APPLICATION OF VISUAL SAMPLE PLAN (VSP) SOFTWARE

CONTENTS

1.0 OBJECTIVE

The objective of this Standard Operating Procedure (SOP) is to provide a process for applying Visual Sample Plan (VSP) software in the development of sampling plans that meet project specific environmental, financial, and logistical objectives. For specific details on running VSP, refer to the software and User's Guide which can be downloaded from the website: http://vsp.pnnl.gov.

A Quality Assurance Project Plan (QAPP) in Uniform Federal Policy (UFP) format describing the project objectives must be prepared prior to deploying for a sampling event. The sampler needs to ensure that the methods used are adequate to satisfy the data quality objectives listed in the QAPP for a particular site.

The procedures in this SOP may be varied or changed as required, dependent on site conditions, equipment limitations or other procedural limitations. In all instances, the procedures employed must be documented on a Field Change Form and attached to the QAPP. These changes must be documented in the final deliverable.

2.0 APPLICABILITY

These procedures apply to the use of VSP in developing statistically defensible sampling plans which meet specified project objectives including optimizing sample numbers and sample/transect placement based on financial and logistical restraints. Statistical training/knowledge is not required to use VSP. It was designed with a visual, graphical and user-friendly interface.

VSP can be used for: developing systematic grid sampling designs (also known as hotspot sampling); identifying confidence levels and sampling numbers required for identification of unacceptable levels of contamination applying "item sampling"; and for contouring contaminant concentrations in soil and sediment. Application of VSP is not limited to these scenarios.

VSP is continuously under development. As new **Sampling Goals** for environmental data-analysis are identified, VSP is updated to handle them. At present, the software can be used for creating and optimizing statistically defensible sampling designs associated with contamination of soil, sediment, groundwater, surfaces, and unexploded ordinance (UXO), as well as contamination within a room or building. It can also compute simple descriptive statistics, linear regression analyses, trend analyses, analysis of variance (ANOVA) and control charts. VSP can generate geostatistical and interpolated spatial contours.

Data quality objectives (DQO) should be agreed upon by the U.S. Environmental Agency (EPA) Environmental Response Team (ERT) Work Assignment Manager (WAM) and the ERT contractor's Task Leader (TL) to ensure optimum use of the software.

3.0 DESCRIPTION

VSP is a free menu-driven software application created under funding from the EPA, U.S. Department of Energy (DOE), U.S. Department of Homeland Security (DHS), U.S. Department of Defense (DOD), Centers for Disease Control (CDC), National Institute for Occupational Safety and Health (NIOSH), and U.K. Atomic Weapons Establishment (AWE). It is based on a systematic planning process and provides the tools and statistical/mathematical methodology to facilitate the selection of the number, type and location of samples collected for a project that will be adequate for meeting project goals. These goals are identified on UFP-QAPP Worksheets 10 and 11.

Each file created in VSP represents a new project; however, multiple sampling designs can be created and over-laid on one site map for a single project. VSP allows the user to run iterative scenarios with adjusted measures of variability and various levels of uncertainty associated with making project decisions. Use of different sampling options can also be chosen and explored (e.g., simple random versus grid sampling) using the iterative functionality.

VSP automatically generates a Statistical Sampling Summary Report that summarizes the sampling objective, design and approach for all/each area of concern. The report includes assumptions that were made, statistical parameters that were entered by the user, statistical/mathematical equations which were used to compute the number of samples, graphs of user's specified confidence/tolerance limits, maps of sampling locations and assessments of sample size versus cost. Guidelines for interpreting the data analysis are also provided.

3.1 Gather Information

When running VSP, the user will be prompted to supply planning information related to the goals and needs of the project. This includes information related to the site, contaminant(s) of concern (COCs), sampling goals, statistical levels of confidence, and project costs. This information is gathered through meetings with the WAM and during scoping meetings.

Prior to applying VSP, information that should be identified and/or agreed upon to facilitate its use includes but is not limited to:

- Physical, spatial, and temporal boundaries of the investigation,
- Depths/strata of contamination that will be examined (soil/sediment contamination),
- Analytes of concern (target population) and their respective action levels,
- Expected levels of contamination per analyte of concern and expected or known variability associated with the analytes of concern as related to the site,
- *Sampling areas* (decision units),
- Statement of specific sampling goals and the environmental questions being asked, also referred to as decision objectives,
- Statistical hypotheses you will be testing (null and alternate hypotheses; includes whether the site will be assumed to be 'dirty' and statistical evidence will be collected to prove otherwise, or if the site will be assumed to be 'clean' and statistical evidence collected to prove otherwise),
- Availability of historic data including historic sample coordinates, the type of sample strategy implemented, and analytical results,
- If historic data are not available determine if there are resources available to conduct a pilot study to estimate variability,
- Levels of statistical confidence to be achieved and/or acceptable limits on error in making a wrong decision,
- Type(s) of sampling to be conducted, probabilistic, judgmental or both,
- Identification of:
    1. Fixed planning/general mobilization and validation costs,
    2. Fixed collection cost per sample, and
    3. Laboratory cost per analysis.

Additionally, locate or have a site/building map generated using planar coordinates. Map files are typically drawing exchange format (.dxf) generated using computer-assisted design (CAD) software or shape files (.shp) generated using geographic information system (GIS) software. VSP has a basic set of drawing tools for creating maps and building plans if needed; however, that is not the primary focus of the application. Typically, it is easier to develop the map in a CAD or GIS application and import it to VSP.

3.2 Sampling Goals

VSP can be used to determine sample sizes and sample placement to support the following sampling goals:

- Compare the average concentration of a contaminant to a fixed threshold such as an action level or clean-up goal,
- Compare the average concentration of a contaminant(s) to averages within reference/background areas,
- Estimate the mean concentration of a contaminant(s) within a defined area,
- Construct a confidence interval on the mean/median,
- Locate a hot spot of contamination in soil/sediment,
- Determine if a percentage of the sampling area is acceptable based on user set criteria,
- Detect a trend (with or without accounting for seasonality),
- Identify sampling redundancy (determine if sample size can be reduced without losing trend or spatial information; reduction of groundwater wells),
- Add sampling locations to reduce spatial uncertainty for geostatistical analyses,
- Compare a proportion to a fixed threshold,
- Compare a proportion to a reference proportion,
- Establish boundaries of contamination,
- Find target areas and analyze survey results (UXOs),
- Conduct post remediation verification sampling (establish confidence in determining if an area is clear of UXOs using transect sampling),
- Remedial investigation for UXO (develop transect survey designs),
- Define sampling strategies within a building (establish percentage of building that meets acceptable criteria; locate a hot spot within a),
- Select the appropriate number of items to be sampled from a population to estimate the proportion of the population that meets specified criteria (item sampling), and
- Develop a non-statistical sampling approach.

3.3 Probabilistic and Judgmental Sampling Options in Visual Sample Plan

VSP supports the development of both probabilistic (statistically based) and judgmental (biased) sampling plans. Projects can combine both types of sampling strategies using professional judgment, historical sampling events and statistical methodology to meet project goals.

3.3.1 Probabilistic/Statistical Sampling

Probabilistic sampling options and corresponding sampling goals currently included in VSP are:

- <u>Simple Random Sampling</u>: Sample locations are chosen using a random number algorithm to generate coordinates where each potential sample location has an equal chance of selection. Can be used to:
    - Estimate a mean,
    - Compute a confidence interval on the mean,
    - Compare an average to a fixed threshold,
    - Compare a site average to a background average,
    - Show that some high percentage of a *sample area* is acceptable according to

specified criteria,
- Compare a site proportion to a fixed threshold, or
- Compare a site proportion to a background/reference proportion.

- Systematic Grid Sampling: Sample locations are based on a regular square, rectangular, or triangular pattern with sample collection at the nodes of the grid built from a randomly generated starting location; used to:
  - Locate hotspots,
  - Determine contaminant distribution in a *sample area*,
  - Estimate mean concentration of a specified COC, and
  - Ensure full coverage of the *sample area*.

- Sequential Sampling (over time): Sample locations are determined in an iterative fashion where an initial set of random samples are collected, the results are entered into VSP and a statistical algorithm run to determine whether additional samples are required to meet the defined objectives; used when:
  - Near-real time analyses can be used for measurement,
  - Multiple, sequential, sampling trips are logistically possible, and
  - The objective is comparison of the average concentration of a COC within a *sample area* to a fixed threshold.

- Collaborative Sampling (Double sampling): Combines the results of many inexpensive field screening measurements with small percentage of expensive laboratory measurements (i.e., XRF and confirmatory laboratory metals analysis) to:
  - Estimate the mean or compute a confidence interval about the mean of a COC or
  - Test if the mean exceeds a fixed threshold.

- Stratified Sampling: The sample area is divided into regions (strata) based on a specified criteria (i.e., distribution of contamination, topography, soil types, etc.) so that heterogeneity of the parameter of concern between regions is greater than heterogeneity within a region; sample locations are divided between stratum based on a formula which places more sample locations in regions of greater heterogeneity. Used to:
  - Estimate a weighted mean across more than one strata or
  - Estimate the mean of each individual stratum.

- Rank-set Sampling: A two-phase sampling design that identifies sets of field locations, utilizes inexpensive measurements or professional judgment to rank locations within a set, and then selects one location from each set for sampling (EPA QA/G-5S EPA 2002 p. 150). VSP provides a tool to determine if rank-set sampling will be a cost-effective approach. Used to:
  - Estimate a mean or
  - Test a hypothesis about the mean.

- Adaptive Cluster Sampling: An iterative sampling methodology where an initial set of *n* randomly selected locations are sampled; if a measurement at a location exceeds a pre-determined threshold value, then 'neighboring' locations are sampled. Used to:
  - Estimate the mean of a population parameter or
  - Define hot spots.

- Multiple Increment Sampling: A form of composite sampling where an optimum number of increments are combined into multi-incremental sample to make decisions on a specific area/decision unit (DU). This strategy can:
    - Result in more representative samples,
    - Decrease variability (standard deviation),
    - Reduce number of analyses,
    - Provide a greater likelihood of locating 'nuggets' of contamination, and
    - Can be used to:
        - Compare site and background means,
        - Compare site mean to a fixed threshold, and
        - Compute a confidence interval on the mean.

- Continuous Transect Sampling: A sampling strategy applied when geophysical or radiological sensors are used to perform continuous sampling along straight lines with the goal of locating a circular or elliptical target. It is typically associated with locating UXOs.

VSP provides guidance through its systematic planning process to assist the user in determining the best sampling strategy or combination of strategies to use to meet their project goals. VSP also provides an **Expert Mentor** (see User Manual for details) to guide the user in the appropriate choice of sampling methodology.

3.3.2 Judgmental/Authoritative/Biased/Targeted Sampling Options in VSP

Judgmental/biased sampling locations can also be placed on a sample map in VSP based on historical or expert knowledge to define specific points of interest or based on a pre-determined sample size. This strategy has also been referred to as targeted sampling. There is no randomization involved in biased sampling and therefore no statistical confidence can be applied to this strategy. Because there is no mathematical or statistical support behind this sampling strategy, it makes the conclusions derived from the sampling results more vulnerable to challenge.

In many instances, judgmental sampling alone may not fulfill the DQO process or provide sufficient information to provide an accurate picture of suspected contamination. It is highly recommended that judgmental sampling be performed in conjunction with and as a supplement to probabilistic sampling. Limited cases where biased sampling alone could be applied include:

- Cases where the project objective is to look for worst case contamination in specified hot areas or along established boundaries, and estimated means are not required;
- An accurate CSM (Conceptual Site Model) has been established based on extensive historical information.

Judgmental sample locations can either be imported from a text (.txt) file or placed on the site map by the user in VSP. To generate biased sampling locations choose **Sampling Goals/Non-statistical Approach** from the menu. Three options are presented: **Predetermined number of samples, Judgment (authoritative sampling),** and **Contour Sampling**. The first option allows the user to choose from a **simple random placement** or a **systematic grid** placement for a pre-determined number of samples. If the user chooses,

simple random placement, the specified number of samples are distributed in a random fashion across the *sample area*, however there is no supporting statistical summary report or graph associated with the sample placement (unless it was used in conjunction with probabilistic sampling) because the number of samples was not based on estimated/known uncertainties. Sample location coordinates are generated and can be exported from VSP.

Similarly, the choice of the **systematic grid** option will generate sample locations at the nodes of the selected grid type (square, rectangular, or triangular) and place them on the *project map*. The grid is built from a simple mathematical formula dividing the *sample area* into the appropriate number of grid cells, which will result in the user specified number of samples. No statistical summary report or graph will be generated by VSP (unless it was used in conjunction with probabilistic sampling) because the basis for sample size was not based on estimated/known uncertainties. Sample location coordinates are generated and can be exported from VSP.

3.4 Operating Visual Sample Plan

Below are the general and basic steps for using VSP. Detailed descriptions of functionality, mapping capabilities, and movement through the VSP menus can be found in the VSP User's Manual. It is assumed that the user has already downloaded and installed the most recent version of VSP from the website http://vsp.pnnl.gov as well as the User Guide.

VSP is a menu-driven software system. The Main Menu bar runs along the top of the screen. When an item is selected from the main menu bar, pull-down menus with options will appear. Below this bar are four separate tool bars with available short cuts: the main toolbar, map drawing toolbar, ranked set toolbar, and room toolbar. When running the mouse over the buttons on the toolbars the user can see a description of what each tool does.

Step 1: Open VSP

Open VSP by selecting **Start>Program Files>Visual Sample Plan>Visual Sample** in Windows. This will open the "Welcome to Visual Sample Plan" screen. Several versions of VSP will be offered in a pop-up window. Choose **General (all-inclusive) VSP**. If you are working on an existing project choose **File>Open Project** from the main menu bar. If you are creating a new project choose **File>New Project**.

In VSP a *Project* refers to the map, report, and sample and cost information associated with one sampling design. This information is contained in one file (filename.VSP).

Step 2: Import the site/building map into VSP

A map of the site or building plan can be imported to VSP in the form of a .dxf file, an ArcView .shp file, or a Geosoft project. A map can also be generated from a list of coordinates imported in the form of a .txt file. VSP requires that map coordinates be in a planar coordinate system so that distances between locations in VSP reflect distances in the real world.

If site/building maps and coordinates are not available, VSP has the tools, which will allow the user to create one. Detailed instructions for using these tools can be found in the VSP User Manual (section 2.2.3, p. 2.4).

To import a map in the form of a .dxf or .shp file, choose **MAP** from the main menu (top row of screen) and then **Load Map from File** from the drop down menu. Locate the map file for import within the directory structure of your computer and double-click on the file. Select which layers you want to import then choose whether or not want to import embedded text from the DXF or SHP files.

To import a Geosoft project file select **File>Import>from Geosoft.**

For both DXF and SHP, VSP assumes that the distances are in feet. If this is not the case, the user must specify a different measurement unit by choosing **MAP>Map Settings,** and then selecting the appropriate unit (meters or inches) or **MAPS>Coordinate System** and change the units by selecting **Change Units**.

<u>DXF Files:</u>
Electronic files with the extension .dxf are produced by CAD systems. They do not contain information regarding the coordinate system, which was used to create the file. When a .dxf file is imported to VSP:

- Polylines, lines, circles and arcs created in .dxf will import to VSP as *map lines* and annotations will import as *text*.
- *Sample areas* have to be defined by the user within VSP.

<u>SHP Files:</u>
Electronic files with a .shp extension are produced by GIS. They do not contain information regarding the coordinate system that was used to create the file; however, VSP can detect if it contains latitude/longitude (lat/long) information. If lat/long information is detected when the file is being imported, VSP will prompt the user to convert the coordinates to a Universal Transverse Mercator (UTM) coordinate system. If this option is not accepted, distances within VSP will not relate to the real world. When a .shp file is imported to VSP:

- Polygons created in a .shp will import to VSP as *sample areas*.
- Polylines and lines will import as *map lines* and sample points in a .shp are ignored by VSP.

<u>TXT Files:</u>
A map can be generated by creating a list of coordinates in another application (spreadsheet or text editor), copying and then pasting them into the **Map View** window. If a spreadsheet was used, there must be a column of X-coordinates followed by a column of Y-coordinates. If using a text editor to create the list of coordinates, the X and Y columns must be separated by a space, tab, or comma. For either type of files, a blank row in the columns indicates the end of one polygon and the start of a new one. When these coordinates are pasted into the **Map View** window, the polygons that are generated will not be recognized as *sample areas*. A map generated using this methodology can be exported as a DXF or SHP file using **Map/Export**.

Save the site/building map by choosing **File>Save project.**

Step 3: Define *sample area(s)* on the *Project* map

Within VSP, a *sample area* is defined as a region from which the user wants to collect samples, often referred to as DU or decision area (DA). Typically, s*ample areas* are bounded regions within the site/building map. Sample size calculations can be computed without a defined *sample area*; however, sample locations cannot be generated or placed on the map without a previously defined *sample area. Project*s can contain a single or multiple *sample areas*.

For each *sample area* you wish to define, choose **Edit>Sample Areas>Define New Sample Area** or press the **New Area** button on the tool bar. A **Color** dialog box will appear. Choose a color for the *sample area*, click on the **OK** button in the right hand bottom corner of the window, and right click with the mouse within the desired closed polygon/sample area; a pop-up window will indicate the *sample area* size and the units it was measured in (e.g., square feet). Click **OK**. The *sample area* will now be shaded with the color you specified.

A second way to define a *sample area* is to place the cursor on each corner of the area and left-click with the mouse. When the area has been fully bounded by left-clicking on all corners, either 1) right-click the last segment of the bounded region, 2) click on the **Finish Area** button of the tool bar, or 3) select **Edit>Sample Areas>Finish New Sample Area** from the main menu bar.

VSP automatically generates a name and a measurement of the area and perimeter for each area. If the *sampling area* is within a building, such as a room, additional parameters are listed such as room height. This information can be accessed by right clicking on the *sample area* on the map. The name of the area can be modified by editing it within the pop-up dialog box.

Sample areas can be selected or deselected by left clicking within their boundaries. *Sample areas* that are selected are viewed as solid colored regions on the map. Deselected areas appear with only the outline in color. Samples will only be located on selected s*ample areas*.

Sample areas can be deleted by first making sure they are selected and then by choosing **Edit>Sample Areas>Delete Selected Sample Areas**.

Step 4: Import historical or other sample locations

If historical data is available such as coordinates of previously sampled locations, with or without corresponding analytical results, it can be applied to the development of the new sampling design or just overlaid onto the site/building map for visualization. Historical data can be imported in the form of a text file as well as from a Scribe file.

<u>Importing Sample Locations from a Text File:</u>
Enter the data so that each line/row represents a single sample location. Columns should be labeled according the following convention and contain the designated information:

Column	Column Header	Information Contained in Column
1	X Coord	X coordinate of sample
2	Y Coord	Y coordinate of sample
3	Label	Name of the sample
4	Value	Reading, analysis result, or quantity associated with the sample
5	Type	("Hotspot", "Random", "Adaptive-Fill", "Manual", "Unknown", "Systematic", "Perimeter")
6	Historical	(True, False)

The file must be saved as a text file in a tab delimited format. Choose **Map>Sample Points>Import** and select the text file. The historical sample locations will be assigned a symbol by VSP and placed within the selected *sample areas*. Right clicking on any sample point will show the Sample Information which VSP has stored for that location.

Importing Sample Locations from a Scribe File:
Sample data from a Scribe database can be imported directly into an existing VSP project. S*ample areas* must be defined and 'selected' (see Step 5) prior to importing the Scribe database. Choose **File>Import>from Scribe**. Select the desired Scribe file from the directory. If more than one site is stored within a single Scribe database file, VSP will prompt the user to select the correct site to import.

Step 5: Select *sample areas*

Select the *sample area(s)* for which you want to create a sampling design.

Select/deselect *sample areas* by clicking the left-mouse button within the bounded *sample area*. *Sample areas* that are filled in with their designated color are 'selected', S*ample areas* which are only outlined in color are 'deselected'.

If there is more than one *sample area* for which a single **Sampling Goal** will be applied, make sure only these *sample areas* are 'selected' and all other *sample areas* are 'deselected'. If desired, these areas can be joined into one *sample area*, even if they are non-contiguous, by making sure they are 'selected' and choosing **Edit**>**Sample Areas**>**Combine** from the menu or by left-clicking with the mouse on the **Combine** button on the tool bar.

Step 6: Define sampling goals

Choose **Sampling Goals** from the main menu.

Select the specific sampling goal for the 'selected' *sample area(s)* from the list. The sampling goal should have been previously determined in the scoping meeting or use the **Expert Mentor** built into VSP.

Follow the series of prompts to the main data analysis window. In some cases, the user will be prompted to select the expected statistical population distribution (normal, non-normal) and/or sampling type (e.g., ordinary, stratified, collaborative, adaptive cluster, etc.).

Click on each tab associated with the **Sampling Goal** and fill in the required information. Most of this information should have been gathered in Step 1. Information that is not available but is required to run the sampling scenario should be estimated using professional judgment and documented as such. VSP provides some help in estimating variability and specifying limits on error.

Select the **Apply** button in the bottom left-hand corner of the top tab. Statistics, sample number calculations and other information will appear as a summary written in red towards the bottom of the top tab of the data analysis window. Samples locations are generated and placed on the site/building map.

At this time, the design process can become iterative. For instance, if sample size is too big, confidence levels can be lowered and/or levels of uncertainty increased. These numbers can be adjusted and then the new information applied. For some sampling goals, different sample types, hot spot sizes and shapes can be explored. The site map, summary paragraph, VSP Report, sample coordinates, and statistical graphs are automatically updated to reflect the changes. Following each iteration:

1. Review the automatic reports and graphs generated by VSP, including the summary report, sample location map, sample coordinates, and statistical graph by choosing the **Quad window** button on the tool bar.
2. Examine the statistical graph to make sure the required confidence levels have been achieved. For many sampling goals, this graph is interactive and will allow the user to change sample sizes, variability, or areas of uncertainty and see the immediate effects in graphic form.
3. Examine sample placement to see if it is reasonable and logistically possible.
4. Make exploratory changes to the *Project* inputs and **Apply**.

Save the *Project*.

If during the iterative process, inputs, which were decided in the Scoping Meeting, had to be adjusted/altered, the Task Leader and the WAM should be contacted to discuss the changes and to verify that these changes are acceptable.

Repeat this process for each additional *sample area*.

When a sampling design has been finalized for all *sample areas*:

- Make sure all *sample areas* are **Selected**,
- Select **View>Report** from the Main Menu or right click on the **Report View** button (on the Tool Bar),
- Save a copy of the VSP Report and copy the report to a Word file by choosing **Edit>Copy** from the main menu and then pasting into an open Word file, and
- Add a summary of any assumptions that were made beyond the ones already included in the VSP Report and document values that were estimated in Step 8 along with the logic used in making the estimations.

3.5 VSP Expert Mentor and Sampling Design Selection

On opening VSP, the user is presented with a list of questions in the main window. These questions are links to instructions and help for applying VSP including a link to the **Expert Mentor**. The questions are a 'quick start' instruction guide for navigating and using VSP. The **Expert Mentor** is a tool, which provides guidance and recommendations to help apply VSP to meet defined DQOs. Current topics include help on using systematic planning, setting up VSP sites, maps and *sample areas*, and sample design selection.

To use the **Expert Mentor** for guidance in selecting the sampling design which will address the environmental questions being asked for a specific site:

1. Open VSP.
2. Choose **Expert Mentor** from the main window.
3. Choose **Sampling Design Selection.**
4. VSP provides a list of possible sampling goals:
   - Estimate or compare averages
   - Locate hotspots if they exist on my site
   - Show that at least y% of all possible samples are acceptable or no more than (1-Y)% are unacceptable
   - Do transect or anomaly sampling/analysis for UXO sites
   - Evaluate trends over time
   - Develop a geospatial contaminant concentration map
   - Evaluate spatial or temporal sample redundancies or inadequacies
   - Explore correlation between multiple analytes
   - Estimate or compare proportions
   - Assess whether the boundary around an area is contaminant free
   - Develop a targeted, purely judgmental sampling scheme
   - Sample items (such as drums, documents, or products)
   - Design a transect survey or develop a spatial map for radiological surveys
   - Compare groups of spatially located samples
5. Left-click on the box of each sampling goal that needs to be addressed for the site.
6. Left-click on the **Next** button at the bottom of the window.
7. Depending on the sampling goals chosen, the user is moved through a series of questions/windows requiring user input until VSP makes a final recommendation of sampling strategies to be applied to the site.

3.6 Uncertainty, Decision Errors and Defining the Gray Region

The objective of an environmental sampling effort is to obtain a representative (unbiased) sample of size *n* from a target population (e.g. surface soil) which will provide an accurate picture of the site (e.g. distribution and levels of arsenic remaining from historical use of pesticides at an orchard) being studied. Obstacles faced in collecting a representative sample include sampling and analytical uncertainty. A measure of uncertainty (standard deviation) can, in theory, be computed for both the sampling and the analytical phase of an environmental project. However, to simplify the statistical assumptions, VSP defaults to a single, combined estimate of standard deviation in its computations, often referred to as the standard deviation of the mean.

3.6.1 Uncertainty

Uncertainty associated with the sampling portion of an environmental response can include:

- Large-scale spatial heterogeneity (e.g., hot spots of contamination),
- Small-scale soil heterogeneity (e.g., differing particle size distributions within the vicinity of sample locations),
- Temporal effects (e.g., seasonal),
- Sample handling, and
- Sampling methodology.

Contributors to analytical uncertainty include:

- Sample preparation,
- Sample dilution,
- Instrument operator,
- Instrument calibration and
- Measurement.

VSP algorithms assume that sample sizes required to accurately meet project objectives will decrease if sampling and analytical uncertainties decrease or if more than one analysis (e.g., field screening and confirmatory laboratory analysis) per sample is used.

Implementation of SOPs for both the sampling and analytical portion of a project can significantly reduce uncertainty. These SOPs should be identified at the required Scoping Meeting and adhered to throughout the sampling and analytical phases of a project.

3.6.2 Decision Errors and the Gray Region

Associated with any environmental sampling *project* there exists an inherent level of variability that can affect the conclusions of a hypothesis test and/or project decision. Hypothesis tests are statistical tests between two competing scenarios; for example the site is dirty versus site is clean or mean levels of COCs on-site are greater than mean levels in a reference area versus levels of COCs on-site are not greater than those in a reference area. These tests begin with defining two hypotheses (scenarios) which will be tested:

1. Null hypothesis (H_o): represents the assumed condition; want to collect evidence to accept this hypothesis; if insufficient evidence is collected, the null hypothesis is rejected and the alternative hypothesis is accepted.

2. Alternative hypothesis (H_A): the motivating hypothesis; if insufficient evidence is collected or the statistical population parameters don't meet the test assumptions this hypothesis will be accepted

Type 1 Error (α, alpha): also known as a false positive is the probability of rejecting the null hypothesis when it is actually true

Type 2 Error (β, beta): also known as a false negative is the probability of accepting the null hypothesis as true when it is false

The Width of Gray Region (Δ, delta) is the distance from the action level to the boundary of the gray region where the gray region represents a range of true means where the consequences for accepting a false decision will be minor. The VSP User's Manual (p. 3.8) explains that "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." The narrower the gray region, the larger the sample size required to meet the DQOs. The gray region is defined based on professional judgment and cost/benefit analysis. A narrow gray region implies the need to determine conclusively whether or not to accept the null hypothesis when the sample mean is close to the action limit. Cases where a narrow gray region would be required would typically involve sites where there is the potential for a significant threat to human health or where clean-up costs would be excessive.

Estimated sampling standard deviation: a measure of variability expected between the samples, which can be obtained from previous sampling events at the site, previous experience with similar sites, or professional judgment. All sampling designs (except judgmental) will require some type of input, which defines the expected variability of the samples.

3.7 Exporting Map and Coordinate Files from VSP

Sample location coordinates generated by VSP can be exported and then imported to a GPS for use in the field. To export coordinates, choose **File** and then **Export…** from the drop down menu. A window will open allowing the user to choose what file format needs to be exported. Follow the prompts to export the file.

3.8 VSP Reports

VSP can generate four types of reports:

1) Map: a map of the sample area including sample points, labels and other information added by the user
2) Graph: a statistical representation of the sampling design which allows the user to perform a sensitivity analysis; can be manipulated to explore statistical confidence associated with increasing/decreasing sample size, accepting more uncertainty, etc. (Refer to the User Manual for details on manipulating these graphs)
3) Coordinates: a list of the sample location coordinates (X, Y) generated by VSP and imported by the User including labels and identification of historical data
4) Statistical Summary Report: a detailed report which includes all the other reports along with assumptions made in generating the sampling design, levels of confidence, number of sample locations, cost of the sampling design, statistical formulas applied and associated references, size of the sample area, etc.

When VSP is utilized for generating a sampling design, the Statistical Summary Report for the final design should be included as an attachment to the site-specific Uniform Federal Policy-Quality Assurance Project Plan (UFP-QAPP) or in Worksheet 17, Sampling Design and Rational, of the UFP QAPP. To save the report, right click on it when it is open in VSP, choose **Report Manager…**, select **Save Current Report to List** (button at top of window), select the file (check box to the right of the file name), choose **Save As…** (right bar of the Report Manager window) and save the file.

- Example 1 in Appendix A: Summary Report for a systematic triangular grid sampling design. This design was used for identification of a hot spot.

- Example 2 in Appendix A: Summary Report for a site divided into multiple sampling areas, applying a random sample design, where the number of samples was designated by the user based on financial limitations of the sampling event.

4.0 RESPONSIBILITIES

4.1 ERT WAM

The ERT WAM is responsible for providing technical expertise and technical direction to the contractor, preparing task orders/work assignments, reviewing deliverables, interacting with the Regional customers and monitoring the financial and administrative management of the project.

4.2 ERT Quality Coordinator

The ERT Quality Coordinator provides quality assurance oversight for all projects and implements/maintains the ERT Quality Assurance Program.

4.3 ERT Contractor Task Leaders

ERT contractor's TLs are responsible for gathering the inputs required for applying VSP. This includes communicating with the WAM to identify the project-specific objectives and determine acceptable levels of confidence/uncertainty, obtaining a suitable .dxf or .shp file, estimating sampling and analytical costs (if a factor in determining sample numbers), and locating available historical data. If the Task Leader is not running VSP, then it is his/her responsibility to ensure the necessary information is conveyed to the designated user of VSP and to be a liaison with the WAM when iterations of confidence, sample numbers and sampling designs are explored. The Task Leader is also responsible for seeking assistance from a Statistician, if needed, to meet the project goals and ensuring the Statistical Sampling Summary Report for the final sampling design is included in the site-specific UFP-QAPP.

4.4 ERT Contractor Direct Report Managers

The ERT contractor's Direct Report Managers are responsible for ensuring adherence to this SOP by ensuring that all deliverables have received an appropriate level of technical review.

4.3 ERT Contractor Statistician

The ERT contractor's Statistician is responsible for reviewing sampling designs produced using VSP and the assumptions made in creating them, to assure their defensibility in meeting project-specific data quality objectives.

4.4 ERT Contractor Quality Assurance/Quality Control Officer

The ERT contractor's Quality Assurance/Quality Control (QA/QC) is responsible for reviewing and approving any revisions to this SOP, and ensuring compliance with this SOP by regularly auditing a percentage of sampling designs prepared using VSP by ERT contractor personnel.

5.0 REFERENCES

Matzke BD, Wilson JE, Newburn LL, Dowson ST, Hathaway JE, Sego LH, Bramer LM, and Pulsipher BA. 2014. *Visual Sample Plan Version 7.0 User's Guide*. PNNL-23211, Pacific Northwest National Laboratory, Richland, Washington. http://vsp.pnnl.gov/documentation.stm

United States Environmental Protection Agency. 2002. Guidance on Choosing a Sampling Design for Environmental Data Collection for Use in Developing a Quality Assurance Project Plan. Office of Environmental Information. Washington, D.C. EPA QA/G-5S. EPA/240/R-02/005. December 2002. https://www.epa.gov/sites/production/files/2015-06/documents/g5s-final.pdf

6.0 APPENDICES

A – Example Statistical Summary Reports

APPENDIX A
Example Statistical Summary Reports
SOP: ERT-PROC-2076-21
June 2021

Example 1.
Statistical Summary Report Generated for a Systematic Grid Sampling Design – Cost Not Included

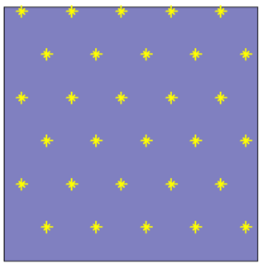
Systematic sampling locations for detecting an area of elevated values (hot spot)

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 (e.g., soil, groundwater, etc.) and how to analyze the samples (in-situ, fixed laboratory, etc.) are addressed in other sections of the sampling plan.

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

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	Detect the presence of a hot spot that has a specified 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 minimum size of hot spot	Algorithm developed by Singer and Wickman (1969)
Input number of samples	30
Type of samples	Point Samples
Number of samples on map [a]	30
Number of selected sample areas [b]	1
Specified sampling area [c]	90000.00 ft^2
Grid pattern	Triangular
Size of grid / Area of grid [d]	58.8566 feet / 3000 ft^2
Total cost of sampling [e]	$0.00

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



Area: Area 1					
X Coord	**Y Coord**	**Label**	**Value**	**Type**	**Historical**
50.5021	40.2757			Hotspot	
109.3587	40.2757			Hotspot	
168.2153	40.2757			Hotspot	
227.0719	40.2757			Hotspot	
285.9286	40.2757			Hotspot	
21.0738	91.2470			Hotspot	
79.9304	91.2470			Hotspot	
138.7870	91.2470			Hotspot	
197.6436	91.2470			Hotspot	
256.5002	91.2470			Hotspot	
50.5021	142.2183			Hotspot	
109.3587	142.2183			Hotspot	
168.2153	142.2183			Hotspot	
227.0719	142.2183			Hotspot	
285.9286	142.2183			Hotspot	
21.0738	193.1897			Hotspot	
79.9304	193.1897			Hotspot	
138.7870	193.1897			Hotspot	
197.6436	193.1897			Hotspot	
256.5002	193.1897			Hotspot	
50.5021	244.1610			Hotspot	
109.3587	244.1610			Hotspot	
168.2153	244.1610			Hotspot	
227.0719	244.1610			Hotspot	
285.9286	244.1610			Hotspot	
21.0738	295.1323			Hotspot	
79.9304	295.1323			Hotspot	
138.7870	295.1323			Hotspot	
197.6436	295.1323			Hotspot	
256.5002	295.1323			Hotspot	

Primary Sampling Objective
The primary purpose of sampling at this site is to detect "hot spots" (local areas of elevated concentration) of a given size and shape with a specified probability, 1-□.

Selected Sampling Approach
This sampling approach requires systematic grid sampling with a random start. If a systematic grid is not used, the probability of detecting a hot spot of a given size and shape will be different than desired or calculated.

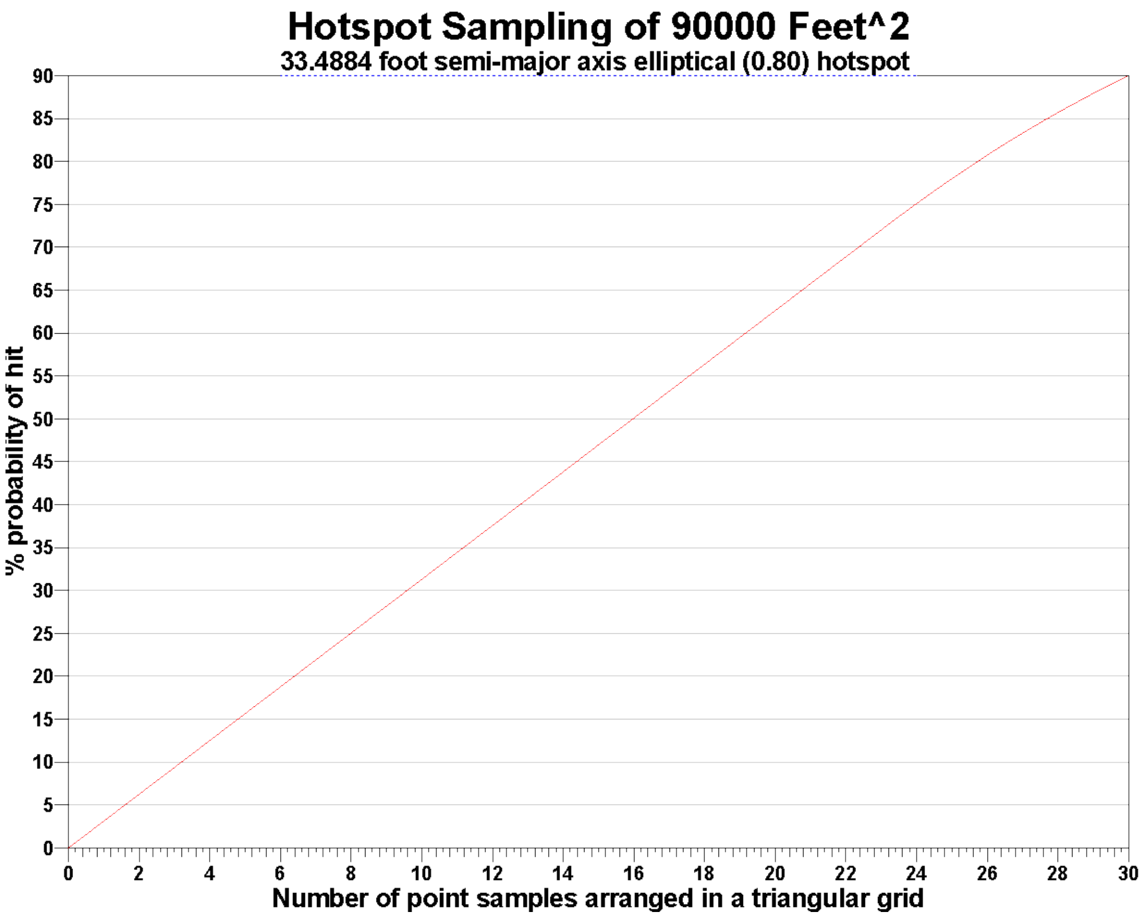
Number of Total Samples: Calculation Equation and Inputs
The algorithm used to calculate the probability of a hit (which makes possible the calculation of the hot spot size or the number of samples) was developed by Singer and Wickman (1969) and Singer (1972) with refinements by Davidson (1995). Gilbert (1987) also discussed hotspot sampling designs. Inputs to the algorithm include the size, shape, and orientation of a hot spot of interest, an acceptable probability of finding a hot spot, the desired type of sampling grid, and the sampling budget. For this design, the smallest hot spot that could be detected was calculated based on the given grid size and other parameters.

The inputs to the algorithm that result in the smallest hot spot that could be detected are:

Parameter	Description	Value
Inputs		
Samples	Number of samples specified by user	30
1-□	Probability of detection	90%
Grid Type	Grid pattern (Square, Triangular or Rectangular)	Triangular
Grid Size	Spacing between samples	58.8566 feet
Grid Area	Area represented by one grid	3000 ft^2
Sample Type	Point samples or square cells	Points
Hot Spot Shape	Hot spot height to width ratio	0.8
Angle	Angle of orientation between hot spot and grid	Random
Sampling Area	Total area to sample	90000.00 ft^2
Outputs		
Hot Spot Size	Length of hot spot semi-major axis	33.4884 feet
Hot Spot Area [a]	Area of hot spot ($Length^2$ * Shape * □)	2818.57 ft^2

[a] Length of semi-major axis is used by Singer-Wickman algorithm. Hot spot area is provided for informational purposes.

The following graph shows the relationship between the number of samples and the probability of finding the hot spot. The dashed blue line shows the actual number of samples for this design (which may differ from the optimum number of samples because of edge effects).



Assumptions that Underlie the VSP Locating a Hot Spot Design Method

1. In the decision area, there is at least one hotspot of the designated size, which is circular or elliptical in shape.
2. The level of contamination that defines a hotspot is well defined.
3. The location of the hotspot is unknown, and if a hotspot is present, all locations within the sampling area are equally likely to contain the hotspot.
4. With a randomly determined starting location, samples are taken on a square, rectangular or triangular (equilateral) grid pattern that covers the decision area.
5. Each sample is collected, handled, measured or inspected using approved methods that yield sufficiently precise measurements.
6. A very small proportion of the surface of the decision area will be sampled. The area sampled by a single sample is much smaller than the hotspot of interest.
7. The sample methodology and sample analysis process is the same for all sample locations.
8. There are no classification errors. If a hotspot is sampled, then contamination is detected (i.e., no false negatives). If an uncontaminated area is sampled, it is not mistakenly identified as a hotspot (i.e., no false positives).

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $0.00, which averages out to a per sample cost of $0.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**30 Samples**
Field collection costs		$0.00	$0.00
Analytical costs (Analyte 1)	$0.00	$0.00	$0.00
Sum of Field & Analytical costs		**$0.00**	**$0.00**
Fixed planning and validation costs			$0.00
Total cost			**$0.00**

Recommended Data Analysis Activities
Post data collection activities generally follow those outlined in EPA's Guidance for Data Quality Assessment (EPA, 2006). The data analysts will become familiar with the context of the problem and goals for data collection and assessment. The data will be verified and validated before being subjected to statistical or other analyses. Graphical and analytical tools will be used to verify to the extent possible the assumptions of any statistical analyses that are performed as well as to achieve a general understanding of the data. The data will be assessed to determine whether they are adequate in both quality and quantity to support the primary objective of sampling.

A map of the actual sample locations will be generated so that the sampling plan and the field implementation may be compared. Deviations from planned sample locations due to topographic, vegetative, or other features will be noted. Their impacts will be qualitatively assessed. If a hot spot is discovered, additional sampling may be performed to determine its size and shape, in which case, the initial assumptions of the sampling design may then be assessed and/or reconsidered.

References
EPA 2006. *Data Quality Assessment: Statistical Methods for Practitioners EPA QA/G-9S*, EPA/240/B-06/003, U.S. Environmental Protection Agency, Office of Environmental Information, Washington DC.

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

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

Singer, D.A. and J.E. Wickman. 1969. *Probability Tables for Locating Elliptical Targets with Square, Rectangular, and Hexagonal Point Nets*. Pennsylvania State University, University Park, Pennsylvania. Special Publication 1-69.

Singer, D.A. 1972. "ELIPGRID: A Fortran IV program for calculating the probability of success in locating elliptical targets with square, rectangular and hexagonal grids." *Geocom Bulletin/Programs* 4:1-16.

This report was automatically produced* by Visual Sample Plan (VSP) software version 7.5.

This design was last modified 4/24/2015 9:48:00 AM.

Software and documentation available at http://vsp.pnnl.gov

* - The report contents may have been modified or reformatted by end-user of software.

Example 2.
Statistical Summary Report Generated for a Random Sampling Design across Two Sampling Areas – Number of Samples was Chosen by the User based on Cost of Sampling

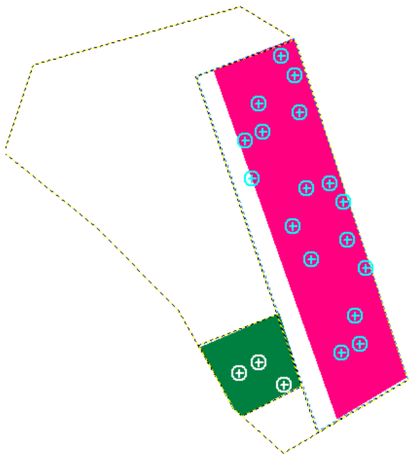
Predetermined Number of Random Sampling Locations

Summary
This report summarizes the sampling design, 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. A figure that shows sampling locations in the field and a table that lists sampling location coordinates are also provided below.

SUMMARY OF SAMPLING DESIGN	
Primary Objective of Design	User Defined
Sample Placement (Location) in the Field	Simple random sampling
User specified number of samples	20
Number of samples on map [a]	20
Number of selected sample areas [b]	2
Specified sampling area [c]	82845.95 m^2
Total cost of sampling [d]	$11,000.00

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



Area: Area 5					
X Coord	**Y Coord**	**Label**	**Value**	**Type**	**Historical**
740913.7107	1991865.8164			Random	
740942.7538	1991883.1099			Random	
740981.4779	1991848.5230			Random	

Area: Area 6					
X Coord	**Y Coord**	**Label**	**Value**	**Type**	**Historical**
740933.2945	1992158.9370			Random	
741069.7830	1992123.4951			Random	
740996.9891	1992314.8815			Random	
740942.3938	1992272.3512			Random	
741087.9814	1991953.3739			Random	
741015.1876	1992144.7603			Random	
741022.0120	1992038.4345			Random	
740949.2182	1992229.8209			Random	
741094.8059	1991910.8436			Random	
741076.6074	1992066.7881			Random	
741003.8136	1992258.1744			Random	
741103.9051	1992024.2578			Random	
740921.9205	1992215.6441			Random	
741067.5082	1991896.6669			Random	
740994.7143	1992088.0532			Random	
741049.3097	1992151.8487			Random	
740976.5159	1992343.2350			Random	

Primary Sampling Objective
The primary purpose of sampling at this site is unknown to Visual Sample Plan. The number of samples may have been calculated in another sampling design in Visual Sample Plan, or may have been calculated externally to VSP. Alternatively, the purpose may be based entirely on professional judgment.

Selected Sampling Approach
Locating the sample points randomly provides data that are separated by many distances, whereas systematic samples are all equidistant apart. Therefore, random sampling provides more information about the spatial structure of the potential contamination than systematic sampling does. As with systematic sampling, random sampling also provides information regarding the mean value, but there is the possibility that areas of the site will not be represented with the same frequency as if uniform grid sampling were performed.

Cost of Sampling
The total cost of the completed sampling program depends on several cost inputs, some of which are fixed, and others that are based on the number of samples collected and measured. Based on the numbers of samples determined above, the estimated total cost of sampling and analysis at this site is $11,000.00, which averages out to a per sample cost of $550.00. The following table summarizes the inputs and resulting cost estimates.

COST INFORMATION			
Cost Details	**Per Analysis**	**Per Sample**	**20 Samples**
Field collection costs		$100.00	$2,000.00
Analytical costs (Analyte 1)	$400.00	$400.00	$8,000.00
Sum of Field & Analytical costs		**$500.00**	**$10,000.00**
Fixed planning and validation costs			$1,000.00
Total cost			**$11,000.00**

Further Recommended Data Analysis Activities
Post data collection activities generally follow those outlined in EPA's Guidance for Data Quality Assessment (EPA, 2000). The data analysts will become familiar with the context of the problem and goals for data collection and assessment. The data will be verified and validated before being subjected to statistical or other analyses. Graphical and analytical tools will be used to verify to the extent possible the assumptions of any statistical analyses that are performed as well as to achieve a general understanding of the data. The data will be assessed to determine whether they are adequate in both quality and quantity to support the primary objective of sampling.

This report was automatically produced* by Visual Sample Plan (VSP) software version 7.5.

This design was last modified 7/27/2016 10:00:59 AM.

Software and documentation available at http://vsp.pnnl.gov

* - The report contents may have been modified or reformatted by end-user of software.