comparing Mean vs. Action Level

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

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

One of the new features of VSP 2.0 is that most of the parameters displayed on the DPC can be changed interactively by moving the lines on the graph, rather than having to change the values in the input dialog box. Table 4.1 describes the interactive features.

Table 4.1. Interactive Graph Features

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

As you change these parameters, you can see the new value of the parameter on the bottom status bar after "watch here for user input." You will notice that changing these values on the graph also changes their values on the other displays, the sampling design is modified, and new samples are placed on the map.

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 4.2 is telling us that for the "Site is Dirty" null hypothesis,

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

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

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

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

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

However, when we reverse the null hypothesis and state it as **H$_0$: True Mean <= Action Level**, i.e., assume "Site is Clean," we see in Figure 4.3 that the gray region where the majority of decision errors occur shifts to the right side of the Action Level. Sites that are dirty now lead to the majority of decision errors. Also note that alpha is now defined for values less than the action level, while beta is defined for values above the upper bound of the gray region.
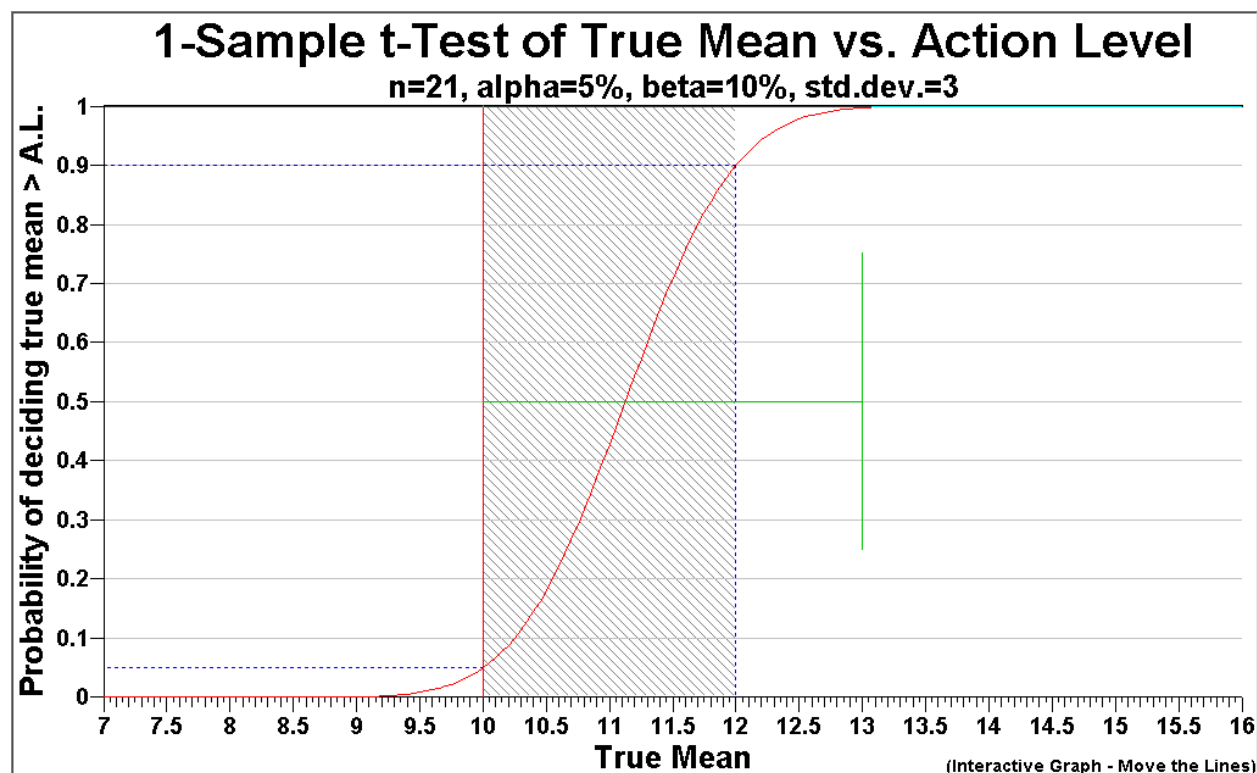
Figure 4.3 Decision Performance Curve for H$_0$: True Mean <= Action Level for Comparing Mean vs. Action Level

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.

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

The display for assessing a confidence interval for a mean differs somewhat from that for comparing an average to a threshold because this is an estimation problem, not a testing problem. As such, there is only one type of decision error rate, alpha. Shown in Figure 4.4 is the Performance Design for a problem where the user specified the width of the confidence interval as **1.0**, the standard deviation as **3**, and a desired **95% one-sided** confidence interval on the mean. We are using a one-sided confidence interval (vs. a two-sided) because we are concerned only about values that exceed the upper bound of the
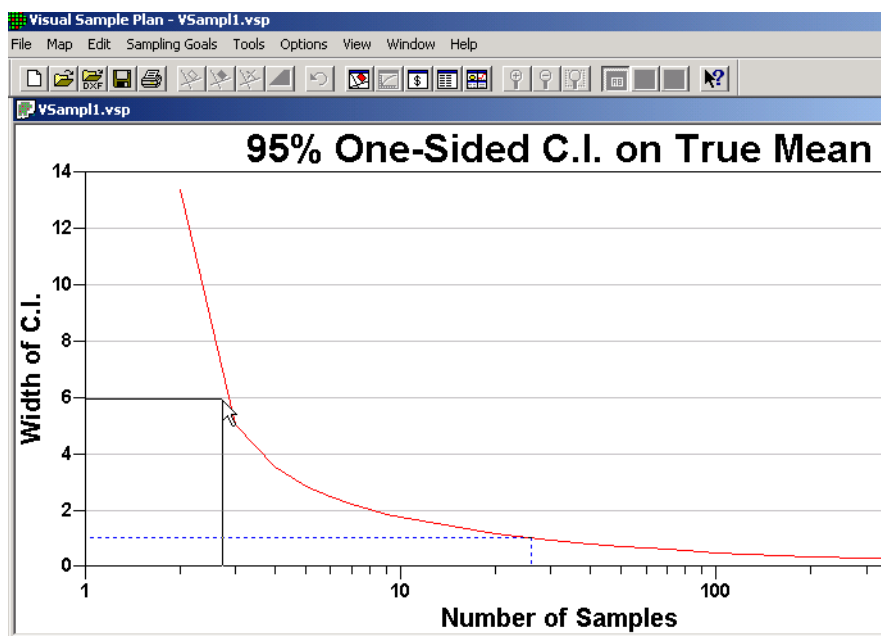
Figure 4.4. Decision Performance Graph for One-Sided 95% Confidence Interval

confidence interval, not values both above the upper bound and below the lower bound. This is consistent with problems in which the mean to be estimated is average contamination, so we are not concerned about values below the lower bound of the confidence interval.

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

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

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

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

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

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

In the example in Figure 4.5, the null hypothesis was set to **True Proportion >= Given Proportion**. As such, beta is the probability of deciding the proportion exceeds the threshold when the true proportion is equal to or less than the lower bound of the gray region. For this problem, we set alpha to **1%** and beta to **5%,** and the lower bound of the gray region to **0.35** (i.e., width of gray region = 0.15). The proportion we want to test against (Action Level) is 0.5. This Action Level is the default for VSP because it is the most conservative. That is, the most number of samples are needed to differentiate a proportion from 0.5 (vs. differentiate a proportion from any other percentage). VSP calculated a sample size of 169. The dashed blue line terminates on the y-axis at 169 samples and on the x-axis at a beta of **5%**. The heading for this graph reminds the user that the one-sample proportion test is the assumed test that will be used for making the decision because the sample size formula in VSP is based on the one-sample proportion test.

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

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

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

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

Note: The standard deviation is the average expected difference between the individual units in a population and the overall mean for that population. It is assumed that both populations (Sample Area and Reference Area) have the same standard deviation. The graph is labeled "2-Sample t-Test" because it is assumed that the two-sample t-test will be used as the statistical test.
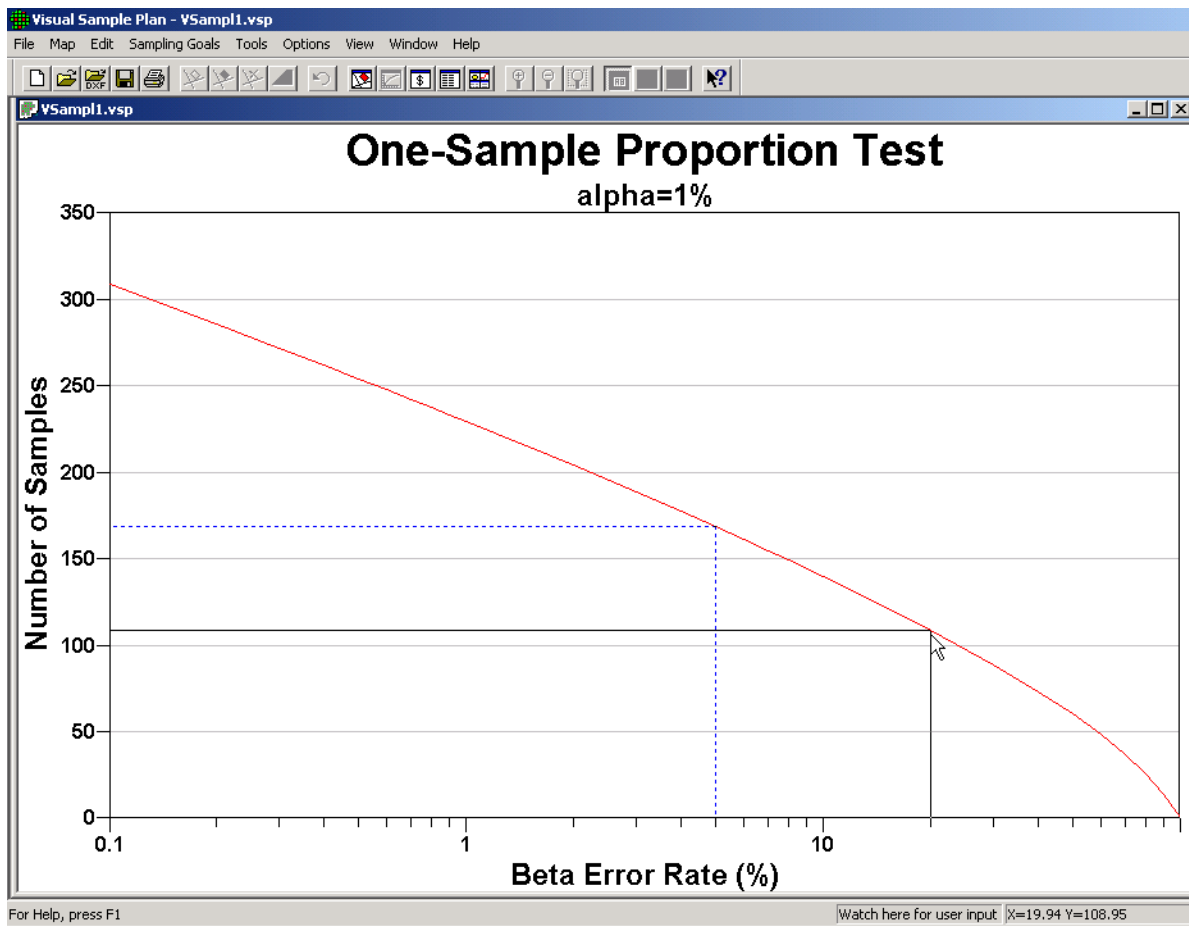
Figure 4.5. Decision Performance Graph for Comparing Proportion to Fixed Threshold

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

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

The MARSSIM WRS test is used when the Sample Area population is symmetrical, the contaminant of concern in the SampleArea is present also in the background (Reference Area), and the contamination is uniformly present throughout the Sample Area. The *Multi-Agency Radiation Survey and Site Investigation Manual* (EPA 1997) contains a discussion of the MARSSIM WRS test. The manual is available from the EPA Internet site at: http://www.epa.gov/radiation/marssim.
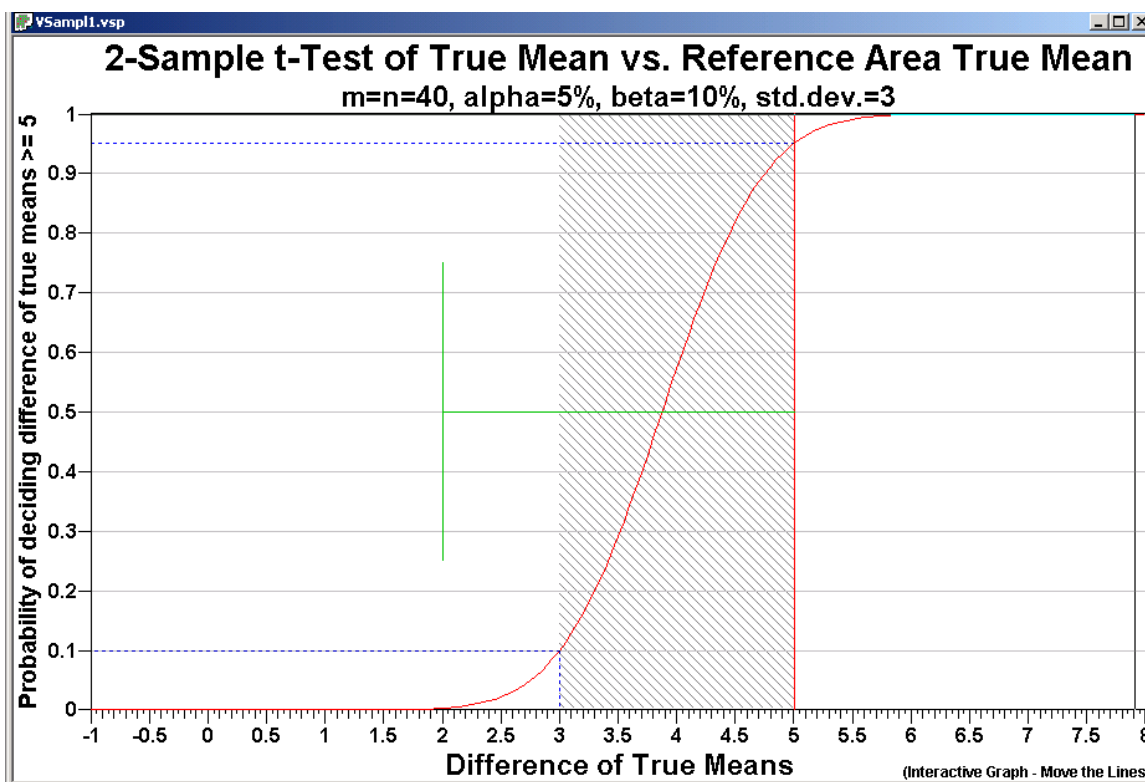
Figure 4.6. Decision Performance Graph for Comparing a Sample Area Mean to a Reference Area True Mean (Parametric Version)

The Wilcoxon rank sum test is discussed in *Guidance for Data Quality Assessment* (EPA 2000b, pp. 3-31 – 3-34). The document can be downloaded from the EPA at: http://www.epa.gov/quality/qa_docs.html. It tests a shift in the distributions of two populations. The two distributions are assumed to have the same shape and dispersion, so that one distribution differs by some fixed amount from the other distribution. The user can structure the null and alternative hypothesis to reflect the amount of shift of concern and the direction of the shift. The verification testing done on VSP shows that the Wilcoxon rank sum test requires slightly higher sample sizes than the MARSSIM WRS test for the same set of inputs, assuming all the appropriate assumptions for each test are met.

The Decision Performance Graphs for the two nonparametric tests are similar to the DPG for the parametric two-sample t-test. In Figure 4.7, we see the DPG for the MARSSIM WRS test using the same inputs as the problem in Figure 4.6. The difference of true means or medians is plotted on the x-axis, and the probability of deciding the difference is equal to or greater than the action level of 5 is shown on the y-axis. For the MARSSIM formulation of the WRS test, the action level is the DCGLw. The lower bound of the gray region is the difference in means or medians where we want to limit the beta error. It can be shown that using the same parameters as in the two-sample t-test, the Wilcoxon rank sum test requires 46 samples in both the Study and Reference Areas while the MARSSIM WRS test requires only 44 samples. The larger sample sizes required for the nonparametric tests reflects the premium that must be paid for not making any assumptions about the underlying distribution of the populations.

MARSSIM WRS Test

m=n=53, alpha=5%, beta=10%, std.dev.=3

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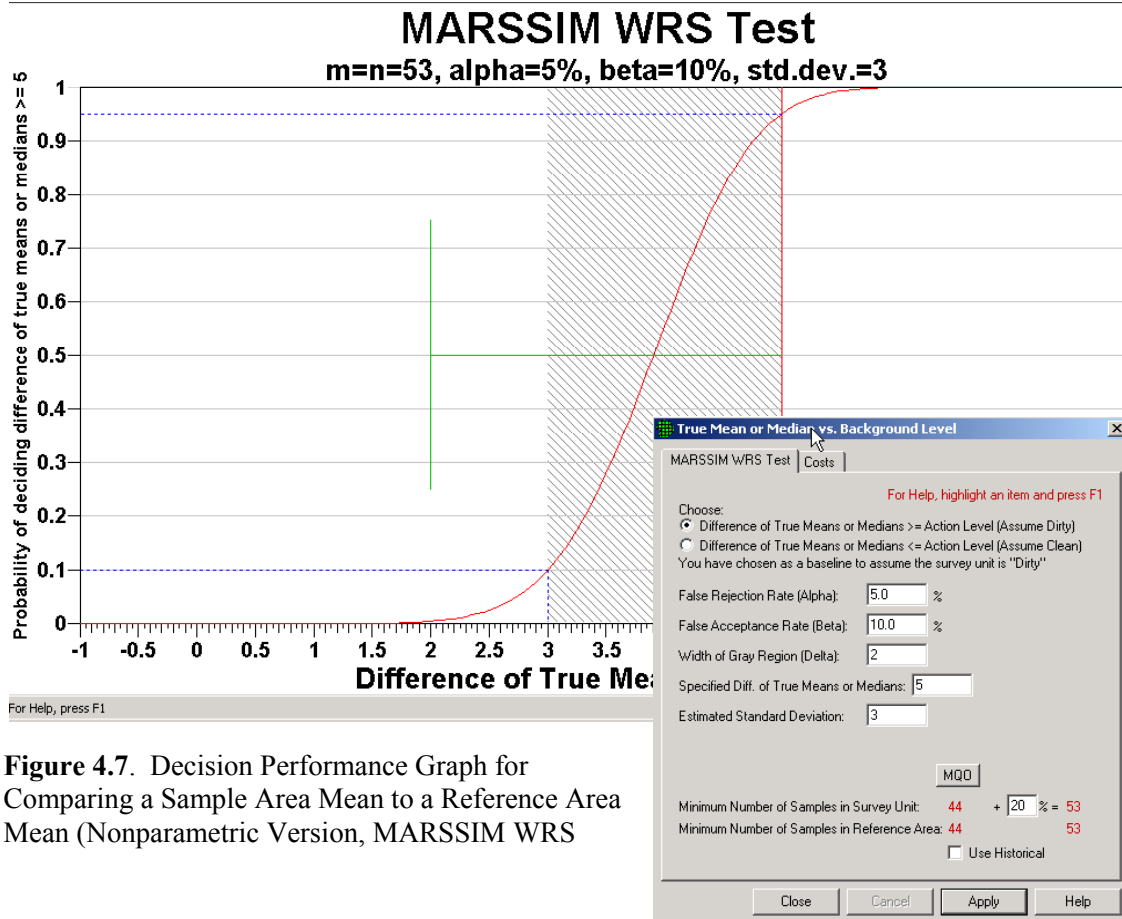
Shown in Figure 4.7, the input dialog for the MARSSIM WRS test allows the user to supply a percent overage to apply to the sample size calculation. MARSSIM suggests that the number of samples should be increased by at least 20% to account for missing or unusable data and for uncertainty in the calculated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

4.3.5 Performance of Design for Sampling Goal for Hot Spot Problem

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

Returning to the problem we laid out in Section 4.1, for the sampling goal of **Sampling Goal > Locating a Hot Spot > Systematic Sampling > Minimize the number of samples**, the Decision Performance Curve is shown in Figure 4.8. This graph is for the Sample Area 2577.6 ft^2 shown in Figure 4.1
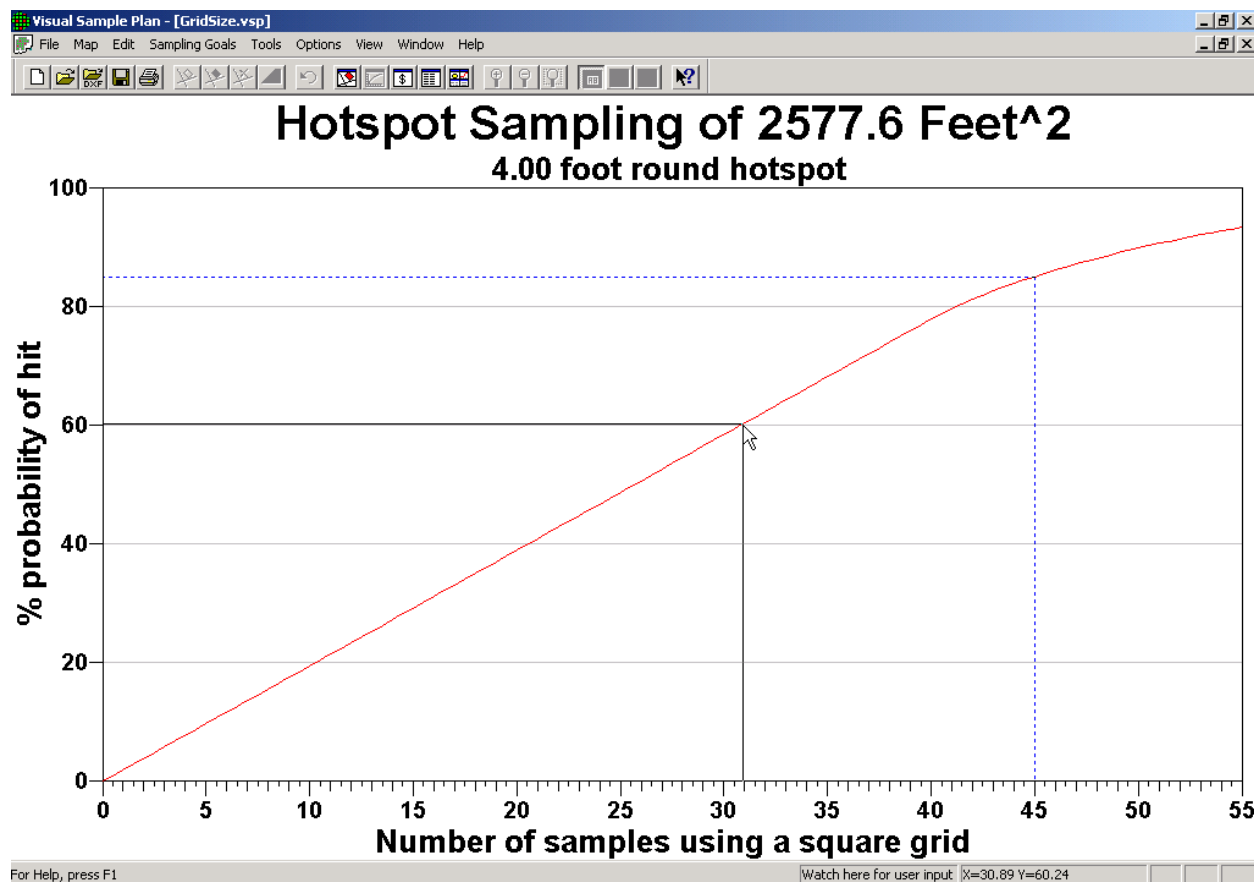
Figure 4.8. Probability of Hitting a Hot Spot vs. Number of Samples

and for finding a **4-ft** round hot spot. The graph shows the desired input of **90%** probability of hitting the circular hot spot of radius 4 ft. and the **51** samples required to achieve this. The dashed blue may show slightly different values, see discussion below.

This example shows how VSP may place a slightly different number of sampling points (nodes) on a map than the exact number calculated. Shown on the performance curve, and displayed in the Report, are anywhere from 45 to 55 samples placed on the map. The difference between the calculated number of sample and the number of samples placed on the map is 1) grid edge effects, 2) adding judgment samples, or 3) selecting or unselecting sample areas. Repeated pressing of the **Apply** button from the dialog box will select a different random starting point for the grid and hence change the number of samples and the probability of a hit associated with that number of samples, as shown by the dashed blue line on the performance curve. This explains why the probability of a hit calculated from the samples placed on the map may not be exactly the probability specified by the user. The report lists the actual number of samples placed on the map. Deselecting the Random Start on the dialog box removes the random

assignment of the grid and keeps the grid fixed with each repeated hit of the Apply button, thus eliminating the sample size shifting problem.

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

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

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

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

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

4.4 Display of the Report

One of the biggest improvements in VSP 2.0 is the expanded report. The Report View for a sampling design is available by either selecting the **REPORT VIEW** button on the toolbar, or by selecting **View > Report** from the main menu. At the current time, VSP 2.0 does not contain all the reports for all the designs. Future versions will have a complete set of reports available. The most frequently used designs have reports available in VSP 2.0.
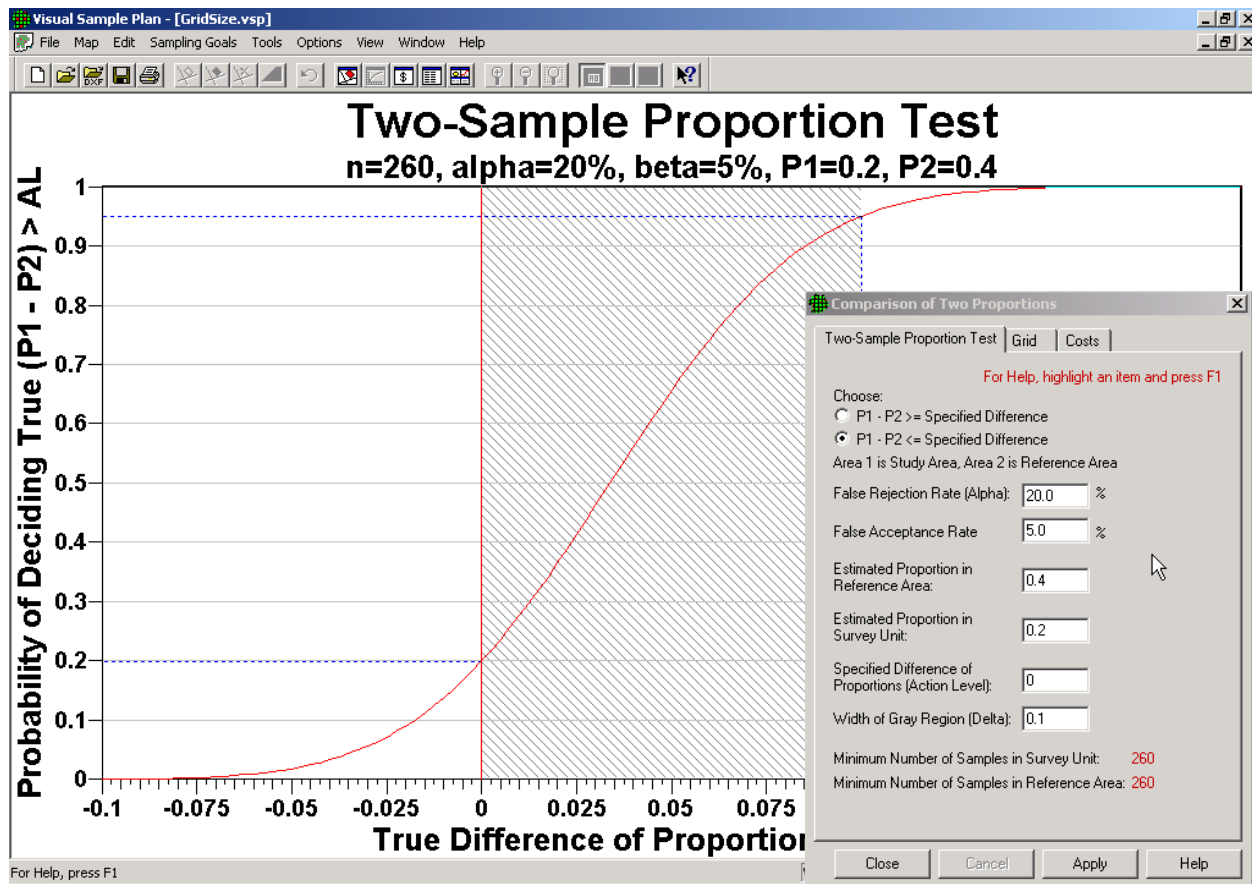
Figure 4.9. Decision Performance Graph for Comparing a Sample Area Proportion to a Reference Area True Proportion. Input dialog box for design shown as an insert.

The goal of the report is to provide the VSP user with a complete documentation of the sampling design selected. This includes the assumptions of each design, sample size formula, inputs provided by the user, a summary of VSP outputs including sample size and costs, a list of samples with their coordinates and labels, a map with areas identified, and a sensitivity table. The reports are suitable for incorporation into a quality assurance project plan or a sampling and analysis plan. Some of the output from VSP, for some designs, is viewable only within the report.

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

Visual Sample Plan - [VSampl1.vsp]

File   Map   Edit   Sampling Goals   Tools   Options   View   Window   Help

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

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

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

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Compare a site mean to a fixed threshold
Type of Sampling Design	Parametric
Sample Placement (Location) in the Field	Simple random sampling
Working (Null) Hypothesis	The mean value at the site exceeds the threshold
Formula for calculating number of sampling locations	Student's t-test
Calculated total number of samples	37
Number of samples on map [a]	37
Number of selected sample areas [b]	1
Sampling area [c]	509779.78 ft^2
Total cost of sampling [d]	$18500.00

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

Area 1					
X Coord	Y Coord	Label	Value	Type	Historical
2265399.3101	10285614.9633		0	Random	
2264951.7999	10285911.3136		0	Random	
2265288.4785	10285633.6746		0	Random	
2265343.2430	10285814.8071		0	Random	
2264945.9484	10285903.5671		0	Random	

Figure 4.10. Report View of the Sampling Goal: Compare Average to a Fixed Threshold, Normality Assumed, Simple Random Sampling

Sensitivity Analysis

The sensitivity of the calculation of number of samples was explored by varying S, β and α table shows the results of this analysis.

Number of Samples		α=5	α=10	α=15
	β=15	167	125	100
S=1.2	β=20	144	105	82
	β=25	126	89	68
	β=15	43	32	26
S=0.6	β=20	37	27	21
	β=25	33	23	18

S = Standard Deviation
β = Beta (%), Probability of mistakenly concluding that μ > action level
α = Alpha (%), Probability of mistakenly concluding that μ < action level

Costs

The total cost of the completed sampling program depends on several cost inputs, some of and measured. Based on the numbers of samples determined above, the estimated total c per sample cost of $500.00. The following table summarizes the inputs and resulting cost

COST INFORMATION			
Cost Details	Per Analysis	Per Sample	37 Samples
Field collection costs		$100.00	$3700.00
Analytical costs	$400.00	$400.00	$14800.00
Sum of Field & Analytical costs		$500.00	$18500.00
Fixed planning and validation costs			$0.00
Total cost			$18500.00

Figure 4.10. (contd)

Reports contain the following sections:

- summary of design – Includes table of VSP inputs and outputs (including number of selected Sample Areas), map with sample locations, list or coordinates of samples and sample labels.

- statement of sampling objective

- description of selected sampling approach and assumptions

- sample size formula along with inputs and parameters used in formula

- Performance Goal Diagram for this design along and assumptions

- sensitivity table showing how sample number changes as parameters change.

- costs of design categorized as field, analytical, and fixed costs.

- general discussion of recommended data analysis activities for how data should be used in the appropriate statistical test to make a decision.

The sensitivity table in the **Report View** is one of the new additions to VSP 2.0. It allows the user to do "what-if" scenarios with VSP input and output. For one sampling goal, the sensitivity table shows how sample size changes with changes in the standard deviation and the two decision error rates, alpha and beta. Different sampling goals and sets of assumptions have different variables and parameters in their sensitivity table. The user can change the variables and range of values shown in the sensitivity table by right-clicking anywhere in the report. A dialog box, as shown in Figure 4.11, comes up where the user can choose which of up to four variables that will be displayed, along with each variable's starting and ending value, and the step-size shown in the sensitivity table. Displayed in the table can be the number or samples, cost, or both.

Figure 4.11. Dialog Box for Changing Variables Displayed, and Range for Variables Shown, in Sensitivity Table in Report View. Shown here is input dialog for sampling goal of compare average to threshold, normality assumed (parametric), simple random sampling.

4.5 Display of Coordinates

The fourth type of display in VSP is the list of coordinates for each sample point on the map. We can see this display by **COORDINATE VIEW** button on the toolbar, or by selecting Main Menu option **View** > **Coordinates**. The x and y coordinates are displayed for each sample point. New to VSP 2.0, also displayed are the sample points label, a value (if entered by the user), the type (e.g., random, systematic, RSS), and a "true/false" indicator of whether or not this sample point is an historical sample. Coordinates are segregated by Sample Area. These coordinates can be copied and pasted into a spreadsheet or word processing file using Main Menu option **Edit** > **Copy**. Figure 4.12 is an example of the Coordinates view.

Area 1

X Coord	Y Coord	Label	Value	Type	Historical
2265399.31	10285614.96		0	Random	False
2264951.80	10285911.31		0	Random	False
2265288.48	10285633.67		0	Random	False
2265343.24	10285814.81		0	Random	False
2264945.95	10285903.57		0	Random	False
2265380.15	10285947.22		0	Random	False
2265472.52	10285951.20		0	Random	False
2264766.31	10285627.64		0	Random	False
2265208.96	10285654.02		0	Random	False
2265319.86	10285589.06		0	Random	False
2265221.53	10286019.53	SS1-101	0	Random	True
2265165.82	10285719.65		0	Random	False
2265166.99	10285424.58		0	Random	False
2265192.29	10285842.13		0	Random	False
2265170.17	10285575.27		0	Random	False
2264926.22	10285943.71		0	Random	False
2265095.74	10285448.71	SS1-104	0	Random	True
2265251.71	10285950.99		0	Random	False
2265242.96	10285986.36		0	Random	False
2265571.36	10285816.68	SS1-100	0	Random	True
2265147.82	10285843.31		0	Random	False
2265441.54	10285682.54		0	Random	False
2265602.09	10285565.63		0	Random	False
2265566.31	10285603.29		0	Random	False
2265112.51	10285451.60	SS1-102	0	Random	False
2265332.08	10285820.85		0	Random	False
2265265.06	10285760.94		0	Random	False
2265125.32	10285899.22		0	Random	False

Figure 4.12. 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. The user can select the **QUAD WINDOW** button from the toolbar for a single keystroke way to display the Quad Window. Figure 4.13 shows the results of the **Quad Window** option.

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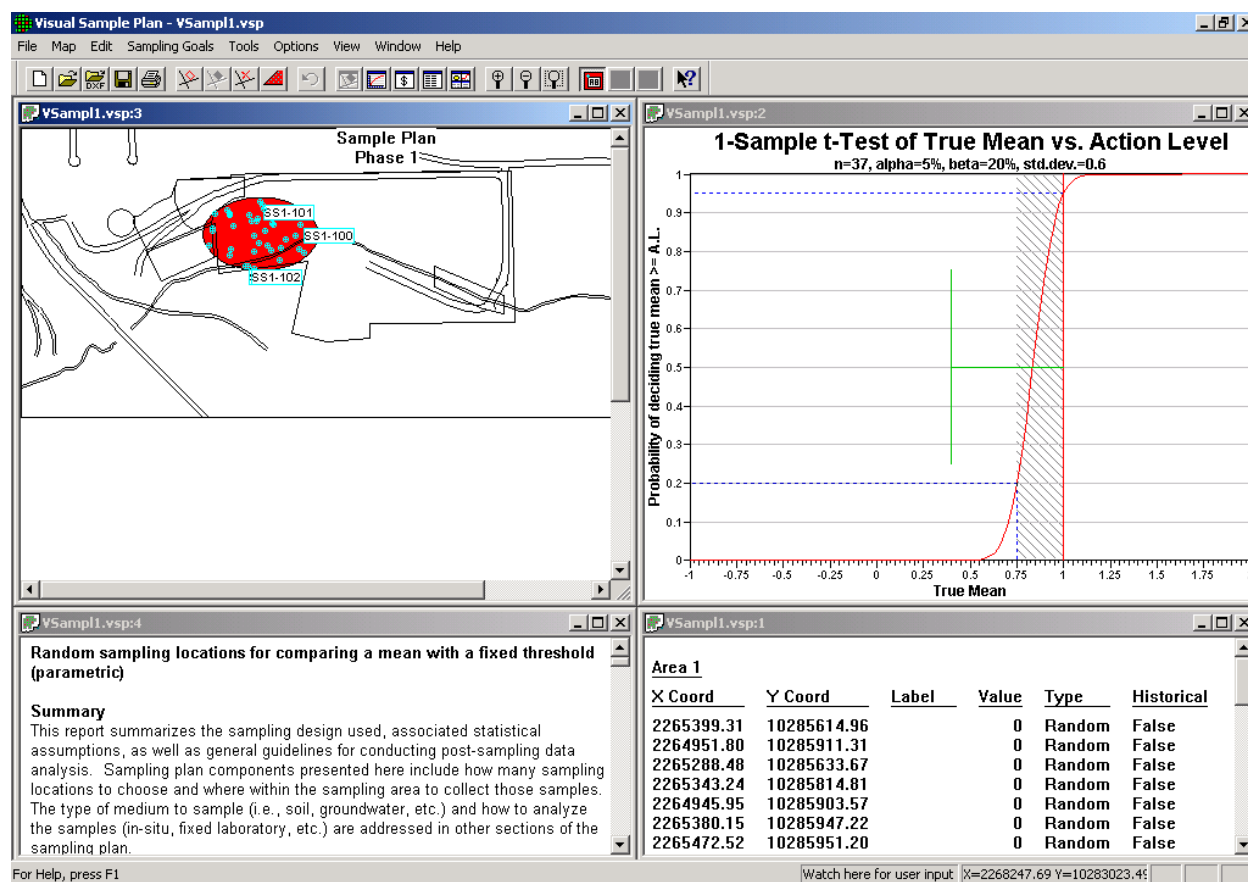

Figure 4.13. Quad Display of Map, Graph, Report, and Coordinates on Same Screen
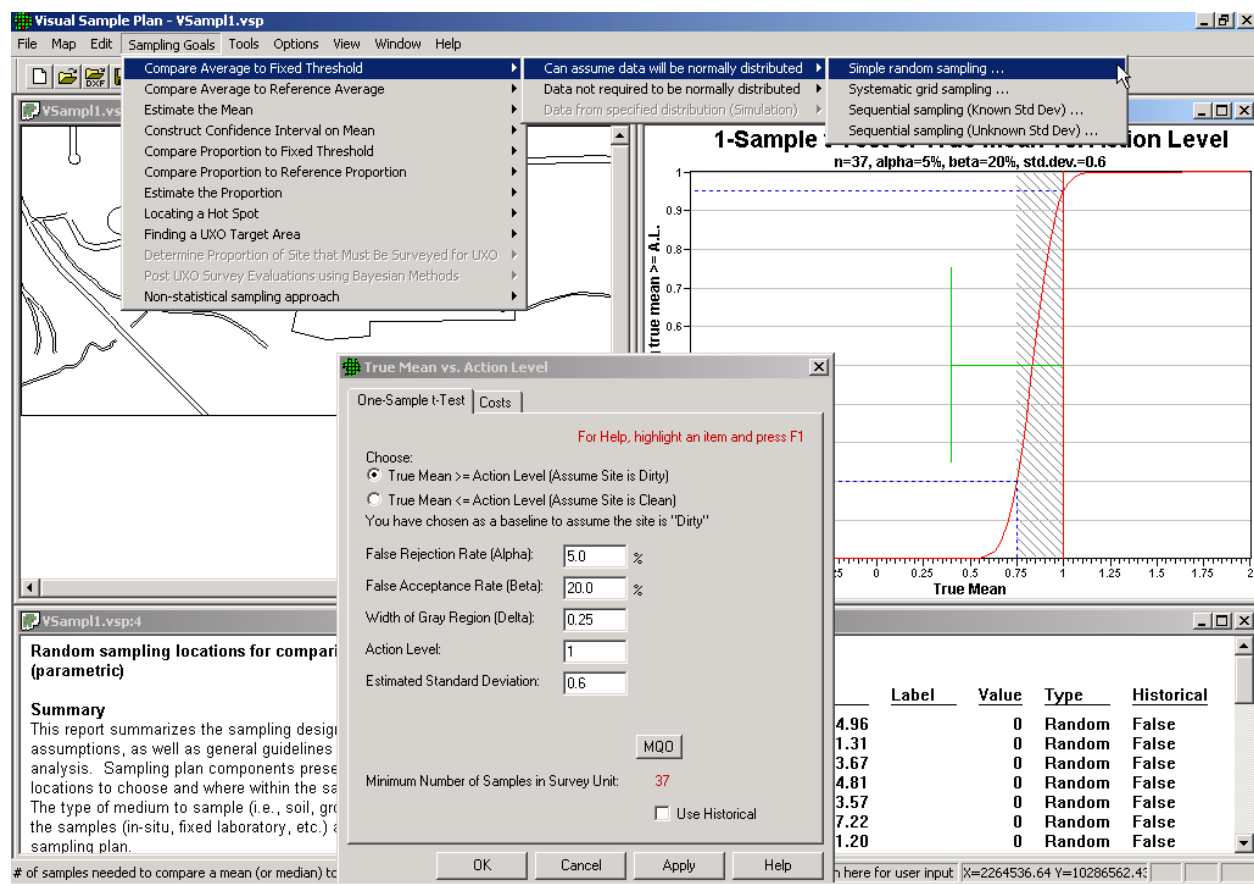
Figure 4.14. Combined Display of VSP Inputs and Outputs

5.0 Extended Features of VSP

VSP 2.0 has many extended features that have not been described so far in this user's guide. In this section, we discuss some key extended features. The beginning user may not need these features, but a more experienced user will find them invaluable. These features expand on VSP's core capabilities. They are useful once a user has identified 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 circle. We next assumed that a sampling-design algorithm not currently in VSP called for $n = 25$ samples. Using option **Sampling Goals > Non-statistical sampling approach > Predetermined number of samples > Simple random sampling**, VSP allocated 6 of the 25 requested samples to the rectangle and 19 to the circle. This is because the circle covers an area approximately three times larger than the rectangle.
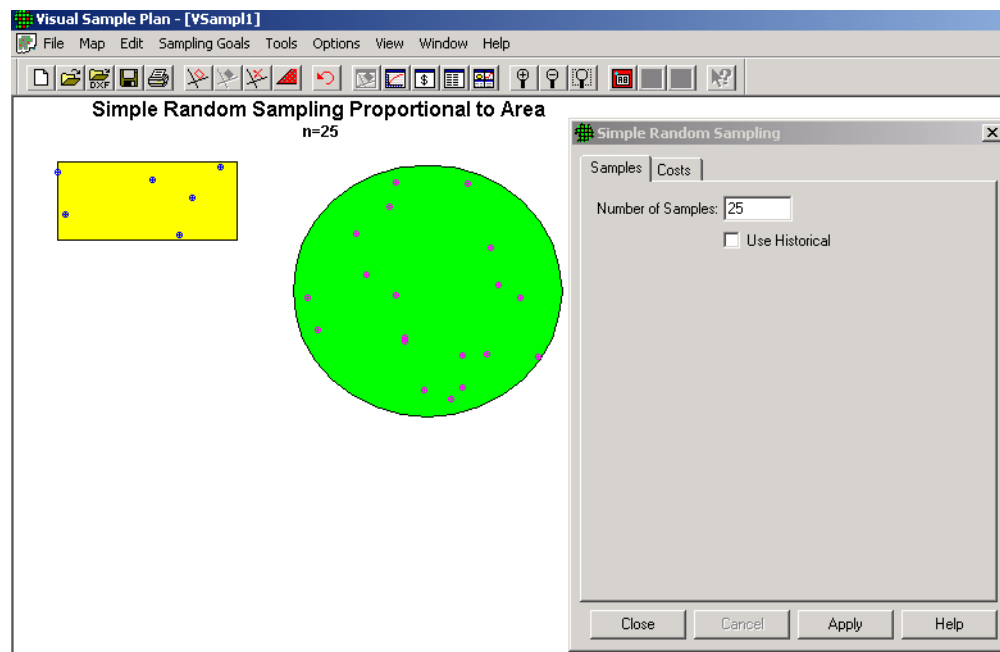

Figure 5.1. Proportional Allocation of Samples to Multiple Sample Areas

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.

5.2 Largest Unsampled Spot

If VSP has generated a sampling design for a Sample Area and you want to know the largest unsampled area, VSP can display this information. The largest unsampled spot is defined as the largest circle that will fit inside a Sample Area without overlapping a sample point. To find this area, use Main Menu option **Tools > Largest Unsampled Spot** > **Find**. This is shown in Figure 5.2. The **SHOW UNSAMPLED SPOTS** button on the toolbar can also be used to find the unsampled areas.
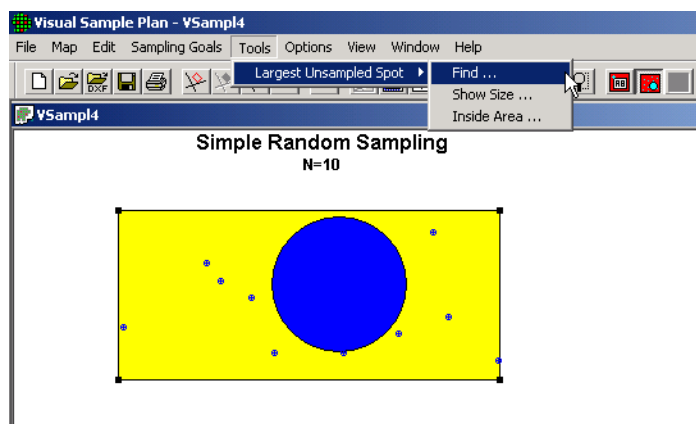
Figure 5.2. **Largest Unsampled Spot** Displayed on Rectangular Sample Area

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 Sample Area.
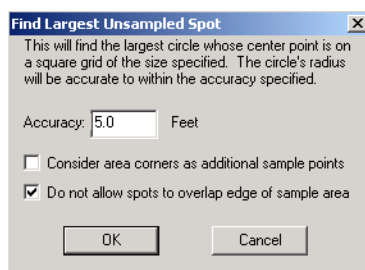
Figure 5.3. Dialog Box for **Largest Unsampled Spot** Allowing Accuracy and Other Options to be Specified

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 27.61 feet. The total area of the circle is 2395.20 square feet.

The other display is the **Inside Area...** dialog box, shown in Figure 5.5. It indicates that 100% of the circle is within the Sample Area. If you choose to allow the largest unsampled spot to overlap the Sample 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 Sample Area.
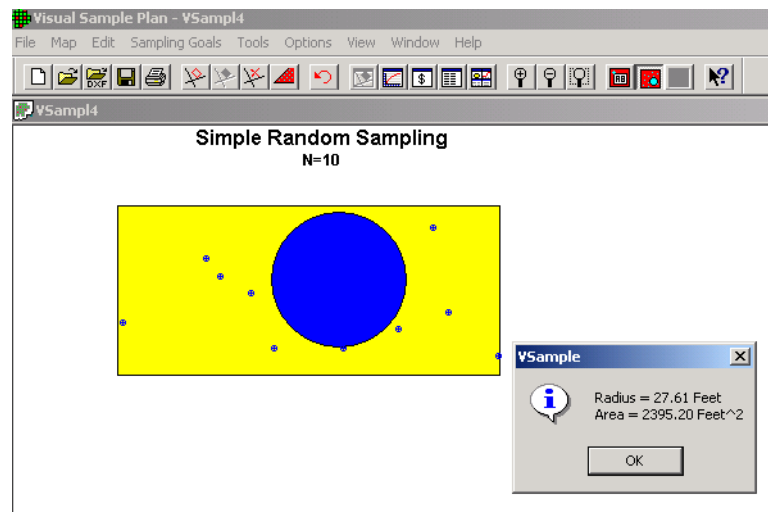
Figure 5.4. Dialog Box for **Largest Unsampled Spot** Showing Size of Circle That Would Fit into Largest Unsampled Area

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

Figure 5.6. Menu for Selecting Type of Random Number Generator

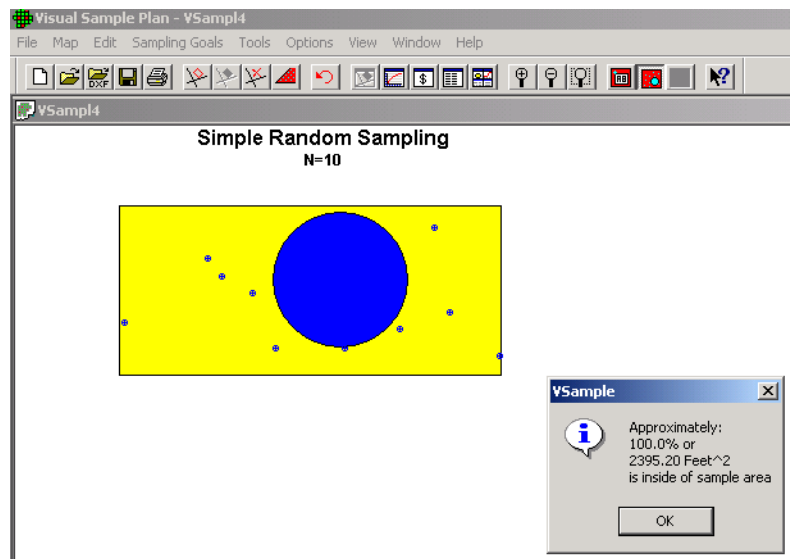The two options are **Pseudo-Random Numbers** and **Quasi-Random Numbers**. The user "toggles" between these two options. This is shown in Figure 5.6. Note that once an option is selected, it remains active until changed. VSP is initialized with the **Pseudo-Random Numbers** option active.
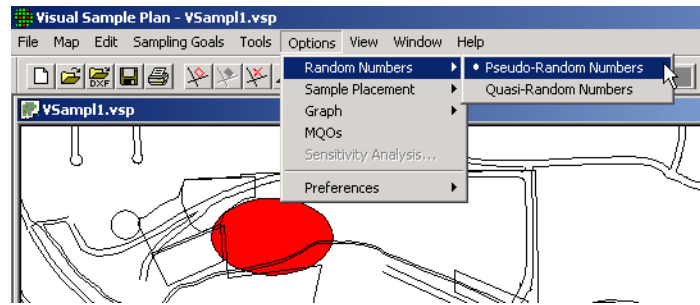
Sampling locations (i.e., the x and y coordinates of the location) chosen with a pseudo-random number generator are not restricted in any way. The first location chosen and the tenth location chosen can be right next to each other or far apart, like throwing darts at a dart board. The locations where 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 *Version 2.0 Visual Sample Plan (VSP) Models and Code Verification* (Gilbert et al. 2002).

If the current sampling design is being added to a study area with existing sampling locations, the quasi-random number generator will have no knowledge of those locations and might by chance put a new sampling location right next to an existing location. See the **Adaptive-Fill** option in Section 5.6 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 *Version 2.0 Visual Sample Plan (VSP) Models and Code Verification* (Gilbert et al. 2002).

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.

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.

Figure 5.7. MQO Input Dialog Box with Default Values Displayed

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.

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**.

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

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

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 **COSTS** 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 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.

Figure 5.10. Cost Input Dialog Box for MQO Option

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.

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

Figure 5.12. MQO Method Comparison Chart

5.5 Judgement Sampling/Manually Adding Samples

VSP allows you to manually add sampling locations to the Sample 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.
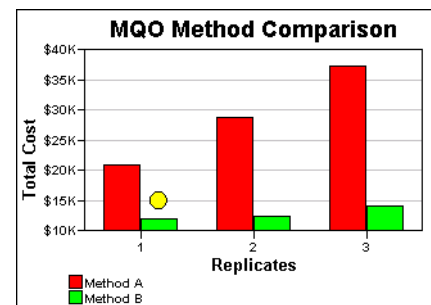
You may manually add sampling locations by selecting **Sampling Goals > Non-statistical sampling approach > Judgement (authoritative) sampling.** A Sample Area must have been defined and selected in order for this option to be available. In Figure 5.13 we selected as our Study Area the large elliptical area in the Millsite.dxf file. After selecting **Sampling Goals > Non-statistical sampling approach > Judgement (authoritative) sampling**, 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.

Note that when **Judgement 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.
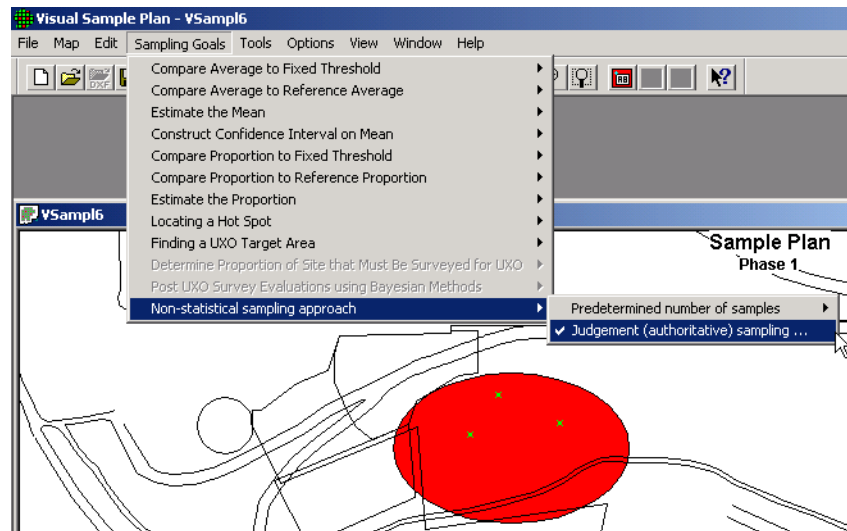
Figure 5.13. **Judgement Sampling** with Three Sampling Locations Added Manually

5.6 Adaptive-Fill Option for Placement of Sample Points

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

VSP usually places new sampling locations using the default option, **Options > Sample Placement > Regular Random.** When **Regular Random** is selected, the sampling locations produced by either of the two random number generators discussed in Section 5.3 are placed in the Sample Area without regard to pre-existing samples. In fact, VSP removes all previous sampling locations prior to placing the new set of sampling locations.

When the **Options > Sample Placement > Adaptive-Fill** option is selected, all pre-existing sampling locations are left in place, and new sampling locations are placed in the Sample Area using an algorithm to maximally avoid 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.
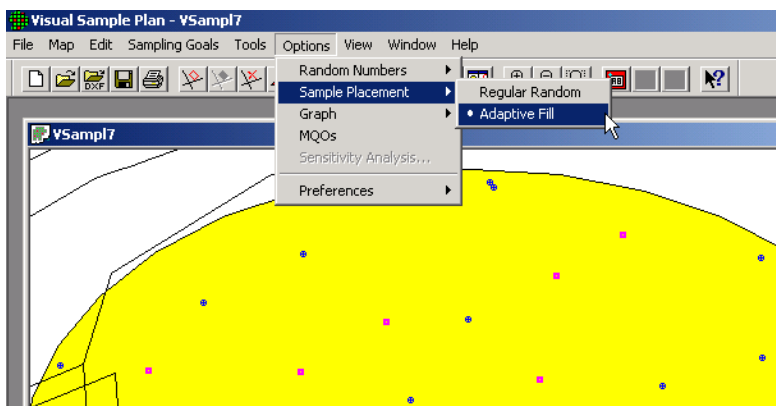
Figure 5.14. **Adaptive-Fill** Option for Sample Placement (Shown Here with Sample Area from Millsite Map)

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 displayed in only 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.
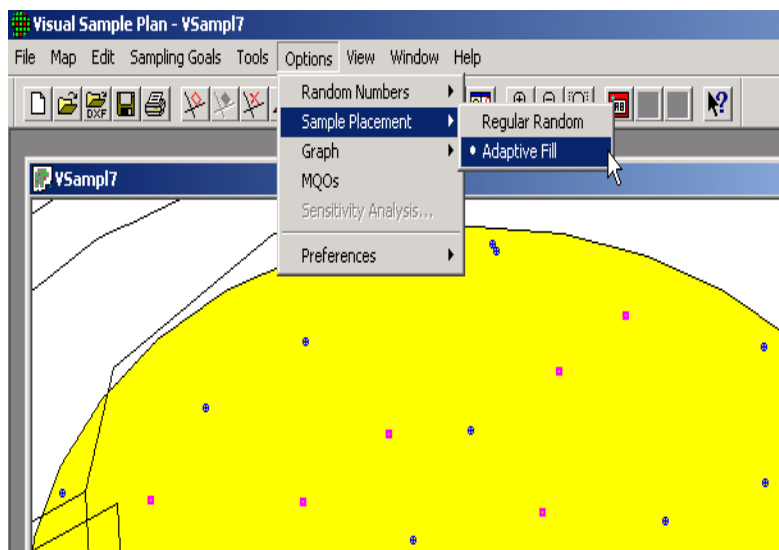


```
adptfill.TXT - Notepad
File  Edit  Format  Help
Area 1
X Coord Y Coord Label   Value    Type     Historical
2264873.4615    10285559.9455            0        Random
2264780.5956    10285719.5641            0        Random
2265427.4867    10285598.9890            0        Random
2265219.5557    10285885.4722            0        Random
2265439.5333    10285810.3616            0        Random
2264921.4536    10285904.2654            0        Random
2265160.2016    10285751.6503            0        Random
2265168.7425    10286032.0601            0        Random
2265249.1045    10286033.7379            0        Random
2265481.7182    10285672.9950            0        Random
2265552.4115    10285841.7924            0        Random
2264953.3272    10285708.4670            0        Random
2265154.6471    10285794.1171            0        Random
2265033.9677    10285958.5216            0        Random
2265192.6278    10285453.7425            0        Random
2265418.4434    10285528.8018            0        Random
2265267.1161    10285685.0950            0        Random
2265550.8324    10285953.9968            0        Random
2264759.6498    10285832.7508            0        Random
2265244.0361    10286039.6867            0        Random
2265287.3445    10285502.0855            0        Random
2265650.1054    10285666.5208            0        Adaptive-Fill
2265105.9763    10285611.1545            0        Adaptive-Fill
2265579.1681    10285526.1472            0        Adaptive-Fill
2265394.6907    10285979.6423            0        Adaptive-Fill
2265003.4862    10285461.9863            0        Adaptive-Fill
2265031.4461    10285825.8491            0        Adaptive-Fill
2265300.3467    10285816.6476            0        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. E. Wilson, R. F. O'Brien, D. K. Carlson, D. J. Bates, B. A. Pulsipher, and C. A. McKinstry. 2002. *Version 2.0 Visual Sample Plan (VSP) Models and Code Verification*. PNNL-13991, Pacific Northwest National Laboratory, Richland, Washington.

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

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

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

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

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

Distribution

No. of
Copies

OFFSITE

D. Bottrell
EM-5/CLVRLF
U.S. Department of Energy
1000 Independence Ave., SW
Washington, DC 20585

B. Murray
FERC/OGC-AL/825NCA
U.S. Department of Energy
1000 Independence Ave., SW
Washington, DC 20585

H. Peterson
EH-412/FORS
U.S. Department of Energy
1000 Independence Ave., SW
Washington, DC 20585

M. Carter
U.S. Environmental Protection Agency
Ariel Rios Building
1200 Pennsylvania Avenue, N.W.
Washington, DC 20460

D. Crumbling
U.S. Environmental Protection Agency
Ariel Rios Building
1200 Pennsylvania Avenue, N.W.
Washington, DC 20460

J. Griggs
U.S. Environmental Protection Agency
National Air and Radiation Environmental
 Laboratory/OAR
540 South Morris Avenue
Montgomery, AL 36115

No. of
Copies

T. Jover
U.S. Environmental Protection Agency
Ariel Rios Building
1200 Pennsylvania Avenue, N.W.
Washington, DC 20460

J. Warren
U.S. Environmental Protection Agency
Quality Assurance Division
1200 Pennsylvania Avenue, N.W.
Washington, DC 20460

W. Wong
U.S. Environmental Protection Agency
Ariel Rios Building
1200 Pennsylvania Avenue, N.W.
Washington, DC 20460

L. Zaragoza
U.S. Environmental Protection Agency
Ariel Rios Building, 5204G
1200 Pennsylvania Avenue, N.W.
Washington, DC 20460

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

M. Chambers
SERDP Program Office
901 N. Stuart Street, Suite 303
Arlington, VA 22203

No. of
Copies

K. Coats
U.S. Army Corps of Engineers
Hazardous, Toxic, and Radiological Waste
Center of Expertise
ATTN: CENWO-HX-C
12565 West Center Road
Omaha, NE 68144-3869

J. Fairbanks
SERDP Program Office
901 N. Stuart Street, Suite 303
Arlington, VA 22203

W. Ingersoll
NAVSEA
1661 Redbank Rd., Suite 104
Goose Creek, SC 29445-6511

D. Koran
CEMP-RT
U.S. Army Corps of Engineers
20 Massachusetts Ave N.W., Rm. 2204
Washington, DC 20314-1000

J. Marqusee
SERDP Program Office
901 N. Stuart Street, Suite 303
Arlington, VA 22203

F. McLean
NAVSEA
1661 Redbank Rd., Suite 104
Goose Creek, SC 29445-6511

No. of
Copies

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

G. Powers
U.S. Nuclear Regulatory Commission
11545 Rockville Pike
2 White Flint North
Rockville, MD 20852

M. Bertoni
1615 M Street N.W., Suite 740
Washington, DC 20036

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

ONSITE

DOE Richland Operations Office

E. M. Bowers	A2-15

17 Pacific Northwest National Laboratory

D. J. Bates	K5-12
D. K. Carlson	K5-12
J. L. Devary	K6-96
R. O. Gilbert (5)	K5-12
N. L. Hassig	K5-12
C. A. McKinstry	K5-12
R. F. O'Brien	K5-12
B. A. Pulsipher (5)	K5-12
J. E. Wilson	K5-12