7. MAKING DECISIONS USING ISM DATA

7.1 Introduction

This section provides guidance on using data generated from ISM samples to make decisions about a DU. Since the data may inform one or more decisions; it is helpful to establish a structured approach to making decisions, referred to here as "decision mechanisms."

"Decision mechanism" refers to the different ways that environmental concentration data can be used to make decisions at a site.

In the context of this discussion, decision mechanisms include the procedures, inputs, and algorithms that are used to aid decision making based on environmental concentration data. The simplest decision mechanism is a comparison of a single measured concentration to an action level. The inputs in this case are the concentration measured in the sample and the action level. The procedure may be, for example, to compare the concentration to the action level to determine whether further sampling, evaluation, or other action is needed. A common example of this type of decision mechanism might be the comparison of individual discrete soil sample results obtained during a CERCLA Preliminary Assessment to Regional Screening Levels (RSLs) for chemical contaminants at Superfund sites. Discrete soil sample results are often compared directly to the RSL benchmarks. Exceedance of a benchmark by one or more discrete soil sample results can be used to identify a contaminant as a COPC.

More complex decision mechanisms may involve procedures that include the use of advanced statistical analysis or numerical models. For example, a surface soil investigation may involve the use of geostatistical modeling or krieging to estimate the distribution and extent of contaminants across a site using high-density discrete soil sampling data. Decision mechanisms may involve a series of procedures that are iterative or progressively more complex.

The specific decision mechanisms that may be needed to make a final decision for a DU should be determined as part of the planning at the start of the investigation as noted below and in Section 3. As discussed below, decision mechanisms that apply to ISM are analogous to decisions with discrete data, and include the following:

- comparison of a summary statistic (e.g., single ISM estimate of the mean, the mean of multiple ISM results, the 95% UCL of multiple results) to an action level
- comparison of results of a quantitative risk assessment which used a summary statistic (typically a 95% UCL) to an acceptable risk range for carcinogens (e.g., 1×10^{-6} excess cancer risk to 1×10^{-5} excess cancer risk) or to an acceptable hazard threshold for noncarcinogens
- comparison of site and background data sets
- combination of data across multiple DUs
- extrapolation of statistics across DUs

One of the primary benefits of ISM sampling is that the volume of media to which a decision will be applied must be determined prior to sample collection. It is also essential to have an understanding of the manner in which ISM samples will be used to make decisions during

project planning. Decision mechanisms must be consistent with the rationale behind the sampling plan design, as discussed in Section 3, and should be based on the following:

- CSM
- goals of the project and end use of the data
- scale of the decision
- requirements for precision, total error, and decision quality
- assumptions of the statistical method(s)
- anticipated and/or measured degree of variability within the DU

Although the primary component of the decision mechanism is the actual procedure, algorithm, or statistical test employed to evaluate the data and make the decision, such variables as the location of the sample, the number of samples or increments involved, and the rationale behind the action level must be considered. The following are important aspects of decision mechanisms that must be included:

- number of, rationale for, and size of DUs and SUs
- number of SUs composing each DU (from one to many)
- number of increments collected to form each ISM sample
- bulk mass of ISM sample
- mass of analytical subsample
- aspects of "correct sampling" (Pitard 1993)
- number of replicate ISM samples in each SU
- particle size reduction or selection (where appropriate)
- statistic calculated
- source, nature, and numerical value of the action level

An ISM-based sampling project should be tailored to the decisions for which the data will be used. Careful planning is the key to ISM data usability.

ISM samples can be used for a number of different applications. The type of decision mechanism employed must be consistent with the type of decision being made. The specific size and location of DUs are guided by site knowledge regarding the spatial distribution of contaminant(s) and the movement or behavior of receptors that may contact different areas of the DU. Estimates of mean concentration provided by ISM can be useful in evaluating risk from direct contact with soil, where a DU is designated to correspond with a presumed or actual exposure area for human health or ecological risk assessment. Likewise, because most soil-to-groundwater leachate models assume a volume of contaminated soil as the source of contamination to the groundwater, estimates of mean concentrations in targeted volumes of soil are directly applicable to assessment of soil concentrations using soil-to-groundwater leachate models. Another useful application of ISM is when multiple decisions must be made for very large volumes of soil, for instance, when large former agricultural fields or dredge piles are intended for future residential uses. ISM has been used for the exploration of concentration gradients because, in the presence of small-scale heterogeneity, ISM provides a better understanding of contaminant distribution than a few widely spaced discrete samples. ISM

samples may also be used over concentric SUs surrounding a suspected source area. Finally, a variety of different strategies may be used in subsurface investigations with ISM samples.

Regardless of the decision mechanism, the standard steps of data quality assessment as discussed in Section 3 apply. After data are collected, it is important to revisit the CSM and determine whether it is supported by the data or should be modified. Methods for statistical analysis of data should be selected based upon the sampling design and project objectives. Key underlying assumptions associated with the statistical test must be identified and determined to be valid for the data to be analyzed.

7.2 Decision Mechanisms

As in discrete sampling, there is no one decision mechanism dictated by the use of ISM sampling; a variety of decision mechanisms are possible. Each decision mechanism has strengths, weaknesses, and assumptions. In some cases, agency requirements will dictate the approach to be used. In other cases, a consensus on the decision mechanisms to be employed needs to be reached among members of the planning team prior to finalization of the sampling plan. This section discusses the benefits and drawbacks as well as the assumptions involved in several decision mechanisms available for ISM sampling. Although decision mechanisms 1–3 are cast in terms of comparison with action levels, the same considerations apply when using ISM data to develop exposure point concentrations for baseline risk assessments.

7.2.1 Decision Mechanism 1: Comparison of One ISM Sample from the DU to the Action Level

The simplest decision mechanism is the comparison of a single ISM sample result for a DU to an action level, which is typically a threshold value derived through risk assessment, regional background estimate, or other regulatory means. Sometimes, more than one set of action levels may apply to a site because they reflect different objectives (e.g., protection of acute and chronic health end points). Because ISM yields estimates of mean concentrations within a DU, it is important to note the spatial and/or temporal scale that was originally intended in the development of the action level.

This decision mechanism is simple and straightforward. The result of the decision is immediately apparent; a failure is indicated if the ISM sample result exceeds the action level. This approach is frequently used with individual discrete samples under the CERCLA preliminary assessment process; however, when the action level is intended to represent a mean concentration for a risk assessment exposure area, it is more logical to compare an estimate of the mean concentration (e.g., 95% UCL) in the DU from an ISM sample to the action level or, similarly, the mean (or 95% UCL) of multiple discrete sample results.

A single ISM sample provides one estimate of the mean concentration, which may be above or below the actual mean concentration in the DU. Unless replicate ISM samples are taken, there is no indication of the extent to which estimates of the mean vary, and consequently, it is difficult to predict how far from the actual mean a single ISM sample result might be. This uncertainty greatly limits the scientific defensibility of this approach. Use of a single ISM result might be acceptable when the estimated mean concentration obtained is much greater than, or much less than, the action level such that even a great deal of error in the mean estimate could be tolerated

Comparison of a **single ISM sample result** to an action level is useful when the ISM result is either well above or well below the action level and error in the estimate of the mean is unlikely to lead to an incorrect decision.

without making a decision error. In this situation, the ISM sample provides confirmation of what may have already been strongly suspected—that the DU clearly passes or fails. However, when the ISM sample result is close to the action level, this decision mechanism is unreliable, and decision errors in both directions are possible (i.e., concluding that the DU fails when the average concentration is in fact below the action level, or vice versa). How big a difference from an action level is big enough to conclude confidently that a DU passes or fails? The problem with this approach is that there is no clear answer. Because only one concentration is available, there is no indication of the potential magnitude of error, and a decision that a concentration difference is large enough that a decision error will not be made is arrived at subjectively. Obviously, uncertainty about making the right decision increases as the ISM sample result gets closer to the action level. Comments from many states suggest that the uncertainty associated with making decisions with only one ISM sample would make this approach unacceptable.

Decision Mechanism 1 example

A single, 5000 ft^2 DU is established across an area suspected to be a former transformer dump site. The objective of the investigation is to determine whether the mean concentration of PCBs in surface soil (0–4 inches bgs) exceeds an action level of 0.22 mg/kg for residential land use. Forty increments of soil are collected using systematic random sampling and combined into a single ISM sample for sample preparation and analysis. The reported concentration of PCBs in the sample is 6.2 mg/kg. The result is substantially higher than the action level, and the planning team comes to consensus that plausible error in estimating the mean would not likely change the ultimate decision that the DU fails.

7.2.2 Decision Mechanism 2: Comparison of the Mean of Replicate Data from the DU to the Action Level

Comparison of the **mean of replicate ISM sample results** to an action level is most useful when quantifying the uncertainty in the mean is not an important element to the decision.

In this decision mechanism, replicate ISM samples are collected in the field from the same DU. These replicates provide a measure of the variability of the entire sampling, preparation, and analytical process. The *mean concentration* of the replicates is calculated and compared to the action level. The mean concentration from replicate samples is likely to be closer to the true mean of the DU than the result from a single ISM sample (see Section 4) and could therefore be considered to provide a more reliable estimate of the mean. There is no assurance, however, that the actual mean concentration has not been underestimated. Consequently, this decision mechanism would not be useful in circumstances where project objectives dictate that estimates of mean concentrations must be conservative (e.g., EPC values in most USEPA human health risk assessments).

Decision Mechanism 2 example

This example is similar to the one provided for Decision Mechanism 1, except three replicate samples are collected from the DU. The reported concentrations of PCBs in the replicates are

0.12, 0.16, and 0.26 mg/kg. The mean concentration of PCBs based on the three replicate samples is 0.18 mg/kg. This does not exceed the action level for residential land use of 0.22 mg/kg, so it appears that no further action may be warranted. Note, however, how close the estimate of the mean is to the action level. This fact may have important implications for decision making. This same example data set is used again in the example for Decision Mechanism 3 to further illustrate these implications.

7.2.3 Decision Mechanism 3: Comparison of the 95% UCL on the Mean of Replicate Data from the DU to the Action Level

Project objectives may specify that the estimate of the mean concentration provided by ISM sampling must be health protective, meaning that there is low chance of underestimating the actual mean concentration within the DU. It is important to recognize that the likelihood of underestimating the mean from any sampling method (discrete, composite, or ISM) increases as the degree of heterogeneity increases. Traditionally, with discrete samples, the concern for underestimating the mean has been addressed by specifying an acceptable level of uncertainty (often 5%) and a method for calculating a conservative estimate of the mean (e.g., a 95% UCL). A similar approach can be used with ISM data as discussed below.

For those accustomed to working with 95% UCL values from discrete data sets, there are some important differences with 95% UCLs from ISM data. As discussed in Section 4, calculation of a 95% UCL for ISM data requires a minimum of three ISM samples. This is fewer than is required for discrete data sets to yield reliable 95% UCL values. Additional ISM replicates increase the performance of the mean estimate (i.e., provides a 95% UCL closer to the actual mean), and although this increases the cost, it may be worthwhile if the site is relatively heterogeneous and the result is anticipated to be close to the action level. A second difference involves what to do if the 95% UCL is higher than any of the individual values used in its calculation. With discrete data sets, the maximum concentration observed is often used as the EPC if it is less than the calculated 95% UCL. With ISM data, the calculated 95% UCL value should always be used as the EPC even if it is higher than any of the individual ISM results. This situation is not uncommon, particularly when the number of replicates is small. In fact, with three replicates, the UCL *always* exceeds the highest individual ISM result.

Two methods for calculating the 95% UCL from ISM data are available: Student's-*t* and Chebyshev. As discussed in Section 4, the choice of method depends on the known or anticipated shape of the probability distribution of contaminant concentrations in the DU. Note that software programs for calculation of 95% UCL values for discrete sample data (e.g., ProUCL) contain algorithms optimized to perform well for discrete data only. They are generally unsuitable for calculation of 95% UCL values for ISM data. A calculator for deriving 95% UCL values for ISM data is provided in Section 4.

Comparison of the **95% UCL on the mean of replicate ISM results** is most useful when the chance of underestimating the true mean must be minimized.

When replicate samples are taken over the entire DU, each is a true replicate and provides a separate estimate of the mean concentration. These estimates can be combined to derive a 95% UCL. Another approach is to divide the DU into SUs and take one ISM from each. The results from each ISM sample (i.e., each SU) can also be combined

to calculate a 95% UCL for the DU. With the latter approach, the ISM samples are not true replicates of the mean throughout the DU in the sense that they provide information on different portions of the DU. Collectively, however, they can provide an unbiased estimate of the mean. The principal disadvantage to this approach is that the UCL often exceeds the true mean by a larger degree than if replicates had been collected across the entire DU. The principal advantage of subdividing the DU for this decision mechanism is that it provides some information on the spatial distribution of contamination. If the DU as a whole fails the comparison with the action level, this spatial information could be valuable if a decision is made to break the DU into smaller DUs for reevaluation. (Note: The single ISM results from each SU would not be adequate to make confident decisions regarding them. Systematic planning would be needed to establish the smaller DUs and resampling would be required.)

Decision Mechanism 3 example

This is similar to the example for Decision Mechanism 2. The same three replicate samples are collected from the DU with reported concentrations of total PCBs of 0.12, 0.16, and 0.26 mg/kg. The 95% UCL of these results is 0.30 mg/kg with the Student's-*t* method and 0.36 mg/kg with the Chebyshev method, both of which exceed the action level for residential land use of 0.22 mg/kg. Therefore, while the sample arithmetic mean is less than the action level (as we saw in the previous example), there is sufficient variability in the results that there is a relatively high likelihood that the true mean exceeds the action level. Uncertainty in the shape of the underlying distribution does not factor into this result, since both 95% UCL methods yield the same conclusion. Options in this situation include deciding that the DU fails or taking more samples to reduce uncertainty and lower the 95% UCL, perhaps to a value below the action level.

7.2.4 Decision Mechanism 4: Comparison to Background

Background data from an appropriate reference area are used to evaluate site data for many environmental projects. With discrete sampling, comparisons between site and background data are generally done in one of two ways: point-by-point comparison of site data to an upper bound of background conditions (e.g., UTL) or distributional comparison using hypothesis tests to determine whether the differences in the central tendency (i.e., mean or median) or upper tails are statistically significant. USEPA guidance on hypothesis testing (e.g., USEPA 2002c, 2007, 2009) was developed with discrete sampling in mind and includes the following elements:

- Set the null hypothesis to state that the central tendency (e.g., mean) for the site distribution is statistically greater than that of the background distribution. This places the "burden of proof" on the data to show that site concentrations are not greater than background and is considered a more conservative (health-protective) approach.

- The use of nonparametric procedures such as the Wilcoxon rank sum (WRS) test relax the assumption of normality but not the assumption of equal variance. Therefore, it is possible that a test outcome is influenced more by differences in variance than by differences in central tendency, for example. For this reason, statistical tests should be accompanied by exploratory graphical analysis (e.g., histograms) to support the overall conclusion regarding background/site comparisons.

- Welch's test (also called Satterthwaite's-*t* or the unequal-variance *t*) is a modified Student's-*t* test that attempts to correct for unequal variances, though it still requires the assumption of normality. Simulations suggest that results are robust to moderate deviations from normality (i.e., moderate asymmetry).

- Both central tendency and upper tail tests should be evaluated to determine whether background and site concentrations are significantly different. A difference in either may suggest significant difference from background. The emphasis on the use of upper tail tests is that it is informative to understand whether subareas of the DU are elevated compared to background.

- Decision errors and determinations of statistical significance are closely tied to sample size and distribution shape, as well as the specified significance level (e.g., $\alpha = 0.01$, 0.05, 0.10, etc). When sample sizes are small for either data set, a formal statistical test may not be appropriate. For example, using WRS with $n = 4$ in both data sets and $\alpha = 0.01$, one can *never* identify a significant difference between two populations. This principle is true no matter what the sample concentrations are, even if all four site measurements are larger than background. WRS requires at least $n = 5$ in a group, or a higher (less-protective) level of significance (e.g., $\alpha = 0.05$ or 0.10).

As discussed in Section 4.4.3.3, ISM results are not suitable for point-by-point comparison with UTLs generated from discrete sample background data because ISM and discrete data sets have fundamentally different characteristics. If background and site data are both generated using ISM, comparisons of central tendencies (e.g., medians) can be made using hypothesis testing, but statistical power to detect differences will be low due to the limited number of replicates in most ISM data sets. Similarly, at least $N = 8$ observations per group is desired before using hypothesis tests to compare upper tails (e.g., quantile test). Nonetheless, hypothesis tests are not the only tool available to determine whether there are important differences between site and background distributions. Simple graphical analysis can provide useful information and serve as a semiquantitative means of comparison.

Decision Mechanism 4 example

Continuing with the example presented in Decision Mechanisms 1–3, five replicate samples are collected from a reference area unimpacted by site contamination for comparison with the site data. The reported concentrations of benzo(a)pyrene from the reference area ISM samples are 0.05, 0.10, 0.12, 0.20, and 0.40 mg/kg. The sample mean and SD of the reference area samples are 0.17 and 0.14 mg/kg, respectively. By comparison, the site sample ISM replicate results are 0.12, 0.16, and 0.26, and the mean and SD are 0.18 and 0.07 mg/kg, respectively. Therefore, the sample means are almost the same, but the SD is greater in the reference area by a factor of 2.

Figure 7-1 provides a graphical comparison of the two ISM data sets using side-by-side dot plots. For context, the action level for benzo(a)pyrene of 0.21 mg/kg is also shown on Figure 7-1. Presenting the information this way, it is clear those concentrations in the reference area exhibit greater variability and that the difference may be partly explained by the difference in sample sizes. If more ISM replicates had been collected at the site, then perhaps more extreme high and low concentrations would have also been observed.

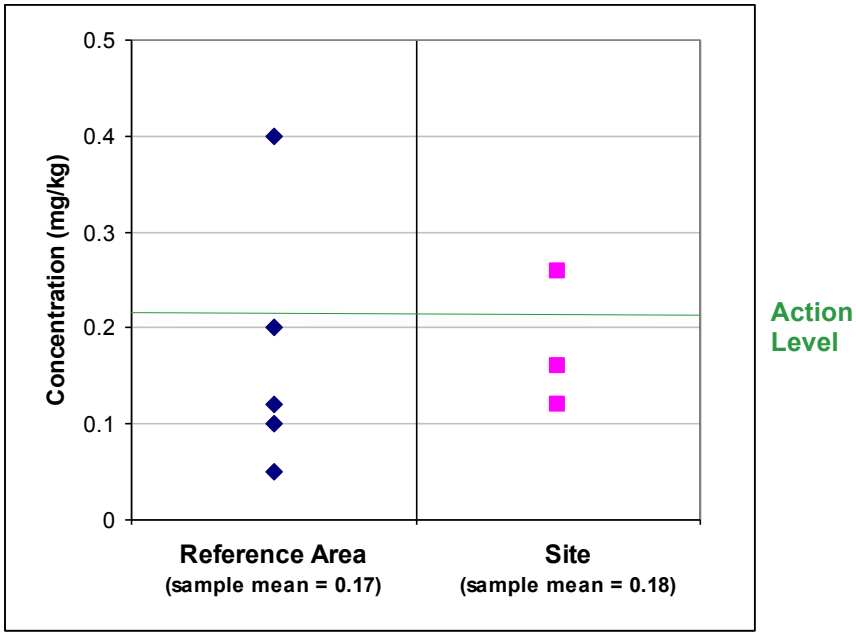

Figure 7-1. Dot plot comparison of background (reference area) and site ISM results.

Since the sample sizes are too small to evaluate a GOF to a normal distribution, a secondary line of evidence may be provided by hypothesis testing (noting the limitations in applying these tests as discussed above). For purposes of this example, a nonparametric WRS test ($\alpha = 0.05$) was applied. Using a one-sided null hypothesis specified as site median less than or equal to the background median, one would not reject the null hypothesis ($p = 0.33$) and, therefore, conclude that the distribution on site is comparable to background. By contrast, using a one-sided null hypothesis specified as site median greater than or equal to the background median, which is consistent with USEPA guidance (e.g., USEPA 2002c, 2007, 2010a), one would again fail to reject the null ($p = 0.77$) and, therefore, conclude that the distribution on site is elevated with respect to background. The result is contingent on the form of the hypothesis test that is selected. Since the latter hypothesis puts the burden of proof on the data to demonstrate that the distributions are comparable, the small sample sizes from ISM data sets very often yield a conclusion that site > background even when the ranges overlap as shown in this example. Therefore, statistical significance should be interpreted with caution.

7.2.5 Decision Mechanism 5: Combining DUs

There are circumstances when it may be advantageous to combine results from two or more DUs into a larger DU. This situation might occur when there are multiple sampling objectives for a given area. For example, delineation of source areas might necessitate creation of several small DUs, while evaluation of risk from exposure is based upon a DU that encompasses two or more of these DUs. DUs can be constructed in such a way as to meet both objectives efficiently if results from smaller DUs can be combined to produce an estimate of the mean for a larger, "super" DU. In constructing the "super" DU, each of the smaller, component DUs is in a sense like a SU. However, all are true DUs in that a decision must be reached for each, based upon one site objective or another.

Another example is a situation in which sampling objectives require assessment of exposure of different receptors or scenarios such that differently sized, superimposed exposure areas must be evaluated. Here again, the ability to combine results from small DUs to estimate mean concentrations for larger DUs would be advantageous.

Combining results from two or more small DUs to estimate the overall mean concentration in a larger combined DU is advantageous when the data must support more than one decision (e.g., overlapping exposure units for ecological and human health receptors).

Operationally, the mechanism requires a stratified sampling plan. The overall mean of the larger DU can be calculated using replicate data from the smaller, component DUs using formulas described in Section 4. These formulas take into account the size of the smaller DUs, weighting their contribution to the larger DU accordingly. The ability to combine DUs extends vertically as well as horizontally; that is, results from DUs from different soil depths can be combined if needed to meet sampling objectives.

Decision Mechanism 5 example

An elementary school is divided into three DUs based on anticipated exposure of students and maintenance workers to soil. The kindergarten children have their own playground that is designated as DU1. The older children have another playground that is designated as DU2. School maintenance workers come in contact with soil from both DUs equally, and their area of exposure is DU3, which consists of DU1 + DU2. DU1 and DU2 are each sampled using systematic random sampling with a total of three ISM samples from each. The results from the six ISM samples are combined, with appropriate weighting as described in Section 4, to derive the average concentration for DU3. The weighting factors applied to each DU result should reflect the assumptions in the CSM.

7.2.6 Decision Mechanism 6: Extrapolating from Sampled to Unsampled Areas

This decision mechanism entails using estimates of the mean obtained from areas where ISM samples are taken to make decisions regarding other DUs that are unsampled. The fundamental assumption made with this mechanism is that the distributions of contaminant concentrations in the unsampled areas are essentially the same as in the sampled areas. The most common rationale for this assumption is that the source of contamination, mechanism(s) of transport, etc. are similar for each of the areas and that these conditions should lead to similar levels of contamination and similar variances. This decision mechanism is typically considered when large tracts of land or large volumes of soil must be assessed with a limited budget.

Extrapolation from a sampled area to an unsampled area requires an assumption that the distributions of contamination in the unsampled areas are sufficiently similar to the sampled areas.

The key to this decision mechanism is confidence that the fundamental assumptions are valid and that there are no significant differences in contaminant distribution among the sampled and unsampled areas. In the absence of data to verify the assumption, that confidence is subjective. There is nothing unique about ISM that enables this extrapolation with reduced uncertainty—the same issue of whether or not to extrapolate exists whether the sampled areas are evaluated with

ISM or discrete samples. Based on feedback obtained in development of this report, this decision mechanism is not acceptable for many states.

A distinction should be made between extrapolation between DUs and extrapolation within a DU. It is sometimes suggested that because there is precedence for using results from discrete samples to make inferences about unsampled areas within a DU, the same uncertainty applies to ISM. In this context, there is a difference between how information from discrete and ISM data may be used. With discrete data, spatial interpolation methods (e.g., geostatistics, inverse distance weighting) or discretization methods (e.g., Thiessen polygons) can be used to provide more reliable estimates of the mean and standard deviation throughout the DU. These methods also have the advantage of using information across DUs (i.e., when a site is split into multiple DUs) to derive estimates of the mean and standard deviation within each individual DU. With ISM, this degree of spatial resolution is lost because the increments are composited, so there is no basis for estimating concentrations in subareas of the DU or for developing a mathematical model that uses data from across the DUs. One exception would be for a site that is divided into many DUs—if a sufficient sample size is available, each estimate of the mean may be considered representative of a portion of the site such that spatial patterns and interpolation method may be explored.

A variation on this approach is to collect replicates in subset of the DUs and extrapolate the estimate of the variance (or the CV) to DUs with a single ISM sample. Although this approach appears to be a less uncertain way to extrapolate findings among DUs, the extent to which the distributions may be comparable across DUs must be considered. The chance that the distributions differ among DUs increases as the number of sources and the complexity of the contaminant transport mechanisms increase. In addition, sites with multiple subareas of elevated concentrations can be expected to introduce inherent variability within and between DUs, making a successful extrapolation of the variance more difficult. In general, the greater the number of DUs where replicate ISM samples are collected, the more likely that the average measure of variance will be representative of DUs with single ISM results (see Section 4.2).

As noted in Section 4, it is unclear whether the appropriate statistic for extrapolation is the SD or CV. The CV is preferred if it can be reasonably assumed or demonstrated that there is a positive correlation between the mean and SD. Based on the proportionality effect, the mean and SD are expected to be positively correlated for positively skewed distributions (Goovaerts 1997). If replicate data are available for multiple DUs, plots of the SD vs. the mean should be developed to explore patterns in the relationship between the sample statistics.

A related situation exists when a DU is subdivided into SUs and only a fraction of the SUs are actually sampled. In this approach, the results from each of the sampled SUs are compared with the action level(s). If all are lower than the action level(s), the entire DU passes. The same assumptions and considerations discussed above apply in this situation as well. If one or more SUs are above the action level, the DU does not pass, and the systematic planning team should be reconvened to plan the next steps, which may include additional sampling.

7.2.7 Decision Mechanism 7: Evaluating Oversized DUs

An oversized DU is one that is larger than can be justified based upon site objectives but is evaluated nevertheless because of practical considerations. While oversizing DUs is strongly discouraged, there are some situations where it is unavoidable. Examples include DUs that are larger than the home range of some of the species of interest in an ecological risk assessment or are larger than the exposure area for some of the receptors/risks of interest in a human health risk assessment, such as when acute exposure to soil is a concern. In this situation, the DU evaluated actually consists of a few to perhaps thousands of smaller, ideally but impractically sized DUs. The problem faced is determining what information the sampled DU can provide concerning concentrations in the smaller sub-DUs.

There are no optimal answers to solve this dilemma, unfortunately, because typical ISM sampling designs are devoid of spatial information on contaminant distribution within the DU. This is not a new problem, as it has been documented in the literature for composite sampling and there are a number of approaches for estimating high-end concentrations within a sampled area. The simplest of these is to multiply the mean value from the composite (or ISM sample) by the number of increments. This approach represents the situation in which all of the contaminant is present in one of the increments. Given the number of increments in a standard ISM design, this approach is *extraordinarily conservative* and can yield quite high values. Given the conservative nature of this method, it is useful only to support "no further action" decisions or decisions to characterize the area further. Other approaches that are less conservative include multiplying the average concentration by the square root of the number of increments or more complicated formulas (for an example, see Barnett and Bown 2002).

Note: A computationally equivalent approach is to use the average concentration but divide the soil *action level* by the number of increments.

7.3 Assessment of Error

It is desirable to seek quantitative information on the potential magnitude of error in ISM data when using those data to make decisions. In all environmental sampling, two basic types of error are produced:

- error associated with the collection of sample(s) in the field
- error associated with the processing and analysis of those sample(s)

The ISM approach, which includes both field and laboratory steps, is intended to minimize the potential error and produce more technically defensible data by specifying the targeted volume and the parameter to be estimated and collecting, subsampling, and processing the sample(s) in accordance with the recommendations of sampling theory. An important component of this process is, to the extent possible, to assess the errors generated in the sampling and analysis scheme from beginning to end (i.e., from collection of soil in the field through the production of an analytical result).

One means of evaluating ISM data is through comparison of the results of replicates, both as taken in the field and in the laboratory. As discussed in Section 3.6, field replicates consist of separate ISM samples taken by the field team from the same area (SU or DU). They are not field

splits—they are collected and processed as separate samples. Laboratory replicates are samples taken from a single ISM sample, usually in the laboratory. They can be taken from the bulk ISM sample at a number of points during sample processing, depending on the process step(s) being evaluated. Replicates taken at the beginning of laboratory processing of the bulk ISM sample are used to evaluate potential overall error resulting from laboratory processing and analysis.

Replicate ISM samples collected from each DU in the field provide a measure of total sampling and analysis error.

Use and interpretation of replicate data depends in part on the decision mechanism being applied. For example, field replicate data allow calculation of 95% UCL values needed for Decision Mechanism 3 and allow statistical comparison of site and background results using hypothesis testing in Decision Mechanism 4. Decision Mechanisms 5–7 can also rely on 95% UCL values calculated from field replicates.

Three or more ISM samples are needed to calculate a defensible 95% UCL.

Replicate data can also be used to calculate an RSD, which is used to evaluate the precision of the data. RSD is a measure of reproducibility of estimates of the mean provided by replicates. Just as the sample mean and standard deviation are estimates of the corresponding population parameters, the sample RSD is an estimate of the ratio of the population parameters. It provides a measure of the total error associated with the data, although not necessarily the accuracy of the estimate. To calculate appropriate statistics, at least three field replicate samples are needed. Ideally, the project team then designates one of these replicates for separation into laboratory replicates. Replicate RSD data are intended to quantify the total error of the measurement system and attribute that error to either field sampling or laboratory procedures.

The precision of ISM data can be quantified from replicate ISM sample results.

The total error is estimated based on the field replicate RSDs. Laboratory error can also be estimated based on the laboratory replicate RSDs. The field sampling component of error can then be estimated by subtracting the laboratory error from the total error. Therefore, the collection of field and laboratory replicates allows the error to be attributed to either the laboratory or the field sampling processes.

High RSD values for the laboratory component indicate potential problems with laboratory subsampling of the bulk ISM sample or other sources of analytical error. In this situation, the source(s) of laboratory error should be investigated and resolved.

High RSD values for the field component can have different implications depending on the decision mechanism being applied. For example, a high RSD (e.g., exceeding 30%–35%) from field replicates, but with acceptable RSDs from laboratory replicates, strongly suggests a substantial degree of heterogeneity in the DU contaminant concentrations. For Decision Mechanism 2, where a simple average of the replicates is used to derive the average concentration, this situation represents a problem. It means that estimates provided by the individual ISM replicates are quite variable and that the estimate of the average for the DU they provide may be unreliable. If the results are close enough to an action level that decision errors are possible, resampling with an increased number of increments may be used to reduce error.

For Decision Mechanism 3, potential error created by heterogeneous concentrations is handled through calculation of the 95% UCL. Simulation studies discussed in Section 4 show, that with appropriate choice of 95% UCL method, conservative estimates of the mean to satisfy sampling objectives for this decision mechanism can be obtained despite high RSD values. This principle applies as well to other decision mechanisms where a 95% UCL is calculated.

A low RSD is not an indication that the mean is accurate or that the 95% UCL exceeds the population mean unless the distribution can be reasonably assumed to be relatively homogeneous.

A low RSD indicates that the field replicates are providing reproducible estimates of the average and generally triggers no additional steps to refine the estimate. However, it must be recognized that RSD is a measure of precision, not accuracy (see Section 4 for addition discussion of these concepts). Thus, an estimate of the average from replicates with a low RSD is not necessarily close to the actual mean. The opportunity for significant error is greatest when the DU is relatively heterogeneous and the replicates by chance give similar results. Unless information on heterogeneity of contaminants within the DU is available, it is difficult to judge whether this situation may have occurred and consequently the degree to which a low RSD should be reassuring. This is certainly an issue for the simple average of replicate data in Decision Mechanism 2. It is also an issue for Decision Mechanism 3 and others where a 95% UCL is calculated. Simulation studies discussed in Section 4 have shown that the UCL does not always ensure that a conservative estimate of the mean is obtained when the RSD is low. That is, when the RSD is low, the mean can be underestimated even by a 95% UCL. In short, a low RSD from field replicates offers information on the reliability of the estimate of average only when the contaminant distribution within the DU is known, or can be confidently assumed, to be relatively homogeneous.

For Decision Mechanism 6, replicates are often collected from a fixed percentage of DUs; however, the selection and number of DUs from which field and laboratory replicates are collected is not a simple matter—there is no one size fits all approach. Therefore, the number of DUs from which replicates are collected must be determined using site-specific considerations. Simply relying on a fixed percentage and arbitrary decisions to select which DUs will have replicates is ill advised.

If budgetary considerations limit the number of samples, field and laboratory replicates should be collected from those DUs that will provide the most useful information. Knowledge of source areas and areas likely to have high or low concentrations should be used to make deliberate choices. If there is a choice between a DU with anticipated high concentrations (i.e., above the action level) vs. one with low concentrations (i.e., close to the action level), the DU with concentrations closest to the action level should be selected for replicate samples. The closer contaminant concentration gets to the action level, the more important replicate statistics are in making a decision. Detection limit may also be a consideration in some situations. DUs with detectable concentrations provide more information than DUs were concentrations cannot be measured.

It is advisable to collect field and laboratory replicates from DUs that are believed to have different characteristics in terms of contaminant distribution, contaminant concentration, sampling design, or sample matrix. When less than 100% of DUs have replicate samples, the RSD (same as CV) from one or more DUs can be applied to similar DUs, subject to the

limitations described for Decision Mechanism 6 above. If different sources of contamination or different release mechanisms are identified, field and laboratory replicates should be collected from each different DU. Furthermore, other factors that may influence the number of DUs with replicates are significantly different soil types that could cause different contaminant distributions and/or sample preparation efficiencies and different numbers of increments in ISM samples.