Measurement Uncertainty in Visual Sample Plan (VSP)

Brent Pulsipher, Richard Gilbert, John Wilson

Pacific Northwest National Laboratory, Richland, Washington.

National Environmental Monitoring Conference, July 2003.

Abstract

Uncertainty is naturally inherent in any environmental measurement. Contributions to uncertainty can come from a variety of sources. When measurements are derived from analysis of field samples, uncertainties in the results can be due to large-scale spatial site variations, small-scale local in-homogeneity, sampling methods, sample handling, sample preparation, sub-sampling, and analytical variations. Each of these components of variation can be broken into additional sub-components that all combine to affect the total uncertainty. Visual Sample Plan (VSP) is a tool for determining the required number and placement of samples to ensure that the resulting data can support a sufficiently confident decision. It is important that users of VSP understand how the uncertainty estimates used in VSP represent the various components of variation described above.

This paper will show how VSP can be used to explore the relative contributions of sampling and analytical uncertainties to the total uncertainty. Using VSP, one can evaluate whether it is better to reduce sampling variations by obtaining more samples or improving the sampling technique verses conducting replicate analyses or using a more precise analytical technique. The Measurement Quality Objectives (MQO) option will be demonstrated and discussed.

Introduction and Background

Environmental characterization, remediation, and monitoring involve sampling of soils, surfaces, air, biota, or other media. Key questions that arise are how many samples are needed to support confident decisions and where should the samples be obtained. Visual Sample Plan (VSP) is a statistically-based software tool that assists in determining the right type, quality, and quantity of samples and recommended sampling locations [2]. VSP has been supported by several offices within DOE, EPA, and DoD.

VSP is based on the Data Quality Objectives (DQO) process [ref] which outlines the process for developing an optimal sampling strategy given specific sampling objectives, decision rules, and tolerance for decision errors. VSP currently supports characterization and assessment of surface soils, subsurface soil layers, sediments, building surfaces, and unexploded ordnance (UXO). Designed for the non-statistician, VSP is a visual, map-based tool that is organized around the possible data uses. Before developing a data-gathering plan, each user must determine what they will do with the data to support their decision-making process. VSP currently supports the following sampling goals or objectives.

- Compare Average Against a Threshold

- Compare Proportion Against a Threshold
- Compare Average or Proportion Against Reference Data (background)
- Find Hot Spot
- Find UXO Target Area
- Demonstrate Low Probability of UXO Presence
- Develop Confidence Interval
- Estimate a Mean
- Delineate Boundary of Contaminated Zone

Several options exist for the type for sampling design that might be appropriate for each of the above sampling goals. Both statistically-based sampling and non-statistical sampling approaches are available including:

- Simple Random Sampling
- Systematic Grid Sampling
- Non-Normal Distributional Approaches
- Sequential Sampling
- Stratified Sampling
- Rank-Set Sampling
- Adaptive Cluster Sampling
- Continuous Transect Sampling
- Judgmental Sampling

VSP has a number of unique, special features to facilitate use and support defensibility. Some of the diagnostic graphics are interactive, allowing immediate evaluation of tradeoffs between data quality objectives (DQO) requirements and costs. Sample locations are automatically displayed on maps and easily output to files that can be transferred to GPS units in support of in-field sampling and analysis activities. With each sampling plan, a 3-5 page report is automatically generated that documents all the site sample area information, the sample locations, map, diagnostic graphics, statistical assumptions, any formulas used, costs, and sensitivity analyses. The sensitivity analysis table is interactive and can be customized for each user. Finally, online help and technical documentation of the statistical methods are also freely available.

Measurement Uncertainty Within VSP

Decisions are usually based on individual or summary values (means or medians) associated with samples. These sample results are intended to represent the true state of nature at a site. However, it is known that the results are never a completely accurate representation of the state of nature. Departures from truth are introduced from several sources including large-scale spatial in-homogeneities across a site, small-scale in-homogeneity (co-located sample differences), sampling systems, sample handling, sample preparations, sub-sampling, and analytical instrument variations. By replicating within each of these possible uncertainty sources, it is possible to obtain estimates of the relative contribution of each source to the total uncertainty using statistical analysis of variance (ANOVA) techniques. Such complete replication is rarely possible in practice due to cost constraints and the desire to optimize resources. Most often, the uncertainties are either all combined into a total standard deviation estimate or are categorized into two groups: sampling variations and analytical uncertainties.

The number of samples required to support confident decisions is dependent on the sampling and analytical uncertainties. Within most of the VSP modules, a key user-specified parameter is the standard deviation. VSP assumes that this is the total combined sampling and analytical standard deviation unless the Measurement Quality Objectives (MQO) option is invoked. The MQO option allows one to input a sampling standard deviation and an analytical standard deviation separately. The total standard deviation (S_{Total}) is determined assuming an additive error model such that

$$S_{Total} = \sqrt{S^2_{Sampling} + S^2_{Analytical}}$$

where $S^2_{Sampling}$ is the sampling uncertainty due to large and small scale soil in-homogeneities and uncertainties introduced from the sampling process and $S^2_{Analytical}$ is the analytical uncertainty due to the process of preparing a sample for analysis and of analyzing the sample.

VSP MQO Example

For illustration purposes, rather than select a specific contaminant of concern, this example will be generic. Suppose that through the DQO process we have determined that we are interested in deciding whether the mean concentration of our particular contaminant of concern is above some action level. For this contaminant of concern, the action level is defined by the regulatory threshold which is 10 ppm. We will assume that the site is dirty unless proven clean. If the true contaminant concentration were 10 ppm or greater, then we want no more than a 5% chance of concluding it is clean (Type 1 error rate of 0.05). We also want no more than a 10% chance of concluding that the site is dirty if the true contaminant concentration were 8 ppm or less (Type 2 error rate of 0.10 and lower bound of gray region at 8). We have also found through similar studies on similar contaminant spread patterns and soils that the total estimated standard deviation is around 3 ppm.

Figure 1 depicts that typical VSP dialog box for this case without exercising the MQO option. Note that 21 samples are required to meet the desired decision error tolerances. Now suppose that the total estimated standard deviation can be decomposed into sampling and analytical components such that the sampling standard deviation is 2.9 ppm and the analytical standard deviation is 0.6 ppm. The total standard deviation is still 3.0 ppm. By selecting the MQO option, the sampling and analytical standard deviations can be input separately (see Figure 2). The sample size required remains the same.

With this option, the user can determine the effect of replicating analyses on each sample might have on overall sampling requirements. Figure 3 illustrates how adding 3 analytical replicates per sample has very little effect on the total number of samples required. This is primarily due to the fact that sampling uncertainties are the major contributor to the total standard deviation. Now suppose that the relative contributions to

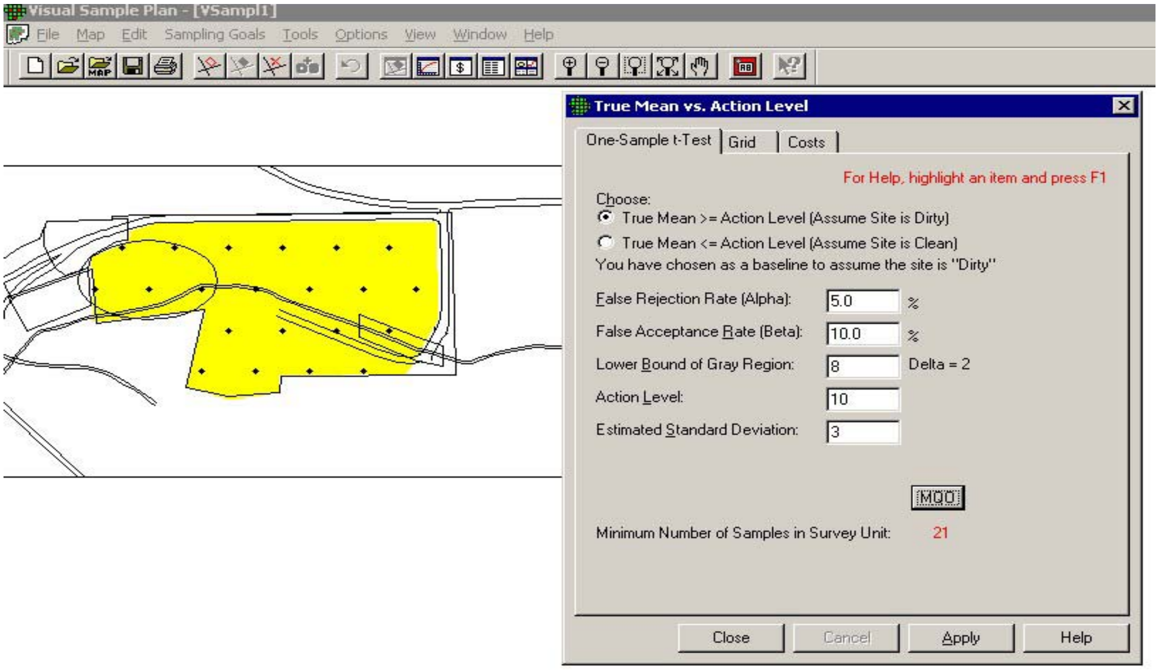

Figure 1. VSP Typical Map and Dialog Box

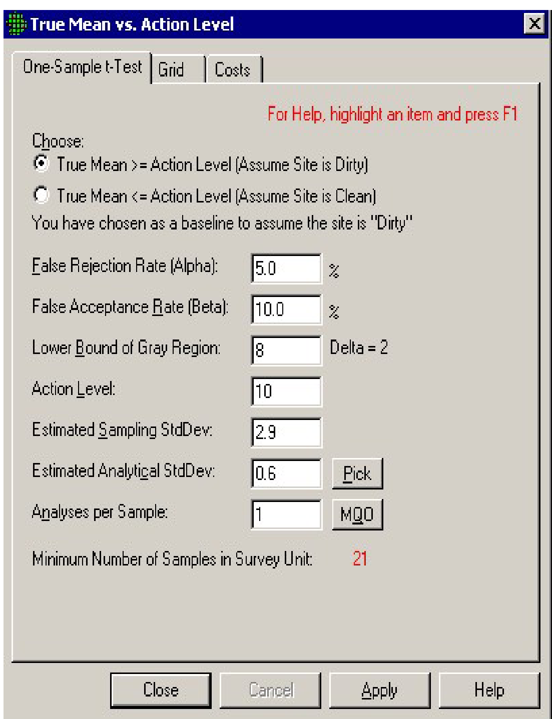

Figure 2. MQO Option Selected

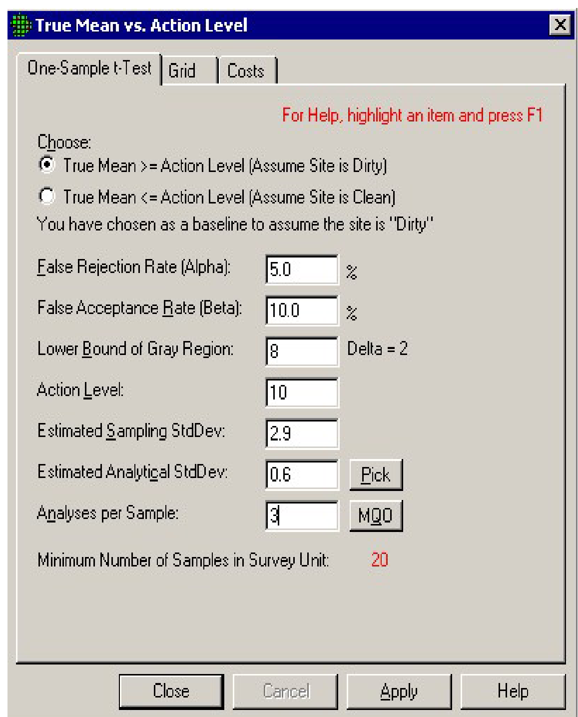

Figure 3. Triplicate Analyses per Sample

total error were reversed such that the analytical standard deviation were 2.9 ppm and the sampling standard deviation were 0.6 ppm. Figure 4 shows that by adding duplicate analyses per sample, the total number of samples required decreases significantly (from 20 to 12). Thus, if the sampling costs were very expensive compared to analytical costs and the analytical standard deviation were large relative to the sampling standard deviation, then duplicating the analyses would provide a significant savings.

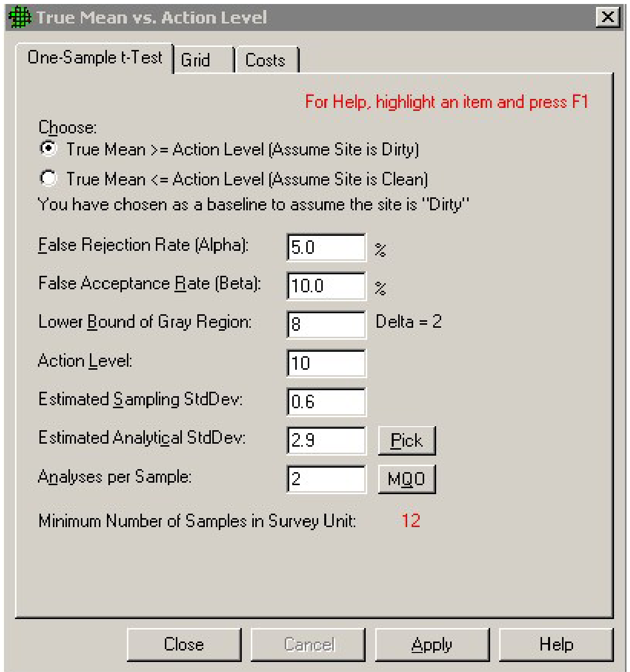

Figure 4. VSP Output With Large Analytical Standard Deviation

Using the VSP MQO Option to Compare Competing Analytical Methods

Given the greater availability of field-able analytical devices, many, including Triad advocates [3], are considering using these cheaper methods as long as decision error tolerances are controlled to acceptable levels. With the VSP MQO option in effect, it is easy to evaluate the tradeoffs between using a very precise but costly analytical method against using a more imprecise but cheaper analytical method assuming there are no detection limit issues.

Using our same example above, suppose we have two competing analytical methods as shown below.

Analytical Method	Sampling Std. Deviation	Cost per Sample	Analytical Std. Deviation	Cost per Analysis
Method A	2.9 ppm	$100	0.60 ppm	$400
Method B	2.9 ppm	$100	3.4 ppm	$35

Using this cost and stand deviation input and assuming a fixed planning and setup cost of $1000, VSP determines which analytical method is recommended and whether analytical replicates are needed. Figure 5 shows that all the required decision error tolerances can be achieved for the least cost by using the less precise, less expensive Method B with duplicate analyses per sample. More samples will be required (32 instead of 21) but the overall cost will be less. In fact, in this example, all DQOs are met with a cost savings of over $5000 by using the less precise analytical method.

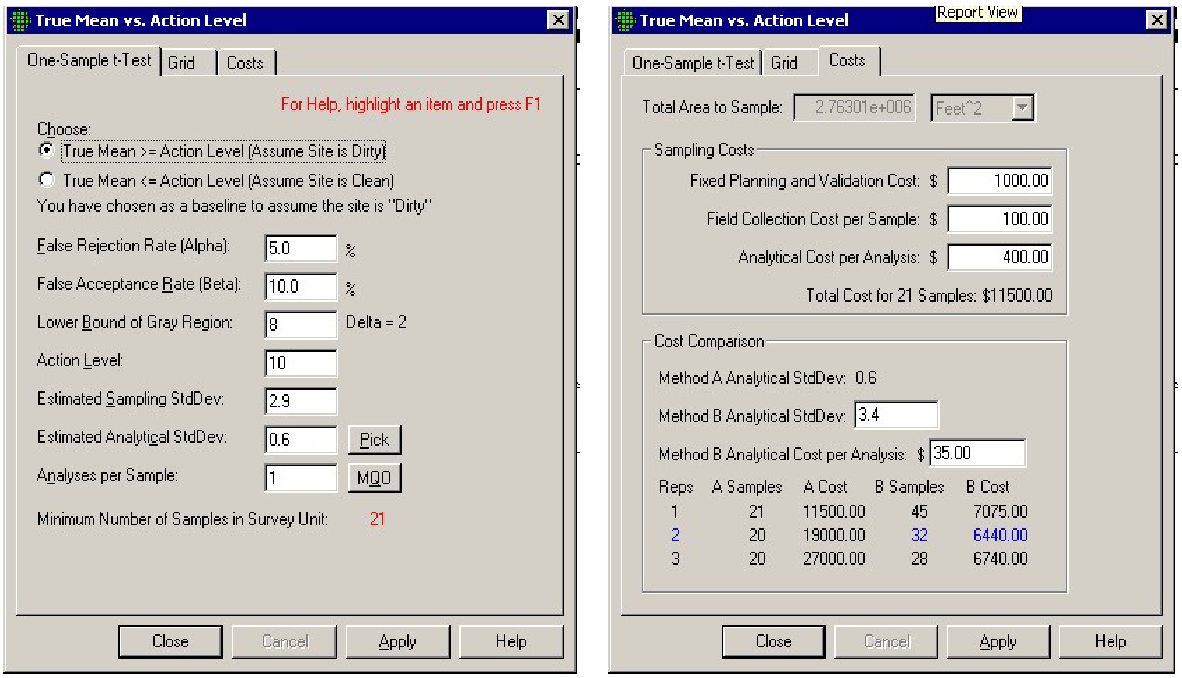

Figure 5. VSP Comparison of Competing Analytical Methods

Summary

With many emerging field deployable analytical systems, one will want to evaluate whether DQOs can be met with less precise but more real time analysis systems. VSP Measurement Quality Objectives options provide one way to evaluate the tradeoffs between competing analytical and sampling methodologies. Although more samples may be required for less precise analytical methods, overall costs may be reduced if the total standard deviation is significantly affected by analytical uncertainties. If local small scale contaminant variations in the soils is a major contributor to the total standard deviation, one could enter the large scale spatial standard deviation and the combined local small scale and analytical standard deviation into VSP to determine whether additional co-located samples or improved sampling methods might minimize costs while maintaining DQOs. VSP is a versatile tool for determining the right type, quality, and quantity of data is required to support confident decisions. VSP software, user's manual, and technical documentation can be downloaded from http://dqo.pnl.gov/vsp.

References

[1] US EPA. 2000a. *Guidance for the Data Quality Objectives Process - QA/G-4*. EPA/600/R-96/055, Office of Environmental Information, US Environmental Protection Agency, Washington, D.C.G4

[2] Hassig, NL, JE Wilson, RO Gilbert, DK Carlson, RF O'Brien, BA Pulsipher, CA McKinstry, DJ Bates. Sept 2002. "*Visual Sample Plan Version 2.0 User's Guide*", PNNL-14002, Pacific Northwest National Laboratory, Richland, WA.

[3] Crumbling, DM. 2001. *Using the Triad Approach to Improve the Cost-Effectiveness of Hazardous Waste Site Cleanups.* EPA 542-R01-016, US Environmental Protection Agency, Washington, D.C.