



U.S. DEPARTMENT OF
ENERGY

PNNL-19869

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

Visual Sample Plan (VSP) Statistical Software as Related to the CTBTO's On-Site Inspection Procedure

TC Pulsipher
SJ Walsh

BA Pulsipher
BD Milbrath

September 2010



Pacific Northwest
NATIONAL LABORATORY

*Proudly Operated by **Battelle** Since 1965*

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.

PACIFIC NORTHWEST NATIONAL LABORATORY

operated by

BATTELLE

for the

UNITED STATES DEPARTMENT OF ENERGY

under Contract DE-AC05-76RL01830

Printed in the United States of America

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

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



This document was printed on recycled paper.

(9/2003)

Visual Sample Plan (VSP) Statistical Software as Related to the CTBTO's On-Site Inspection Procedure

TC Pulsipher
SJ Walsh
BA Pulsipher
BD Milbrath

September 27, 2010

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

Pacific Northwest National Laboratory
Richland, Washington 99352

Executive Summary

In the event of a potential nuclear weapons test the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) is commissioned to conduct an on-site investigation (OSI) of the suspected test site in an effort to find confirmatory evidence of the nuclear test. The OSI activities include collecting air, surface soil, and underground samples to search for indications of a nuclear weapons test. These indicators include radionuclides and radioactive isotopes Ar and Xe. This report investigates the capability of the Visual Sample Plan (VSP) software to contribute to the sampling activities of the CTBTO during an OSI.

VSP is a statistical sampling design software, consistent with the systematic planning philosophy of the data quality objectives process. VSP modules have been developed to support environmental remediation and contamination detection problems for several US government agencies including the DOE, EPA, DoD, CDC, NRC, and DHS among others. This report outlines a number of VSP sample designs and statistical evaluations, which may be pertinent to the work undertaken during an OSI. The current VSP modules deemed most applicable or potentially applicable (with significant modifications) include

- Hot Spot Sampling
- Combined Random And Judgment Sampling
- Multiple Increment Sampling
- Radiological Transect Surveying
- Stratified Sampling (with modifications)
- Collaborative Sampling (with modifications)
- Adaptive Cluster Sampling (with modifications)

Each of these VSP modules is described and the applicability to OSI activities is examined. The use of such designs may enable canvassing a sample area most effectively, provide a measure of confidence that radionuclides are not present if unobserved, and allow investigators to refocus resources in other areas of concern.

Although some existing VSP modules may be directly applied to meet OSI needs, based on the unique aspects associated with an OSI (time constraints, phased sampling approach, multiple sample types, wide area vs. small area sampling, etc.), it is recommended that a holistic sampling design and statistical evaluation strategy be developed that meets all needs and objectives while dealing with every contingency that might arise. We envision that such a strategy should be documented and supported by a detailed decision flowchart that would identify the most appropriate, defensible sampling designs that should be employed that will result in greatest detectability while meeting feasibility requirements. In concert with this comprehensive sampling strategy, VSP tools and methods can be identified, developed, or modified to provide the OSI team with optimal sampling designs/locations that will meet all objectives at all stages of the investigation.

We propose creating a VSP module designed specifically to meet the sampling needs of the CTBTO's OSI process. VSP demonstrates the technical leadership necessary to implement such a large-scale sampling approach using a statistical framework and approach.

Contents

1 Introduction.....	4
1.1 Description of CTBT Sampling Problems	4
1.2 Visual Sample Plan Background.....	5
2 VSP Sampling Scenarios	7
2.1 Sampling Design Scenarios in VSP Pertinent to the CTBT.....	7
2.1.1 Hot Spot Sampling	7
2.1.2 Combined Random and Judgment Sampling	13
2.1.2.1 Presence/Absence Compliance Sampling	13
2.1.3 Multiple Increment (Composite) Sampling.....	17
2.1.4 Radiological Transect Surveying	20
2.2.1 Stratified Sampling.....	24
2.2.2 Collaborative Sampling.....	25
2.2.3 Adaptive Cluster Sampling	25
2.2.4 Tolerance Limits	28
3 Summary and Conclusions	30
4 Future Work / Recommendations	31
References	32

List of Figures

Figure 1: VSP view of sampling area, result graphs, automated report, and sample coordinates...	6
Figure 2: Initial Samples in a Region Highly Suspect to Contain a Detectable Hot Spot	9
Figure 3: Graphic Showing the Hot Spot Sample Design	10
Figure 4: Graphic Indicating Initial Judgment Samples	11
Figure 5: Largest Hot Spot that Would Not be Detected by Judgment Samples	11
Figure 6: Updated Sample Design, Conditioned on Initial Judgment Samples	12
Figure 7: Compliance Sampling Dialogue	15
Figure 8: Graphic Illustrating Placement of Judgment Samples.....	16
Figure 9: Graphic Illustrating the Placement of Both Judgment and Other Samples	17
Figure 10: Composition of an MI Sample	18
Figure 11: Multiple Increment Sampling Dialog Box	20
Figure 12: Multiple Increment Sampling - Group Increments in Localized Square Area	20
Figure 13: Radiological Survey Transects – Survey Pattern Options.....	21
Figure 14: Radiological Survey Transects - Transect Spacing on Sampling Area	22
Figure 15: Radiological Survey Transects – Geostatistical Methods to Map Survey Results	22
Figure 16: Identified and Delineated Hot Areas from a Radiological Survey	23
Figure 17: Screen shot of Multiple Sampling Areas Stratified by their Estimated Variability	25
Figure 18: Adaptive Cluster Sampling – Initial Samples Dialog.....	26
Figure 19: Adaptive Cluster Sampling – Grid Size and Number of Neighbors.....	27
Figure 20: Upper Tolerance Limit Design.....	29

1 Introduction

The purpose of this paper is to document Visual Sample Plan's (VSP) current capabilities and how they may be applied or adapted to facilitate the inspection phase of the Comprehensive Nuclear-Test-Ban Treaty (CTBT). A brief description of the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) on-site inspection (OSI) follows. A detailed description of several statistical sampling methods highlighting potential applicability to the OSI is the main focus of the paper. We conclude with recommendations to enhance VSP for applications more consistent with the objectives of the OSI.

1.1 Description of CTBT Sampling Problems

Under the verification regime of the CTBT, an on-site inspection (OSI) may be conducted at a location where a suspected nuclear explosion occurred if approved by the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO). The inspection area can be as large as 1,000 km², the location of which would likely be determined in part by seismic analysis from the CTBTO's International Monitoring System (IMS). The treaty limits the number of inspectors to 40 people and the inspection duration to 60 days, with the possibility of an additional 70 days if approved. Several methods for visual/aerial inspection and the monitoring of possible aftershocks would be initially employed during the OSI to further narrow the inspection area to a few smaller regions of interest. The OSI concentrates on sampling of soil, groundwater, and air to search for radioactive noble gases (radioactive Ar and Xe isotopes) and other radionuclides indicative of a nuclear explosion. The detection of such radioisotopes is considered to be the "smoking gun" that inspections seek. Anticipated paths to the surface for the radionuclides from an underground nuclear test include faults and other natural geological features; and radioactive noble gases could be detected near fractures or trees with deep roots. Thus visual/aerial inspection combined with knowledge of the geological features would lend itself to an initial judgment sampling approach.

The sampling approach can be categorized into two levels of sampling, namely, as a "Wide Area Sampling Problem" (~1000 km²) and a "Fine Sampling Problem" (~1 km²). The wide area sampling problem consists of determining the initial sampling locations across the 1,000 km² area considering multiple expert opinions that may or may not be driven by visual/aerial inspection and/or natural geological features. Reduction of the area of interest from 1,000 km² to smaller regions of interest is the primary goal of the wide area sampling problem.

In the fine sampling problem the inspection team (IT) will likely perform a gamma radiation survey before collecting soil samples. Under section 6.6.50 of the Model Text of the draft OSI Operational Manual (CTBT/WGB/TL-18/40), "in situ radionuclide analysis can provide information on the presence and identity of specific radionuclides distributed as localized 'hot spots' or relatively uniformly." After detection of a radiation anomaly from the gamma radiation survey it is believed that a "collection of soil samples will yield the most pertinent information on the presence of radionuclides of OSI interest." (6.6.58) Hand-held gamma radiation monitors may assist to locate possible soil sampling locations. It may be appropriate to take a composite sampling approach in areas of relatively low radioactivity. (6.6.61)

The fine sampling problem assumes the team may find N1 number of "hot spots" or smaller regions of interest where environmental sampling as stated in various sections of the Model Text (e.g. 6.6.51) or shallow drilling for seeping radioactive noble gases would help verify the presence of radionuclides listed in Annex 1.4. These smaller regions of interest may be areas of size approximately 1 km². In the case of sampling for noble gases time allows for 15-20 drill holes strategically and spatially dispersed across each of the N1 suspected hot spots to collect the samples.

The current implementation of the CTBT suggests sampling of two media (soils and air) in a somewhat sequential and adaptive manner to determine the presence/absence and possible locations of surface radionuclide hot spots or noble gases seeps from underground. Surface soil sampling and the collection of air samples are two of the sampling schemes now included in the CTBT recommended sample plan. Air samples will be taken above ground but most of the effort is expected to focus on underground samples from drilled wells to detect radioactive noble gases gradually reaching the surface through atmospheric pumping. Such underground soil air samples would be taken continuously in anticipation of a result above background. For surface soil samples, the expectation is that if no radionuclides are detected upon arrival, none would be detected in the future.

Although the CTBT is not currently in force, the capabilities needed are being developed under the auspices of the Provisional Technical Secretariat (PTS) of the CTBTO. Though the IMS is fairly robust, the OSI capability is at this point much less developed. This includes not only the needed equipment but also the understanding of the necessary concept of operations (con-ops) for the many activities that must happen during an OSI, such as radionuclide sampling. Tools and methods are needed to help determine the number and optimal placement of soil and air samples that will optimize detectability while balancing against time and cost constraints. These statistical-based methods would also allow quantification of decision confidence and uncertainties. Without the development of tools to provide more sophisticated, technically defensible sampling plans, one would have to rely on a purely judgmental approach based on expert opinion only or on a simple grid-based sampling scheme and quantification of decision confidence would be difficult, especially when no radionuclide or noble gases are detected.

In the next section we provide some background on Visual Sample Plan (VSP), a PNNL developed software tool that helps one develop an optimal sampling approach and perform statistical evaluations for a variety of applications. Descriptions of several existing VSP modules that are either directly applicable or adaptable to the OSI wide area or fine sampling problems are provided. Other sampling approaches/methods not yet fully developed in VSP are also discussed. It appears that VSP could aid in optimizing sampling approaches while quantifying decision confidence and uncertainties.

1.2 Visual Sample Plan Background

The Visual Sample Plan (VSP) software facilitates decision making for problems that require sampling from buildings or geographic regions. VSP contains sampling designs and data analyses, which are built around the Data Quality Objectives process (DQO). Depending on the application, VSP helps investigators choose a sampling design appropriate for their problem, helps determine the number of samples and sampling locations under a specific DQO, enables statistically defensible and confident decisions, and performs statistical and data quality assessments (diagnostics) in support of the decision making process.

VSP v 1.0 was released in 2001 after nearly 9 years of research and development of its core tools by researchers at Pacific Northwest National Laboratory (PNNL) and Oak Ridge National Laboratory (ORNL). The current version of VSP (6.0) was released in June 2010 and culminates research, development, evaluation, and testing supported by the following sponsors: US EPA – Office of Solid Waste and Emergency Response and Office of Environmental Information; US DOE – Office of Health, Safety, and Security, Office of Legacy Management, and Office of Environmental Management; US Department of Defense – Navy, Strategic Environmental Research and Development Program, and Environmental Security Technology Certification Program; US DHS – Technical Support Working Group and Chemical and Biological Countermeasures, UK Atomic Weapons Establishment; and US Centers for Disease Control – NIOSH. Nearly 5,000 individuals at various government, academic, and private contracting agencies currently use VSP. VSP is free, public domain software.

VSP adheres to the systematic planning philosophy of the DQO process, which contains seven steps:

1. State the problem
2. Identify the goals of the study (what decisions need to be made?)
3. Identify information inputs
4. Define the boundaries of the study
5. Develop the analytical approach (create decision rules)
6. Specify performance or acceptance criteria (tolerance for decision errors)
7. Develop the plan for obtaining data (create the sampling design)

VSP uses as input information derived from the first 5 DQO steps and then facilitates user-friendly implementation of steps 6-7 for the non-statistician. VSP's sampling options are heavily driven by the EPA's G-5S guidance and include, but are not limited to the following: Simple Random Sampling, Systematic Grid Sampling, Sequential Sampling, Collaborative Sampling, Stratified Sampling, Rank-Set Sampling, Adaptive Cluster Sampling, Continuous Transect Sampling, Judgmental Sampling, and Combined Judgment and Probabilistic Sampling.

VSP allows the user to import a map and/or image of varying formats to use as the template to create sampling areas of interest, from which samples will be obtained. Several coordinate systems are supported and the sample locations can be output for use in GPS systems. VSP also automatically creates a report, which is populated based on the sampling goals and area of interest, complete with tables, graphics, sampling design results, statistical assumptions, analysis, and conclusions. This is particularly useful in making statements regarding DQO methods, sampling schemes, and interpreting results. These features are demonstrated in **Figure 1**, a screen shot of the sampling area, results, automatically generated report, and sample location coordinates. In the remainder of this report we illustrate and describe some VSP sampling methodologies that may be pertinent to the CTBT inspection problem.

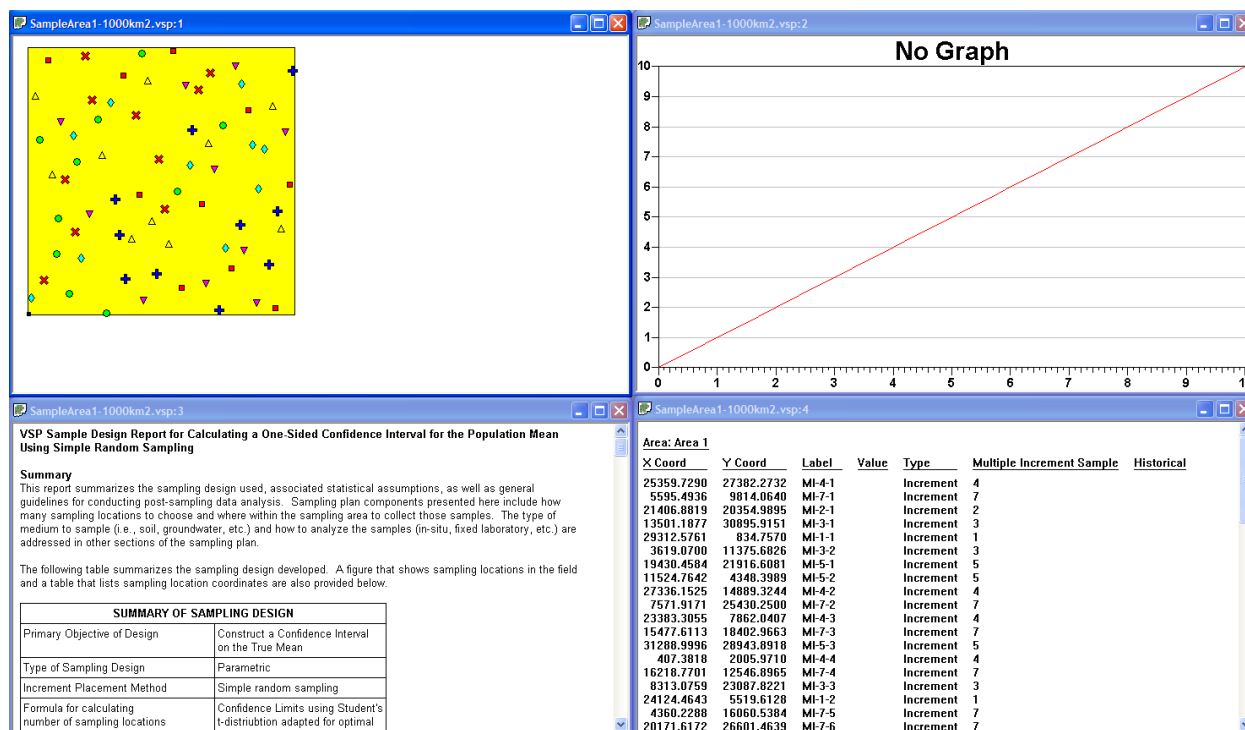


Figure 1: VSP view of sampling area, result graphs, automated report, and sample coordinates

2 VSP Sampling Scenarios

In this section we present and describe a partial set of VSP's sampling design modules. The designs presented in this section are identified as having some relevance to the CTBT sampling scenario. The intent of these illustrations is to help communicate how they may be useful for the CTBTO's OSI activities. We describe the motivation for applying each methodology, give a basic description of the methodology, describe the assumptions invoked when applying the particular sampling methodology, illustrate some of the capabilities of the methodology, and describe how it may be applicable to the CTBT sampling scenario. The provided illustrations should not be considered an exhaustive representation of VSP's sampling designs and capabilities.

2.1 Sampling Design Scenarios in VSP Pertinent to the CTBT

2.1.1 Hot Spot Sampling

Hot spot sampling is applicable when it is necessary to determine with a specified high probability that no hot spots of a specified size and shape exist in the study area. A hot spot is defined as a local contiguous area that has concentrations that exceed a threshold value, that is, that a constituent of concern (COC) can be measured at some level as to distinguish the measurement as elevated from background. Initially, the conceptual site model should be developed and used to hypothesize where hot spots are most likely to be present. *Initially, the regions which are hypothesized to contain the hot spots are sampled and it is determined if the hot spot exists at these regions.* If no hot spots are found by sampling at the most likely locations, then VSP can be used to construct a sampling design over a systematic grid. The attributes of the grid (size, layout, location) are a function of user specified inputs, for example: size of suspected hot spot, desired probability of detecting a hot spot of specified size, cost constraints, etc. Samples (measurements) are made at each node of the systematic grid.

There are four objectives to which the hot spot sampling approach may be applied. Note that under each objective, the design computes an unknown quantity of interest (e.g. the smallest hot spot size that may be detected under the current sample grid in number 4.) along with the sample design. Outputs of these objectives are as follows:

1. The number and spacing of samples on the systematic sampling grid required to achieve a specified high probability that at least one of the samples will fall on a circular or elliptical hot spot of the specified size.
2. The probability that at least one of the samples collected at the nodes of the specified systematic sampling grid will fall on a circular or elliptical hot spot of specified size.
3. The probability that at least one of the samples will fall on a hot spot of the specified size given that the spacing between nodes of the systematic sampling grid is the minimum that can be achieved with project funding (cost constrained).
4. The smallest size circular or elliptical hot spot that will be detected with specified high probability by sampling at the nodes of the systematic sampling grid.

For conclusions to be considered valid and reliable, the sampling scenario must conform to the assumptions upon which the hot spot sampling design is constructed. These assumptions are as follows:

1. The shape of the hot spot of concern is circular or elliptical.
2. The level of contamination that defines a hot spot is well defined.

3. The location of the hot spot is unknown, and if a hot spot is present, all locations within the sampling area are equally likely to contain the hot spot.
4. Samples are taken on a square, rectangular or triangular grid pattern.
5. Each sample is collected, handled, measured or inspected using approved methods that yield unbiased and sufficiently precise measurements.
6. A very small proportion of the surface being studied will be sampled (the sample is much smaller than the hot spot of interest).
7. Sample locations are independent of the measurement process.
8. The systematic grid is placed at a randomly determined starting place to cover the surface area of interest.
9. There are no classification errors (if a hot spot is sampled, it is not mistakenly overlooked or an area is not mistakenly identified as a hot spot).
10. A hot spot is detected if any part of the hot spot lies within a grid cell sample.

VSP contains a suite of GIS and CAD tools that allow one to import geo-located maps and construct graphical representations of the region to be sampled. These tools enable the construction of highly detailed maps, which can be tied to the real world. We illustrate this capability through construction of a simple arbitrary sample region. Before using the hot spot sampling module to generate a sample design, let us consider that we have prior information that a hot spot likely exists in the lower left region of the sample area. Thus we initially allocate, say, 12 judgmentally located samples here to see if we can detect the hot spot. We present a VSP screen shot of the initial samples (the sampling area is approximately 31 km × 31 km, or an area of 1,000 km²) in Figure 2. If we did not detect the hot spot in this region we may wish to canvas the rest of the sampling region for a hot spot of specified size. This may be accomplished by using VSP to create a hot spot sampling design.

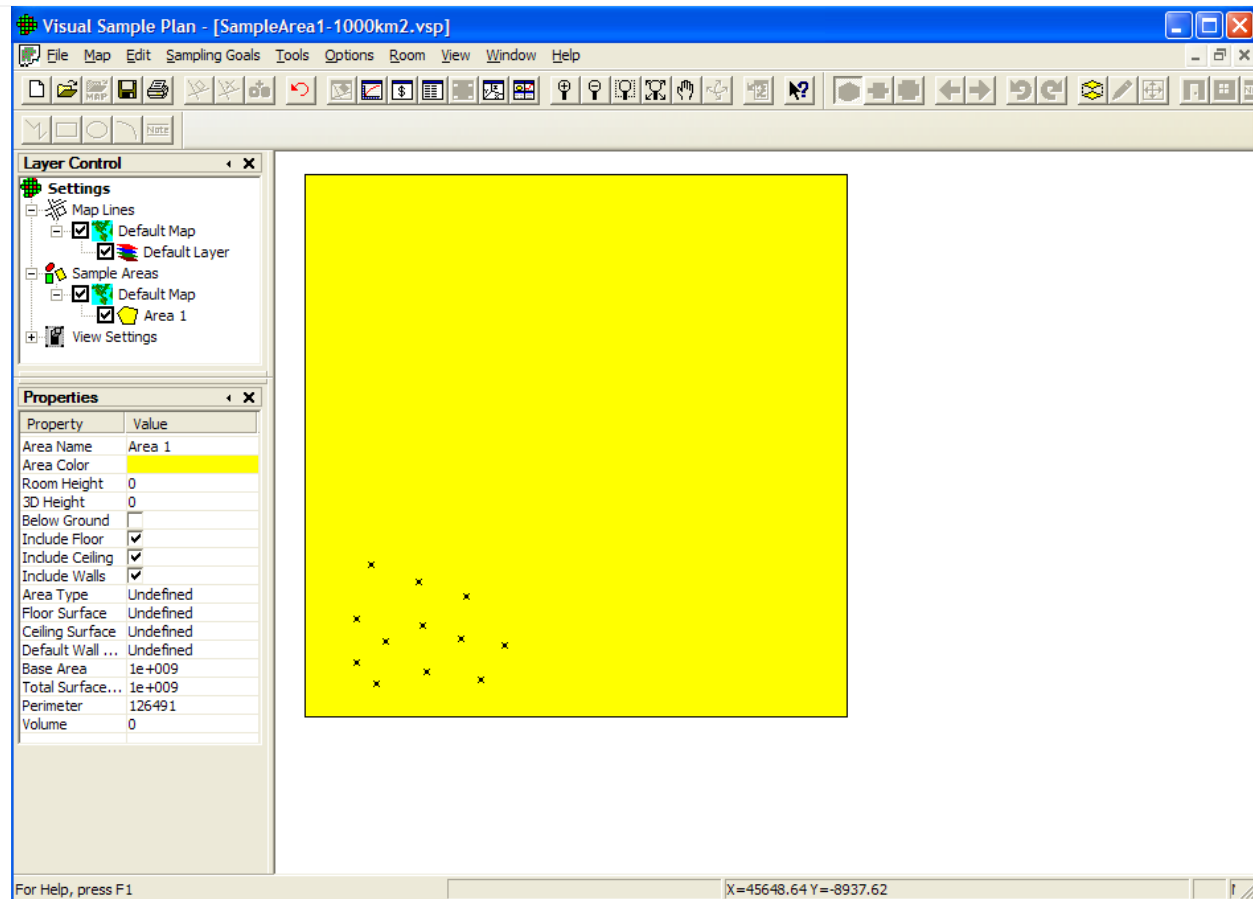


Figure 2: Initial Samples in a Region Highly Suspect to Contain a Detectable Hot Spot

Recall that VSP can build this design while solving for one of three free variables: grid spacing conditioned on total cost, probability of finding a hot spot of specified size, and hot spot size. Let's say that we wish to construct a sampling grid that has a 95% chance of detecting a circular hot spot with a 7,000 ft radius. We note that this approach would disregard the samples already taken in the lower left corner of the sample region (since we did not detect the hot spot) and will canvas the entire region. Application of the sample design results in a randomly placed regular grid of sample points. We illustrate the new sample design in Figure 3. Again, this scenario may be run under constrained resources (cost) or constrained probability of detecting a hot spot (e.g. we needed to be 100% guaranteed that we would detect a hot spot).

VSP may also use a current sampling design to allocate new samples under some specified criteria. For example, consider we initially placed some judgment samples over the sample region. We present a graphic of our judgment samples in Figure 4. Given these judgment samples VSP can identify the largest hot spot that would not be detected by existing samples. As a first step we can ask VSP to show visually the largest hot spot that would not be detected by existing samples. A graphic showing the largest hot spot not detected by existing samples is displayed in Figure 5.

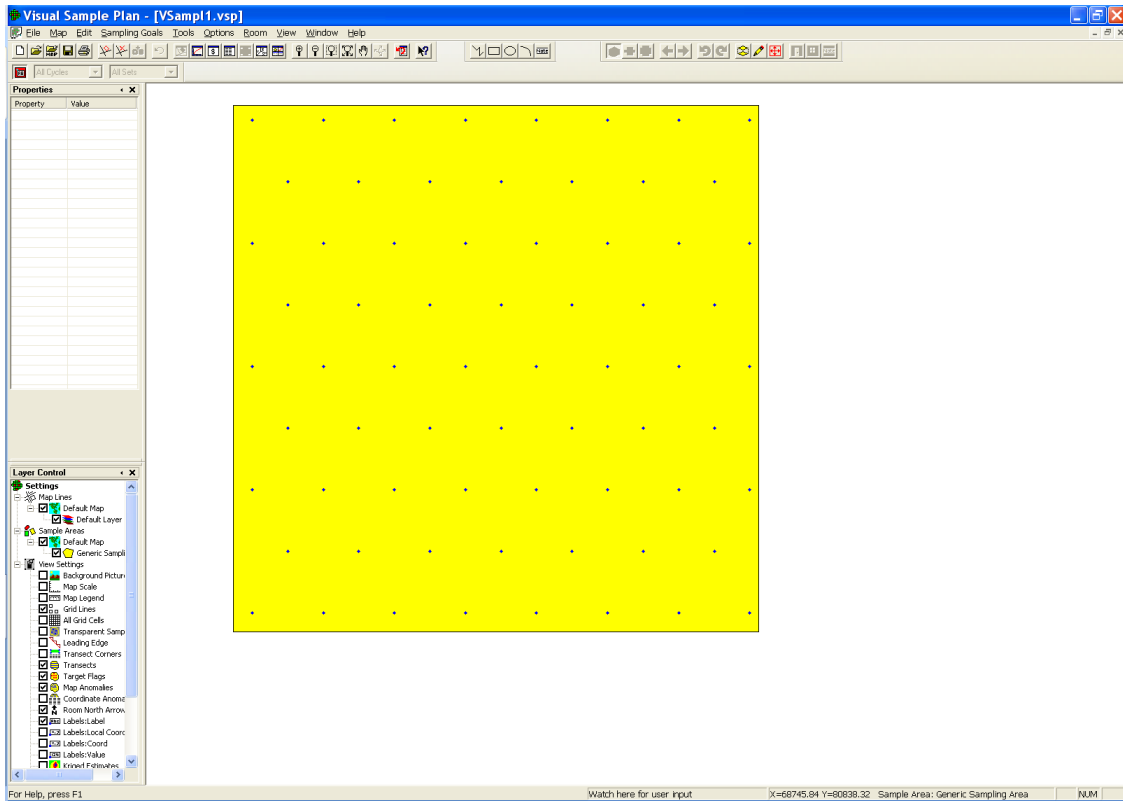


Figure 3: Graphic Showing the Hot Spot Sample Design which can detect a 7,000 ft radius circular hot spot with a 0.95 probability

We might then consider placing a sample at the center of the largest unsampled region and perhaps iterate that process of placing a sample at the center of the largest unsampled region until the largest unsampled region was of acceptable size. Another approach considers conditioning on existing samples and overlaying a new sample set such that there would be no unsampled areas of, say, 9000 ft. VSP will condition on already placed samples and overlay new samples whence the specified criteria. VSP will also illustrate the next largest area that was not sampled under the design. An example of this computation is presented in Figure 6. This approach may be conducted iteratively.

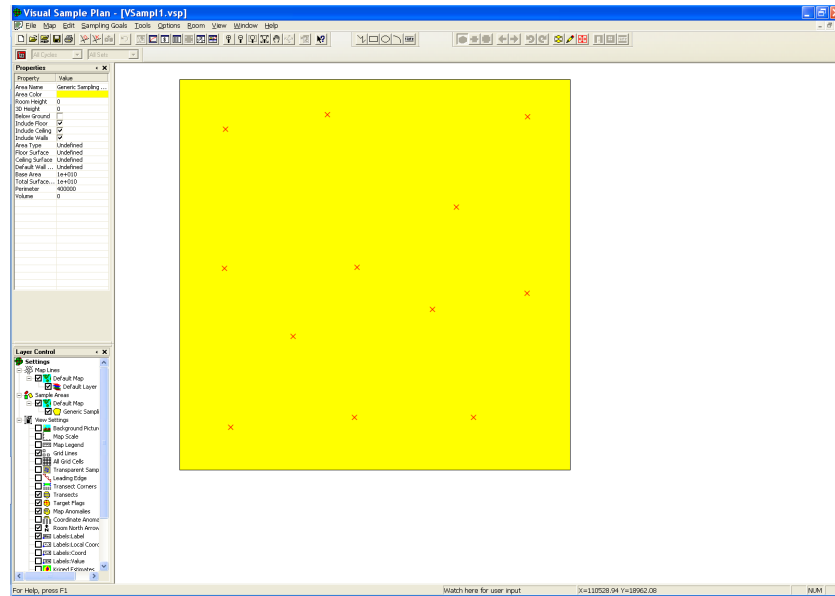


Figure 4: Graphic Indicating Initial Judgment Samples

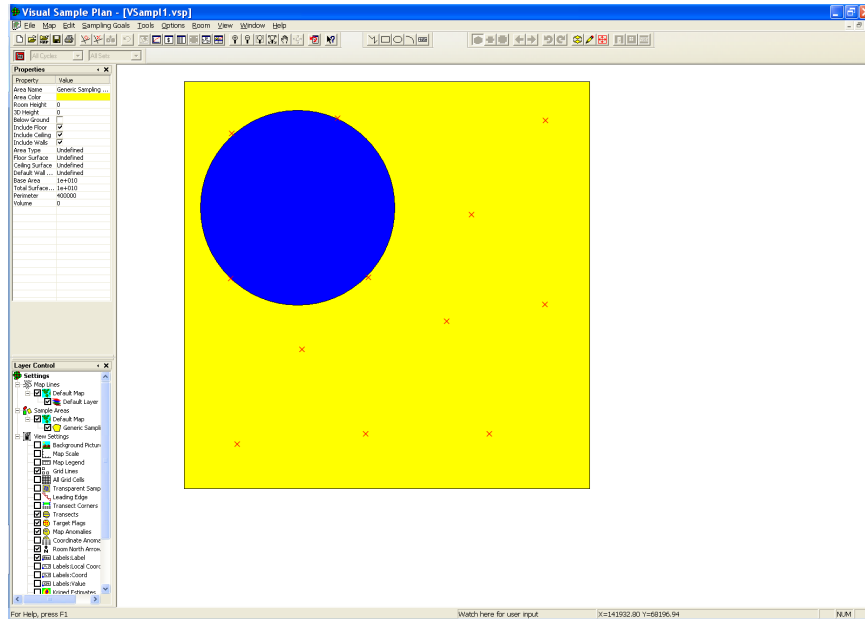


Figure 5: Largest Hot Spot that Would Not be Detected by Judgment Samples

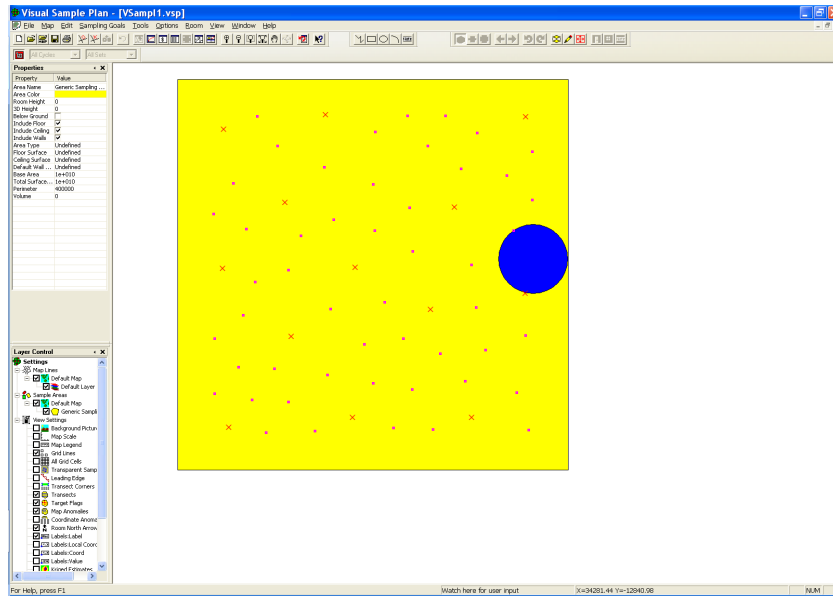


Figure 6: Updated Sample Design, Conditioned on Initial Judgment Samples, which will detect a hot spot area of 9,000 ft radius, next largest unsampled area (blue)

In this section we have illustrated some basic capabilities of the hot spot sampling design in VSP. The illustration and description of the hot spot design is not exhaustive; more capability for the design exists and can lead to more complicated application of the design. A synopsis of this design follows:

1. The hot spot sampling design is useful if the quantity of interest can be considered to be in an elliptical (or circular) area over the sample region.
2. The hot spot design can be constructed to detect a hot spot of specified size with some large probability.
3. The hot spot design can be constructed under cost constraints, and will compute the smallest hot spot size that could be detected under the cost constraints.
4. Judgment samples can be placed initially and conditioned upon to update the sampling design. This can be an iterative process; one may identify the next largest unsampled region and place a single sample, or, one may specify the size of a hot spot that must be detected and VSP will compute sample locations to accomplish this conditioned on current samples.

Applicability to OSI:

The hot spot sample design is a method to canvas a region (both for the wide area and fine sampling problems). Likely during an OSI there will be a few initial areas of interest as identified by visual observation where surface soil samples could be quickly obtained. Aerial- or ground-based gamma radiation surveys may identify potential hot spot locations of elevated radioactivity that would be sampled to determine if the radioactivity was caused by radionuclides indicative of a nuclear explosion. If neither of these preliminary sampling locations provided sufficient detection, VSP could provide a sampling plan to maximize the remaining sampling effort, rather than just rely on a strict grid-approach. Input would include the existing samples already taken balanced with the expected size of nuclear test hot spots as derived from underground nuclear explosion data from the Nevada Test Site for events that leaked. In the case of underground soil sampling for noble gases in a small area, after a few sampling locations are determined based on favorable geological features, VSP could advise on the best placement of the remaining sampling locations.

2.1.2 Combined Random and Judgment Sampling

Combined judgment and random (CJR) sampling can be used on binary presence/absence data or with quantitative measurements. We will discuss each of these in turn.

2.1.2.1 Presence/Absence Compliance Sampling

The objective of presence/absence compliance sampling is to verify, with some high probability, that a large percentage of the decision area (or population) is acceptable, based on 100% of the prescribed number of samples being unacceptable. Similarly, the objective could be to ensure a high probability that at least one sample detected unacceptable conditions if in fact at least X% of the total area is unacceptable.

A hypergeometric model is applied in this design and requires that each sample result be categorized as binary, such as:

1. The presence of a particular quality
2. A sample result being acceptable or unacceptable as defined by an action level threshold (i.e. labeling a quantitative measure as detect vs. non-detect)
3. Contamination being detected/not-detected

In compliance sampling for binary measures, the sampling area is divided into non-overlapping equal-size grid cells of specified size. The size of the grid cell should correspond to the footprint of the sampling methodology, that is, the grid cells should be areas that can be considered completely canvassed by the sampling. If more than one sampling methodology is to be employed in a decision area, the size of the grid cell should be chosen to match the sampling methodology with the smallest footprint. Compliance sampling is especially suited for use in decision areas where very few of the grid cells may be unacceptable. If at any time during the sampling process, one of the samples is found to be unacceptable, the decision area is declared to be unacceptable and no further samples for this design need be taken (unless further delineation of the unacceptable area is desired). If this occurs, it may be desirable to implement a hot spot or geospatial sampling plan to characterize the extent of the unacceptable locations or items.

Given some specified criteria, this design will grid up the decision area, compute the number of samples required, and randomly distribute them across the sample region. The underlying driver for a compliance sampling design is what is commonly referred to as a tolerance statement. Specifically, the tolerance statement is in the form of:

We are X% confident that at least Y% of the grid cells are acceptable (not contaminated).



Thus if none of the measured grid cell samples are unacceptable, the above confidence statement reflects the appropriate statistical conclusion. The number of samples required is determined such that the confidence statement will be valid.

To illustrate this design, let's consider the 1,000 km² region to be the decision area and the grid cell size of interest to be 1 km². We can now specify inputs to initialize the design. To facilitate exposition, a screenshot of the input dialogue box for this design is included in Figure 7. The dialogue in Figure 7 shows that we have specified:

1. None of the measured grid cells can be unacceptable (contaminated),
2. We wish to incorporate prior belief into this design,
3. We wish to include judgment samples in this design.

The dialogue also shows that there are a total of 900 grid cells in the population and that we chose to sample 20 of the grid cells based on our own judgment at locations that we believe most likely to be contaminated. Further, we have specified a 0.50 probability that a judgmental sample is acceptable (this design can give us extra credit, or weight, for judgment grids being clean, but in our problem we suspect our judgment cells to be contaminated so we do not wish to bias the results, thus the 0.50 probability). Also, we have specified that a judgment sample is twice as likely to be contaminated as a randomly selected grid cell. Next, in the dialogue we observe our specified confidence criterion and state “We desire 95% confidence that at least 99% of the decision area is acceptable”.

With the design criteria specified, we next need to distribute the judgmental samples. Figure 8 depicts a hypothetical placement of these sample locations. We may now compute the number of random samples required to attain the desired confidence statement (VSP will also optimally distribute them for us). In order to attain the desired confidence, VSP indicates that we must take an additional 195 random grid samples, the distribution of these samples is shown in Figure 9. Further, VSP indicates that we must sample 215 locations (195 random and 20 judgment) in order to achieve the desired confidence.


Presence / Absence Sampling


Presence / Absence | Sample Placement | Costs | Data Analysis | Analytes

None of the grid cells in my sample can be unacceptable
I want to account for prior belief in my design
I want to include judgment samples in my design

My decision area contains a total of 900 grid cells. Pick Judgment Grid Cells
I will sample 20 grid cells based on my own judgment.
Before any sampling takes place, I expect there is a
Unknown (50%) (50.000 %) probability that a judgmentally selected grid cell is acceptable.
I believe that a judgmentally selected grid cell is 2 times more likely to be unacceptable than a randomly selected grid cell.
I want 95.000 % confidence that at least 99.000 % of the decision area is acceptable.
Number of grid cells that must be randomly selected and examined: 195
If all 20 judgmentally selected grid cells and all 195 randomly selected grid cells are acceptable, then I am 95% confident that at least 99% of the decision area is acceptable.
If any of the judgment or random samples are unacceptable, then the confidence statement above no longer applies and I must assume that at least some portion of the decision area is unacceptable.
☒ Calculate number of samples ☐ Calculate % confidence based on number of samples

Close Cancel Apply Help

Figure 7: Compliance Sampling Dialogue

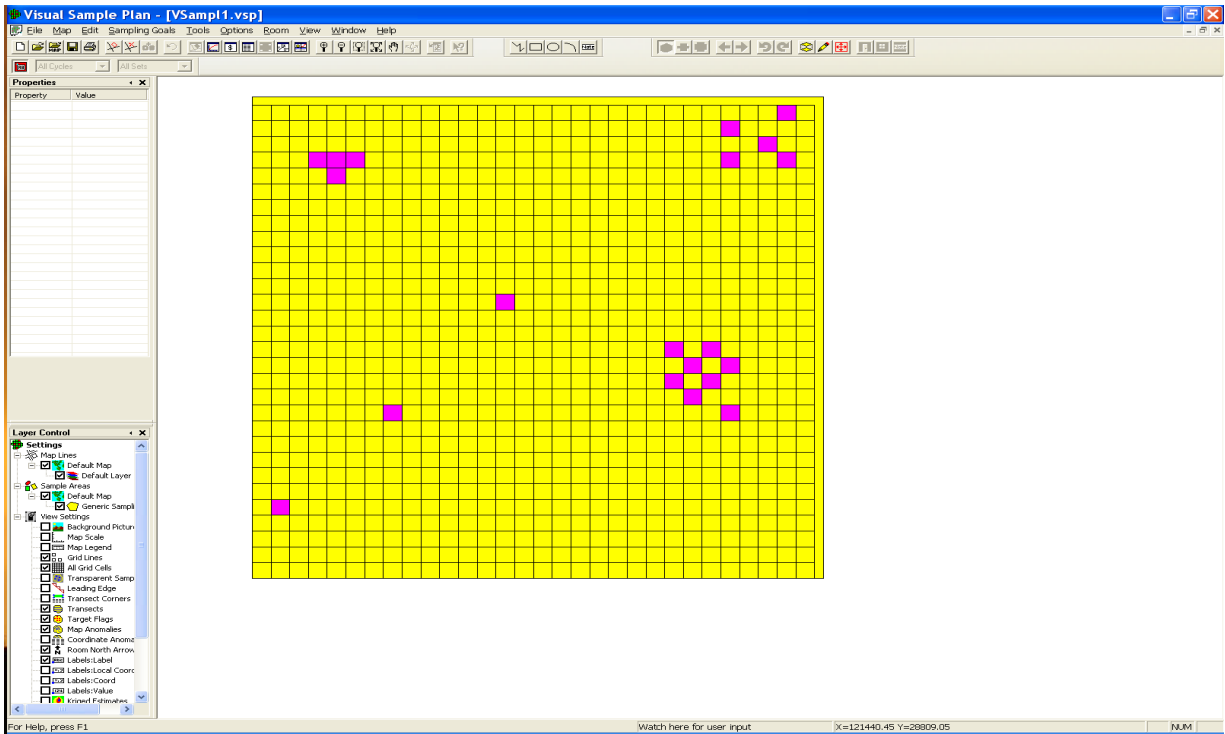


Figure 8: Graphic Illustrating Placement of Judgment Samples (pink)

We would then sample and analyze according to the selected grids. If no grids were observed as contaminated we could conclude that *we are 95% confident that at least 99% of the grid cells are acceptable*. If even one grid is observed as contaminated, the confidence statement no longer holds and we may wish to use the detected contaminations in another sampling design (such as hot spot sampling) in order to better characterize the contamination, however, for the CTBTO inspection problem detecting the radionuclide is sufficient to declare detection of a nuclear test.

To conclude, we note that while we constructed the design with a specified confidence statement and calculated the number of samples required to attain the confidence we may specify other inputs and have the design compute the unspecified ones for us. For example, we may do the converse. Perhaps we are constrained as to the number of samples we can take due to time or money resources. We could specify the number of samples that we could take, and the design would compute the confidence statement attainable under a fixed number of samples. Of course, there is more capability within VSP, but this example is intended to provide an initial exploration of these options.

Applicability to OSI:

The greatest benefit from CJR compliance sampling is the ability to place a statistical inference statement in combination with an existing design. Currently experts may be able to make informed judgment sampling decisions, but lack a full understanding of how much confidence could be placed in a statement regarding the number of samples taken and in absence of radionuclide detection. VSP can assist the IT to determine appropriate sampling locations given the number of samples constrained by time or resources, or given a desired statistical inference statement such as “we are X% confident that at least Y% of the site is compliant”. CJR compliance sampling seems reasonable to apply to surface soil sampling over a gridded region.

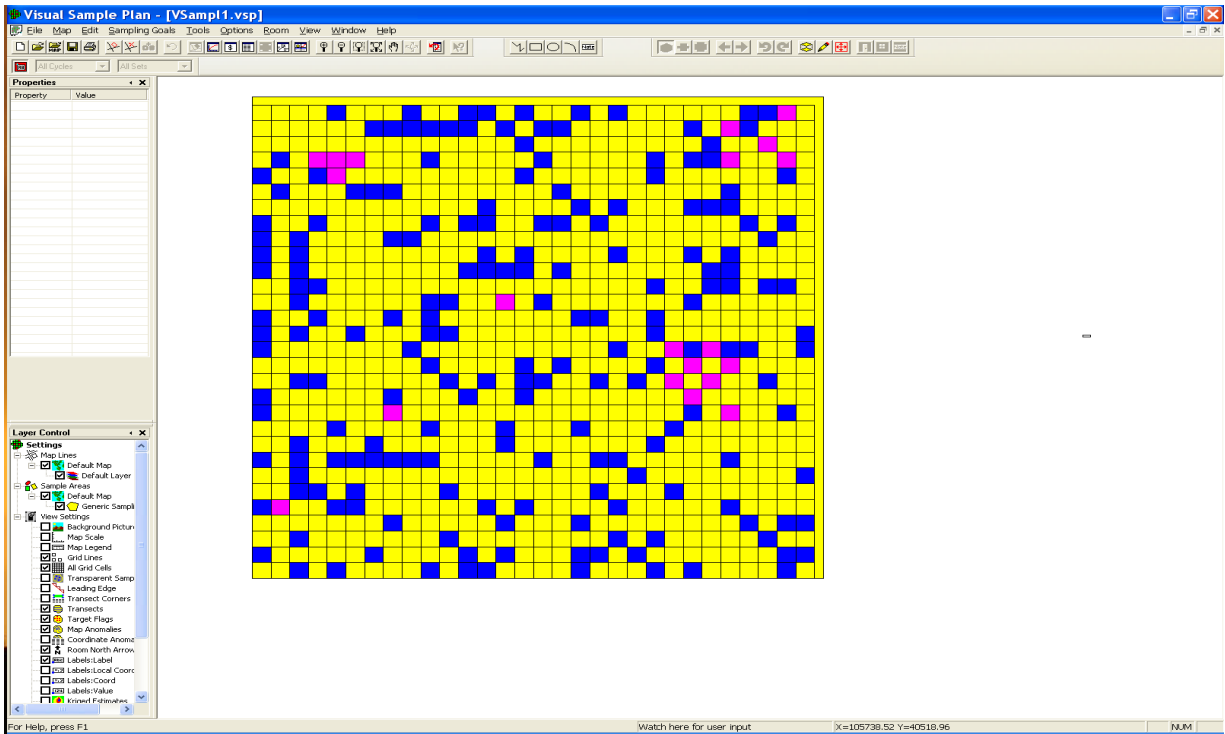


Figure 9: Graphic Illustrating the Placement of Both Judgment Samples (pink) and Other Samples (blue) needed to meet the objective

2.1.3 Multiple Increment (Composite) Sampling

Environmental sampling using multiple increment (MI), incremental sampling (IS) or composite, sampling is one approach that can overcome expensive analytical costs and provide more representative samples, by combining/mixing small samples together. If collection of field samples is relatively inexpensive/quick and analytical laboratory work is expensive, then MI sampling may allow researchers to appropriately sample from a very large area while minimizing the expensive analytical costs. The tester would combine, or mix n increments into r groups (or r MI samples) and then measure each MI sample to get the average concentration of the combined increments. The increments may be taken from any sampling scheme, such as grid, random, etc.

MI sampling is particularly useful for several reasons. MI sampling is useful when one desires to sample a large area and analysis is expensive. MI allows one to combine multiple increments and only analyze one MI sample. The second most obvious benefit of MI sampling is to make inference about a large area without having to exhaustively analyze many samples from that area. MI sampling also has a better chance of selecting nuggets of contamination, thereby potentially increasing hot spot detectability, and providing more representative samples. One other benefit of MI sampling is a strong reduction in the number of non-detects and a reduction in the variability (assuming quality blending). The MI sampling approach may be well suited for surface soil sampling under the CTBT.

VSP allows for estimating the minimum number of samples necessary to calculate the mean within a pre-conceived margin of error for a given confidence level. VSP also allows for comparison of the mean of a site to the background mean or a given threshold. Several statistical assumptions are made regarding the formulas used to compute the number of samples.

1. The MI sample mean is normally distributed (this happens if the data are roughly symmetric or the total number of increments is more than 30; for extremely skewed data sets, additional samples may be required for the sample mean to be normally distributed).
2. The standard deviation estimates, are reasonable and representative of the population.
3. The population values are not spatially or temporally correlated.
4. The sampling locations will be selected probabilistically.
5. The process of mixing increments in together in each MI sample is very good.
6. The number of MI samples must be two or greater.

After the data are collected, the first three assumptions should be assessed. The fourth assumption is met due to using a random start to select the sample locations. Limitations to the value of MI sampling arise when variation due to the blending/mixing process is so large that it overwhelms any cost advantage of combining the samples (the fifth assumption).

We have chosen just one of the MI sampling modules (Confidence Interval) to illustrate the VSP MI sampling tools. The VSP user must select from several possible options and input certain key parameters including: One-sided Confidence Interval vs. Two-sided Confidence Interval, the desired Confidence Level, the Estimated Increment Standard Deviation, the Estimated Analytical Subsample Standard Deviation, the number of Analytical Subsamples per MI Sample, and the number of Increments per MI Sample. For demonstration purposes only, we have chosen values for those input parameters. In practice these parameter inputs would depend on site information, measurement technologies employed, and desired data quality objectives. Figure 10 shows how n increments make up a single MI sample upon which two analyses are performed.

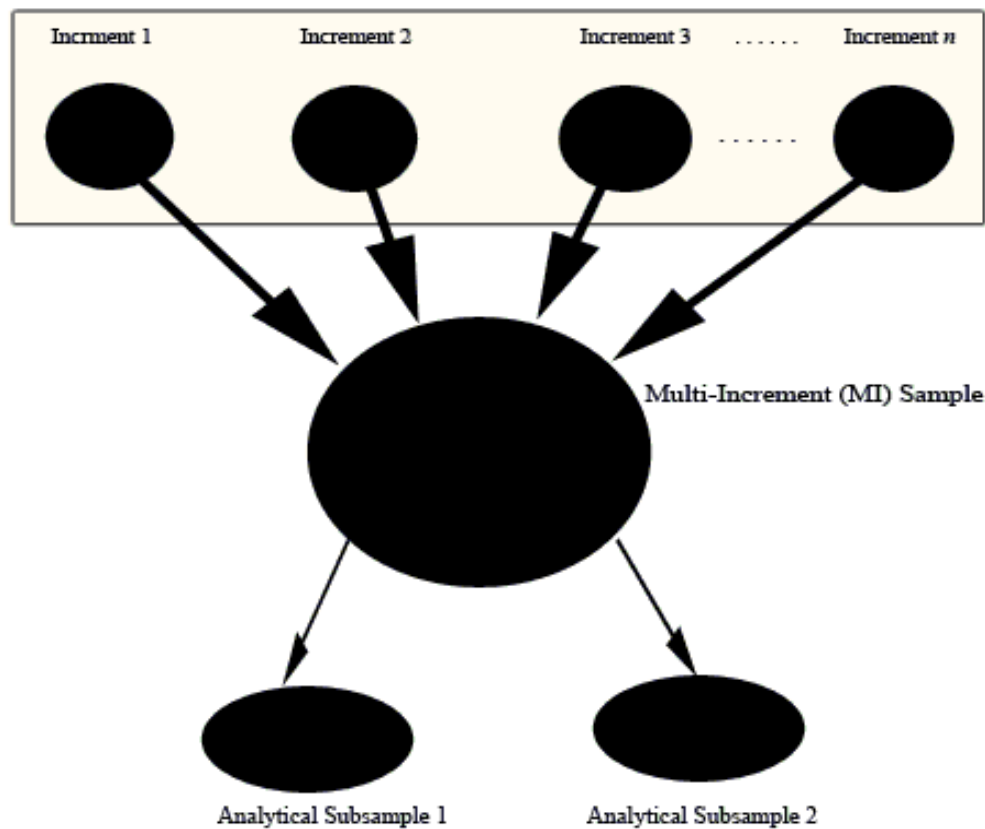


Figure 10: Composition of an MI Sample

Figure 11 and Figure 12 show two different tabs of the actual VSP dialog box for MI sampling mentioned above. The information in red displays the results calculated given the assumptions and user inputs from the other options and tabs. Once the user has selected an appropriate set of inputs they can view the placement of samples in the sampling area by selecting the “Apply” button. In this case, 13 MI samples have been placed randomly across the sampling region of interest using a feature found in the “Sample Placement” tab. These samples are grouped as 13 sets of 5 increment samples in a square localized area, though the option certainly exists to place the increments over the entire site. The researcher would then collect and blend each set of 5 samples. One MI sample would be taken from each of the 13 blended samples and analyzed.

Confidence Interval on True Mean - Multiple Increment Sampling

Confidence Interval | Sample Placement | Costs | Analytes

Choose:

- ☒ One-sided Confidence Interval
- ☐ Two-sided Confidence Interval

Analyte:

Confidence Level: 95.0 %

Maximum acceptable width of confidence interval: 1

Estimated Increment StdDev: 3

Estimated Analytical Subsample StdDev: 2

Analytical Subsamples per MI Sample: 2 Optimal

Increments per MI Sample: 5 Optimal

Minimum Number of MI Samples for Analyte 1: 13

Minimum Number of Multiple Increment (MI) Samples: 13

If 2 analyses are performed on each multiple increment sample composed of 5 increments, then 13 of these multiple increment samples are required to achieve 95% confidence that the true mean lies within the confidence interval width of the estimated mean.

A total of 65 increment samples will be placed on the map.

Close Cancel Apply Help

Figure 11: Multiple Increment Sampling Dialogue Box

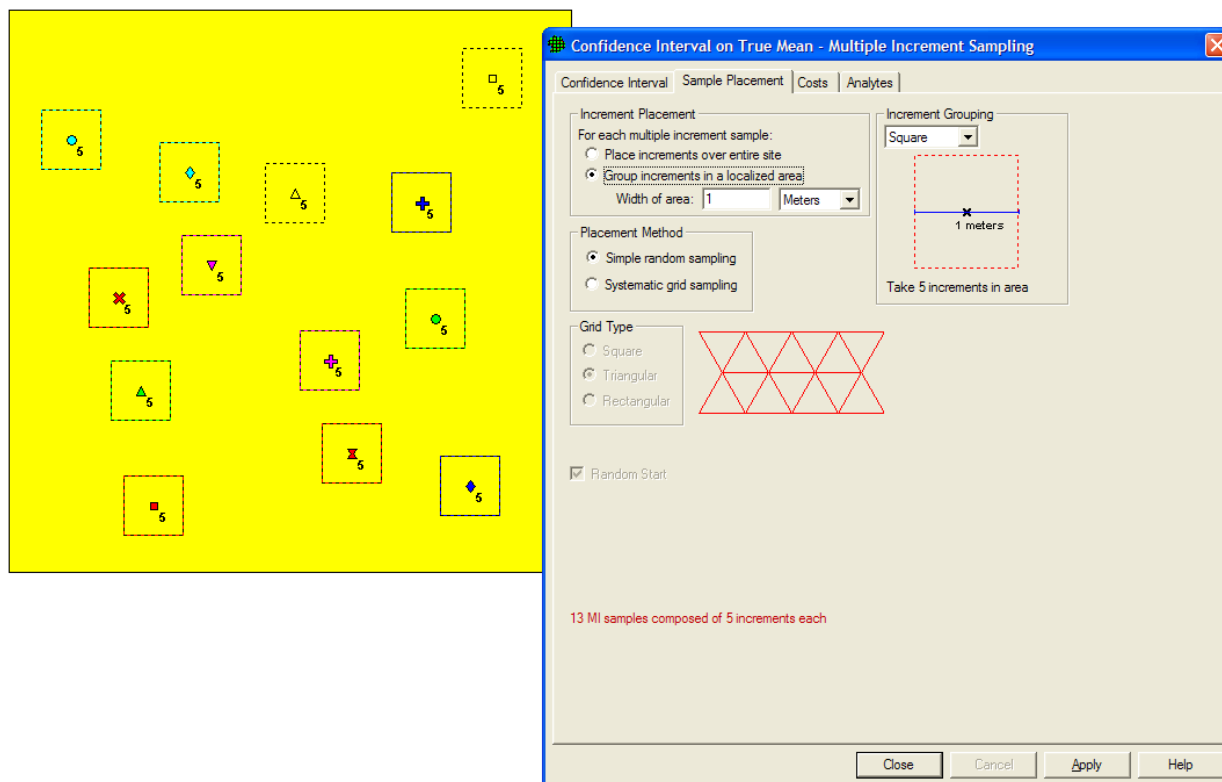


Figure 12: Multiple Increment Sampling - Group Increments in Localized Square Area

Applicability to OSI:

Collecting surface soil samples attempting to canvass a large area under some cost constraints (resources, time, human capital, etc.) make MI sampling a potentially important methodology for the OSI. MI sampling may help the IT to make inference about a large area without having to exhaustively analyze many samples from the area. MI sampling has a better chance of selecting nuggets of contamination, detecting hot spots (if dilution effects aren't too great), and providing more representative samples. MI sampling can also reduce the number of non-detects under certain conditions and reduce the variability (assuming quality blending). Last, we note that this method is applied to data of ratio or interval level of measure and that inference is made to the mean contamination level in the population.

2.1.4 Radiological Transect Surveying

VSP can design radiological transect surveys to find radionuclide hot spots when surveys are performed on the ground or in the air for a region of interest and under user-specified cost constraints. The software can also evaluate meandering transects and their ability to find target areas/hot spots, even if the survey wasn't carefully designed or failed to follow the original design. Collected and analyzed data can be spatially represented using common geostatistical analyses in VSP.

Radiological transect surveys directly apply to the wide area sampling problem by optimally surveying the site (1,000 km²) and possibly identifying smaller areas of interest (2 km²), which is one form of hot spot detection. In this case, hotspot detection via radiological transect surveying following VSP

guidelines helps the user determine the number and spacing between transects needed to have a high probability of traversing a suspected hot area of a given size. VSP also employs geostatistical kriging algorithms to create an interpolated map of the radionuclide concentrations across the site.

Figure 13 shows a VSP dialog box allowing the user to specify the size and shape of a potential hot spot. VSP will use the given transect width and sampling area dimensions to provide the user with the transect paths needed to ensure a high probability of traversing/detecting a hot spot of that size and shape. The transect width is defined (in this case at 10 m) and the hot spot size/shape of interest is a circle with radius 750 m (area of approximately 1,767,000 m²). As seen in Figure 14, VSP makes a statement in **red** suggesting we will always (100% of the time) traverse this sized hot spot given the transect width and number of transects evenly spaced over the sampling site.

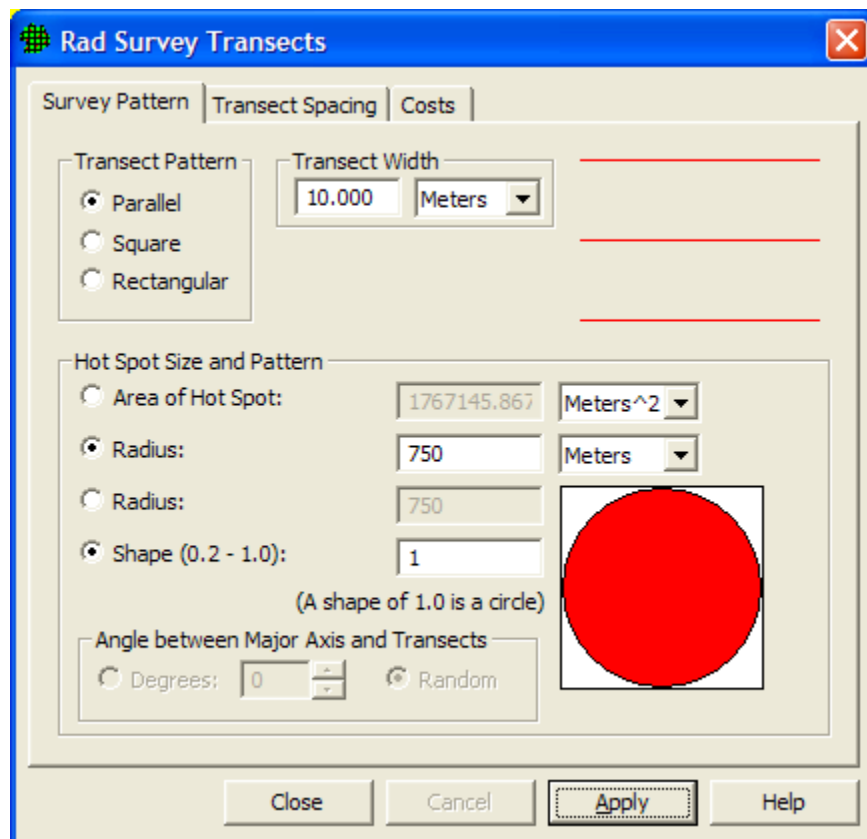


Figure 13: Radiological Survey Transects – Survey Pattern Options

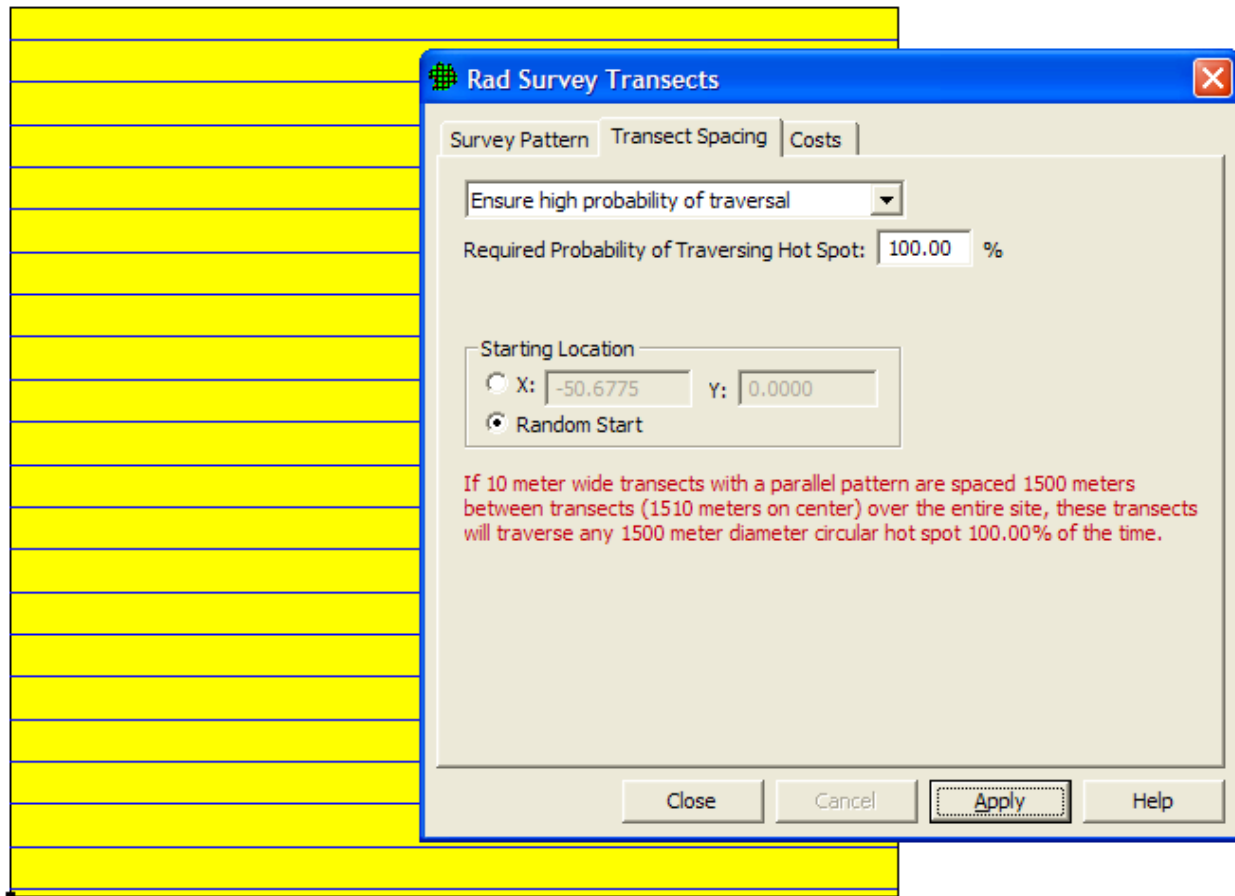


Figure 14: Radiological Survey Transects - Transect Spacing on Sampling Area

An example of the type of maps obtainable from these radiological surveys are shown in Figure 15 and Figure 16 and below. These are displayed to give the reader an appreciation of the graphical capabilities of the geostatistical features highlighting results from the radiological survey transects. Note the use of the underlying map to help visualize transect and hot spot locations.

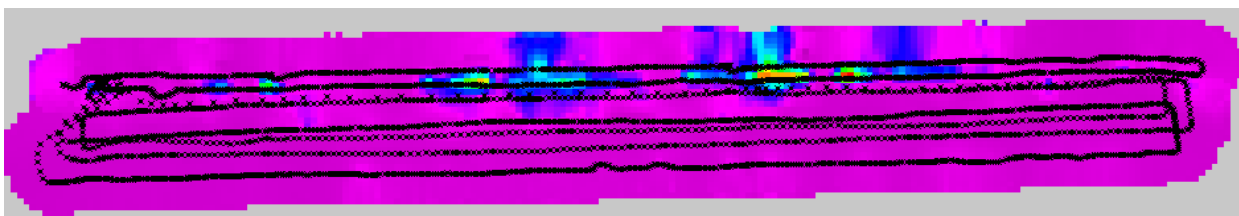


Figure 15: Radiological Survey Transects – Example of Geostatistical Methods to Map Survey Results

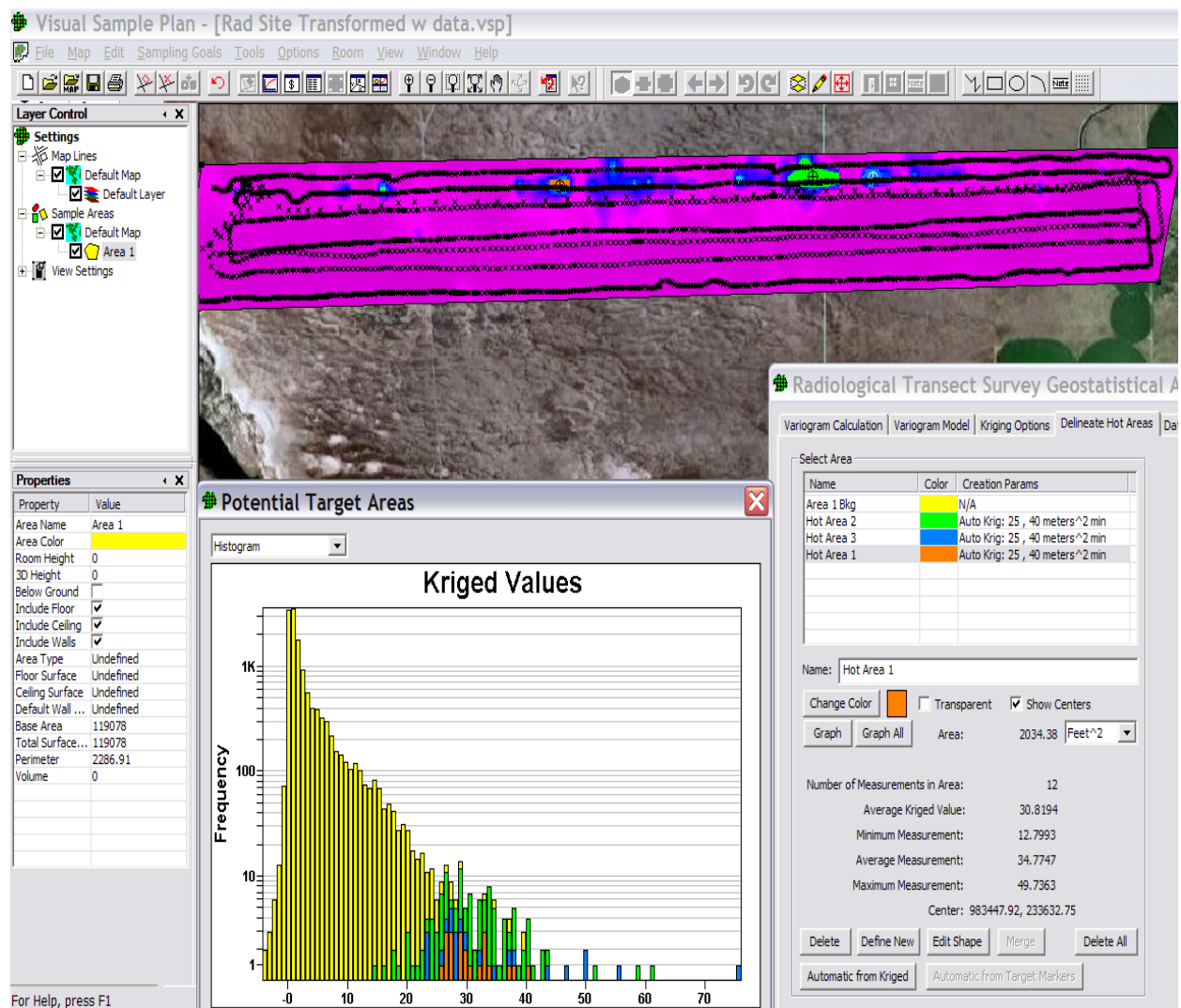


Figure 16: Identified and Delineated Hot Areas from a Radiological Survey with Diagnostics About Each Hot Area

Applicability to OSI:

Any gamma radiation survey work could benefit from a radiological transect survey design and geostatistical evaluation in VSP. It seems that the CTBTO already has the technology in place to do radiological surveys; however, it appears that the CTBTO could benefit from a comprehensive design and analysis system provided by VSP. Data collected from a statistical design or collected haphazardly in areas influenced by expert opinion or visual inspection could be visualized appropriately using traditional geostatistical spatial modeling applications, which are available in VSP. Aerial- or ground-based surveying can be designed and analyzed.

2.2 Other Sampling Methods That May Be Applicable to an OSI with Significant Modification or Under Certain Circumstances

Several other sampling modules in VSP might shed some light on other sampling approaches that could have some pertinence to an OSI. These methods would probably require significant modifications or be combined with other more pertinent methods to make them relevant to OSI needs but the underlying framework and concepts might be adaptable. These include stratified sampling, collaborative sampling, and adaptive cluster sampling. Spatial geostatistical analysis methods also would be applicable to most sampling results as a way to visualize interpolations between spatially correlated samples. Each of these extraneous sampling methods is described briefly in the following subsections.

2.2.1 Stratified Sampling

Stratified sampling is appropriate when the site consists of varying subsets where the variability within each subset is small relative to the variability between each subset. For example, these subsets could be based on soil types, layers, or other distinguishing characteristics. Three methods exist in VSP for calculating the number of samples needed using a stratified sampling method. The method deemed most applicable to the CTBT applications would be to “Minimize Cost for Required Standard Deviation of Sample Mean”. If a subset of the overall site requires further examination, we can perform stratified sampling in these areas to estimate the site mean. This seems most effective if each sub-area has a different standard deviation – meaning we’ll sample more from sub-areas with high variability and sample less from those sub-areas containing low variability.

Note that the current implementation of VSP assumes that cost is the driving force behind stratified sampling. A simple example is shown in Figure 17. Although this does not seem to apply directly to CTBT, a combined VSP approach that uses both judgment sampling and stratified sampling, where the strata are based on likelihood of radionuclide presence, is being developed for the Department of Homeland Security.

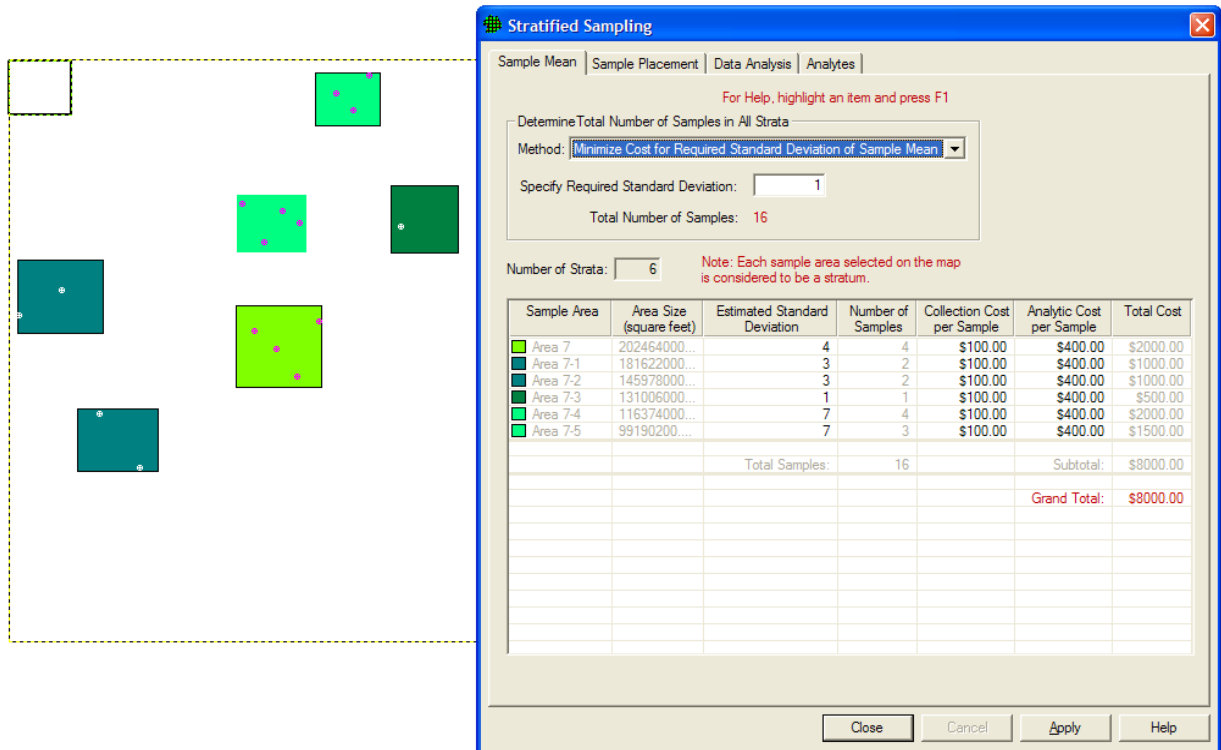


Figure 17: Screen shot of Multiple Sampling Areas Stratified by their Estimated Variability

2.2.2 Collaborative Sampling

A collaborative sampling design is used to find a cost effective solution when two measurement techniques are available. Often one of these measurements requires expensive analytical resources, but the other is inexpensive (though sometimes considered not as accurate). This sampling approach is particularly helpful in the reduction of the number of samples needed to make strong statistical inference statements when using expensive sampling equipment or expensive analysis resources. The tool becomes more cost effective if the relationship between the two sampling methods is linear and the correlation is high, and if the difference in expense between the two measurement methods is great. VSP determines the number of samples (n) required to be analyzed using the inexpensive measurement method and a small subset of those n samples to be analyzed using the expensive measurement method. Our ability to show a potential application of this sampling design methodology to the CTBT problem may be limited by our understanding and knowledge of the OSI procedures, technology, and equipment. Anytime field samples can be taken and analyzed more efficiently (less expensive, more quickly, easier) using other technologies the approach of collaborative sampling may be appropriate. The current VSP methodology is appropriate for comparing means against some threshold of concern but the methods could be developed for the case when the objective is to have a high probability of detecting radionuclides or noble gases if they are present.

2.2.3 Adaptive Cluster Sampling

Adaptive cluster sampling excels in situations where the characteristic of interest is sparsely distributed but highly aggregated, or clustered. Examples of such situations include hot spot or mineral investigations, soil remediation, hazardous waste site characterizations, pollution concentrations, and

epidemiology of rare diseases. Quick turnaround of analytical results greatly benefits adaptive clustering by keeping the iterative sampling moving.

Adaptive cluster sampling begins by dividing the sample area into a grid of sampling units. Using the dialog boxes, a researcher then inputs the desired confidence level, estimated standard deviation, and action level for the measured analyte. Initial samples are then selected using a probability-based design with the primary objective of estimating the mean and standard deviation of the contamination across the site. Additional neighboring samples are selected when an initial sample exceeds some pre-specified condition or a characteristic of interest is present. Each iteration adapts by sampling the neighbors of the interesting previous samples. This makes the sampling scheme both adaptive (iterative sampling) and clustered (selection of neighbors). The dialog box for adaptive cluster sampling is shown in Figure 18 and Figure 19. Eight neighbors would look like an orange cell surrounded completely by eight yellow cells, similar to the four neighbor example in Figure 19.

The image shows a software dialog box titled "Adaptive Cluster Sampling for Estimating a Mean". It has four tabs: "Number of Initial Samples", "Grid Size & Follow-Up Samples", "Costs", and "Analytes". The "Number of Initial Samples" tab is selected. Inside the dialog, there is a "Choose:" section with two radio buttons: "One-sided Confidence Interval" (which is selected) and "Two-sided Confidence Interval". Below this is an "Analyte:" section containing four input fields: "Confidence Level" (set to 95.0 %), "Threshold for triggering additional samples" (set to 10), "Maximum acceptable width of confidence interval" (set to 1), and "Estimated Standard Deviation" (set to 2). At the bottom of the dialog, there are two lines of text: "Minimum Number of Samples for Analyte 1: 13" and "Minimum Number of Samples in Survey Unit: 13". The number 13 is displayed in red. At the very bottom of the dialog are four buttons: "Close", "Cancel", "Apply", and "Help".

Figure 18: Adaptive Cluster Sampling – Initial Samples Dialogue

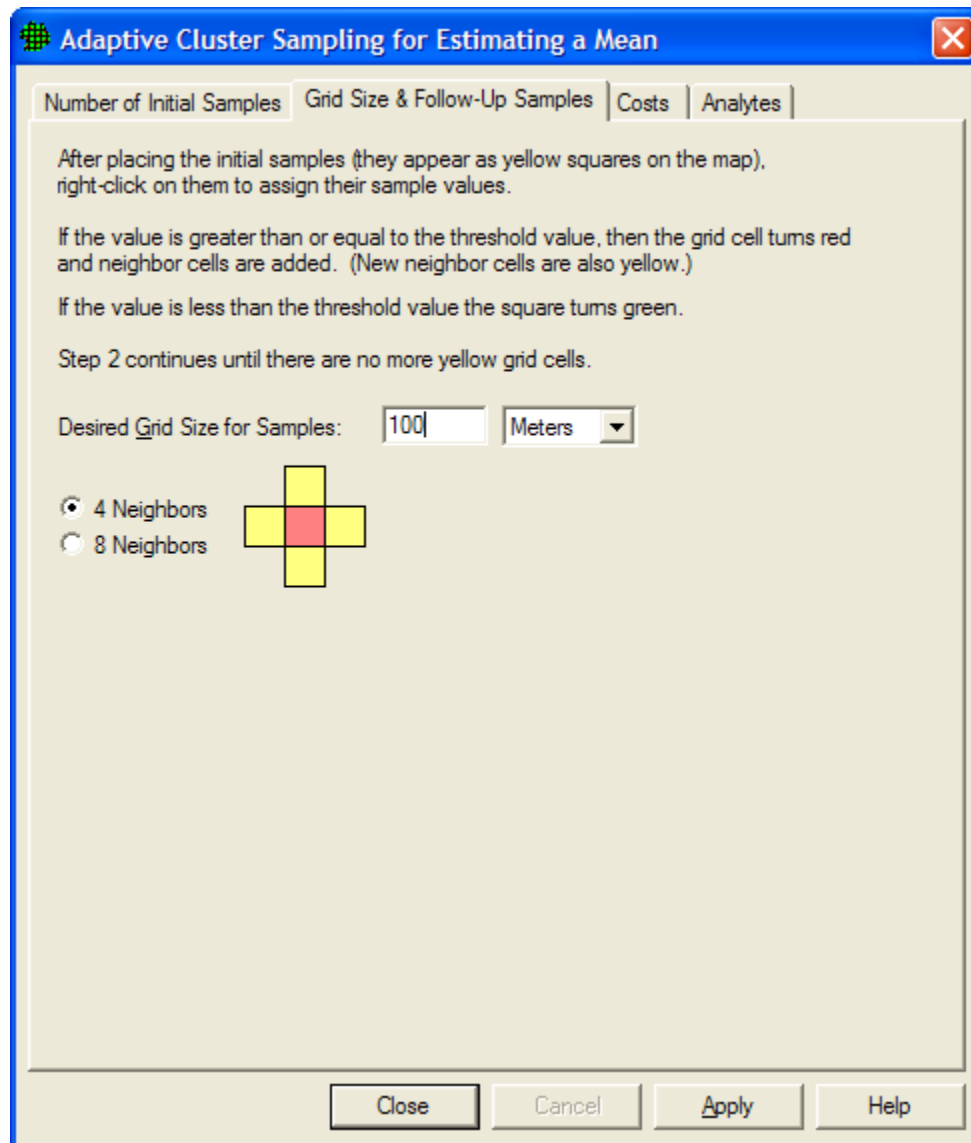


Figure 19: Adaptive Cluster Sampling – Grid Size and Number of Neighbors

The statistical assumptions made to compute the number of samples are:

1. The sample mean is normally distributed
2. The variance estimate, s^2 , is reasonable and representative of the population being sampled
3. The estimate of the sample mean is reasonable and representative of the population being sampled
4. The field locations that will be the initial units are selected using simple random sampling.

The first two assumptions will be assessed in a post data collection analysis. The third assumption is valid because the estimate of the mean will be an unbiased estimate.

Adaptive cluster sampling is typically used most often for constructing a confidence interval about the mean but upon detection of a potential hot spot, it adapts to delineate the hot spot boundary – which is not of direct interest to an OSI. The iterative adaptive approach is interesting and may need to be combined with other sampling methods to more directly provide benefit to the CTBTO's sampling goals.

2.2.4 Tolerance Limits

If the decision area may be considered as consisting of a population of continuous measures (as opposed to presence/absence measures), then it is possible to construct a sample design that tests whether a specified area is contaminated with radionuclides at levels greater than some action limit (the action limit may represent a detect threshold or a limit of “acceptable” contamination as specified by regulations or domain knowledge). This design computes a confidence interval around some percentile of the population and this is commonly referred to as a *tolerance limit*.

As with other designs, this design is built around a confidence statement, requires some inputs, and can solve for the unspecified input. The inputs are

1. The required fraction of the population to be less than the action level
2. The required confidence level
3. The action level
4. Minimum number of samples to attain required confidence.

We provide a screen shot of the dialogue box in Figure 20 for exposition. The sample design can distribute samples randomly or on a random grid. Also, this design may be constructed under cost constraints and will compute the maximum confidence that may be attained.

Given data, the design will construct an upper 1-sided confidence interval around the 99th percentile of the distribution (this is the tolerance limit). If the tolerance limit is less than the action limit we may reasonably conclude that the decision area is not contaminated.

Nonparametric Upper Tolerance Limit

Nonparametric Upper Tolerance Limit

Sample Placement

Costs

Data Analysis

Decision Rule: If the Upper Tolerance Limit (UTL) is \geq to the action level, conclude that the decision unit is contaminated; otherwise the decision unit is not contaminated.

Required fraction of the population to be less than the action level: %

Required confidence Level: %

Action Level:

Minimum number of samples required:

If the maximum of the 299 sample values (the nonparametric UTL) is less than the action level, then you will be able to conclude with 95% confidence that at least 99% of the values in the population are less than the action level and therefore that the decision unit is uncontaminated.

☒ Calculate number of samples

☐ Calculate % confidence based on number of samples

☐ Calculate the fraction $<$ action level based on number of samples and % confidence

Close

Cancel

Apply

Help

Figure 20: Upper Tolerance Limit Design

3 Summary and Conclusions

In this report we have identified a number of VSP modules and sample designs that may be of immediate use to the CTBTO's OSI sampling activities. Some of the most pertinent designs include: Hot Spot Sampling, Compliance Sampling, Multi-increment Sampling, and Radiological Survey Transect Sampling. We have provided an in depth, but not exhaustive exploration and demonstration of how these designs may be applied and what inferences and conclusions can be gleaned from them. We have also briefly described a number of other VSP sampling modules that may be of interest to the inspection team if modified and/or combined with other sampling approaches.

Aside from direct VSP modules that may be applicable to the CTBT OSI, we have identified a potential impact and contribution of using VSP, namely, that *VSP's sampling designs are constructed under a statistical inference statement*. This encourages a sampling design to not just declare a conclusion (such as detected contamination or not), but also quantify the uncertainty with said conclusion ("we are 95% confident that ..."). This is a potentially distinguishing characteristic of VSP relative to current OSI operations.

A statistical design may be useful to help allocate resources more efficiently, but the statistical design also helps researchers determine variation/uncertainty in the estimates and conclusions. For example, instead of canvassing an entire area, perhaps there is some risk that may be tolerated allowing the canvas of a portion of an area and conclude that contamination is not detected with 95% confidence. Using a statistical design approach may potentially enable canvassing a larger area of interest in the time window specified by CTBT.

A combination of several statistical methods in VSP may greatly benefit the OSI activities by providing a more fluid, comprehensive approach to sampling design, analysis, and statistical inference/conclusions. More description of such a comprehensive approach follows under Recommendations.

4 Future Work / Recommendations

Although some existing VSP modules may be directly applied to meet OSI needs, based on the unique aspects associated with an OSI (time constraints, phased sampling approach, multiple sample types, wide area vs. small area sampling, etc.), it is recommended that a holistic sampling design and statistical evaluation strategy be developed that meets all needs and objectives while dealing with every contingency that might arise. We envision that such a strategy should be documented and supported by a detailed decision flowchart that would identify the most appropriate, defensible sampling designs that should be employed that will result in greatest detectability while meeting feasibility requirements. In concert with this comprehensive sampling strategy, VSP tools and methods can be identified, developed, or modified to provide the OSI team with optimal sampling designs/locations that will meet all objectives at all stages of the investigation.

Some augmentation to the VSP software would enable it to better fit the CTBT sampling procedures. For example, it would be constructive to apply a sampling method encompassing both an aerial-based gamma radiation survey and a ground- or surface-based gamma radiation survey. These surveys would be performed in two distinct sampling schemes currently, though it would be interesting and more powerful to combine them into one sampling design and analysis considering the technologies and sampling procedures are the same.

The inspection team would benefit from a more comprehensive version of their already adaptive and sequential ad-hoc sampling approach. Tailoring VSP to allow somewhat simultaneous collection of samples from multiple sampling schemes would improve on current on-site inspections. As mentioned previously, the greatest benefit of VSP as applied to this problem is that of approaching the design and analysis using statistical inference. Statistical inference statements assist the IT in defending their analytical results with probability statements and greater clarity.

We propose creating a VSP module designed specifically to meet the sampling needs of the CTBTO's OSI process.

References

CTBT/WGB/TL-18/40 (2009) *Model Text for the Draft OSI Operational Manual: Basis for the Third Round of Elaboration*.

Matzke BD, NL Hassig, JE Wilson, RO Gilbert, BA Pulsipher, LL Nuffer, ST Dowson, J Hathaway, CJ Murray, and LH Sego. 2007. *Visual Sample Plan Version 5.0 User's Guide*. PNNL-16939, Pacific Northwest National Laboratory, Richland, Washington.