PNL-9408

Statistical Evaluations of Current Sampling Procedures and Incomplete Core Recovery

P. G. Heasler
L. Jensen[a]

March 1994

Prepared for
the U.S. Department of Energy
under Contract DE-AC06-76RLO 1830

Pacific Northwest Laboratory
Richland, Washington 99352

(a) Westinghouse Hanford Company
Richland, Washington.

Summary and Conclusions

This document develops two formulas that describe the effect of incomplete recovery on core sampling results for the Hanford waste tanks. The formulas evaluate incomplete core recovery from a worst-case (i.e., biased) and best- case (i.e., unbiased) perspective. A core sampler is unbiased if the sample material recovered is a random sample of the material in the tank, while any sampler that preferentially recovers a particular type of waste over others is a biased sampler. There is strong evidence to indicate that the push-mode sampler presently used at the Hanford site is a biased one.

The formulas presented here show the effects of incomplete core recovery on the accuracy of composition measurements, as functions of the vertical variability in the waste. These equations are evaluated using vertical variability estimates from previously sampled tanks (B110, U110, C109). Assuming that the values of vertical variability used in this study adequately describe the Hanford tank farm, one can use the formulas to compute the effect of incomplete recovery on the accuracy of an average constituent estimate. To determine acceptable recovery limits, we have assumed that the relative error of such an estimate should be no more than 20%.

Unbiased Sampler

If the sampler is unbiased, Equation 11 is used to determine the relative error associated with incomplete recovery. If we desire the relative error in constituent estimates from a tank to be no more than 20%, and if the vertical variability is low to moderate, then segment samples with recovery rate (p) as low as p=15% may be acceptable. On the other hand, when the vertical variability is large, the segment recovery should be at least p=80%.

This means that, for an unbiased sampler, the current recovery rule of at least p=85% per segment is reasonable for all levels of vertical variability. Furthermore, if relative errors larger than 20% were acceptable for constituent estimates, then the segment recovery rate threshold could be decreased below 85%. For example, if a relative error of 30% in the concentration estimates is acceptable, then a recovery rate for an unbiased sampler as low as p=50% could be tolerated.

Biased Sampler

The behavior of a biased sampler is described by Equation 19 with respect to incomplete segment recovery. For a fixed recovery rate p, the biased sampler produces a much higher relative error in the concentration estimates, as compared to an unbiased sampler.

If we desire the relative bias in constituent estimates to be below 20%, then a recovery of at least p= 40% is required even when the vertical variability is low. If the vertical variability is moderate then a recovery of at least p=85% is required. Finally, when the vertical variability is high, the segment recovery must be near p=100%.

Since the average historical recovery rate is p=70%, one can see that sampler bias is an important issue for the Hanford tank characterization program. This recovery rate may produce a substantial bias in analyte concentration estimates when the vertical variability is moderate or high.

It is also important to note that the effects of sampler bias are not properly reflected in the measures of uncertainty (variance, confidence bounds) that are calculated from the sampling data. Without a properly designed calibration experiment, one will never know how poorly a biased sampler is doing. This is in direct contrast to an unbiased sampler. Even when an unbiased sampler has recovery problems, the variance statistics calculated from the data will give an accurate description of the sampler's uncertainty.

The current sampling data gives only indirect information as to the sampler's present bias, and it is difficult to express this information in quantitative form. To quantify the actual bias of the samplers, a drilling experiment using a standard material has to be designed; or the results from a core sampler must be compared to a more reliable waste sampling method. If such experiments or comparisons are properly designed, the magnitude of the bias can be estimated.

The results of this analysis illustrate an obvious fact: that a biased sampler requires more stringent limits on segment recovery p than an unbiased sampler.

Contents

List of Figures

List of Tables

1 Introduction

Core sampling of Hanford's waste tanks was halted in May 1993 because the core sampler was not recovering waste samples adequately. Several core samples had substantially less than 50% recovery. The sample recovery goal was at least 85% of each segment (a core sample consists of multiple disjoint segments). An investigation into the core sampling procedures indicated additional problems; e.g., lack of quality control and sampler bias. A program to improve the sampling equipment and procedures was initiated. The Tank Characterization Advisory Panel, an external advisory body, was assembled to provide suggestions to the sample recovery group.

1.1 Sampling Recovery, Bias, and Accuracy

This report presents formulas that relate core sampling bias and sample accuracy to percent recovery. These formulas were developed to evaluate the effect of incomplete sample recovery on the final estimates of waste concentration. The validity of the formulas in this document relies upon basic assumptions concerning waste heterogeneity and sampler bias. For example, if the waste is homogeneous, any sample is satisfactory and incomplete recovery is not a problem. If it is assumed that the core sampler is unbiased, a recovery rate as low as 50% may be sufficient. Alternatively, if the waste is heterogeneous or if the core sampler is biased, then nearly 100% of each core sample is required.

The appropriate method to determine which set of assumptions is correct is to design a set of experiments that will quantify sample bias as a function of the spatial variability in the tank waste. These experiments are beyond the scope of this report. They should be part of the core sampler certification program.

An additional question this investigation addresses is "Is the goal of at least an 85% recovery rate per segment appropriate?" This goal was established in 1993. The original recovery rule of 80% was established in 1986 ([1], Appendix S). The original rule was based upon the maximum amount of each segment that could be obtained on a routine basis, and upon some statistical requirements.

1.2 Biased and Unbiased Core Samplers

A core sampler is an unbiased sampler if the sample material recovered is a random sample of the material in the segment. In this case, the mean of the chemical analysis data from the (complete or incomplete) segment is an unbiased estimate of the mean composition of the waste for the entire segment. If the core sampler does not recover a random sample of waste, it is defined as a biased sampler.

If the core sampler is unbiased, the effects of incomplete recovery translate into an increase in the variance of the mean concentration of an analyte in the waste; i.e., an increase in the width of the confidence interval on mean concentration. That is, a valid level of confidence can be assigned to the composition of the tank waste estimated from incomplete core sample data obtained by an unbiased sampler. If the sampler is unbiased, classical statistical methods can be used to address the uncertainty in the estimates of waste composition.

This is not the case with a biased sampler. In this case, estimates of the composition of the tank waste, based upon biased incomplete core samples, cannot be assigned a valid level of uncertainty. For example, if the sampler selectively recovers a certain type of waste, then the variability in the mean concentration of an analyte will reflect only the variability in that type of waste and not the tank in general. If the sampler is biased, it is not appropriate to use standard variance and confidence interval formulas to describe the level of uncertainty associated with estimates of the tank composition. Data from a biased sampler will typically provide an optimistic view of sampling uncertainty.

To estimate the magnitude of the bias, an experiment has to be performed that uses a biased sampler in a waste of known composition, or compares the biased sampling method against an unbiased method. The evaluation of the sampling bias is a difficult problem.

2 Segment Recovery from Tank Sampling Campaigns

The following is the definition of the core or segment sample recovery rate:

> Percent recovery is calculated as the ratio of the volume (or length) of waste actually recovered to the volume (or length) of waste expected.

Tables 1 and 2 give recovery rates for the two tanks most extensively sampled while Table 4 presents recovery rates for a recently sampled tank (1992). Core segment recovery rates are also available from 18 tanks that were sampled in 1986 and 1987. Table 3 presents a summary of the recoveries achieved from the cores and segments taken during the 1986-87 sampling campaign. During this campaign, segment samples were obtained from liquid or wet sludge using a push mode. If crust or hard sludge was encountered, samples were obtained using the rotary mode. Since rotary mode sampling could be used during this campaign, the reported recovery rates may be better than those achievable today.

The segment recovery data in Tables 3 through 4 are not complete. Other waste tanks have been sampled in the past years, but the recovery rates listed in these five tables are indicative of the recovery problem.

Figure 2 provides a summary of the segment recovery data given in Tables 3 through 4. This figure displays the proportion of segments that exhibited a percent recovery less than or equal to a certain value. From this plot, one can see that about 50% of the segments had a recovery of 100% and 20% of the segments had a recovery of 0%, with the recovery rate for the remaining segments uniformly distributed between 0 and 100%. The overall average recovery is 68%.

Table 1: Recoveries from Tank B-110 Sampling in 1989

		% Recovery				
Core	Riser	Seg 1	Seg 2	Seg 3	Seg 4	Seg 5
1	7	20%	100%	100%	100%	100%
2	7	91%	100%	100%	100%	100%
3	5	0%	100%	100%	100%	100%
4	1	45%	100%	95%	100%	100%
9	3	0%	97%	100%	100%	100%
10	3	42%	100%	100%	100%	100%
11	3	78%	84%	100%	100%	100%

(Information abstracted from [2])

Table 2: Recoveries from Tank U-110 Sampling in 1989

		% Recovery			
Core	Riser	Seg 1	Seg 2	Seg 3	Seg 4
5	19	0	0	75	85
6	17	0	27	70	35
7	7	50	80	30	40
12	2	0	21	65	60
13	2	15	37	80	40
14	9	83	80	100	85
15	8	25	85	70	15

(Information abstracted from [3])

Table 3: Recovery Percentages from 86-87 Tank Sampling

Tank/ Core	Segments 1	2	3	4	5	6	7	8	Comments
C-102/1	100								Problems with sampler
2	0	100	68	63	79	0	16	65	
C-103									100% Overall
C-104									100% Overall
C-105									100% Overall
C-106/1	100	100	100	100					6 in. crust layer
2	100	100	100						
A-102									100% Overall
A-103									100% Overall
A-104									0%, Overall
A-106									100% Overall
TY-101/1	0	72	78						51% Overall
TY-102/1	82	67							78% Overall
2									0% Hard waste
TY-103/1	62	52	100	94					82% Overall
2									100% Overall
3	100	39	66	27					45% Overall
4									32% Overall
5	58	68	16						51% Overall
6	34	18	63						39% Overall
TY-104/1									0% Overall
2									0% Overall
3									52% Overall
4									78% Overall
5									51% Overall
6									100% Overall
TY-105/1	100	100	0	0	0				Hard layer after 1 in.
TY-106/1									0% Overall
2									50% Overall
3									50% Overall
4									0% Overall
5									0% Overall
6									0% Overall
7									47% Overall
BX-104/1	100	82	100						91% Overall
BX-105/1									54% Overall
2									100% Overall
SX-109/1	0	–	60	100					50%, Waste is dry salt cake

Table 4: Recoveries from Tank C-112 Sampling in 1992

Core	Segment	Recovery	Comments
34 Upper	92-001	87%	0.5 in. solids
34 Lower	92-002	74%	14 in. solids
35 Upper	92-003	0%	Valve open;no sample
35 Lower	92-004	34%	3 in. solids
36 Upper	92-005	64%	8.5 in. solids
36 Lower	92-	90%	17 in. solids

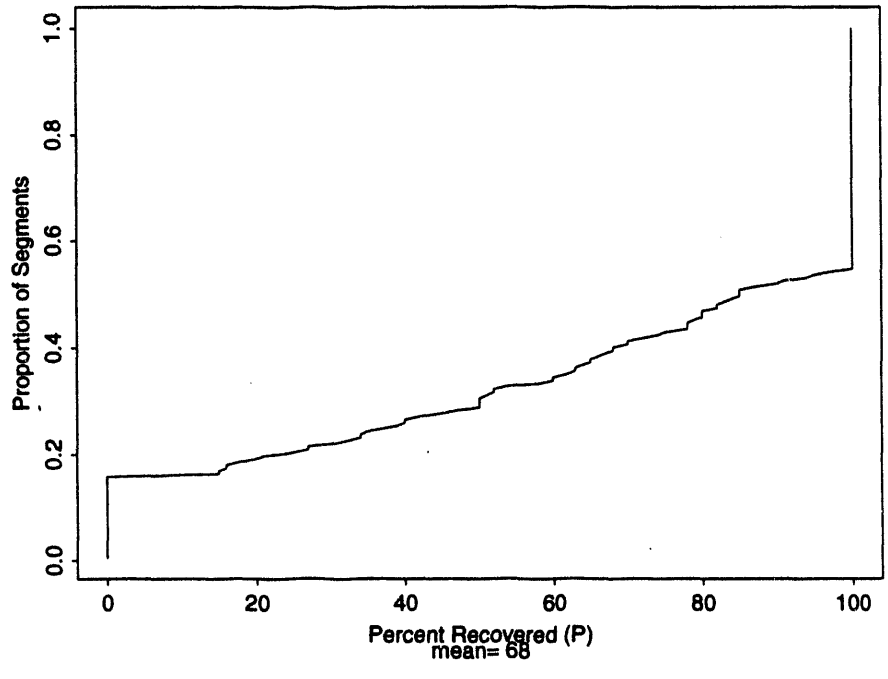


Figure 1: Proportion of Segments Exhibiting a Recovery $\leq P\%$

3 Models and Formulas to Evaluate Incomplete Recovery

Let $C(z)$ represent a quantitative property of the waste in the tank at depth z within a particular riser. This property may be a physical property (e.g., density) or a chemical or radiological property, (e.g., concentration of Pu). Currently, about 50 to 100 properties are measured in a tank by core sampling. Throughout this discussion, we will focus on the effects of incomplete recovery on the measurement of one property. The reader must realize that a complete evaluation of the effects of incomplete recovery must involve all important properties.

If recovery is complete, the core will produce a complete profile of the property between the 0 level (at the top of the waste layer) and level T (at the bottom of the tank). Incomplete recovery deletes some portions from the interval $[0, T]$, so that $C(z)$ will be available only for a subset H of the interval. When recovery is complete, the average concentration obtained for the core after compositing is given by

$$A_C = \frac{1}{T}\int_0^T C(z)dz \tag{1}$$

(The subscript "C" stands for complete). The quantity A_C represents the true core composite average, which provides a very good estimate of tank contents, particularly when the waste is layered. When recovery is partial, the average concentration obtained for the core is

$$A_P = \frac{1}{\ell(H)}\int_{z\in H} C(z)dz \tag{2}$$

(The subscript "P" stands for partial.) The term $\ell(H)$ is the length of the set H. Given the current sampling objectives, the principal question we are interested in is this: How close is A_P to the "true" average, A_C?

The magnitude of the difference between A_C and A_P can be measured by several related quantities. When the sampler produces unbiased results, A_C can best be compared to A_P by utilizing the mean squared error (MSE), or a related unitless quantity, the relative root mean square error (RRE). When A_P is biased, the relative bias (RB) is a natural statistic to evaluate. Therefore, the objective will be to calculate the following three quantities for several partial recovery scenarios. The MSE:

$$MSE = E(A_P - A_C)^2 \tag{3}$$

the relative root MSE:

$$RRE = \frac{\sqrt{E(A_P - A_C)^2}}{E(A_C)} \tag{4}$$

and the relative bias:

$$RB = \frac{E[A_C - A_P]}{E(A_C)} \tag{5}$$

(The symbol $E(\cdot)$ represents the expectation operator).

A knowledge of the vertical heterogeneity exhibited by the property of interest is necessary to determine these statistics. We will assume that $C(z)$ is a stationary stochastic process so that a complete description of the vertical heterogeneity in the tank is given by a function called the auto-covariance function. The auto-covariance function is defined by

$$\gamma(s) = Cov(C(z), C(z+s)) \tag{6}$$

This function simply describes how highly correlated two points in a core, separated by a distance s, are to each other. If this function could be estimated directly from present sampling data, it would be a simple matter to evaluate the effects of incomplete recovery.

Equations 3, 4, and 5 are not the only way incomplete recovery might be evaluated. One could also ask how incomplete recovery affects one's ability to estimate the complete profile, $C(z)$, when some portions of the profile are missing. This is a classic interpolation problem, and several solutions are available [4]. This formulation of the incomplete recovery problem might be relevant to scanning measurement technologies, which may be utilized in the future.

3.1 Unbiased Sampling

In this section, we will make some assumptions concerning vertical variability, so that a conservative formula for the unbiased case can be developed. To simplify the problem, assume that the material recovered comes from the contiguous interval $[0, pT]$, where p represents the proportion recovered. This assumption is justified because it produces a worst-case incomplete recovery scenario for the mathematical models utilized. Given this simplification, the MSE and RRE become functions of p, and we will write $MSE(p)$ and $RRE(p)$ to make this relationship explicit.

Given a particular spatial variability defined by $\gamma(s)$, it can be shown that the incomplete recovery mean square error is

$$\begin{aligned} MSE(p) &= \left(\frac{1-p}{pT}\right)^2 \int_{-pT}^{pT} (pT - |s|)\gamma(s)ds \\ &\quad + \frac{1}{T^2} \int_{-(1-p)T}^{(1-p)T} ((1-p)T - |s|)\gamma(s)ds \\ &\quad - \frac{1-p}{T^2} \int_0^T \nu(s)\gamma(s)ds \end{aligned} \tag{7}$$

where the function $\nu(s)$ is defined by the formula:

$$\nu(s) = Min(2T - s, 2pT + s) - Max(2pT - s, s) \tag{8}$$

This rather complex equation can be simplified considerably when a worst-case form of variability is assumed. If the profiles $C(z)$ are assumed to have a spatial variability resembling "white noise," (i.e., $\gamma(z) = \sigma_0^2\delta(z)$, where $\delta(z)$ represents a Dirac delta function), then the formula becomes

$$\begin{aligned} MSE(p) &= \frac{(1-p)^2}{pT}\sigma_0^2 + \frac{1-p}{T}\sigma_0^2 \\ &= \frac{1-p}{pT}\sigma_0^2 \end{aligned} \tag{9}$$

This formula is quite useful because it contains only one unknown parameter, σ_0^2, and this parameter can be related to a quantity that can be calculated from existing sampling data. Suppose segments of length S were recovered during a sampling campaign, the average concentrations were measured, and a segment-to-segment variability of σ_{seg}^2 was computed. Then an estimate for σ_0^2 in the above formula would be $\sigma_0^2 = \sigma_{seg}^2 S$. If this is substituted into the formula for MSE, one obtains

$$MSE(p) = \sigma_{seg}^2 \frac{(1-p)S}{pT} \tag{10}$$

or if a relative root mean square error is desired, the result is

$$RRE(p) = \rho_{seg}\sqrt{\frac{(1-p)S}{pT}} \tag{11}$$

where ρ_{seg} is the segment-to-segment relative standard deviation (σ_{seg}/μ, where μ is the mean concentration) calculated from the sampling data.

For example, a segment sample in tank B-110 is 19 inches long and a core is approximately 81 inches long. The segment-to-segment relative standard deviation (RSD) for Sr-90 is $\rho_{seg} = 12\%$. This means that, for strontium, the relative root mean squared error is

$$RRE(p) = 0.12\sqrt{\frac{19(1-p)}{81p}} \tag{12}$$

In other words, if only 50% of a core were recovered ($p = 50\%$), the RRE would be approximately 6%. Consequently, a 95% confidence interval on the mean concentration of strontium would be approximately $\pm 12\%$ about the estimated mean. If the recovery were 85%, the RRE would be 2.4%. This is a small error. If the segment-to-segment RSD is increased from 12% to 100%, then recovery percentages of $p = 50\%$ and $p = 85\%$ produce RREs of 48% and 20%, respectively. In this case, the recovery rate goal of 85% might not be sufficient.

The preceding analysis illustrates the effect of vertical heterogeneity on incomplete recovery data. The RRE for incomplete recovery is proportional to the segment-to-segment relative standard deviation (RSD). If the RSD is small (say 10%), then one can tolerate a small recovery proportion. If the RSD is large (say 100%), then the recovery proportion should be close to 100%. These conclusions are based upon the assumption that the sampler always obtains a random sample of the waste; i.e., the sampler is unbiased.

3.2 Biased Sampling

To develop a set of formulas that describes the consequences of a biased sampler, assume the sampler does not recover waste when $C(z)$ is above a threshold q. In other words, the set of recovered material, H is defined as

$$H = \{z \,:\, C(z) < q\}. \tag{13}$$

This set H is random. The relative root MSE is complicated, and cannot be evaluated in closed form. The relative bias, however, can be evaluated. Since

$$MSE = Var + Bias^2 \tag{14}$$

the relative bias also serves as a lower-bound estimate to the relative root MSE.

The bias will be evaluated assuming that **exactly** $p\%$ of the core is recovered. That is, the amount recovered is not a random variable and $pT = \ell(H)$. The bias is equal to

$$\begin{aligned} BIAS &= E[A_C - A_P] \\ &= E[\frac{1}{T}\int_0^T C(z)dz - \frac{1}{pT}\int C(z)I_H(z)dz] \\ &= E[\frac{p-1}{pT}\int_0^T C(z)I_H(z)dz + \frac{1}{T}\int_0^T C(z)I_{H^c}(z)dz] \\ &= \frac{p-1}{pT}\int_0^T E[C(z)I_H(z)]dz + \frac{1}{T}\int_0^T E[C(z)I_{H^c}(z)]dz \end{aligned} \tag{15}$$

The bias will be most pronounced if $C(z)$ displays a heavy-tailed distribution. The log normal distribution has this property and can be justified for other reasons. Assume that $C(z)$ is log-normally distributed with parameters (μ_0, σ_0). Under these distributional assumptions, the recovery threshold is $q = \mu_0 + Z_p\sigma_0$ and the expectations in Equation 15 evaluate to:

$$\begin{aligned} E[C(z)I_H(z)] &= \frac{1}{\sqrt{2\pi}\sigma_0}\int_{-\infty}^{q} e^z \exp(-\frac{1}{2}\frac{(z-\mu_0)^2}{\sigma_0^2})dz \\ &= \exp(\mu_0 + \frac{1}{2}\sigma_0^2)\Phi(Z_p - \sigma_0) \end{aligned} \tag{16}$$

and

$$\begin{aligned} E[C(z)I_{H^c}(z)] &= \frac{1}{\sqrt{2\pi}\sigma_0}\int_{q}^{\infty} e^z \exp(-\frac{1}{2}\frac{(z-\mu_0)^2}{\sigma_0^2})dz \\ &= \exp(\mu_0 + \frac{1}{2}\sigma_0^2)\Phi(\sigma_0 - Z_p) \end{aligned} \tag{17}$$

In these formulas, the function $\Phi(\cdot)$ represents the standard normal cumulative distribution function, and Z_p the p'th quantile ($\Phi(Z_p) = p$). If these results are substituted into Equation 15, one obtains a workable formula for bias:

$$BIAS = \exp(\mu_0 + \frac{1}{2}\sigma_0^2)\left(\frac{p-1}{p}\Phi(Z_p - \sigma_0) + \Phi(\sigma_0 - Z_p)\right) \tag{18}$$

Since the mean of a log-normal distribution with parameters μ, σ^2 is $\exp(\mu + \frac{1}{2}\sigma^2)$, the relative bias (RB) is

$$RB(p) = \Phi(\sigma_0 - Z_p) - \frac{1-p}{p}\Phi(Z_p - \sigma_0) \tag{19}$$

3.2.1 Relating Formula Parameters to Segment-to-Segment RSD

The parameter σ_0^2 is a function of the variance of $C(z)$ (this variance is the autocovariance $\gamma_0 = \gamma(0)$) and the mean concentration μ. The relationship is

$$\sigma_0^2 = ln(1 + (\gamma_0/\mu)^2) = ln(1 + (\rho_0)^2) \tag{20}$$

The variance γ_0 is the point-to-point vertical variability of the waste in the tank and ρ_0 is the point-to-point RSD. At present, point-to-point variability cannot be measured. The smallest measurable variability is the segment-to-segment variability.

To determine the point-to-point variability from the segment-to-segment variability, one must make some assumptions about the shape of the autocovariance function, $\gamma(z)$. To do this, assume that the waste in the tank is a layer-cake of randomly ordered layers α inches thick. Then the autocovariance, $\gamma(z)$ is given by;

$$\gamma(z) = \gamma_0 Max(0, 1 - |z/\alpha|) \tag{21}$$

where z is measured in inches. The typical sampling data which produces average concentration in a segment of length S would have a segment-to-segment variability of

$$\begin{aligned} \sigma_{seg}^2 &= \frac{2\gamma_0}{S^2}\int_0^S (S-z)\gamma(z)dz \\ &= \begin{cases} \frac{\gamma_0\alpha}{S}\left(1 - \frac{\alpha}{3S}\right) & \text{for } \alpha < S \\ \gamma_0\left(1 - \frac{S}{3\alpha}\right) & \text{Otherwise} \end{cases} \end{aligned} \tag{22}$$

The above formula also produces a relationship between the segment-to-segment RSD, ρ_{seg} and the point-to-point RSD, ρ_0. The relationship between the RSDs is given by

$$\rho_{seg} = \begin{cases} \rho_0\sqrt{\frac{\alpha}{S}\left(1 - \frac{\alpha}{3S}\right)} & \text{for } \alpha < S \\ \rho_0\sqrt{\left(1 - \frac{S}{3\alpha}\right)} & \text{Otherwise} \end{cases} \tag{23}$$

Assume a layer thickness $\alpha = 2$ inches. A 2-inch thickness is chosen because it is believed that typical layers in the tank are from 2 to 4 inches thick (This is the thickness one processing batch might create). Two inches may be justified on another ground also: it is the diameter of the sampler. A layer of this size would have a good chance of being pushed away during sampling, and would produce the poorest (worst-case) results.

If, for example, the segments are 19 inches long, then ρ_{seg} is divided by

$$\sqrt{\frac{2}{19}\left(1 - \frac{2}{3 \cdot 19}\right)} = 0.319 \tag{24}$$

to obtain ρ_0. Applying the formula to the strontium results obtained from tank B-110 ($\rho_{seg} = 12\%$), produces a point-to-point RSD $\rho_0 = 37.6\%$. If this value is substituted into Equations 19 and 20, the following relative biases result:

- 50% recovery on log-normally distributed $C(z)$ produces a relative error of 28%
- 85% recovery on log-normally distributed $C(z)$ produces a relative error of 11%

If the point-to-point RSD is increased to 100%, much larger relative errors are produced. For log-normally distributed data, the results for $\rho_0 = 100\%$ are:

- 50% recovery produces a relative error of 59%
- 85% recovery produces a relative error of 32%

3.2.2 An Alternative: A Binary Distributional Model

An alternate model describing the consequences of a biased sampler is also possible. This model makes the assumption that the waste is made up of two types of material (binary distribution). The first type is accurately gathered into the sampler, but the second type is never recovered by the sampler. The concentration of the first type of waste is C_1, while the concentration of the second type is C_2. If p percent of the waste is recovered in the sampler, then $1 - p$ percent of the waste is of the second type, so the true concentration of the waste is

$$\mu = pC_1 + (1-p)C_2 \tag{25}$$

But the measured value is C_1. Therefore the relative bias is

$$RB(p) = \frac{(1-p)(C_2 - C_1)}{pC_1 + (1-p)C_2} \tag{26}$$

Since the relative standard deviation for such a case is given by

$$\rho = \frac{\sqrt{p(1-p)}(C_2 - C_1)}{pC_1 + (1-p)C_2} \tag{27}$$

the formula can also be expressed in terms of the RSD as follows:

$$RB(p) = \rho_0 \sqrt{\frac{1-p}{p}} \tag{28}$$

which is very similar to the unbiased-case formula (Equation 11).

4 Variability Information from Tank Sampling Campaigns

During 1989, two tanks (B-110 and U-110) were sampled for a pilot study. One of the objectives of this study was to estimate the degree of variability in the data associated with tank sampling and chemical measurement. The sources of variability under study included;

- Laboratory Measurement Variability: The variability exhibited by the laboratory analytic procedure.
- Homogenization Variability: The variability exhibited between the top and bottom sections of a homogenized core segment.
- Sampling Variability: The variability exhibited between replicate core samples taken through the same riser at nearly the same location.
- Vertical Spatial Variability: The variability exhibited between segments within a core.
- Horizontal Spatial Variability: The variability exhibited between riser locations.

The variabilities measured in this pilot study are relevant to the recovery rate problem. They provide one of the best descriptions of the capabilities of the present core sampling procedure. Another tank, C-109, was recently sampled and analyzed on a quarter segment level, which provided the highest resolution information on vertical variability to date (see Table 7). Data from these three tanks will be discussed in this section.

Sampling variability describes the repeatability of the core sampling procedure and how closely the core/segment samples resemble the waste in the tank. An ideal sampler would have zero sampling variability; i.e., core samples taken at the same location would be identical. However, it should be noted that a small sampling variability does not necessarily indicate the sampler is operating acceptably; it is possible for a biased sampler to have zero sampling variability. In this pilot study, sampling variability was quite large, apparently because the sampler displaced the waste as it took a sample.

Vertical spatial variability is a required input for the incomplete recovery formulas developed in the previous section, and is the most important value we need to extract from previous sampling data. Unfortunately, information concerning vertical variability is limited in the current data. Very limited segment-level measurements were analyzed for U-110, because of the recovery problems experienced during sampling, so hardly any vertical variability information is available from this tank. Also, the top segment was usually not available in B-110, so the vertical variability listed for this tank does not include the most variable layer. Since B-110 was a fairly homogeneous tank to begin with, the vertical variability shown for this tank in Table 6 is probably smaller than one would expect to see in a "typical" Hanford tank.

The other sources of variability can serve as a useful comparison to any uncertainties caused by incomplete recovery. A very justifiable way to set an incomplete recovery threshold is to calculate the uncertainty caused by incomplete recovery, and compare it to the other sources of measurement variability. If the uncertainty caused by incomplete recovery is a small percentage of the total, then the threshold is appropriately set.

Table 5 presents a summary of the core composite variability determined from tanks B-110 and U-110. These variabilities include all sources listed above, and serve as benchmark values to compare against the errors caused by incomplete recovery. For an analyte which has a mean concentration RSD of approximately 40%, a relative bias (or root mean square error) due to incomplete recovery of 15% would be considered a secondary problem. On the other hand, if the mean concentration RSD were about 10%, a relative bias of 15% would be the most dominant error. Given the range of mean concentration RSDs experienced in B-110 and U-110, it would seem prudent to limit the relative bias (or root mean square error) due to incomplete recovery to no more than 20%.

Table 6 presents the sample-replicate RSDs for certain selected analytes from B-110. The sample-replicate RSD measures the variations between replicate core samples (i.e., core samples taken at the same location). The sample-replicate RSD was generally the largest source of variability in the B-110

Table 5: Summary of B-110 and U-110 Mean Concentration RSDs from Core Composite Measurements

	B-110			U-110		
	Min	Mean	Max	Min	Mean	Max
Anions	8%	14%	18%	20%	40%	62%
Cations	6%	36%	104%	19%	64%	200%
Radnuc	8%	72%	171%	17%	35%	48%
Overall	7%	41%	98%	19%	46%	103%

RSDs include all sources of variations

Table 6: B-110 Segment and Sample Replicate RSDs

Constituent	RSDs		
	Segment	Sample	Total
Metals:			
Al	0%	132%	132%
Ca	42%	27%	50%
Fe	6%	5%	8%
Na	4%	2%	5%
Pb	0%	112%	112%
U	100%	31%	105%
P	7%	5%	9%
Mean	23%	45%	60%
Anions:			
Cl	0%	0%	0%
NO2	52%	0%	52%
NO3	12%	0%	12%
PO4	14%	0%	14%
SO4	5%	0%	5%
Mean	17%	0%	17%
Radnuc:			
Alpha	24%	64%	68%
Sr90	52%	0%	52%
Cs137	11%	0%	11%
Mean	29%	21%	44%

Min. RSD=0%
Mean RSD=42%
Max. RSD=132%

Table 7: C-109 Quarter-Segment RSDs

Constituent	Statistics (ppm) DOF	Mean	Std.Dev.	RSD
Metals				
Al	5	73850	52332	71%
Ca	5	19033	9258	49%
Fe	5	21789	17019	78%
Na	5	82333	23097	28%
P	5	16833	8923	53%
Pb	3	3118	4842	155%
U	4	11139	3056	27%
Anions				
CN	5	6672	1826	27%
Cl	5	781	129	16%
NO2	5	41278	7838	19%
NO3	5	42111	7052	17%
PO4	5	19978	16138	81%
SO4	5	8011	1521	19%
Radnuc				
Cs137	5	759	271	36%
Sr90	5	943	1752	186%

RSD: Min=16% Mean=57% Max=186%
DOF= Degrees of Freedom

data. This indicates that the sampler did not produce identical cores from the same location. This large core-replication error may be caused by the sampler "mixing" the waste layers in the tank as the core is taken. Therefore, the segment-to-segment RSDs listed in Table 6 may be too small. To determine a more realistic segment-to-segment variability, the segment RSD and the sample-replicate RSD were combined (square root of the sum of squares) to obtain the values listed in the "Total" column.

If the RSDs in the Total column in Table 6 are used to evaluate vertical variability, then 10% would be considered a small RSD, 42% is typical, and 130% is large. These variabilities describe the differences that occur between 19-inch segments. To apply the formulas developed in the last section, one is interested in vertical variability on a smaller scale than this.

In tank C-109, segments were divided vertically into quarter segments (4.5 inches). The variability between quarter segments is a better approximation of the point-to-point variability, referred to in Section 3.2.1, than is the between-segment variability. Table 7 presents the quarter-segment variability determined for this tank. Note that the variability is somewhat larger than that reported for B-110. These results indicate that for the type of waste in C-109, a small vertical variability would be 10%, typical would be 50%, and large would be 130%. These values for vertical variability are suggested for comparison to the error due to incomplete recovery, calculated from the formulas in Section 3.

5 Applications of the Incomplete Recovery Formulas

In this section, the formulas developed for incomplete core recovery are used to quantify a segment recovery goal, for both biased and unbiased sampling. The error in concentration estimates due to incomplete recovery is a function of the vertical variability present in the waste. In Figure 2, the concentration error is plotted along the vertical axis as relative error RRE(p), or as relative bias RB(p). Core recovery is plotted along the horizontal axis as a percent recovery p. The vertical variability is expressed as the segment-to-segment relative standard deviation RSD. Based on data from tanks B-110, U-110, and C-109, three levels of vertical variability were defined: low (10%), moderate or typical (50%), and high (130%).

The length of a core, needed to evaluate the incomplete recovery error for an unbiased sampler, is assumed to be 80 inches.

5.1 Relative Error for an Unbiased Sampler

If the sampler is unbiased, Equation 11 expresses the relative (root mean square) error associated with incomplete recovery. This equation is graphed in the top half of Figure 5.1. Let the relative error in the estimates of the concentration of an analyte be fixed at 20%. If the vertical variability is low to moderate (RSD of 10% to 50%), then segment samples with recoveries as low as p=15% may be acceptable. These values were extrapolated from the results given in Figure 5.1. If the vertical variability is large (RSD=130%), the segment recovery should be at least p=80%.

That is, an unbiased sampler operating with the current recovery rule of p=85% should produce concentration estimates with relative errors (RREs) less than 20%. In this case, the errors due to incomplete recovery are automatically incorporated into the mean concentration RSDs; after the samples are taken, the effect of incomplete recovery will be known (for an unbiased sampler).

This figure also shows that if large relative errors in the concentration estimates are acceptable, then the required amount of sample recovered per segment (p) decreases. For example, if the allowable relative error in the concentration estimates is increased to 30%, then a recovery rate for an unbiased sampler as low as p=50% is acceptable, regardless of the magnitude of the vertical variability.

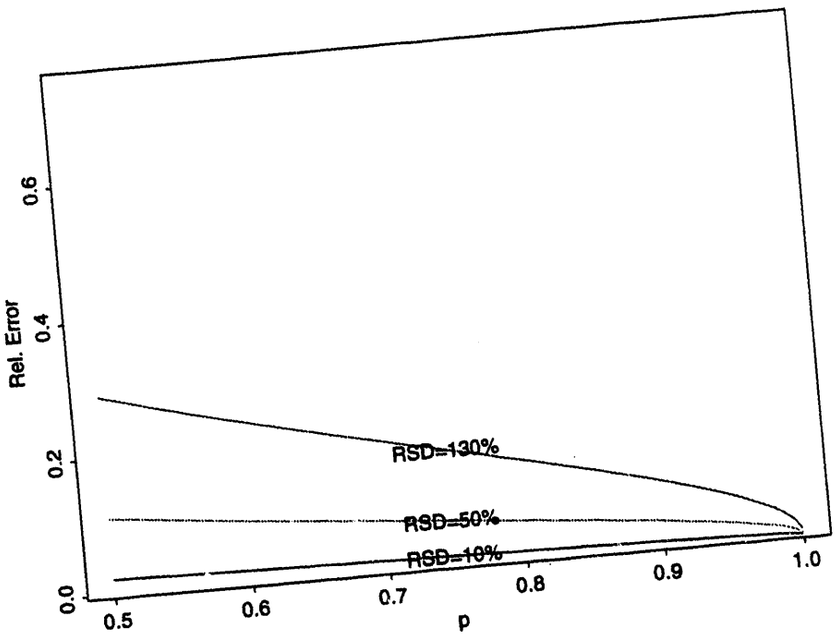
5.2 Relative Error for a Biased Sampler

If the sampler is biased, then Equation 19 expresses the relative bias due to incomplete segment recovery. This equation is graphed in the bottom half of Figure 5.1. It is obvious from this figure that for a fixed recovery rate p, the biased sampler results in a much higher relative error in the concentration estimates, as compared to an unbiased sampler. This result is not surprising.

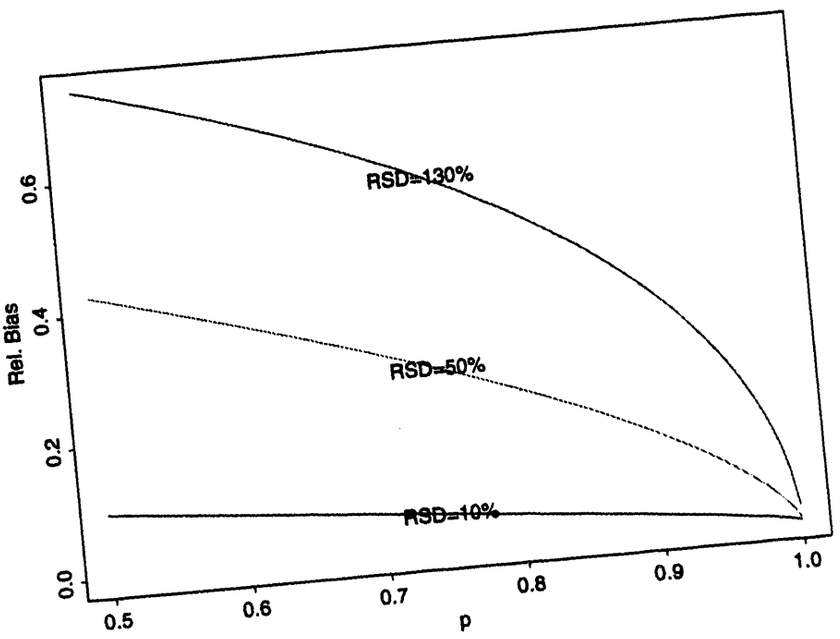
Suppose the relative bias in the estimates of the concentration of an analyte must be less than 20%. In this case, a moderate vertical variability (an RSD of 50%) will require a recovery rate of at least p=85% to meet the 20% threshold. And if the vertical variability is high, the recovery must be near p=100% to meet the threshold. Since the average historical recovery rate is p=70%, one can see that substantial bias may be present in previous concentration estimates. The historical recovery rate would not be considered acceptable for a biased sampler.

This also means that, for a biased sampler, the current recovery rule of p=85% per segment (at least) is reasonable only for low and moderate degrees of vertical variability. In contrast to an unbiased sampler, errors in the concentration estimates due to incomplete core recovery are not present in the core-to-core variability calculated from sampling data. If the sampler is biased, the variability estimates we compute from the data are optimistic.

This figure also shows that if a large relative bias in the concentration estimates is acceptable, then the required amount of sample recovered per segment (p) decreases. For example, if the allowable relative bias in the concentration estimates is increased to 40%, then a recovery rate for a biased sampler as low as p=50% is acceptable, provided the vertical variability is low to moderate. As before, if the vertical variability is large, then a recovery rate near p=100% is required.



a) Unbiased Sampler



b) Biased Sampler

Figure 2: Relative Error as a function of Proportion Recovered

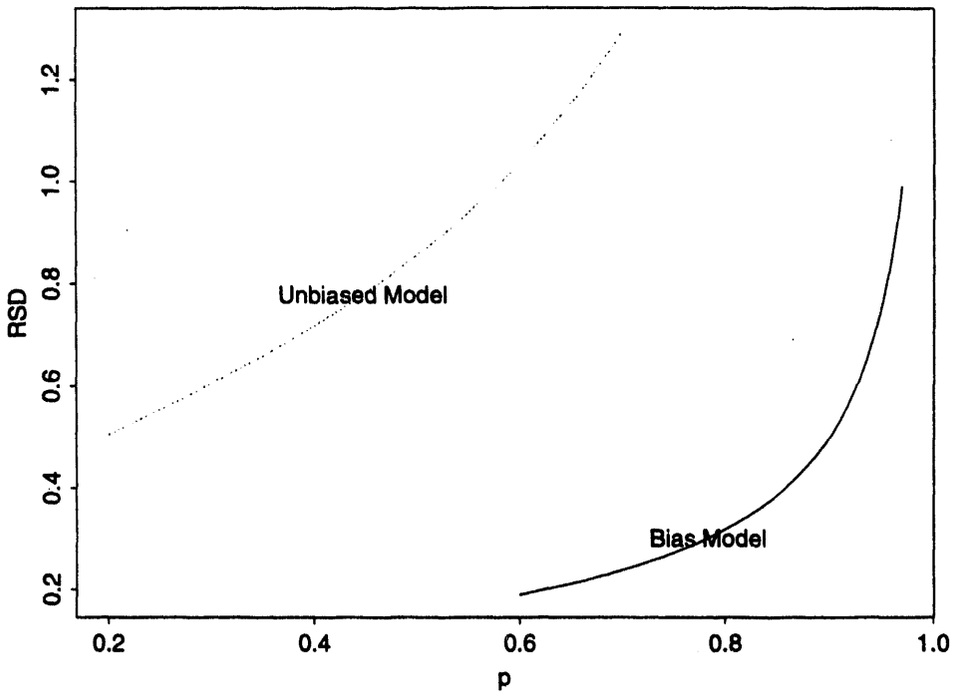


Figure 3: Segment Recovery Thresholds (p) Required to Maintain a 15% Relative Error or Bias for Various Vertical Variabilities (RSD)

5.3 Comparing Biased and Unbiased Sampling

Figure 5.3 presents two curves that are useful for general planning. These curves provide limits or threshold values for percent recovery (p) as a function of the vertical variability (RSD). It is assumed that the relative error (i.e., relative bias or root mean square error) in the analyte concentration estimates is limited to no more than 15%. From the figure, it is obvious that a biased sampler requires more stringent thresholds for segment recovery (p) than an unbiased sampler.

To use this figure, choose a reasonable limit on vertical variability (RSD) for the material being sampled, and then determine the segment recovery proportion (p) required. This recovery value is the threshold for core sampling associated with that degree of vertical variability. A reasonable starting value for the vertical variability RSD is 100%. If the sampler is biased, then a recovery rate (p) close to 100% is required. If the sampler is unbiased, then a recovery rate (p) near 60% is sufficient.

Equations 11 and 19 can be used to generate curves similar to those in Figure 5.3, for various values of the relative error or relative bias in the analyte concentration estimates.

6 References

[1] Morgan, L.G., W.W. Schulz, M.R. Adams, and K.W. Owens. 1988. *Summary of Single-Shell Tank Waste Characterization: 1985 to 1987*, WHC-EP-0075, Westinghouse Hanford Company, Richland, Washington.

[2] Heasler, P.G., R.J. Serne, C.M. Anderson, P.D. Whitney, and D.B. Baird. 1993. *Statistical Evaluation of Core Samples from Hanford Tank B110*, PNL-8745, Pacific Northwest Laboratory, Richland, Washington.

[3] Brown, T.M. and L. Jensen. 1993. *Tank Characterization Report for Single-Shell Tank 241-U-110*, WHC-EP-0643, Westinghouse Hanford Company, Richland, Washington.

[4] Box, G.E.P. and G.M. Jenkins. 1976. *Time Series Analysis Forecasting and Control*, Holden-Day, Oakland California.

Distribution

No. of
Copies

OFFSITE

2 DOE/Office of Scientific and Technical Information

K. T. Lang
U.S. Department of Energy, Headquarters
EM-36/329 Trev
1990 Germantown Road
Trevion Building
Germantown, MD 20874-1290

S. M. Blacker
Vice President, Environment
MACTechnical Services Co.
8310 Center Brook Place
Alexandria, VA 22308

M. H. Campbell
MACTechnical Services Co.
8310 Center Brook Place, B1-42
Alexandria, VA 22308

ONSITE

2 DOE Richland Operations Office

J. M. Clark	A4-02
T. Noble-Dial	A4-02

No. of
Copies

ONSITE

9 Westinghouse Hanford Company

H. Babad	R2-78
M. L. Bell	T6-16
T. M. Brown	R2-12
C. S. Haller	R2-12
L. Jensen	T6-07
D. A. Reynolds	R2-11
L. M. Sasaki	R2-12
W. I. Winters	T6-50
W. F. Zuroff	S6-85

23 Pacific Northwest Laboratory

C. M. Anderson	K7-34
J. W. Buck	K6-96
P. G. Heasler (10)	K7-34
B. A. Pulsipher	K7-34
F. M. Ryan	K7-70
J. L. Scott	P7-34
P. D. Whitney	K7-34
T. M. Wood	K6-47
Publishing Coordination	
Technical Report Files (5)	

DATE
FILMED
5/17/94
END