Contract No:

This document was prepared in conjunction with work accomplished under Contract No. DE-AC09-08SR22470 with the U.S. Department of Energy.

We put science to work.™

Savannah River National Laboratory™

OPERATED BY SAVANNAH RIVER NUCLEAR SOLUTIONS

A U.S. DEPARTMENT OF ENERGY NATIONAL LABORATORY • SAVANNAH RIVER SITE • AIKEN, SC

Statistical Evaluation of Tank 12 Sampling Options

E.P. Shine

June 2014

SRNL-STI-2014-00263, Rev. 0

SRNL.DOE.GOV

Statistical Evaluation of Tank 12 Sampling Options

E.P. Shine

OPERATED BY SAVANNAH RIVER NUCLEAR SOLUTIONS

1.0 APPROVALS/TASK TECHNICAL REQUEST IDENTIFICATION

Author: E. P. Shine	**Signature:**	**Organization:** **SRNL-CS-ACES**	**Date:**
Technical Reviewer: S. H. Reboul	**Signature:**	 **SRNL-E&CPT-ACP**	
Technical Reviewer: T. B. Edwards	**Signature:**	**Organization** **SRNL-CS-ACES**	**Date:**
Responsible Manager: M. K. Harris	**Signature:**	**Organization:** **SRNL-CS**	**Date:**
Manager: F. M. Pennebaker	**Signature:**	**Organization:** **SRNL-E&CPT**	**Date:**
Customer: J. P. Pavletich	**Signature:**	**Organization:** **SRR–RM&A**	**Date:**
Customer: M. J. Mahoney	**Signature:**	**Organization:** **SRR–RM&A**	**Date:**
Customer: J. E. Occhipinti	**Signature:**	**Organization:** **SRR–Engineering**	**Date:**

Task Technical Request Title: SRNL Evaluation of Tank 12 Sampling Options	**TTR Number:** **G-TTR-H-00007** **Revision: 0**	**TTR Date:** **4-28-14**

LIST OF REVISIONS

Revision Number	Summary of Changes	Date
0	Initial Issuance	June 2014

2.0 INTRODUCTION

Savannah River Remediation (SRR) is preparing Tank 12 for removal from service and has issued a scope of work to the Savannah River National Laboratory (SRNL) under a Task Technical Request (TTR) G-TTR-H-00007 (Savannah River Remediation, 2014) to provide statistical support for the Sample Location Determination Report (SLDR) and the Tank-Specific Sampling and Analysis Plan (TSAP). In response SRNL has issued a Task Technical and Quality Assurance Plan (TTQAP) (Shine, 2014) outlining the strategy for assessing sampling options.

This report documents the recommendations for sampling analyses following the plan in the TTQAP. The general aim is to evaluate a specific set of sampling options in the TTR from the perspective of statistical uncertainty.

The target population for the sampling plan is the residual material in Tank 12 that is found as a crystalline encrustation on the upper portions of the internal cooling coils and the sludge found on the tank floor. The floor material is comprised of an undisturbed mound behind the valve house piping and a thin layer of material elsewhere. The first two columns of Table 1 reproduce the preliminary volume estimates from the TTR. The last two columns display the relative volume in percent of the total residual material and as a percent of the residual material on the floor and in the mound.

Table 1. Preliminary Volume Estimates for Tank 12 Residual Materials

Location of Residual Material	Approximate Volume (gallons)	Relative Volume (%)	Relative Volume without Coils (%)
Encrustations on cooling coils	400	28.6%	
Tank floor (outside the mound)	780	55.7%	78%
Mound behind valve house	220	15.7%	22%

The Liquid Waste Tank Residuals Sampling Analysis and Program Plan (LWTRSAPP) (Pavletich, 2013) prescribes the general approach to design sampling plans for residual material in SRS liquid waste tanks. The sampling strategy documented in the LWTRSAPP outlines a method to allocate samples to strata based on the relative volumes of the material present in the strata. This is a frequently used strategy and is known as “allocation proportional to size” (of strata) in the general sampling literature (see (Cochran, 1976), (Sukhatme & Sukhatme, 1970), and (Thompson & Seber, 1996)). For example, if a total of 15 primary samples are to be located in Tank 12, then according to the guidance in the LWTRSAPP, the number of primary samples to be obtained from the mound behind the valve house piping would be $15.7\% \text{ of } 15 \text{ samples} = 2.26 \rightarrow 3 \text{ samples}$[1], the number of primary samples to be obtained from the tank floor area would be $55.7\% \text{ of } 15 = 8.36 \rightarrow 9 \text{ samples}$, with the remaining 3 samples allocated to the residual material on the cooling coils. This allocation does not appear very realistic since the undisturbed mound

[1] The symbol $\rightarrow$ is interpreted as “rounded to” the nearest multiple of 3. In this example, numbers of samples in each stratum must be rounded to the nearest multiple of three so that the same number of samples from each stratum can be placed into each of the three composite samples.

of material behind the valve house piping is expected to be more heterogeneous than the material on the tank floor that should be well-mixed due to the waste treatment and removal activities.

An allocation strategy to minimize sampling uncertainty for a fixed budget (or minimize costs for a fixed sampling uncertainty) considers heterogeneity of the material within strata as well as the relative volume of material within strata[2] and is referred to as Neyman allocation. Neyman allocation of samples to strata is equivalent to allocation proportional to size when the material heterogeneity is the same within all strata and the variable cost of a sample is the same in every stratum.

Table 2, also from the TTR, lists a series of sampling schemes that are to be evaluated. Sampling Options 1, 2, and 3 specify several different sample allocation schemes that differ in the number of samples taken from the floor and the number of samples taken from the mound material. Note that Sampling Options 1, 2, and 3 specify that each sample from the cooling coil material will be obtained at a different elevation.

Table 2. Possible Compositing Schemes Using Different Numbers of Mound Samples

Composite Sample	Aliquots Used for Compositing Each Analytical Sample					Mound Sampling Details
Option 1	Coil*	Floor	Floor	Floor	Mound A1	3 samples; 3 different locations and depths
Option 2	Coil*	Floor	Floor	Mound B2	Mound B1	6 samples; 3 different locations and 2 different depths at each location
Option 3	Coil*	Floor	Mound C3	Mound C2	Mound C1	9 samples; 3 different locations and 3 different depths at each location
Option 4	Use more than 5 samples per array. The material (coil, mound, and floor) for the extra samples used for compositing to be evaluated for impact to the sampling uncertainty.					

*Each sample would be from a different elevation on the coil. It is assumed that at least 3 cooling coil samples will be obtained. The cooling coil samples will be incorporated into the composite samples.

[2] The relative costs of sampling in different strata can also be considered, but it is not a design parameter in this study.

It is assumed that at least 3 distinct cooling coil samples will be obtained, and Sampling Options 1, 2, and 3 infer that all cooling coil material is considered to be from a single stratum. Sampling Option 4 uses more than 5 primary samples per composite sample, permitting the cooling coil material to be partitioned into different strata based on process history.

The general evaluation strategy is to first produce a baseline result assuming that all strata have the same material heterogeneity. This is the result that would be obtained by applying the LWTRSAPP strategy. Following this, a sensitivity analysis will be performed using a series of different relative material heterogeneities for the mound, floor, and cooling coil strata.

Note that the aim of this task is to provide a ranking of the sampling options based on expected statistical uncertainty. A precise prediction of the statistical uncertainties for each sampling option depends on material heterogeneity, and sampling and analytical uncertainties (for certain matrices) that have not been quantitated.

3.0 Sample Allocation Strategy

Primary samples refer to samples of residual material obtained directly from the tank. The LWTRSAPP prescribes a total of 15 primary samples in order to achieve a representative selection from the population of residual material. The population of residual material is partitioned into non-overlapping regions called strata based on distinctive features such as mounds and shallow layers of material on the tank floor, and material adhering to or contained in other locations. Sample allocation refers to the distribution of these primary samples to the various strata.

Table 3 lists design parameters that are generally considered in determining the sample allocation. The LWTRSAPP strategy is based only on the relative volume[3] of residual material in each stratum. This strategy is commonly used in situations where the variation of a response (analyte concentration) is assumed to be the same within each stratum. It is also sometimes used where there are a large number of responses with widely different spatial distributions and none of the responses is much more important than any of the others. For Tank 12, one of the strata contains the residual material in an undisturbed mound, so there is a rationale to believe that the mound material will be more spatially heterogeneous than the layer of material deposited on the floor after cleaning operations ceased.

[3] The density of the residual material and, thus the mass of the residual material, will not be determined until the samples have been collected and measured for physical parameters. Consequently, the sample allocation aspect of the sampling design is determined by volume, but once the physical parameters of the samples have been measured in the laboratory, the composite samples are formed from the primary samples based on mass.

Table 3. Design Parameters for Sample Allocation

Parameter	Description
Stratum Size	Relative volume of residual material in the strata: $p_i, i=1,2,\ldots,k$.
Stratum Variation	Spatial heterogeneity, sampling standard deviation: $S_i, i=1,2,\ldots,k$.
Stratum Sampling Cost per Sample	Variable cost per sample in different strata: $c_i, i=1,2,\ldots,k$.

3.1 Sample Allocation when Random Sampling is Performed within All Strata

The formula for sample allocation is derived assuming that the random sampling is performed within each stratum. The detailed results are given in Appendix A and are summarized here. The objective is to minimize the sampling uncertainty for a fixed budget $c_{T.}$ In this report, the variable costs will be assumed to be the same in every stratum. This implies that the total number of samples, n_o, is set by the budget. The total volume of material collected in the samples from each stratum is assumed to be a negligible fraction of the population of material in that stratum. These simplifying assumptions lead to the following formula to determine the number of samples to collect from each of the *k* strata.

$$n_i = n_0 \frac{p_i S_i}{\sum_{j=1}^{k} p_j S_j}, i=1,2,\ldots,k, \qquad (1)$$

where the proportion of the volume of residual material from stratum *i* is p_i, and the standard deviation from sampling in stratum *i* is $S_i, i=1,2,\ldots,k$.

The LWTRSAPP assumes that the true sampling standard deviations $S_i, i=1,2,\ldots,k$, are equal. This leads to the formula for allocation proportion to (stratum) size,

$$n_i = n_0 \frac{p_i}{\sum_{j=1}^{k} p_j}, i=1,2,\ldots,k, \qquad (2)$$

The material from the primary samples will be consolidated into 3 composite samples, so the sample sizes determined from Eqns (1) and (2) should be rounded to the nearest multiple of 3.

3.2 Sample Allocation when Stratified Random Sampling is Performed within the Mound Stratum

In Section 3.1, three strata were discussed: the cooling coils, the floor, and the mound. In Section 3.2, the mound stratum will be further partitioned into nonoverlapping horizontal layers in order to obtain stratified random samples from horizontal layers of mound material. The change from random sampling to stratified random sampling for the mound material can be viewed as just adding more strata to the overall structure described in Section 3.1. For example, if the mound were partitioned into two layers, there now would be 4 strata: the cooling coils, the floor, the upper mound, and the lower mound. If the mound were to be partitioned into three nonoverlapping layers, then there would be 5 strata: the cooling coils, the floor, the upper mound, the middle mound, and the lower mound. An advantage of this perspective is that all of the formulas from Section 3.1 can still be used, except the number of strata, *k*, would now be 4 or 5 instead of 3. The assumption is made that the sampling standard deviation, S_{M_L} , is the same within all *L* layers, but may vary in value based on the number of layers, *L*. Typically, it may be expected that S_{M_L} will stay the same or decrease with an increase in *L*. If $L = 1$, $S_{M_L} = S_M$, the sampling standard deviation when the mound is randomly sampled. The details are given in Appendix B.

The sample allocation formulas are as follows.

$$
\begin{aligned}
n_i &= n_0 \frac{p_i S_i}{\sum_{\substack{j=1\\ j\neq M}}^{k} p_j S_j + p_M S_{M_L}}, i = 1, 2, \ldots, k, \text{except for M (the Mound)} \\
n_M &= n_0 \frac{p_M S_{M_L}}{\sum_{\substack{j=1\\ j\neq M}}^{k} p_j S_j + p_M S_{M_L}},
\end{aligned} \tag{3}
$$

Note that Eqn (3) will be used iteratively to determine the number of primary samples. The procedure is illustrated by example in Section 4.2

4.0 Application of the Allocation Schemes to Tank 12

The sample allocation formulas are applied to the Tank 12 sampling options in this section. First, the procedure of Section 3.1 is used to assess the performance of the sampling options when random sampling is used to obtain the primary samples from the mound, then the procedure of Section 3.2 is used to assess the performance of the sampling options when stratified random sampling is used to obtain the primary samples from the mound.

4.1 Tank 12 Sample Allocation when Random Sampling is Performed within All Strata

Tank 12 has three strata: the mound (M), the floor (F), and the cooling coils (C). It is assumed that the number of cooling coil samples is limited to $n_c = 3$. This leaves $n_o = 15 - 3$ or 12 samples to be allocated between the mound and floor strata.

Define $\lambda = \frac{S_M}{S_F}$ to be the ratio of the mound to floor sampling standard deviations. Both S_M and S_F reflect the total variability for concentrations among samples from their respective strata. The value of lambda should be based on an engineering estimate and should consider the spatial heterogeneity of the material and the performance of the sampler.

Applying Eqn (1) to the mound and floor strata yields the following.

$$n_F = n_0 \frac{p_F S_F}{p_F S_F + p_M S_M} = n_0 \frac{p_F}{p_F + \lambda p_M} \text{ and}$$
$$n_M = n_0 \frac{p_M S_M}{p_F S_F + p_M S_M} = n_0 \frac{\lambda p_M}{p_F + \lambda p_M}. \quad (4)$$

The relative stratum volumes, p_F and p_M, were given in the right column of Table 1. An engineering estimate of λ, the ratio of the sampling standard deviation for the mound material to the sampling standard deviation for the floor material, is needed to begin the process sample allocation. Figure 1 plots the mound sample size, n_M, vs. lambda (λ) for Tank 12. As the ratio of the mound to the floor standard deviations increases, the graph indicates that the number of samples allocated to the mound increases (and the number of samples allocated to the floor decreases). Since the primary samples will be consolidated into three composite samples or primary samples each, the allocation of samples to the mound and the floor should be in multiples of three.

Under the LWTRSAPP guidance, $\lambda = 1$. The mound sample size for $\lambda = 1$ is displayed in Figure 1 as the solid red circle, showing that the LWTRSAPP recommends 3 samples from the mound based on only its relative volume.

Three composite samples have been proposed for Tank 12. Each composite sample requires material from primary samples obtained from every stratum. Consequently, the number of primary samples from each stratum should be a multiple of 3, and sample allocation results needed to be rounded to the nearest multiple of 3 samples. Sample allocation results below 4.5 are rounded to 3 primary samples, those between 4.5 and 7.5 are rounded to 6 primary samples, and those greater than 7.5 are rounded to 9 primary samples. Based on these midpoints, the next paragraph determines a set of rules that determines the number of primary samples per stratum based on the value of λ.

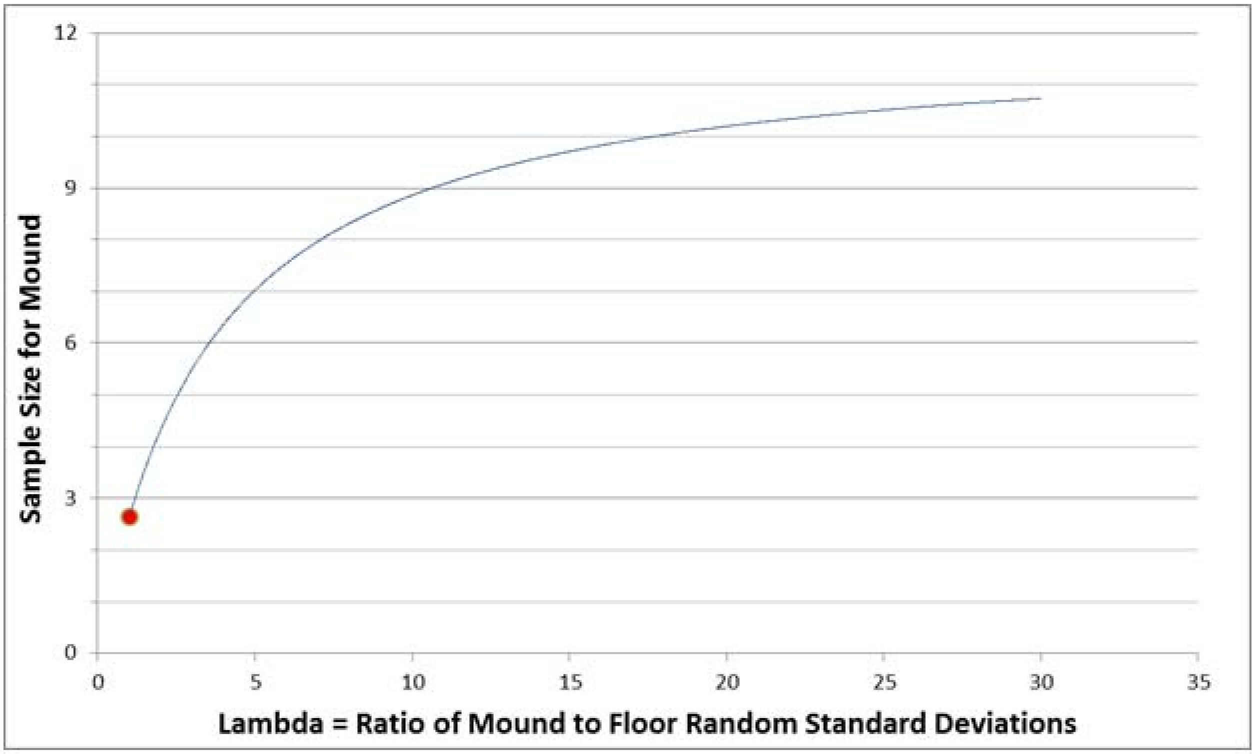

Figure 1. Plot of Mound Sample Size vs. Ratio of Mound Standard Deviation to the Floor Standard Deviation

Consider that the midpoint between 3 and 6 samples is 4.5, and that the midpoint between 6 and 9 samples is 7.5. Substituting $n_M = 4.5$ for the number of mound samples and the relative volumes $p_F = 78\%$ and $p_M = 22\%$ into Eqn (3), the corresponding value of lambda, $\lambda = 2.1$. Similarly, $n_M = 7.5$ yields $\lambda = 5.9$. Therefore, an engineering estimate of λ below 2.1 indicates the number of primary samples from the mound should be 3 (and the floor and cooling coil primary samples should be 9 and 3, respectively), an estimate of λ between 2.1 and 5.9 suggests 6 primary samples from the mound (and the floor and cooling coil primary samples should be 6 and 3, respectively), and an estimate of λ over 5.9 indicates 9 primary samples from the mound (and the floor and cooling coil primary samples should both be 3).

These sample allocation results when the mound is randomly sampled (no layers) are summarized in Section 5.0.

4.2 Tank 12 Sample Allocation when Stratified Random Sampling is Performed within the Mound Stratum

Tank 12 has three major strata: the mound (M), the floor (F), and the cooling coils (C). It is assumed that the number of cooling coil samples is limited to $n_c = 3$. This leaves $n_o = 15 - 3$ or 12 samples to be allocated between the mound and floor strata. Further, the mound will be substratified into L horizontal layers.

Previously, S_M and S_F were defined as the sampling standard deviations for the mound and the floor materials, respectively, when random sampling was performed. The term $\lambda = S_M / S_F$ retains its definition as ratio of the sampling standard deviation for the mound material to the sampling standard deviation for the floor material.

Define $\delta_L = \dfrac{S_{M_L}}{S_M}$, for $L = 1, 2,$ and 3 layers, where S_{M_L} is the sampling standard deviation within a single layer of a mound with L layers. Obviously, $\delta_L = \dfrac{S_{M_L}}{S_M} = \dfrac{S_M}{S_M} = 1,$ when $L = 1$. The design parameter, $\lambda_L^{(st)} = \delta_L \lambda = \dfrac{S_{M_L}}{S_M} \cdot \dfrac{S_M}{S_F} = \dfrac{S_{M_L}}{S_F}, L = 1, 2, 3,$ is used in the sample allocation formulas when the mound samples are obtained by stratified sampling (st). The sample allocation formulas are as follows.

$$n_F = n_0 \frac{p_F S_F}{p_F S_F + p_M S_{M_L}} = n_0 \frac{p_F}{p_F + \delta_L \lambda p_M} = n_0 \frac{p_F}{p_F + \lambda_L^{(st)} p_M} \text{ and}$$
$$n_M = n_0 \frac{p_M S_M}{p_F S_F + p_M S_M} = n_0 \frac{\delta_L \lambda p_M}{p_F + \delta_L \lambda p_M} = n_0 \frac{\lambda_L^{(st)} p_M}{p_F + \lambda_L^{(st)} p_M}. \quad (5)$$

Figure 1 in Section 4.1 is appropriate for the mound sample size, n_m, but the horizontal axis is now $\lambda_L^{(st)} = \delta_L \lambda,$ the ratio of the sampling standard deviation within a single layer of the mound with L layers to the sampling standard deviation for the entire mound (no layers).

Unlike the sample allocation for random sampling in Section 4.1, Eqn (5) must be used in an iterative fashion. The process is depicted in the flowchart in Figure 2. An important characteristic of the formula for n_M in Eqn (5) is that the value of $\lambda_L^{(st)} = \delta_L \lambda$ depends on the number of layers, L. The procedure for a stratified mound or a layered design begins by estimating the parameter, λ, the ratio of the sampling standard deviation for the mound material to the sampling standard deviation for the floor material under random sampling. An initial estimate of the number of layers, L_0, must be chosen, and then δ_L, which depends on the number of layers, is estimated. The subscript "0" on L indicates that it is the initial choice. The parameter $\lambda_L^{(st)} = \delta_L \lambda$ is calculated, and entered into the formula for n_M in Eqn (5).

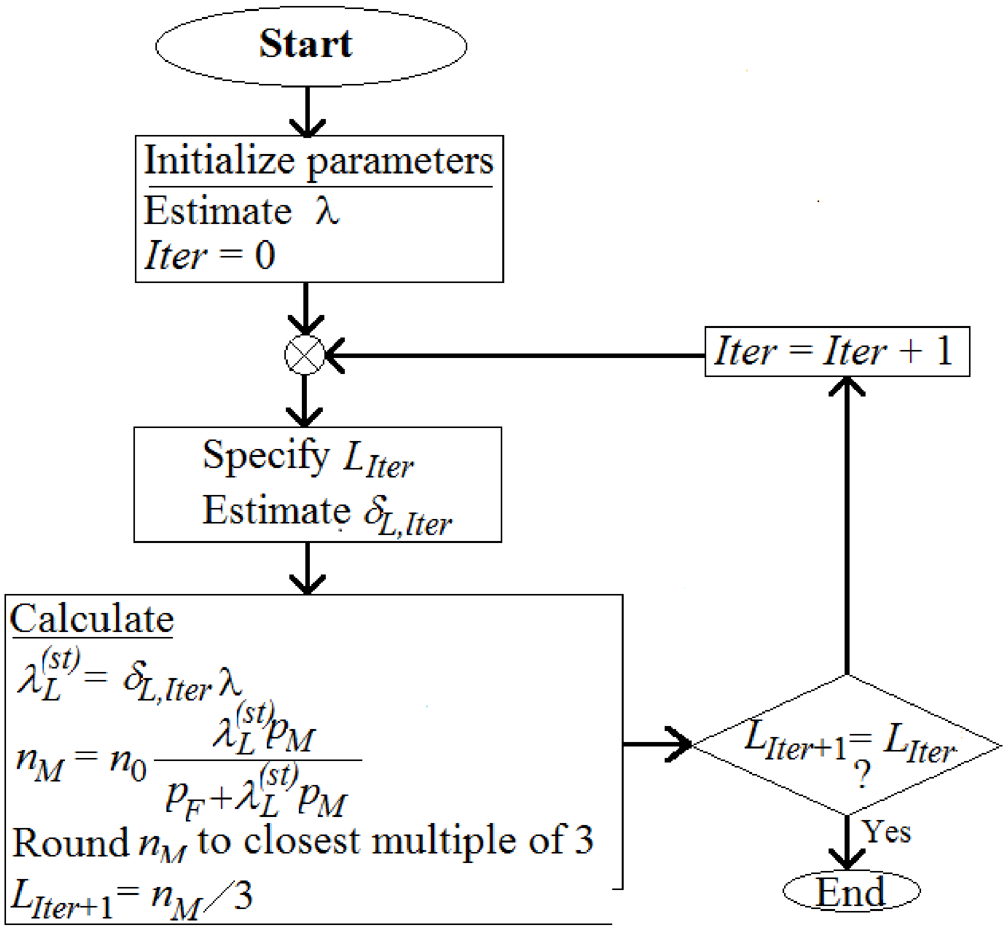

Figure 2. Flowchart to Determine the Sample Allocation When the Mound is Stratified into Layers

Since the mound has 3 primary samples per layer, the number of layers in the mound, $L_1 = n_M/3$, can be inferred from n_M. The value of n_M is confirmed to be the choice that minimizes the sampling uncertainty if $L_1 = L_0$.. If $L_1 > L_0$, then increase the number of layers in the mound from L_0, or if $L_1 < L_0$, then decrease the number of layers in the mound from L_0, and then estimate δ_L for the updated number of layers and repeat the process until the number of layers has been confirmed.

Consider this example. The ratio of the sampling standard deviation of the mound material to the sampling standard deviation for the floor material is estimated to be $\lambda = 8$. The number of mound layers initially is chosen to be $L_0 = 3$. Based on 3 layers, the ratio of the sampling standard deviation within a single layer to the sampling standard deviation for the entire mound estimated to be $\delta_L = 0.6$. Then the ratio of the sampling standard deviation within a single layer of the mound material to the sampling standard deviation of the floor material is $\lambda_L^{(st)} = \delta_L \lambda = 0.6 \cdot 8 = 4.8$.

Calculate the number of primary samples to be obtained from the entire mound as

$$n_M = n_0 \frac{\lambda_L^{(st)} p_M}{p_F + \lambda_L^{(st)} p_M} = 12 \frac{4.8 \cdot 0.22}{0.78 + 4.8 \cdot 0.22} = 6.9 \rightarrow \text{Rounds to 6 and } L_1 = \frac{6}{3} = 2.$$

Since $L_1 = 2 < L_1 = 3,$ reduce the number of layers in the mound to 2. Since there are now fewer and wider layers, δ_L is reestimated to be 0.70, since the wider layers are thought to be more heterogeneous than in the 3-layer scheme. Then the ratio of the sampling standard deviation within a single layer of the mound material to the sampling standard deviation of the floor material is $\lambda_L^{(st)} = \delta_L \lambda = 0.7 \cdot 8 = 5.6.$

Recalculate the number of primary samples to be obtained from the entire mound as

$$n_M = n_0 \frac{\lambda_{(st)} p_M}{p_F + \lambda_{(st)} p_M} = 12 \frac{5.6 \cdot 0.22}{0.78 + 5.6 \cdot 0.22} = 7.3 \rightarrow \text{Rounds to 6 and } L_2 = \frac{6}{3} = 2.$$

Now $L_2 = L_1,$ so it is confirmed that there should be $n_M = 6$ primary samples taken from the mound, arranged in 2 layers of 3 primary samples each in order to minimize the sampling uncertainty. The number of primary samples to be obtained from the cooling coils is 3, and the number of primary samples to be obtained from the floor material is

$$n_F = n_0 \frac{p_F}{p_F + \lambda_{(st)} p_M} = 12 \frac{0.78}{0.78 + 5.6 \cdot 0.22} = 4.65 \rightarrow \text{Rounds to 6.}$$

5.0 Results and Conclusions

Sample allocation for residual sampling in Liquid Waste tanks is currently described in the LWTRSAPP (Pavletich, 2013). This scheme is based on the relative volume of residual material in each stratum. If the basis for this procedure is extended to include stratum variability as well, then a more realistic allocation can be made when a stratum, such as the Tank 12 mound, is suspected as having greater spatial heterogeneity than the residual material in the floor stratum. The sample allocation for random sampling depends on an engineering estimate of design parameter $\lambda,$ and the sample allocation for stratified random sampling depends on engineering estimates of design parameters λ and δ_L, that are described below. Therefore, this report provides a method to assess the sampling options based on engineering estimates of the design parameters.

In the case of random sampling within every stratum, a design parameter lambda, $\lambda,$ was defined as the ratio of the mound stratum to the floor stratum sampling standard deviations. An engineering estimate of the approximate value of lambda is required to use this sample allocation procedure which is summarized in Unlike the case when the mound is randomly sampled, the case of stratified sampling in the mound may require iteration. An initial number of layers, $L_0,$ for the mound must be chosen. Engineering estimates of two design parameters, λ and δ_L, are required, and then $\lambda_L^{(st)} = \delta_L \lambda$ is calculated and used to enter Table 4 (instead of λ itself that was used when the mound was randomly sampled (no layers)).

Table 4. For example, if λ, the ratio of the standard deviation of the mound material to the floor material, is estimated to be 10, then it is recommended to obtain 9 mound samples, 3 floor samples, and 3 cooling coil samples in order to minimize the sampling uncertainty for a mean concentration of an analyte.

Unlike the case when the mound is randomly sampled, the case of stratified sampling in the mound may require iteration. An initial number of layers, L_0, for the mound must be chosen. Engineering estimates of two design parameters, λ and δ_L, are required, and then $\lambda_L^{(st)} = \delta_L \lambda$ is calculated and used to enter Table 4 (instead of λ itself that was used when the mound was randomly sampled (no layers)).

Table 4. Number of Primary Samples from Each Stratum

Random Mound Sampling, λ, Stratified Mound Sampling, $\lambda_L^{(st)}$	**Sampling Option**	**Number of Mound Samples**	**Number of Floor Samples**	**Number of Cooling Coil Samples**
Below 2.1	1	3	9	3
Between 2.1 and 5.9	2	6	6	3
Over 5.9	3	9	3	3

The result obtained from Table 4 for the number of primary samples in the mound can be used to compute an updated number of layers for the mound, $L_1 = n_M / 3$. If $L_1 = L_0$, then the best allocation has been found to minimize the sampling variance. If L_1 does not equal L_0, then the number of layers chosen for the mound needs to be adjusted, and then process is repeated until the number of layers does not change. A flowchart of this process and an example are presented in Section 4.2.

6.0 Bibliography

Baker, R. A., Edwards, T. B., Harris, S. P., Shine, E. P., & Watson, H. L. (2014). *Verification & Validation for Commercial Statistical Packages Utilized by Applied Computational Engineering and Statistics, Revision 2.* SRNL Report B-VVR-A-00002, Revision 2, Computational Sciences Section. Aiken: Savannah River Nuclear Solutions (SRNS), LLC.

Cochran, W. G. (1976). *Sampling Techniques* (Second ed.). New York, NY: John Wiley & Sons, Inc.

Davis, G. R. (2014). *Tank 12 Cooling Coil Scale Material Sampling Evaluation.* SRR Interoffice Memorandum SRR-CWDA-2014-00046. Aiken, SC: Savannah River Remediation (SRR), LLC.

Graybill, F. A. (1976). *Theory and Application of the Linear Model.* North Scituate, MA, USA: Duxbury Press.

Pavletich, J. P. (2013). *Liquid Waste Tank Residuals Sampling and Analysis Program Plan.* SRR Interoffice Memorandum SRR-CWDA-2011-00050, Rev. 2. Aiken, SC: Savannah River Remediation (SRR), LLC.

SAS Institute, Inc. (2014). JMP Statistical Discovery Software Version 11.1.1. Cary, NC, USA.

Savannah River Remediation. (2014). *SRNL Evaluation of Tank 12 Sampling Options.* SRR Liquid Waste Task Technical Request G-TTR-H-00007. Aiken: Savannah River Remediation (SRR), LLC.

Shine, E. P. (2014). *Task Technical and Quality Asurance Plan for Statisical Evaluation of Tank 12 Sampling Options.* Savannah River National Laboratory Report SRNL-RP-2014-00426, Computational Sciences Section. Aiken SC: Savannah River Nuclear Solutions (SRNS), LLC.

Sukhatme, P. V., & Sukhatme, B. V. (1970). *Sampling Theory of Surveys with Applications* (Second, Revised Edition ed.). Ames, IA, USA: Iowa State University Press.

Thompson, S. K., & Seber, G. A. (1996). *Adaptive Sampling.* New York: John Wiley & Sons, Inc.

Appendix A

Allocating Primary Samples in Stratified Sampling when The Sampling is Random within Every Stratum

In stratified random sampling, the population is partitioned into non-overlapping strata, and random sampling is performed within each stratum. A primary sample defines a portion of the residual material collected directly from the tank. Primary samples can be individually analyzed for physical and chemical properties, or material from groups of primary samples can be consolidated with material from certain other primary samples to form a composite sample that will be analyzed for physical and chemical properties.

Suppose there are *k* strata. The relative mass in stratum *i* is defined as

$$p_i = \frac{m_i}{m_0}, \tag{A1}$$

where m_i is the mass of residual material in stratum *i*, *i* = 1, 2, ..., *k*, and $m_0 = \sum_{j=1}^{k} m_j$ is the total mass of the residual material in all strata. The general formula for the variance of a stratified mean with *k* strata is

$$V\{\bar{y}_{strat}\} = \sum_{i=1}^{k} p_i^2 V\{\bar{y}_i\}. \tag{A2}$$

Letting f_i be the fraction of the mass[4] that is collected in the samples from stratum *i*, the variance of the sample mean $\bar{y}_i$ for stratum *i* is

$$V\{\bar{y}_i\} = \left(1 - f_i\right)\frac{S_i^2}{n_i}, i = 1, 2, ..., k. \tag{A3}$$

The finite population correction factor (*fpc*) is $fpc = 1 - f_i$. When the *fpc* is near one in all strata, say greater than 0.9 (90%), then it can usually be ignored in the derivation. When random sampling is performed within every stratum, the variance of the stratified mean is

$$V\{\bar{y}_{strat}\} = \sum_{j=1}^{k} p_j^2 \left(1 - f_j\right)\frac{S_j^2}{n_j}, \tag{A4}$$

[4] $f_i =$ total mass collected in all samples from stratum $i / m_i, i = 1, 2, \ldots, k$.

Appendix A - continued

based on Eqns (A2) and (A3). Suppose that sampling has a fixed cost c_0 plus a variable cost per sampled unit, where the cost per sampled unit in stratum i is c_i, i = 1, 2, ..., k. The total cost[5] of sampling is

$$c_T = c_0 + \sum_{j=1}^{k} c_j n_j \ . \tag{A5}$$

The total number of primary samples and the allocation of the sampling units to the strata can be derived by fixing the budget c_T to obtain the minimum variance of the stratified mean $V\{\bar{y}_{st}\}$ or fixing the required level of precision (by fixing $V\{\bar{y}_{st}\}$) and minimizing the budget c_T. This derivation fixes the budget, but either approach leads to the same allocation of units. Using the method of Lagrange multipliers, the function to be minimized is

$$obj = \sum_{j=1}^{k} p_j^2 \left(1 - f_j\right) \frac{S_j^2}{n_j} + \eta \left[\sum_{j=1}^{k} c_j n_j + \left(c_0 - c_T\right) \right], \tag{A6}$$

where η is the Lagrange multiplier. The set of k+1 partial derivatives is

$$\begin{aligned} &\frac{\partial obj}{\partial n_i} = -\frac{p_i^2 S_i^2}{n_i^2}\left(1 - f_i\right) + \eta c_i = 0, i = 1, 2, \ldots, k \text{ and} \\ &\frac{\partial obj}{\partial \eta} = \sum_{j=1}^{k} c_j n_j + \left(c_0 - c_T\right) = 0. \end{aligned} \tag{A7}$$

The first k equations define,

$$n_i = p_i S_i \sqrt{\frac{1 - f_i}{\eta c_i}}, i = 1, 2, \ldots, k. \tag{A8}$$

Substituting Eqn (8) into the last relationship in Eqn (A7) yields

$$\sqrt{\eta} = \frac{\sum_{j=1}^{k} \sqrt{c_j \left(1 - f_j\right)} p_j S_j}{\left(c_T - c_0\right)}. \tag{A9}$$

[5] Not every sampling problem has a cost function of this form. For example, if there is a cost associated with setting up the sampling apparatus over a tank riser and groups of samples are selected at each riser location, then there would be a cost per riser in addition to a fixed cost and a cost per sample.

Appendix A - continued

Substituting Eqn (A9) into each of the first *k* relationships in Eqn (A7) produces the optimum number of primary samples to be obtained from stratum *i* is

$$n_i = \frac{\left(c_T - c_0\right) p_i S_i \sqrt{\dfrac{1-f_i}{c_i}}}{\sum_{j=1}^{k} \sqrt{c_j \left(1-f_j\right)} p_j S_j}, i = 1, 2, \ldots, k \ . \tag{A10}$$

The minimum total sample size n_0 required for random sampling is

$$n_0 = \sum_{i=1}^{k} n_i = \frac{\left(c_T - c_0\right) \sum_{i=1}^{k} p_i S_i \sqrt{\dfrac{1-f_i}{c_i}}}{\sum_{j=1}^{k} p_j S_j \sqrt{c_j \left(1-f_j\right)}}, i = 1, 2, \ldots, k. \tag{A11}$$

If the variable cost of sampling is the same in every stratum, then $c_i = c, i = 1, 2, \ldots, k,$ Eqn (A11) does not offer a solution for the total number primary samples. Instead, the fixed total cost of sampling in Eqn (A5) immediately yields

$$c_T = c_0 + \sum_{j=1}^{k} c_j n_j = c_0 + c \sum_{j=1}^{k} n_j = c_0 + c n_0 \to n_0 = \frac{c_T - c_0}{c} \ . \tag{A12}$$

as the total number of primary samples based on a fixed budget c_T, and the number of primary samples to allocate to stratum *i* is

$$n_i = \frac{\left(c_T - c_0\right) p_i S_i \sqrt{1-f_i}}{c \sum_{j=1}^{k} p_j S_j \sqrt{\left(1-f_j\right)}} = n_0 \frac{p_i S_i \sqrt{1-f_i}}{\sum_{j=1}^{k} p_j S_j \sqrt{\left(1-f_j\right)}}, i = 1, 2, \ldots, k \ . \tag{A13}$$

When the $fpc = 1 - f_i$ is sufficiently near one to be ignored in all strata and the variable cost of sampling is the same in all strata, the number of primary samples in stratum *i* is

$$n_i = n_0 \frac{p_i S_i}{\sum_{j=1}^{k} p_j S_j}, i = 1, 2, \ldots, k \ . \tag{A14}$$

Eqn (A14) is the basis of results in Section 4.1.

Appendix A - continued

If the cost of sampling and the variance of sampling within all strata is the same, then $c_i = c$ and $S_i = S, i = 1, 2, \ldots, k,$ then the allocation of primary samples to stratum *i* is

$$n_i = n_0 \frac{p_i S \sqrt{1 - f_i}}{\sum_{j=1}^{k} p_j S \sqrt{(1 - f_j)}} = n_0 \frac{p_i}{\sum_{j=1}^{k} p_j} \sqrt{1 - f_i}, i = 1, 2, \ldots, k. \quad \text{(A15)}$$

When the $fpc = 1 - f_i$ is sufficiently near one to be ignored in addition to the cost of sampling and the variance of sampling being the same in all strata, then Eqn (A16) is the policy of allocating units proportional to stratum size (mass) that is in the current LWTRSAPP,

$$n_i = n_0 \frac{p_i S}{\sum_{j=1}^{k} p_j S} = n_0 \frac{p_i}{\sum_{j=1}^{k} p_j}, i = 1, 2, \ldots, k. \quad \text{(A16)}$$

Appendix B

Allocating Primary Samples in Stratified Sampling when the Mound contains Substrata

Treating the layers in the mound as additional strata, the variance of the mean concentration of an analyte in the tank residual material is as follows.

The total variance is now

$$V\{\bar{y}_w\} = \sum_{\substack{i=1 \\ i\neq M}}^{k} p_i^2 (1-f_i)\frac{S_i^2}{n_i} + \sum_{\ell=1}^{L}\left(\frac{p_M}{L}\right)^2 \left(1-f_{M_\ell}\right)\frac{S_{M_L}^2}{3} \approx \sum_{\substack{i=1 \\ i\neq M}}^{k} p_i^2 (1-f_i)\frac{S_i^2}{n_i} + p_M^2 (1-f_M)\frac{S_{M_L}^2}{3L}, \quad \text{(B1)}$$

where the first sum is over all strata except the mound (M), and the second sum is over the L distinct horizontal layers in the mound. The sampling standard deviation for sampling within a layer of the mound is S_{M_L}. The term f_{M_ℓ} is the fraction of the mass in all of the primary samples obtained from Layer ℓ of the mound to the population mass of Layer ℓ in the mound, ℓ = 1, 2, ..., L, and f_M was previously defined to be the fraction of the mass contained in all samples from the mound to the population mass of the mound.

$$n_i = n_0 \frac{p_i S_i}{\sum_{\substack{j=1 \\ j\neq M}}^{k} p_j S_j + p_M S_{M_L}}, i=1,2,\ldots,k, i\neq M \text{ (the Mound) and}$$
$$n_M = 3L, \text{ 3 primary samples per layer} \quad \text{(B2)}$$

with just the floor (F) and the mound material, Eqn (B9) reduces to the following.

$$n_F = n_0 \frac{p_F S_F}{p_F S_F + p_M S_{M_L}} = n_0 \frac{p_F S_F}{p_F + p_M \lambda_L^{(st)}} \text{ and}$$
$$n_M = 3L = n_0 \frac{p_M \lambda_L^{(st)}}{p_F + p_M \lambda_L^{(st)}} \Rightarrow L = \frac{n_0}{3}\frac{p_M \lambda_L^{(st)}}{p_F + p_M \lambda_L^{(st)}}. \quad \text{(B3)}$$

Eqn (B3) is used iteratively to find a solution. A flowchart detailing these steps is presented in Figure 2 in Section 4.2.