APPENDIX C
Sampling Strategies

C-1. Introduction. As addressed in USACE's Technical Project Planning—Phase I, project technical staff must consider which sampling strategy is appropriate for the current project phase (EM 200-1-2). It is not necessary to apply the same strategy throughout all phases of a project's life cycle. Frequently, early screening sampling may employ a simple strategy, and subsequent phases may require more complicated strategies, using data results from previous phases. Whenever possible, it is best to use available site knowledge in developing a sampling strategy.

C-1.1. Although there are many sampling approaches, this Appendix presents a discussion of the most commonly employed strategies, which are:

- No sampling.
- Judgmental sampling.
- Random sampling.
    - Simple random sampling.
    - Stratified random sampling.
    - Systematic and grid sampling.
- Ranked set sampling.
- Composite sampling.
- Adaptive sampling.

C-1.2. The first two strategies are qualitative; the remaining strategies are probabilistic. In the latter, statistics may be used to estimate sample characteristics such as mean, standard deviation, and uncertainties. Whether performing on-site, field, or off-site laboratory analysis, the sampling design requires equal consideration. For further insights into environmental sampling, see Gilbert (1987) and EPA/600/R-96/084.

C-2. No Sampling. It may be possible to establish the absence of human health or environmental risk without any sampling. There are three criteria necessary to create a quantifiable risk: i) a chemical release to the environment; ii) a pathway of exposure; and iii) an exposed population. If any of these conditions are not satisfied, a risk does not exist and sampling is not required.

C-2.1. Historical quantitative and qualitative information available during the early stages of a project's life cycle may be adequate for site closure without sampling. Qualitative data are typically not as expensive to collect as quantitative data and may be more informative than quantitative data for answering questions about hazardous, toxic, and radioactive waste sites.

C-2.2. Historical qualitative and quantitative data hold an array of site information useful in reaching a conclusion. The reliability and applicability of historical data and qualitative information (such as interviews with site personnel and photographs) should be evaluated. For example, have historical chemical data been gathered using comparable methods? Is the set of material safety data sheets complete and current? Do toxicity data derived from studies demonstrate adequate quality control? Are engineering drawings pre-construction or "as-builts"? Statistical techniques are often critical to assessing the usability of quantitative historical data, particularly when incorporating historical data into more recent data sets. Simple descriptive statistics (such as the mean, standard deviation, and range) and statistical plots (such as box-and-whisker plots) are useful for qualitative comparisons of different data sets (Appendices G and J). Quantitative statistical comparisons are also frequently appropriate. For example, it may be desirable to compare the mean or variance of a prior data set to a recent data set (Appendix M). When quantitative statistical comparisons are made, the data should also be evaluated to verify that they satisfy the underlying assumptions of the statistical tests (for example, random sampling and adequate numbers of samples).

C-3. Judgmental Sampling. Perhaps the most common sampling strategy is judgmental sampling (also known as targeted or biased sampling). As the name implies, this sampling strategy relies upon the investigator's knowledge and experience. Judgmental sampling is the selection of samples without a statistical design, that is, without any randomization. It can be useful when good documentary data are available and when it is done by an experienced professional with technical expertise. Judgmental sampling is frequently used to target high-contaminant concentrations or worst-case site conditions, such as the collection of samples in visibly stained soils. The underlying rationale for this approach is that, if contamination were not detected (or detected at acceptable levels) in the areas of the site that would have been most impacted by site-related waste handling activities, then acceptable levels of contamination could be assumed in the remaining portions of the study site. However, if unacceptable levels of contamination were detected, the results would be inappropriate for evaluating site-wide average concentrations. An example of judgmental sampling is presented below to illustrate a common *improper* use of the sampling technique.

C-4. Case Study 1—Judgmental Sampling, Ordnance Demolition Area.

C-4.1. The project team used judgmental sampling to obtain a worst-case estimate of explosive residues in surface soils associated with an ordnance demolition area. They did this by sampling where activities historically occurred, specifically targeting stained soils, pits, and debris-laden areas. The team collected background samples and compared group means and variances. They found a statistically significant increase in on-site concentrations relative to the background samples for several explosive residues, concluded that the entire site was contaminated with explosives, and scheduled the area for further investigation and remediation.

C-4.2. In this case, it was incorrect for the project team to compare judgmental non-randomized data sets in a *statistically quantitative* manner. This problem is common in using historical data. One of the primary assumptions in conducting any statistical analysis is that data were obtained in a random fashion. The fact that the on-site samples were biased toward areas of known or suspected high concentration increased the probability that the on-site average concentration would exceed background, potentially leading to biased conclusions. Either the initial round of sampling should have been performed randomly or new samples should be randomly collected and submitted for analysis prior to concluding the presence of site-wide contamination. Alternatively, it might be possible to stratify the site in such a manner that the judgmental samples are representative of only select portions of the entire study area. See Section II of Chapter 3 for further discussion of comparing on-site to background concentrations.

C-5. Random Sampling. The term random sampling encompasses a set of unbiased techniques to choose locations from which to sample at a site. Random sampling has the advantage that its lack of bias allows for robust statistical calculations. However, random sampling is not the same as arbitrary sampling; it does not mean "sample in any manner." The sampling design must be such that every portion of the population possesses an equal opportunity of being selected in the sample. Therefore, when implementing a random sampling design, planners must define and consider the entire population. Both the spatial and temporal boundaries of the environmental population must be well-defined, as instructed in EPA/600/R-96/055, QA/G-4. Samples may need to be collected randomly, not just horizontally across a study area, but vertically as well. Likewise, a continuing waste stream would be sampled randomly in time. Three forms of random sampling are discussed in this paragraph: simple random sampling, stratified random sampling, and systematic random sampling. EPA Quality Assurance QA/G5-S*, Guidance for Choosing a Sampling Design for Environmental Data Collection,* describes the three random sampling methods in detail.

C-5.1. *Simple Random Sampling*. In simple random sampling, sample locations are selected using random numbers. Every possible set of locations has an equal chance of being selected. For example, a simple random sample from a group of liquid waste drums may be taken by numbering all the drums and randomly selecting numbers from that list. Simple random sampling does not presuppose any information regarding the spatial distribution of the likely contamination at the site, other than assuming that no spatial correlation exists. Samples are collected at random from the study area without consideration for factors such as suspected disposal activities, debris locations, spills, or other spatial control on contamination.

C-5.1.1. The major advantages of simple random sampling are that i) it provides statistically unbiased estimates of the mean, proportions, and variability; ii) it is easy to understand and use; and iii) sample size calculations and data analysis are simple to do.

C-5.1.2. The disadvantages of simple random sampling are as follows.

C-5.1.2.1. The environmental population must be relatively homogeneous for simple random sampling to be effective. In particular, major spatial or temporal trends should not exist. Simple random sampling would be inappropriate if localized areas of high contamination or hot-spots exist. Because every portion of the site has an equal opportunity of being selected, if hot-spots constitute a small portion of the total study area, it is likely that random sampling will fail to detect them. Under these circumstances, random sampling will give undue weight to the less contaminated portions of the site.

C-5.1.2.2. It is possible that, by random chance alone, the sample points will be clustered within a small portion of the study area and will not reliably characterize (e.g., owing to heterogeneity) the entire study area.

C-5.1.2.3. Random sampling is often less efficient and, as a result, more expensive than other sampling designs because it requires more samples to obtain the same result. It is most viable when the target population or study area is small. The analytical costs may be offset by the streamlined sampling design, which requires less research than judgmental sampling.

C-5.2. *Stratified Random Sampling*.

C-5.2.1. In stratified sampling, the target population is separated into non-overlapping subpopulations, or strata, that are expected to be relatively homogeneous. Strata may be chosen on the basis of spatial or temporal proximity of the units or on the basis of existing information or professional judgment about the site or process. For instance, if an exposed population is likely to contact only surface soil rather than all soil, then the site could be divided into a surface soil stratum and subsurface soil stratum. Once the strata are defined, each stratum is randomly sampled. This approach allows the project team to focus on areas of greatest concern while retaining the benefits of a random sampling plan. Some examples of stratification at a hazardous waste site include different soil types, depth within an aquifer or surface water body, or separate waste ponds used at different times in site history.

C-5.2.2. Stratified random sampling can be a very effective approach to site characterization. If there is less variation within each subpopulation than in the target population as a whole, stratified random sampling can be more efficient than simple random sampling. Other advantages of this design are that it has potential for achieving greater precision in estimates of the mean and variance, and that it allows computation of reliable estimates for population subgroups of special interest. In fact, a well-constructed stratified sampling plan is the best alternative in most instances where judgmental sampling plans are now employed.

C-5.3. *Systematic Random Sampling*. In systematic sampling, samples are taken at regular intervals in time or space, i.e., along some sort of grid. An initial location or time is selected at random, and subsequent samples are collected at regular spatial or temporal intervals. The sampling scheme retains its random characteristic as long as the initial sampling location or time is randomly, not arbitrarily, selected.

C-5.3.1. Systematic sampling methods are used to search for hot-spots and to infer means, percentiles, or other parameters. They are also useful for estimating spatial patterns or trends over time. These designs provide practical and easy methods for designating sample locations and ensure uniform coverage of a site, unit, or process. One significant benefit of a systematic design is that it generally ensures that some samples from each possible subgroup within a population will be selected.

C-5.3.4. There are two approaches to grid sampling. One may select a particular grid pattern and sample at every node within the grid. Although it is common for sampling plans to specify a square grid pattern, there are a variety of patterns that can be used, often to some advantage in terms of cost or efficacy. Grid blocks may be squares, rectangles, triangles, parallelograms, pentagons, hexagons, or other polygons, depending upon the application. Alternatively, one may randomly pick a starting point in a grid and then collect samples in some logical pattern (for example, move south two blocks and east three blocks). When the edge of the grid is encountered, the pattern starts again on the opposite side of the grid.

C-5.3.5. One can immediately see that such an approach could be very expensive. This type of sampling is often reserved for situations where the analytical cost is low, or where the area to be covered is quite large, as in the estimation of lead analysis over a firing range using a portable x-ray fluorescence (XRF) spectrometer. An important consideration is the size of the individual blocks within the grid or the distance between grid lines.

C-5.4. *Hot-Spot Sampling*. Searching for a hot-spot is a special case where grid spacing may be estimated using information about the suspected hot-spot size and shape. Hot-spots may be located on two-dimensional surfaces or in three-dimensional volumes. For volumes, a three-dimensional grid is generated via the extension of a pair of two-dimensional grids.

C-5.4.1. This method relates the likelihood of successfully locating hot-spots based on their assumed size, shape, and orientation. The acceptable probability of not finding a hot-spot (β) must be specified at the outset. This value must be decided upon by the project team depending on the degree of risk associated with not identifying the hot-spot. Gilbert (1987) provides graphs (called nomographs) that correlate the shape of the hot-spot with the acceptable probability of not finding the spot and the length of the hot-spot divided by the required grid spacing. Table C-1 provides a summary of the nomographs for square and triangular grids. Users will need to interpolate, reference the original citation, or use a conservative set of values in applying this table to individual studies.

C-5.4.2. As mentioned above, to determine the grid spacing (G) for a hot-spot, assumptions must be made about its size and shape (Figure C-1). The shape is represented by the factor (S), defined as the width (W) of the elliptical target spot divided by the expected length (L). If the expected shape is a circle, S is equal to 1. If S is an ellipse, S is less than 1, but greater than 0. If S is unknown, planners may choose to assume that the hot-spot is a narrow elliptical shape, i.e., S

is 0.5 or less. This assumption is conservative. Accommodating a narrower target shape results in denser grid spacing.

Table C-1.
Tabulation for Hot-Spot Grid Spacing

For Square Sampling Grids—Values Listed Are *L*/*G*										
	S									
β	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
0.0						1.00	0.80	0.77	0.74	0.70
0.1				1.00	0.83	0.74	0.68	0.62	0.58	0.55
0.2				0.87	0.77	0.68	0.62	0.58	0.53	0.51
0.3			0.93	0.78	0.69	0.62	0.57	0.53	0.49	0.47
0.4			0.85	0.72	0.64	0.58	0.53	0.49	0.47	0.44
0.5		0.94	0.77	0.65	0.57	0.51	0.48	0.44	0.42	0.40
0.6		0.83	0.68	0.58	0.51	0.47	0.43	0.41	0.39	0.37
0.7	1.00	0.71	0.58	0.50	0.44	0.41	0.38	0.35	0.33	0.31
0.8	0.78	0.56	0.44	0.49	0.35	0.32	0.30	0.28	0.27	0.26
0.9	0.57	0.39	0.32	0.29	0.27	0.25	0.23	0.21	0.20	0.19
1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
For Triangular Sampling Grids—Values Listed Are *L*/*G*										
	S									
β	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
0.0					0.94	0.81	0.74	0.66	0.60	0.57
0.1				0.90	0.78	0.69	0.62	0.57	0.52	0.50
0.2			0.95	0.80	0.70	0.62	0.57	0.52	0.49	0.47
0.3			0.87	0.73	0.63	0.57	0.52	0.48	0.46	0.43
0.4		1.00	0.79	0.67	0.58	0.53	0.48	0.45	0.42	0.40
0.5		0.86	0.69	0.59	0.52	0.48	0.43	0.41	0.39	0.37
0.6		0.75	0.61	0.52	0.47	0.42	0.39	0.37	0.35	0.32
0.7	0.94	0.84	0.52	0.44	0.40	0.37	0.33	0.31	0.30	0.28
0.8	0.75	0.52	0.41	0.37	0.32	0.30	0.28	0.27	0.24	0.22
0.9	0.51	0.36	0.30	0.25	0.22	0.20	1.90	1.80	1.70	1.70
1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

β = probability of missing the hot-spot
W = width of elliptical hot-spot
L = length of the semi-major axis (radius of a circle)
G = grid spacing

S =*W*/*L* (ratio of width to length of hot-spot)
S = 1 is a circle
S = 0.1 is a narrow ellipse
L/*G* = a dimensionless value

C-5.4.3. Based on an estimate of the length of the target hot-spot, we may define the value (*L*), which is one-half of the long axis of the ellipse. In the case of a circular hot-spot (*S* = 1), this is equivalent to the radius of the circle. Finally, the nomographs presented as Table C-1 may be used to determine the appropriate grid spacing (expressed in terms of L/G), based on the values of S and β.

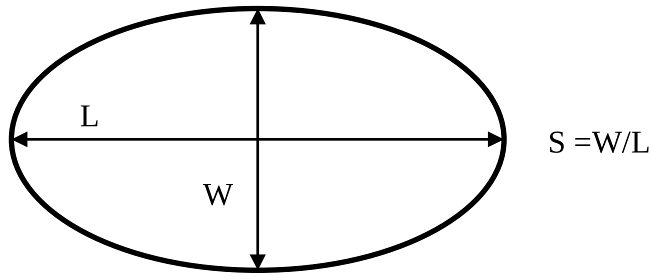

Figure C-1. Elliptical hotspot.

C-5.4.4. The effectiveness of the hot-spot sampling method depends on the accuracy of existing site-specific information. Without prior knowledge, it is difficult for planners to estimate the shape and dimensions of the anticipated hot-spot. In practice, this information is rarely known with confidence, and hot-spot spatial dimensions are often determined on the basis of economic considerations rather on the basis of pre-existing information on site conditions. The required number of samples depends greatly upon the assumed dimensions of the hot-spot. Planners should do a number of calculations, varying the shape and estimated size of the hot-spot. If the resulting grids are similar and differences in sample design relatively minor, then planners may feel more confident about the methodology applied to the site.

C-5.4.5. The hot-spot mathematical procedure may also be applied in reverse; if grid spacing and presumed hot-spot size and shape are known, the probability of having missed a hot-spot (of some specified size) may be determined. Thus, site investigation managers may be able to convey to regulators the level of certainty that no problems were missed, within reasonable expectations. By applying the nomographs and solving for different variables, a researcher can answer such questions as the size of a hot-spot likely to be found by a given grid spacing, and the probability of not finding a hot-spot based on a given grid spacing. The following case study compares sampling strategies for a site with a hot-spot.

C-6. Case Study 2—Comparing Random Sampling Strategies at a Site with a Hot-Spot. Table C-2 illustrates examples of the three random sampling approaches at a generic site and the differences in descriptive statistics that might influence a manager's decisions related to the site. The three different sampling plans are applied to the same data set: Plan A is simple random sampling, Plan B is stratified random sampling, and Plan C is systematic and grid sampling. The site is represented by a 9-by-9 grid with the 3 right-most grid columns divided by a heavy solid line indicating a hot-spot, and the lower left 12 cells a secondary hot-spot (applicable to Plans B and C only). For Plans B and C the largest group of cells is Group 1; the lower left corner is Group 2; and the right three columns make up Group 3. The number in each cell represents a generic analytical result, had a sample been collected from every cell. A collected sample is represented by a shaded cell. The systematic sampling (Plan C) was determined by using a set pattern beginning at a randomly selected first location. (This is not obvious from the pattern of shaded cells.)

C-6.1. For this example, assume that decisions will be based on a 2-stage comparison criterion: values less than 5 units require no action; values greater than 5 units but less than 50 units require further remedial investigation but no immediate action; and values greater than 50 units indicate an immediately dangerous condition requiring an emergency removal action.

C-6.2. The three sampling plans are judged against a hypothetical sampling of every cell across the site. In this case, the following are determined:

C-6.2.1. Total number of samples, $N = 81$.

C-6.2.2. Summation of all results, $S = 1708.1$.

C-6.2.3. Total population average, $\mu = 21.09$.

C-6.3. For Plans B and C, the following are determined for the entire populations of each group:

Group 1:	$n_1 = 42$	$S_1 = 23.4$	$\mu_1 = 0.56$
Group 2:	$n_2 = 12$	$S_2 = 47.7$	$\mu_2 = 3.98$
Group 3:	$n_3 = 27$	$S_3 = 1,637$	$\mu_3 = 60.63$

C-6.4. Note that population mean may be viewed as a weighted mean calculated from each group population mean:

$$\mu = \sum_i (n_i / N)\,\mu_i = \sum_i w_i \mu_i$$

$w_1 = 42/81$, $w_2 = 12/81$, $w_3 = 27/81$

C-6.5. For Plans B and C, a total of nine samples are randomly selected. (For example, for the nine samples collected for Plan B, two are from Group 1, four are from Group 2, and three are from Group 3.) The mean of the population mean (i.e., entire set of 81 samples) is estimated by calculating the sample mean of each group and weighting them:

$$\bar{x} = \sum_i (n_i / N)\,\bar{x}_i = \sum_i w_i \bar{x}_i$$

Table C-2.
Comparison of Random Sampling Method Results

Plan A: Simple Random Sampling

0.26	0.24	0.74	0.95	0.25	0.34	94.18	20.16	61.90
0.97	0.54	0.13	0.18	0.17	0.48	5.40	13.39	19.79
0.97	0.30	0.72	0.09	0.48	0.79	55.28	55.10	94.98
0.82	0.03	0.95	0.72	0.22	0.81	29.31	1.26	72.37
0.52	0.66	0.48	0.83	0.92	0.43	78.73	84.02	77.05
2.82	1.45	1.24	0.52	0.69	0.47	89.00	98.76	83.54
3.14	8.24	8.48	0.55	0.11	0.85	76.71	96.91	84.19
7.18	1.68	0.96	0.74	0.47	0.86	42.95	16.94	72.67
5.84	3.73	2.98	0.65	0.99	0.51	96.66	52.85	62.86

Plan A	S_i	$\bar{x}_i$	$\bar{x} = \sum w_i \bar{x}_i$
All nine samples	193.84	21.54	N/A

Plan B: Stratified Random Sampling

0.26	0.24	0.74	0.95	0.25	0.34	94.18	20.16	61.90
0.97	0.54	0.13	0.18	0.17	0.48	5.40	13.39	19.79
0.97	0.30	0.72	0.09	0.48	0.79	55.28	55.10	94.98
0.82	0.03	0.95	0.72	0.22	0.81	29.31	1.26	72.37
0.52	0.66	0.48	0.83	0.92	0.43	78.73	84.02	77.05
2.82	1.45	1.24	0.52	0.69	0.47	89.00	98.76	83.54
3.14	8.24	8.48	0.55	0.11	0.85	76.71	96.91	84.19
7.18	1.68	0.96	0.74	0.47	0.86	42.95	16.94	72.67
5.84	3.73	2.98	0.65	0.99	0.51	96.66	52.85	62.86

Plan B	S_i	$\bar{x}_i$	$\bar{x} = \sum w_i \bar{x}_i$
All nine samples	210.22	23.36	N/A
Group 1	2.06	0.51	
Group 2	3.12	1.56	
Group 3	205.04	68.35	
			23.28

Plan C: Systematic and Grid Sampling

0.26	0.24	0.74	0.95	0.25	0.34	94.18	20.16	61.90
0.97	0.54	0.13	0.18	0.17	0.48	5.40	13.39	19.79
0.97	0.30	0.72	0.09	0.48	0.79	55.28	55.10	94.98
0.82	0.03	0.95	0.72	0.22	0.81	29.31	1.26	72.37
0.52	0.66	0.48	0.83	0.92	0.43	78.73	84.02	77.05
2.82	1.45	1.24	0.52	0.69	0.47	89.00	98.76	83.54
3.14	8.24	8.48	0.55	0.11	0.85	76.71	96.91	84.19
7.18	1.68	0.96	0.74	0.47	0.86	42.95	16.94	72.67
5.84	3.73	2.98	0.65	0.99	0.51	96.66	52.85	62.86

Plan C	S_i	$\bar{x}_i$	$\bar{x} = \sum w_i \bar{x}_i$
All nine samples	244.75	27.19	N/A
Group 1	2.26	0.45	
Group 2	3.73	3.73	
Group 3	238.76	79.59	
			27.31

SUMMARY

Grouping	Population Mean#	Simple $\bar{x}_A$	Stratified $\bar{x}_B$	Systematic $\bar{x}_{C\,\#}$
Group 1	0.56	—	0.52	0.45
Group 2	3.98	—	1.56	3.73
Group 3	60.63	—	68.35	79.59
Entire Grid	21.09	21.54	23.28	27.31

Notes:
Shading indicates a sampled grid location

C-6.7. To assess each sampling plan, the mean concentrations determined from the limited sampling to those for the entire site data set are compared. Simple random sampling (Plan A) provides the best estimate of the overall population average. However, it is fairly limited in identifying the best course of action for the underlying strata in that it suggests that the entire population is subject to additional investigation or action. Another shortcoming is that none of the random sampling designs identified the "secondary hot-spots" in Group 2; that is, none of the samples selected in Group 2 (the shaded cells) exceed 5. Stratified sampling (Plan B) resulted in better data for decision-making because data were obtained for all three groups, although some of the group mean estimates are rather poor. In the systematic plan (Plan C), each stratum is represented in the statistics at a frequency roughly equal to its portion of the whole. (The ratio of the total number of cells for Groups 1, 2, and 3 is approximately 5:1:3, the ratio of the number of samples collected for each group.) Had the presence of underlying strata been unknown, the systematic plan would have given the best indication of potential problems at the site.

C-7. Systematic Sampling Over Time. Systematic sampling can also be applied when the parameter of interest is expected to vary over time. This one-dimensional scheme is sometimes called periodic sampling and is quite simple. Divide the span of time under examination into an arbitrary number of "blocks" (e.g., 20 intervals) and, having calculated an appropriate number of samples for the application, simply divide the number of samples required into the number of blocks available. This gives the time between samples. The starting time is chosen randomly. (Note that the same strategy may be used to establish the distance between grid lines, where the intervals would be measured in units of distance rather than time.) In general, the greater the variability in the parameter being measured is, the greater the number of samples required for the required degree of confidence.

C-8. Ranked Set Sampling. As stated in EPA QA/G5-S: "Ranked set sampling is an innovative design that can be highly useful and cost-efficient in obtaining better estimates of mean concentration levels in environmental media." The technique typically entails the use of two analytical methods, a "definitive" method (e.g., a fixed laboratory method) and a "screening" method (e.g., a field method). Usually, the cost of the screening method is significantly less than that of the definitive method, while the analytical quality of the definitive method exceeds that of the screening method. Ranked set sampling is a two-phase sampling design. It first identifies sets of field locations and uses inexpensive measurements to rank locations within each set; next, it selects one location from each set for analysis by the definitive method. Only a brief overview of this sampling technique is presented in this Appendix. The reader is referred to the EPA QA/G5-S guidance document for a more detailed discussion and illustration of rank set sampling.

C-8.1. For a "balanced design," m sets of m samples (at total of m^2 samples) are initially analyzed using professional judgment or some screening method. The field samples in each set are then independently ranked (e.g., from highest to lowest). The first ranking sample (the highest sample) is selected from the first set, the second highest ranking sample is selected from the second set, and so forth, until m samples are selected for analyses using the definitive (i.e., more accurate and expensive) analytical method. The process is repeated r times, giving a total of $m^2 r$ field analyses and mr definitive analyses.

C-8.2. One of the best reasons for applying ranked set sampling is its ability to provide samples from across the distribution of values at the site. This, in turn, creates a better estimate of the population mean and improves the performance of various other statistical tests, especially those that entail distributional assumptions. A wide variety of field screening tools can be used to supplement the professional judgment of the samplers and, in certain circumstances, can even be used later as definitive data, assuming good correlation with fixed laboratory results is achieved. Paragraph C-9 illustrates a practical application of ranked set sampling.

C-8.3. Relative to simple random sampling, this design results in a more representative sample, and therefore leads to more precise estimates of the population parameters. A large number of screening analyses increases site coverage, and the ranking information from the screening analyses reduces the required number of definitive analyses relative to the number that would be required from a random sampling design. Therefore, the ranked set sampling approach has the added benefit of typically being less expensive than a simple random sampling approach. Because preliminary data are used to ensure representative samples are collected, the variability among the samples is better controlled and the number of samples required to make a probabilistic decision with the same degree of confidence is reduced.

C-8.4. However, there are several limitations to ranked set sampling. The screening and definitive methods must be strongly correlated with one another. In addition, the cost of the definitive analyses compared to the cost of the ranking procedure used for the field methods must be relatively large for the approach to be cost-effective. One should consider whether two phases of sampling is cost-effective relative to a more standard sampling method and whether it is technically feasible given project resource constraints. Finally, the statistical computations to be performed on the resulting data set are more complex relative to those used for a simple random sampling design.

C-9. Case Study 3—Ranked Set Sampling. The project team used field screening test kits on a grid established over a wide area to characterize an ordnance demolition area. Using the information from the field screening, the team was able to stratify the site into three areas: i) a region requiring no remediation; ii) an area clearly requiring remediation and for which samples at depth were required to provide volume estimates; and iii) an area requiring additional study with definitive methods to establish the need for remediation or no further action. Definitive samples were then collected to distinguish the various explosives and their daughter products that the test kit could not resolve. These results were then used to better estimate the average concentration of individual explosives within the various strata, and to serve as confirmation samples for the test kits. The definitive samples helped correlate low-, mid-, and high-range concentrations in each area. Thus, the screening data were used to select locations for definitive samples to ensure more representative mean concentrations within each area.

C-10. Composite Sampling. Composite sampling is the physical averaging of environmental samples in a manner that yields an accurate and representative estimate of environmental conditions, usually at a reduced cost. It involves physically combining and homogenizing two or more environmental samples (referred to as "grab" samples, and called "subsamples" in this context) to form a new sample referred to as a composite sample. Compositing is used when the mean is primarily of interest (i.e., because the process is a physical averaging) and information on the spatial or temporal variability of contamination is not needed (i.e., because this information is lost unless the subsamples can be reanalyzed). Tables C-3 and C-4 suggest circumstances under which compositing can be useful. Various sampling designs may be used to select subsamples to be mixed together into composites.

Table C-3.
Objectives of Composite Sampling—Fundamental Cases

1. Objectives that rely on composite sampling	a.	Estimating a population (or stratum) mean for a continuous variable (e.g., analyte concentration)*
	b.	Estimating proportion of population exhibiting some trait
2. Objectives that rely on composite sampling and retesting protocols	a.	Classifying sampling units as having or not having some trait such as being in a hot-spot or from a contaminated cell
	b.	Identifying the sampling unit with highest value of some continuous measure (e.g., concentration), or identifying sampling units in the upper percentiles

* In general, information on variability and spatial or temporal patterns is lost when compositing is used for this objective; however, in some cases, some information on patterns can be acquired.

Table C-4.
Criteria for Judging Benefits of Composite Sampling

Criterion or Objective	Composite sampling is likely to be beneficial if…
1. Analytical costs	Analytical costs are high relative to sample acquisition/handling costs.
2. Analytical variability	Analytical variability is small relative to variability of the target population.
3. Analytical sensitivity	Concentrations of relevance are much larger than detection and quantitation limits.
4. Representativeness	Compositing does not affect sample integrity (expect no chemical reactions/interferences or analyte losses from volatility) or result in safety hazards. Individual samples can be adequately homogenized.
5. Objective is to estimate population mean (See 1a in Table 2-3)	Information on individual samples is not important. Information on associations is not important. Criteria 1, 2, and 4 are met.
6. Objective is to estimate proportion of population with a trait (See 1b in Table 2-3)	Composite has trait if individual sample does. Likelihood of misclassification is small. Trait is rare. Criteria 1, 2, 3, and 4 are met.
7. Objective is to classify samples as having/not having a trait (See 2a in Table 2-3)	Composite has trait if individual samples do. Likelihood of misclassification is small. Retesting of aliquots (grab samples) for each composite sample is possible. Trait is rare. Criteria 1, 2, 3, and 4 are met.
8. Objective is to identify the sample(s) with the highest value (See 2b in Table 2-3)	Measurement error is negligible. Retesting of aliquots from individual samples is possible. Criteria 1, 2, 3, and 4 are met.

C-11. Compositing Fluids. A typical application of compositing fluids is in creating a representative sample when one or another condition, tied to contaminant mass or concentration, varies over space or time. National Pollutant Discharge Elimination System (NPDES) monitoring provides a classic case in point.

C-11.1. The fundamental objective for this type of compositing is to develop a single sample that accurately represents the whole area or time under consideration. The alternative entails greatly increased sampling and analysis costs and agreement on an acceptable mathematical approach to combining the individual sample results. Table C-5 examines a variety of compositing approaches linked to particular circumstances. Paragraph C-12 illustrates an example of flow-proportioned compositing.

C-11.2. Another classic use of compositing fluids is in sampling stack emissions. When a fluid (or gas in the case of stack emissions) flows through a pipe, the fluid does not move at a uniform speed across the diameter of the pipe. Friction with the interior surface of the pipe causes fluids near the casing to move more slowly than at the center. Thus, when measuring mass per unit volume per unit of time, *isokinetic* sampling is applied. In this case, subsamples are collected across the diameter of the pipe for identical time intervals, along with a measure of the flow rate at the individual locations. Using this information, the engineer can balance concentration against the flow rate to yield an accurate estimate of the average mass discharged from the stack (or pipe) over time.

Table C-5.
Compositing Methods

Method No.	Sampling Mode	Compositing Principle	Comments	Disadvantages
1.	Continuous	Constant sample pumping rate	Practicable but not widely used	Yields large sample volume; may lack representativeness for highly variable flows
2.	Continuous	Sample pumping rate proportional to stream flow	Not widely used	Yields large sample volume but requires accurate flow measurement equipment
3.	Periodic	Constant sample volume, constant time interval between samples	Widely used in automatic samplers and widely used as manual method	Not most representative method for highly variable flow or concentration conditions
4.	Periodic	Constant sample volume, time interval between samples proportional to stream flow	Widely used in automatic sampling but rarely used in manual sampling	Manual compositing from flow chart
5.	Periodic	Constant time interval between samples; sample volume proportional to total stream flow since last sample	Not widely used in automatic samplers but may be done manually	Manual compositing from flow chart
6.	Periodic	Constant time interval between samples; sample volume proportional to stream flow at time of sampling	Used in automatic samplers and widely used as manual method	Manual compositing from flow chart

After: EPA 600/4-82-029

C-12. Case Study 4—Flow-Proportioned Compositing. At a manufacturing facility in Ohio, an existing NPDES permit called for the facility to collect a single, three-part, equal-weight composite sample monthly. The facility operated three shifts. Production on all three shifts was essentially the same, although the bulk of maintenance activities took place on the second shift. Three grab samples, one from each shift, were composited at the laboratory prior to analysis.

C-12.1. A change in business climate led to a reduction in demand such that the midnight to 8 a.m. shift was canceled and the 4 p.m. to midnight shift was reduced by roughly two-thirds. The facility manager asked that the overall effect the change in shifts would have on discharge rates be assessed in preparation for permit renewal negotiations. For this case study, only the nitrate data are considered. The following analysis was performed:

Original flow–shift 1	200,000 gal/day*	New flow	200,000 gal/day
Original flow–shift 2	200,000 gal/day	New flow	70,000 gal/day
Original flow–shift 3	200,000 gal/day	New flow	5,000 gal/day

C-12.2. Historical composite results for the previous year were as follows:

Jan	0.48	Average	0.38 mg/L†	
Feb	0.12	Variance	0.20 mg/L	
Mar	0.26			
Apr	0.34	Current Permit Limit		2.5 lb/day‡
May	0.48	EPA Proposed New Limit		1.0 lb/day
Jun	0.31			
Jul	0.47			
Aug	0.46			
Sep	0.13	Assuming average concentration does not change		
Oct	0.40			
Nov	0.16	Under Equal Volume sampling, lb/day = 1.9		
Dec	0.20	Under Flow Proportioned sampling, lb/day = 0.87		

C-12.3. Thus, the new permit limit will be acceptable if the permit also incorporates a change in the compositing method.

C-13. Compositing Solids. Generally speaking, solids and, in particular, soils are composited to estimate the concentration of a contaminant over large areas, or when the granular or globular nature of the contaminant of concern (e.g., explosives, PCB oils) can provide false estimates of concentration from individual measurements because of excessive heterogeneity in the individual samples. Other applications are also possible. Compositing can also be used to assess the proportion of samples that meet a specific condition and, with retesting of a small subset of original locations, can also be used to locate rare events (like hot-spots) where too many individual samples would be required. For example, at a site with very few historical data, 12 composite samples of 4 subsamples each may be analyzed for a long list of possible contaminants. If only one sample contains only a few contaminants of concern, then further investigation is limited to those contaminants and in only four small areas. Exhaustive testing of the 48 original discrete samples was not necessary, and further study of most of the site is precluded. As extensive mixing of the subsamples is required to form a representative composite, composite sampling is not generally applied to samples when volatile organic compounds (VOCs) are of particular interest.

* gal/day = gallons per day
† mg/L = milligrams per liter
‡ lb/day = pounds per day

C-14. Adaptive Sampling. *Adaptive sampling* designs are typically used to characterize the extent of contamination using multiple sampling events; they rely upon cost-effective field methodologies with rapid turn-around time. The results of an initial sampling event are used to modify the selection of future sampling locations for the study area. *Adaptive cluster sampling* is useful when the characteristic of interest is sparely distributed through the site. Adaptive cluster sampling could be used for a study area that contains mostly low-level or negligible contamination but also isolated pockets of high-level contamination (i.e., hot-spots). This is illustrated in Figure C-2. As stated previously, under these circumstances, a random sampling design would not be the optimum approach (as the hot-spots could remain undetected).

C-14.1. Three major elements characterize adaptive cluster sampling. First, a set of sampling locations is initially determined. Though there may be insufficient data to support firm conclusions overall, information may exist that suggests particular areas of the site are clean or contaminated. The result is an initial conceptual model for the site. For example, a grid is placed over the geographical area of interest, where each cell of the grid represents a potential sampling unit (location). A subset of all the potential sampling units is selected for sampling. Figure C-2 illustrates the use of random sampling for the selection of the initial sampling event. Second, a decision rule for each sampling unit must be established. If the contaminant of interest exceeds the decision limit, additional sampling is required "near" the sampling unit (i.e., adjacent sampling units are sampled). Third, the "neighborhood" of each sampling point (i.e., the area required for additional sampling) must be defined. Several additional stages of sampling are designated on Figure C-2. The symbol "X" denotes the neighboring sampling units that were sampled. (Note: In the example illustrated in Figure C-2, one area of contamination was missed.) The decision rule and additional sampling are repeatedly applied until contamination is not detected above the decision limit for each sampling unit. This results in a "mapping" of contaminants as illustrated in the final stage in Figure C-2, where the extent of "hot-spots" is delineated using a large number of sample units. The shaded areas in Figure C-2 represent "hot-spots" (i.e., area in which contamination exceeds the decision limit).

C-14.2. Adaptive sampling and analysis plans (SAPs) provide a cost-effective alternative to traditional sampling designs. Adaptive SAPs are based on field analytical methods allowing for rapid sample turnaround and field-based decision support to guide the sampling program. One objective of adaptive SAPs is to support removal actions.

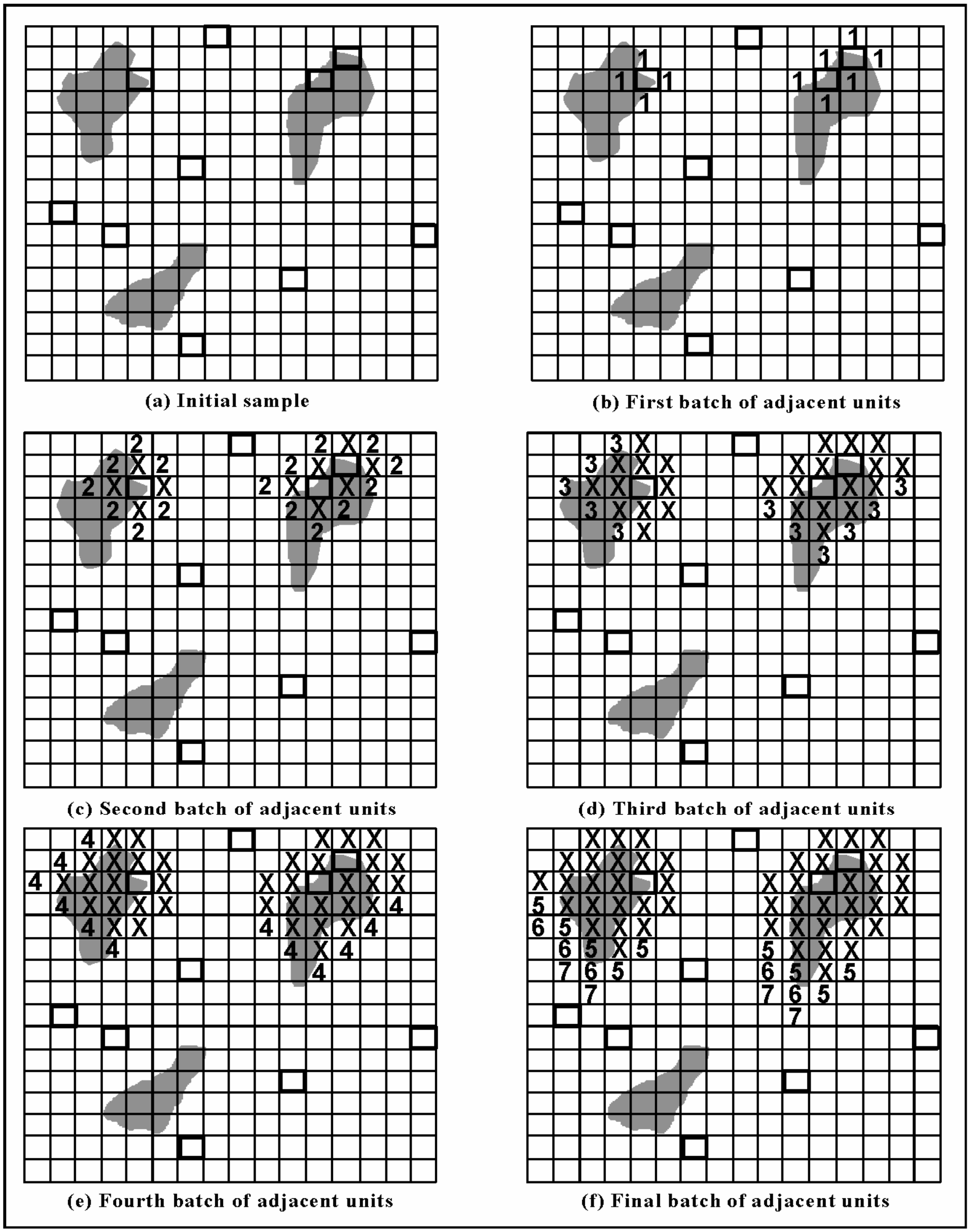

Figure C-2. Population grid with initial and follow-up samples and areas of interest. From EPA QA/G-5S.

C-14.3. Traditional approaches to designing and executing a removal action have relied on "digging to the design line" and then taking confirmation samples. The static work plans that have accompanied these efforts have specified the number and location of samples. Often, however, the design lines have been at best rough approximations of the real extent of contamination, resulting in either extensive under- or over-removal of soils. In both cases, the economic impacts have been significant. An important factor in establishing the design line is the site cleanup levels. Cleanups should be implemented so that concentrations left at the site meet the cleanup goal to a predetermined level of certainty, with the level of certainty agreed upon by the design team and regulators.

C-14.4. Adaptive SAPs rely on field analytical methods to generate sample results quickly enough to have impact on the course of the sampling program. They are based on dynamic work plans that specify the logic of how sampling numbers, locations, and analyses will be determined as the program proceeds. They also rely on rapid, field-level decision-making. Adaptive SAPs require: i) field analytical methods that are appropriate for the types of contaminants expected at a site; and ii) a means for supporting decision-making in the field that is appropriate for the goals of the program.

C-14.5. Rapid field decision-making requires qualitative and quantitative decision support. Qualitative decision support means having technical staff equipped with an accurate understanding of the sampling progress. Large adaptive SAPs can produce hundreds of samples per day. Managing, integrating, and displaying the sample information pose a serious logistical challenge that can interfere with program process if not adequately addressed. A typical adaptive SAP includes some type of field- or web-based database system along with a Geographic Information System for data display to help with logistics and visualization.

C-14.6. Quantitative decision support for adaptive SAPs that delineate removal areas requires the ability to estimate contaminant extent based on sampling results, determine the uncertainty associated with those results, predict expected values from previous sampling, and identify new removal locations based on that information.

C-14.7. The adaptive sampling scheme presented in Figure C-2 may be applied to contamination removal actions as well. In such an application, each sample is used to determine whether soil removal (i.e., excavation) is necessary, and the areal (and volumetric) extent of soil needing removal can be established via such sampling techniques.

C-14.8. The adaptive SAP design and implementation process for guiding removal actions follows these steps.

C-14.8.1. Sampling location decision points forming a regular grid are laid across the site. Each sample decision point is so named because at each sampling location, the following decision must be made: will this point be removed or left in place? For instance, if the petroleum hydrocarbon concentration at this location exceeds an action level, it will be excavated from the site. An action level serves as the criterion for differentiating among decision points that can be considered clean and points that must be treated as contaminated. Because the acceptable level of uncertainty is very important to the design of the adaptive SAP, it must be determined prior to sampling or before the program begins (i.e., during the data quality objective development process), with mutual agreement from all the stakeholders involved with the site.

C-14.8.2. Based on professional judgment and historical information available for the site, a probability is initially assigned to each decision point; namely, the likelihood contamination at that location is greater than some action level.

C-14.8.3. As sample results become available, the probabilities for each of the decision points are updated with actual data. The site is then divided into three regions: i) the portion of the site (decision points) where the probability that contamination exceeds the action level is low (this region is accepted as clean with perhaps only minimal confirmatory sampling); ii) the portion of the site where the probability of contamination is so high that confirmatory sampling is unnecessary; and iii) the portion of the site where there is neither a high nor low probability of contamination above the action level, i.e., the gray area where there is significant uncertainty whether the presence or absence of contamination is greater than the pre-determined action level. Indicator kriging (Appendix Q) may be a powerful tool for such an application.

C-14.8.4. Predetermined decision rules are applied. There may be several alternative decision rules that can be used to drive the sampling process. Additional sampling may need to be done for the gray areas, especially if the removal action is desired to lower overall site risk. The decision rules should tend to produce a sampling program that works its way around suspected areas of contamination. The decision rules should also tend to produce a sampling pattern that starts from areas of suspected contamination and works its way outward to the boundary where removal can cease.

C-14.9. Regardless of the decision rule used, the process is the same. Sampling locations are selected that have the greatest opportunity to provide the most benefit in the context of the selected decision rule. After results are obtained, the extent of contamination is re-estimated along with the number of uncertain decision points remaining, and a decision is made where additional removal is justified until no such locations remain.

C-14.10. Figure C-3 shows the adaptive sampling plan process, and Paragraph C-15 illustrates a practical application of an adaptive SAP.

C-15. Case Study 5—Argonne's Adaptive Sampling and Analysis Program. The U.S. Department of Energy's (DOE's) Argonne National Laboratory developed the following case study.

C-15.1. Oil and gas producers may save millions of dollars in cleaning up soils contaminated with naturally occurring radioactive materials by applying an on-site soil sampling and analysis method developed by the U.S. DOE's Argonne National Laboratory.

C-15.2. Naturally occurring radioactive material accumulates when the production of oil and natural gas from underground reservoirs transports small quantities of radium to the surface. Over time, the radium—usually radium-226 and, to a lesser extent, radium-228—can concentrate in pipe scale and sludge deposits, which in turn can contaminate soil and equipment.

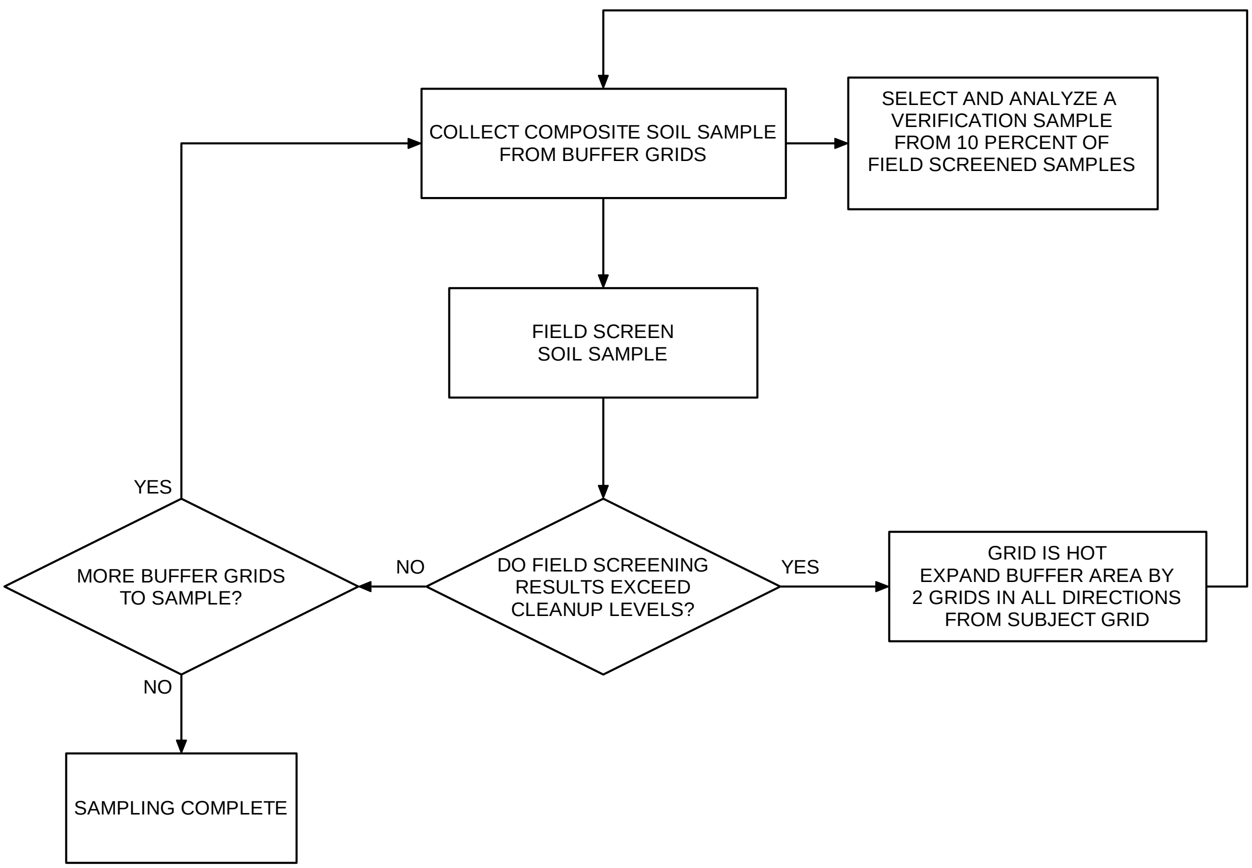

Figure C-3. Adaptive sampling plan flow chart.

C-15.3. The traditional approach to cleaning up such sites involves complicated soil sampling techniques and shipping these samples to off-site laboratories for analysis—a time-consuming and costly process. But a recent demonstration has shown that Argonne's adaptive SAP can dramatically reduce the time and money needed to characterize and remediate sites contaminated with naturally occurring radioactive materials. Adaptive SAP combines real time data collection techniques with in-field decision-making for faster and more precise characterization of a site. It was first used successfully for faster and cheaper cleanup of radioactive contamination at DOE sites.

C-15.4. The demonstration was conducted on a 3.5-acre site at Lease Management, Inc., in Mt. Pleasant, Michigan. Pipe salvaged from nearby oil and gas production sites was stacked there prior to being cleaned and reconditioned. Contaminated scale on the outside of the pipes had fallen off during handling and from exposure to the elements. As a result, soils across the pipe yard had varying levels of radium-226 concentrations.

C-15.5. First, scientists walked over the site with a portable global positioning system and a hand-held gamma ray detection device to map surface gross activity levels. The scientists then used a commercial technology called the RadInSoil™ meter to develop a relationship between gross activity values and radium-226 activity concentrations. State guidelines are based on these activity concentrations. With the field data, researchers then used unique Argonne-developed techniques to determine where soil concentrations of contaminants exceeded regulatory standards and would need to be excavated for disposal. To confirm the presence of radium-226, scientists used a tripod-mounted, camera-like device called a High Purity Germanium gamma spectroscopy system that directly measures radium-226 concentrations in surface soils. With use of the results from adaptive SAP, decisions on excavating contaminated soil for disposal can be made immediately. It took 4 days to characterize and remediate the Michigan site.

C-15.6. The average cost for soil disposal ranges from about $100 to $200 per cubic yard, so keeping soil volumes to an absolute minimum is very important. The goal is to be as precise as possible in digging up dirt for disposal so one doesn't take anything clean away or leave anything above cleanup standards behind.

C-15.7. For sites contaminated with naturally occurring radioactive materials, it is estimated that using adaptive SAP for site characterization costs only 10% of a more traditional approach. In the Michigan demonstration, the use of adaptive SAP is expected to save the site owner at least $36,000 in disposal costs.