Advanced Design Application & Data Analysis for Field-Portable XRF

A Series of Web-based Seminars Sponsored by
Superfund's Technology & Field Services Division

Session 7

Q&A for Session 6
Module 6.2 – Dynamic Work Strategies Part 2

6.2-1

How To . . .

- Ask questions
  - » "?" button on CLU-IN page
- Control slides as presentation proceeds
  - » manually advance slides
- Review archived sessions
  - » http://www.clu-in.org/live/archive.cfm
- Contact instructors

6.2-2

Notes

- When you registered, you were directed to this seminar's specific URL, which is the front page of today's seminar. The Front Page of the web cast contains a short abstract of today's session. We have also included pictures and short biosketches of the presenters. Please note the presenters' email addresses are hotlinked on that page in case you have any questions for one of them after today's presentation.
- For those of you joining us via the phone lines, we request that you put your phone on mute for the seminar. We will have Q&A sessions at which point you are welcome to take your phone off mute and ask the question. If you do not have a mute button on your phone, we ask that you take a moment RIGHT NOW to hit *6 to place your phone on MUTE. When we get to the question and answer periods you can hit #6 to unmute the phone. This will greatly reduce the background noises that can disrupt the quality of the audio transmission.
- Also, please do not put us on HOLD. Many organizations have hold music or advertisements that can be very disruptive to the call. Again, keep us on MUTE. DO NOT put us on HOLD.
- Also, if you experience technical difficulties with the audio stream, you may use the ? icon to alert us to the technical difficulties you are encountering. Please include a telephone number where you can be reached and we will try to help you troubleshoot your problem.

- Instructor contact information:

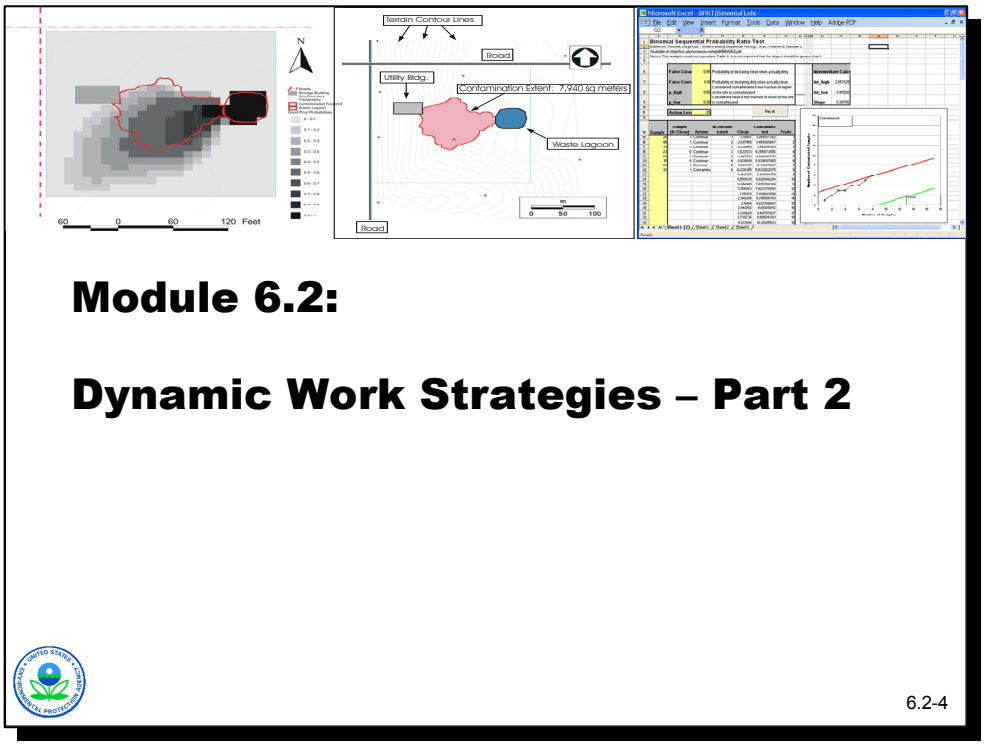
Deana Crumbling, U.S. EPA
Phone: (703) 603-0643
Fax: (703) 603-9135
E-mail: crumbling.deana@epa.gov

Robert Johnson, Argonne National Laboratory
Phone: (630) 252-7004
Fax: (630) 252-3611
E-mail: rlj@anl.gov

Stephen Dyment, U.S. EPA
Phone: (703) 603-9903
Fax: (703) 603-9135
E-mail: dyment.stephen@epa.gov

Q&A For Session 6 – Dynamic Work Strategies Part 1

6.2-3



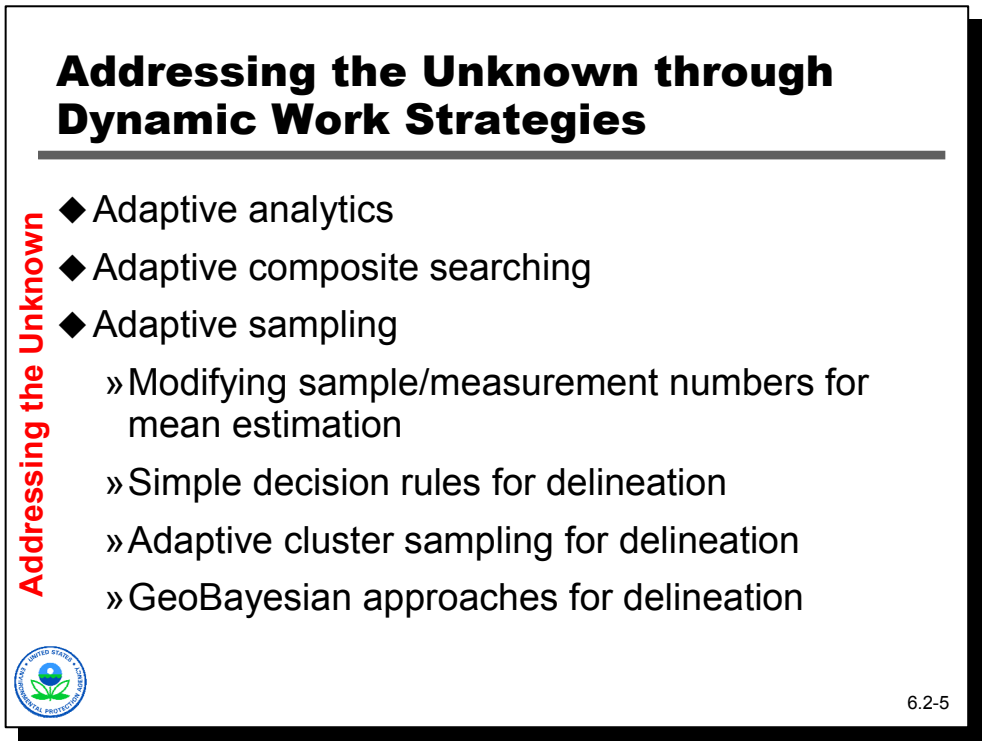
Module 6.2:

Dynamic Work Strategies – Part 2

6.2-4

Notes

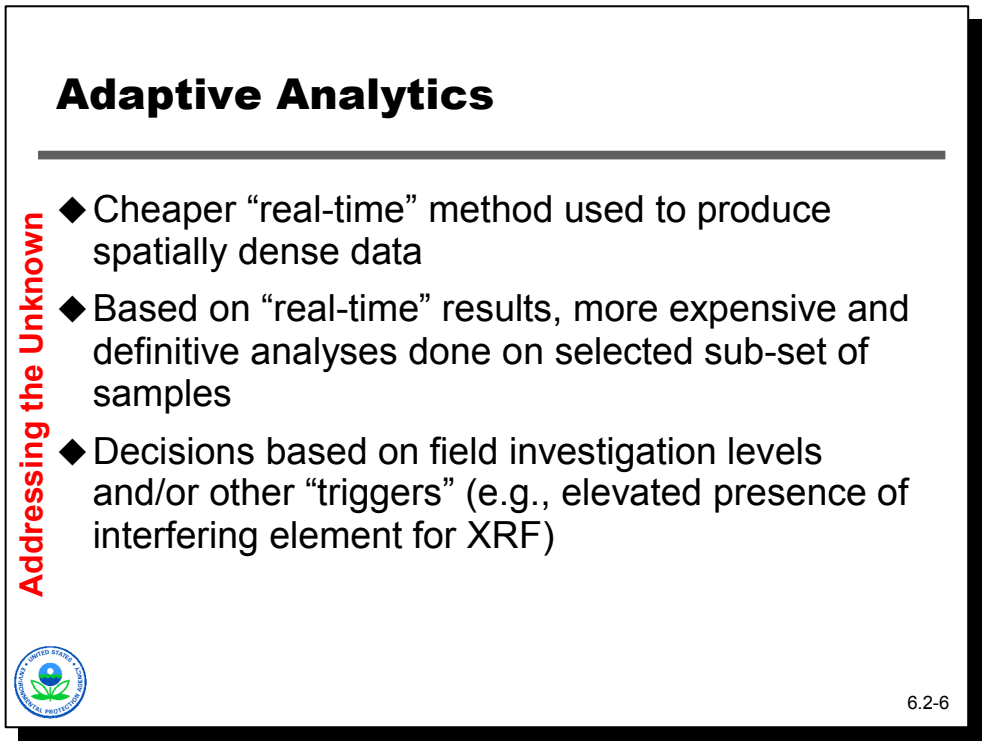
- This module is a continuation of the discussion of dynamic work strategies.



Notes

Addressing the unknown through dynamic work strategies: This module covers the following topics:

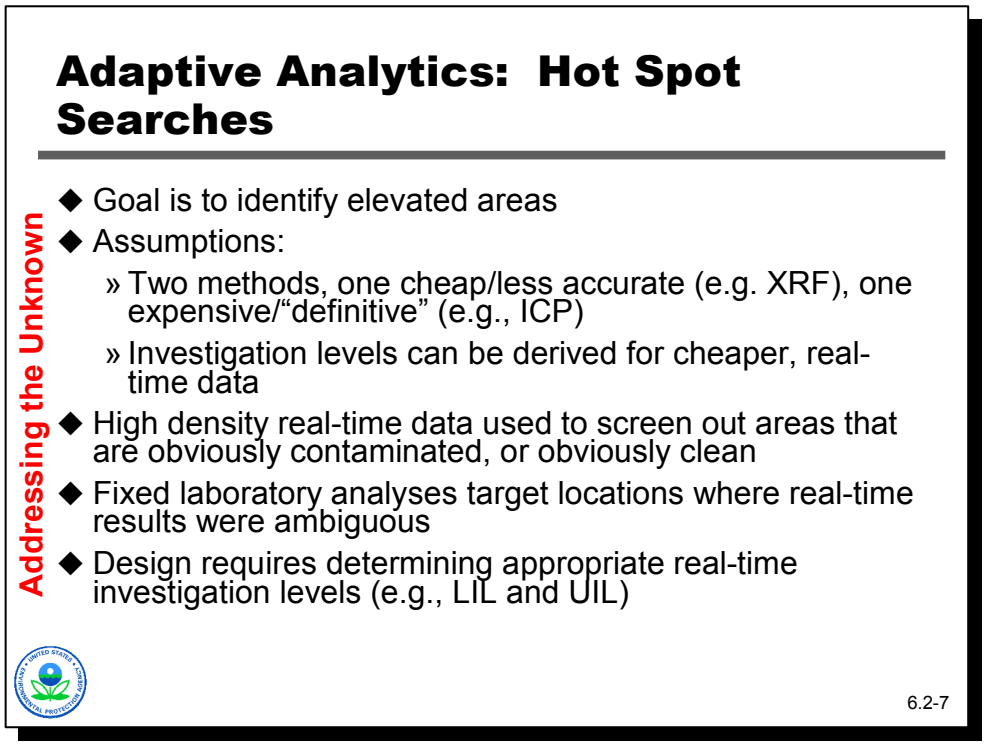
- **Adaptive analytics**
- **Adaptive composite searching**
- **Adaptive sampling**
  - » Modifying sample/measurement numbers for mean estimation
  - » Simple decision rules for delineation
  - » Adaptive cluster sampling for delineation
  - » GeoBayesian approaches for delineation



Notes

At the end of the last module we discussed the use of field investigation levels (lower investigation levels and upper investigation levels) to help with real-time data decision-making.

- **Cheaper "real-time" method used to produce spatially dense data:** Adaptive analytics uses less expensive "real-time" methods (such as the XRF) to produce data that provides dense coverage of the decision unit.
- **Based on "real-time" results, more expensive and definitive analyses are done on selected subset of samples:** The real-time data results are reviewed to guide the selection of a subset of samples for analysis by a laboratory method. The laboratory results are used to investigate real-time results that are of particular concern.
- **Decisions based on field investigation levels:** The decision as to whether to send a sample off for confirmatory laboratory analysis is driven by field investigation levels that are applied to the real-time results. These investigation levels guide decision-making.



Notes

Making use of adaptive analytics is one approach for combining collaborative data sets and dynamic data collection strategies. In this case the goal would be to identify elevated areas or delineate contamination. Sampling locations are fixed. The dynamic dimension of this type of program stems from the ability to select from different analytical techniques as work progresses.

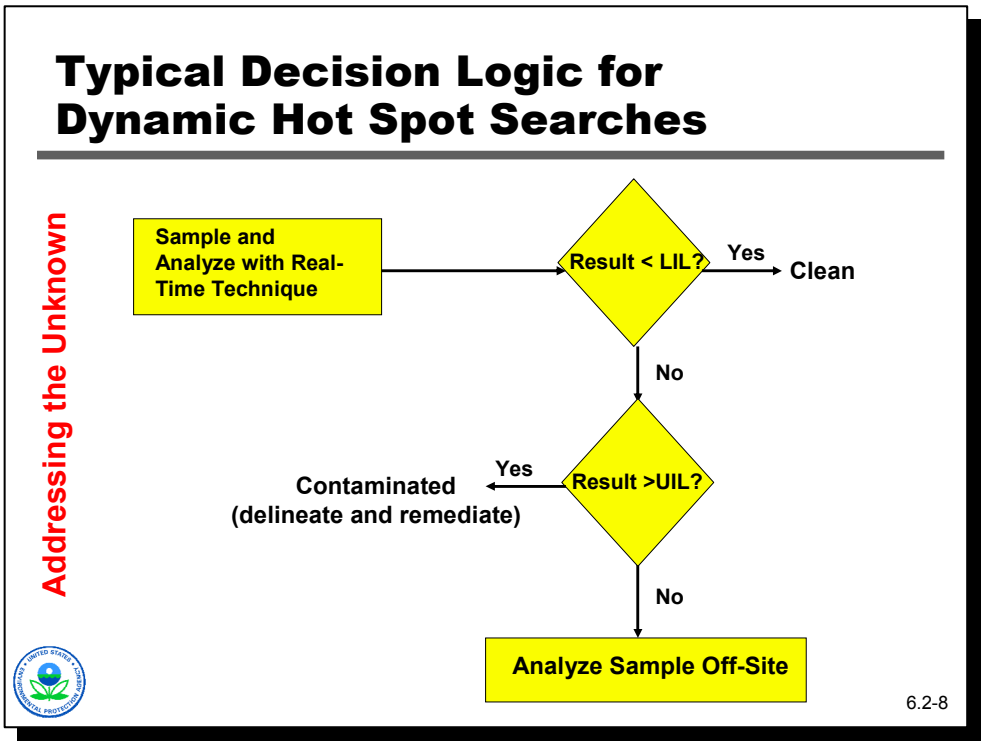
- **Goal is to identify elevated areas:** The goal of a hot spot search is to identify those areas of a site or decision unit that contain elevated levels of contamination that are significantly higher in concentration than other areas of the site. Areas with higher levels of contamination pose greater risks to human health and the environment and may need to be treated differently than other areas.

- **Assumptions:** The adaptive analytics approach assumes there are two methods available, one real-time method such as an XRF that provides data at a low cost but that is not highly accurate and another method such as ICP that is expensive but provides accurate data. This approach also assumes that investigation levels (lower investigation level and upper investigation level) can be derived for the less expensive real-time method.

- **High density real-time data used to screen out areas that are obviously contaminated, or obviously clean:** The real-time method is used to take many measurements (typically systematically) across the decision unit, creating a dense picture of contamination levels. The data is first used to screen out areas that are obviously contaminated, or obviously clean.

- **Fixed laboratory analyses target locations where real-time results were ambiguous:** The real-time data that is ambiguous, or that is between the established LIL and UIL, are targeted for fixed laboratory analyses.

- **Design requires determining appropriate real-time investigation levels (e.g., LIL and UIL):** This approach requires that the LIL and UIL for the real-time measurement levels be established. The LIL and the UIL define the obviously clean and obviously dirty areas and the ambiguous areas. Recall that the last module provided an example of how a LIL or UIL might be selected.



Notes

- This flow chart shows the decision logic for dynamic hot spot searches. Many samples are analyzed using a real-time method systematically across an area. Results that are less than the LIL indicate an area is clean of hot spot concerns. Results that are greater than the UIL indicate an area is contaminated at "hot spot" levels. Results in between the LIL and UIL are ambiguous and samples are sent off-site for laboratory analyses.

An Example…

Addressing the Unknown

- Looking for contaminated sediment layer
- Uranium used as a proxy for primary COC (which is not measurable by real-time technique)
- UIL and LIL derived for XRF and uranium
- Every six inch interval of 3 ft cores screened by XRF
  - » If all intervals < LIL, core considered clean
  - » If one or more intervals > UIL, core considered impacted
  - » If one or more intervals between LIL and UIL (but none above UIL), highest interval sent for lab analysis

6.2-9

Notes

- Here is an example of this logic at work. This site has contaminated sediment concerns. The contaminated sediment layer, when it is present, can exist at varying depths (i.e., close to the surface or at the surface in some areas, but buried deeper in other areas). The purpose of the GeoProbe work was to identify areas where contaminated sediments were a concern.
- The primary COC was a contaminant that was not amenable to real-time techniques. Fortunately, however, historical data indicated it was collocated with elevated uranium. Uranium is something that can be easily measured by XRF.
- A UIL and LIL were derived for the XRF and uranium based on a review of historical data. The LIL was selected so that if XRF uranium results were below that value, there was little chance the primary COC was present at levels of concern. The UIL was selected so that if XRF uranium results were above that value, there was a high probability that the primary COC was present at levels of concern.
- GeoProbe cores were systematically placed across the area of interest, with coring done down to a depth of 3 feet. Each six inch interval of the each core was screened by XRF. If all of the XRF uranium results were below the LIL, the conclusion was that there were no risk concerns at that location. If at least one XRF uranium result was above the UIL, then the assumption was that the contaminated sediment layer was present. If one or more XRF uranium results were above the LIL, but none were above the UIL, the core interval with the highest XRF uranium reading was selected and sent off-site for laboratory analysis.

When are Adaptive Analytics Cost-Effective for Searches?

Addressing the Unknown

- The "smaller" the "unclear" zone, the better the performance
- The greater the difference is between background and the action level, the better the performance
- The greater the difference between the action level and average contamination concentration, the better the performance
- Best case: no follow-up fixed laboratory sampling required beyond QA/QC needs
- Worst case: every sample requires follow-up laboratory analysis

6.2-10

Notes

- **The "smaller" the "unclear" zone, the better the performance:** Adaptive analytics will be more cost effective when the difference between the LIL and the UIL is small. Fewer samples results will be between the LIL and the UIL that require subsequent laboratory analysis.
- **The greater the difference is between background and the action level, the better the performance:** Adaptive analytics will be more cost effective when the action level is much greater than the background level because more sample results will be below the LIL and clearly defined as "clean."
- **The greater the difference between the action level and average contamination concentration, the better the performance:** Adaptive analytics will be more cost effective if there is a large difference between the action level and the average concentration present because it will be less likely that the real-time method will yield a result between the LIL and UIL.
- **Best case:** In the best case, the real-time technique can be relied upon without additional follow-up using fixed laboratory sampling except for that required for quality assurance/quality control.
- **Worst case:** In the worst case, the real-time technique yields useless data and every sample requires follow-up laboratory analysis.

Cost/Benefit Calculation

Addressing the Unknown

- The break-even cost analysis for adaptive analytics data collection is:

$$C_{rt}/C_f < (N_{rt} - N_f)/N_{rt}$$

where

  - » C_{rt} = cost of real-time,
  - » C_f = cost of laboratory analysis,
  - » N_{rt} is the # of real-time analyses, and
  - » N_f is the expected number of confirmatory laboratory analyses

6.2-11

Notes

- The equation provided in the slide can be used to determine whether the use of adaptive analytics is justified from a cost perspective. The cost analysis is straightforward…if the ratio of per sample real-time costs to laboratory costs is less than the fraction of the total samples that is expected to need off-site laboratory analysis, then there will be a cost savings. Of course there may be other reasons to use a real-time technique apart from simple cost savings…for example to quickly identify problem areas so that they can be delineated, something more difficult to do with a sampling strategy based solely on off-site laboratory analysis.

Adaptive Analytics: Mean Estimation

Addressing the Unknown

- Goal: Identify areas of concern and estimate their mean concentration
- Assumptions:
  - » Two methods, one cheap/less accurate (e.g., XRF), one expensive/"definitive" (e.g., ICP)
  - » Non-parametric relationship available between cheaper, real-time data and fixed lab data
- Cheaper, lower quality analytical data identifies areas of concern…data used to estimate number of more expensive analyses required
- More expensive, higher analytical quality data used to estimate average concentrations

6.2-12

Notes

- **Goal:** The idea is that lots of cheap, lower analytical quality data can be used to identify areas of concern, and then limited sampling with more expensive, higher analytical quality data can provide definitive information about those areas (e.g., estimate average contamination concentrations for an area).
- **Assumptions:** This approach assumes there are two methods, one of which is inexpensive and not highly accurate (e.g. XRF) and another which is expensive but accurate (e.g., ICP). The only requirement for the cheaper technique is that it has sufficient detection capabilities to confidently identify areas or situations that would be of concern.
- **Cheaper, lower quality analytical data identifies areas of concern . . . data used to estimate number of more expensive analyses required:** From a dynamic work plan perspective, the results from cheaper, "real-time" methods can be used to determine which areas requires more definitive sampling, and how many samples should be used.
- **More expensive, higher analytical quality data used to estimate average concentrations:** The off-site laboratory analyses are used to estimate the average concentrations within the decision unit.

An Example....

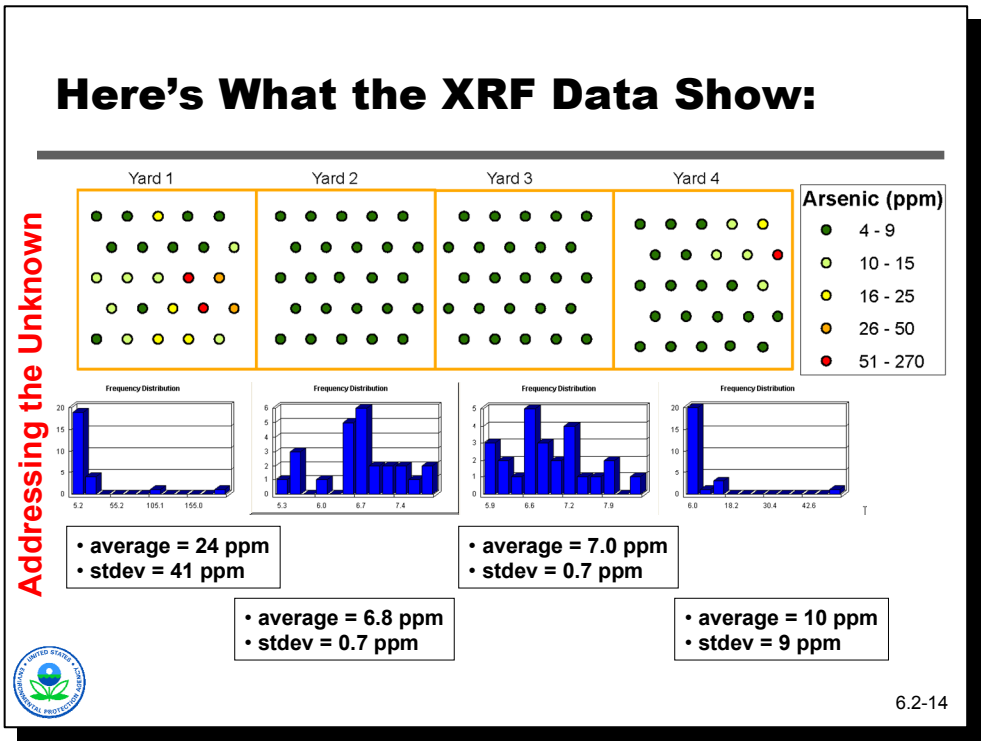
Addressing the Unknown

- 4 residential backyards screened by XRF for arsenic w/ action level of 25 ppm averaged over yard
- Regulator insists final release decision be based on lab data
- Use XRF to determine:
  - whether each yard is likely above or below action level, and
  - if below, how many lab samples are required to statistically show it?

6.2-13

Notes

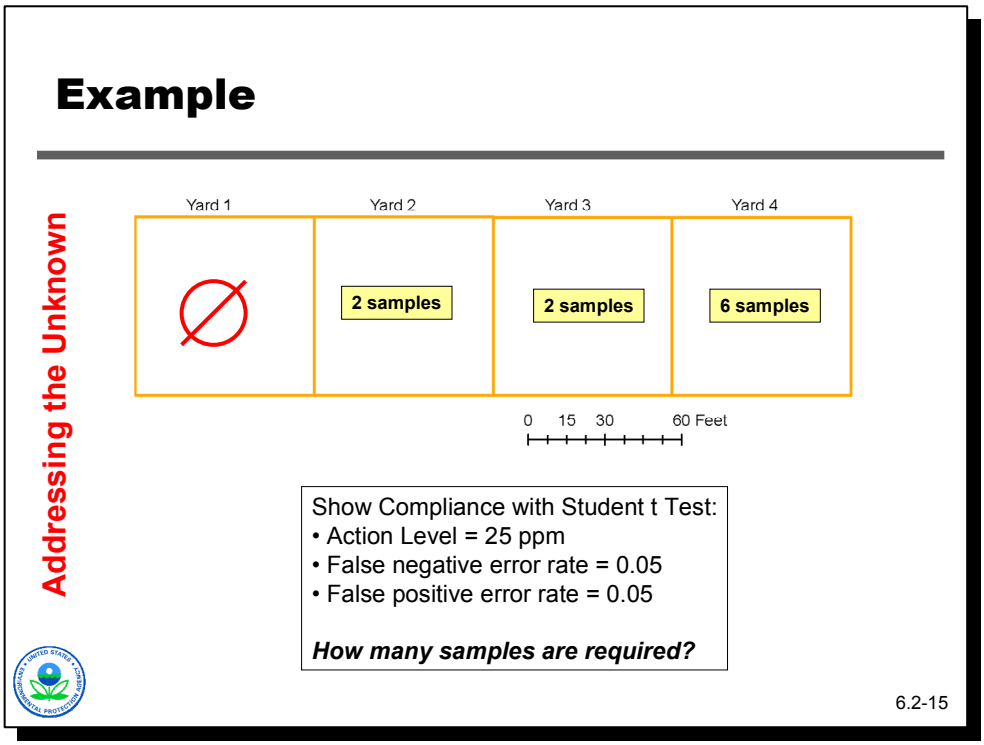
- Here is an example of mean estimation. In the case of this project, XRF data were deemed by the regulator involved as not sufficient to establish that individual yards met arsenic action levels for release purposes. However, in situ XRF data could be used to quickly get a sense for whether a yard was a candidate for closure (or conversely was going to require remediation), and if it was ready for closure to identify how many “definitive” laboratory samples would be required to statistically establish that the 25 ppm requirement had been met.



Notes

- The map shows the four yards and the results of 25 in situ XRF readings systematically spread across each yard, color coded by arsenic values (4 – 9 is definitely background). Below each yard is a histogram of those 25 results along with the observed average and standard deviation (note that none of these, including the background yards, look particularly "normally" distributed).

- As should be clear from these XRF data, two yards are not impacted, while the other two appear to be impacted at varying levels. In the case of the yard furtherest to the left, the average is very close to the 25 ppm standard, indicating it probably is not a candidate for release. In the case of the yard furtherest to the right, the average is still well below 25 ppm despite the impacts, suggesting that with enough laboratory samples it may be able to be released.



Notes

- The XRF data allow estimation of the average arsenic concentration in each yard and the variability in arsenic concentrations that is present. That information, in turn, allows for customization of the number of discrete samples sent off to the laboratory for each yard to demonstrate compliance with the cleanup criteria, assuming a Student t test would be used to make that determination. Note that the yard to the left is not a candidate for closure…the XRF data suggests it would be futile to try release this yard. Note too that the number of samples required varies significantly from yard to yard.

Adaptive Analytics Also Pertinent to QC

Addressing the Unknown

- Heavy lifting for decision-making done using real-time techniques (i.e., XRF)
- Adaptive analytics used to target real-time technique QC
- Criteria developed that flag real-time samples as candidates for off-site lab analysis in support of QC

6.2-16

Notes

- These same concepts also apply to the use of adaptive analytics for QC purposes. Typically when collaborative data sets are used, a fixed percentage (e.g., 5% or 10%) of samples are sent off-site for more definitive laboratory analyses. Often these samples are either randomly identified, or are sent at specific intervals (e.g., after every 10 or 20 samples collected). The problem with this is that catching analytical problems this way is a real hit-or-miss affair.

  By understanding the ways that a real-time method such as an XRF might go "bad" (e.g., particular soil matrices, presence of other contaminants that interfere, extremely high or low results, etc.), then the project team can be much smarter in designing criteria that flag real-time samples as candidates for off-site laboratory analysis, and stand a much better chance of catching and correcting "problems" before they jeopardize the outcome of the field effort.

Example: Arsenic and Lead

Addressing the Unknown

- Site has known arsenic problems with arsenic decision-making driver
- Data collection primarily based on field-deployed XRF targeting arsenic
- Elevated lead also present, potentially at levels interfering with XRF arsenic readings
- Decision rule: if XRF lead > 10x arsenic result, require ICP analysis on sample

6.2-17

Notes

- Applying the XRF to an arsenic problem when elevated lead is present illustrates this. It would be most efficient to make decisions based on the XRF, but the project team needs to be wary of potential problems that the lead might introduce. The simple decision rule in the case for weeding out samples where our XRF arsenic data might be "bad" is to send off every sample for ICP analysis when the lead concentration is greater than ten times the arsenic concentration reported by the XRF.

Adaptive Compositing Strategies for Searching

Addressing the Unknown

- Goal: Identify elevated areas (looking for contamination)
- Assumptions:
  - » Contamination is believed to be spotty
  - » Action level is significantly greater than background levels
  - » Sample acquisition/handling costs are significantly less than analytical costs
  - » Appropriate methods exist for sample acquisition and aggregation
- Aggregate samples (single or multi-increment) into composites
- Design requires determining the appropriate number of samples to composite, and developing decision criteria for composites that indicate when analyses of contributing multi-increment samples are necessary

6.2-18

Notes

The next topic picks up where the previous module left off. Recall that the last module drew the distinction between multi-increment sampling and adaptive compositing. It talked about the use of multi-increment sampling for estimating average concentrations in decision units. This module will now discuss the use of adaptive compositing strategies when looking for the presence of contamination (e.g., hot spots).

- **Goal:** The goal of adaptive compositing is to cost-effectively identify areas of elevated concentrations of contamination.

- **Assumptions:** These strategies are primarily applicable to situations where action levels are significantly greater than background levels, contamination is presumed to be spotty, sample acquisition and handling costs are significantly less than analytical costs, and appropriate methods exist for sample acquisition and aggregation. So this may not be appropriate for chasing VOC contamination. Care would need to be taken for mercury. In the case of XRF, if the XRF is capable of addressing all COCs, then it would also not be useful (we would simply screen every sample with an XRF).

- **Aggregate samples into composites:** These types of dynamic work strategies typically aggregate samples (single or multi-increment) into larger composites, and then develop investigation levels for the larger composites that would indicate when analysis of the contributing samples is necessary.

- **Design requires determining the appropriate number of samples to composite, and developing decision criteria for composites that indicate when analyses of contributing multi-increment samples are necessary.**

Recipe for Adaptive Compositing

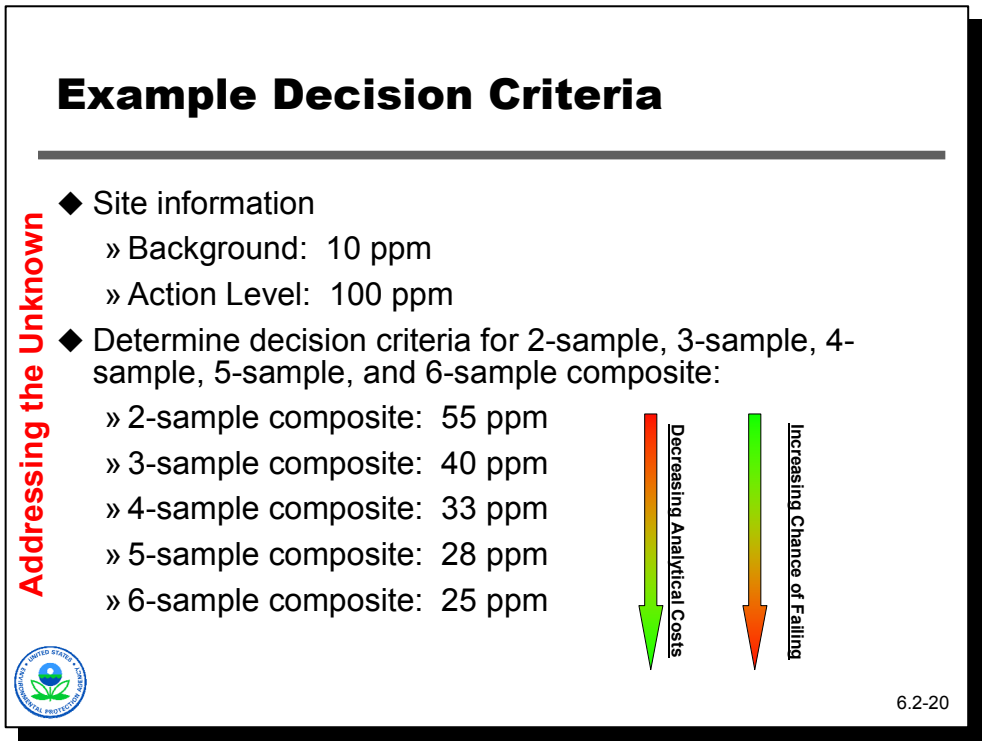
Addressing the Unknown

- Determine appropriate number of samples to composite and resulting decision criteria
- Decision criteria = (action level - background) / (# of samples in composite) + background
- Sample and split samples. Use one set of splits to composite and save other set.
- If:
  - » composite result < decision criteria, done
  - » composite result > decision criteria, analyze splits contributing to composite

6.2-19

Notes

- **Determine appropriate number of samples to composite and resulting decision criteria:** The appropriate number of samples to composite will be discussed in a later slide.
- **Decision criteria = (action level – background) / (# of samples in composite) + background:** The decision criteria is the level below which analysis of the samples in the composite is not required and above which analysis of samples in the composite is required. The following is a simple example of how an investigation level could be derived. Suppose the background for a sample is 10 ppm and the action level 50 ppm. If a composite sample is formed from four different samples, then the worst case scenario would be 3 of the samples are at background levels, and one around the action level. In this case, the composite result is 20 ppm. As long as a composite result in this instance is below 20 ppm, there is confidence that none of the contributing samples is above 50 ppm.
- **Sample and spit samples:** One set of splits should be used to create the composite and the other set of splits should be archived in case individual analyses are necessary.
- **If:** Sampling proceeds in accordance with the following:
  - » Composite result < decision criteria, done
  - » Composite result > decision criteria, analyze splits contributing to composite
  - » To continue with the example, if a composite result is above 20 ppm, then analyzing the four contributing samples identifies which individual samples are above the action level (if, in fact, any are).



Notes

The following is another example of how decision criteria are determined, and how they are affected by the number of samples contributing to each composite.

◆ **Site information**: The site information for this example is:

» Background: 10 ppm

» Action level: 100 ppm

◆ **Determine decision criteria for 2-sample, 3-sample, 4-sample, 5-sample, and 6-sample composite:** Using the equation

(action level – background) / (# of samples in composite) + background

» 2-sample = (100 – 10) / (2) + 10 = 55 ppm

» 3-sample = (100 – 10) / (3) + 10 = 40 ppm

» 4-sample = (100 – 10) / (4) + 10 = 33 ppm

» 5-sample = (100 – 10) / (5) + 10 = 28 ppm

» 6-sample = (100 – 10) / (6) + 10 = 25 ppm

Note that as the number of samples contributing to each composite increases, the cost of analytical analyses will fall for the sampling program. However, as the number of samples contributing to each composite increases the decision criteria also falls, making it more and more likely that we might fail a particular composite and have to analyze each of the splits from the samples originally contributing to the composite.

When is Adaptive Compositing Cost-Effective?

Addressing the Unknown

- The "spottier" contamination is, the better the performance
- The greater the difference is between background and the action level, the better the performance
- The greater the difference between the action level and average contamination concentration, the better the performance
- Best case: no composite requires re-analysis
- Worst case: every composite requires re-analysis

6.2-21

Notes

- **The "spottier" contamination is, the better the performance:** Adaptive compositing is most cost effective when the contamination is spotty. With spotty contamination, fewer "dirty" samples will be combined with relatively "clean" samples, reducing the number of composites above the decision criteria and reducing individual subsamples that need to be analyzed.

- **The greater the difference is between the background and the action level, the better the performance:** The decision criteria will be higher the greater the difference between background and the action level. The higher the decision criteria level is above background, then the greater the likelihood that composite sample results will be below the decision criteria.

- **The greater the difference between the action level and average contamination concentration, the better the performance:** If the average concentration is close to the action level then it will be likely that many composite samples will also be close to the action level and over the decision criteria. If the average concentration is much lower than the action level, fewer composite samples will exceed the decision criteria.

- **Best case:** In the best case, all composite results will be below the decision criteria and no composite will require re-analysis. This would also mean that no contamination above action levels was found.

- **Worst case:** In the worst case, every composite requires re-analysis. This would also mean that contamination is either not spotty, or the average concentration is close to the action level, or the background is close to the action level.

Cost/Benefit Calculation

Addressing the Unknown

- Worthwhile as long as:

$$F_f < 1 - 1/N_c$$

where:

» N_c = number contributing to composite

» F_f = fraction of composite samples "failing" (results above decision criteria)

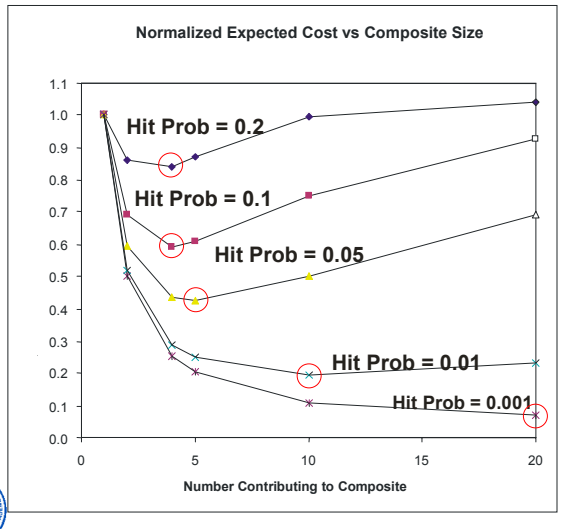
6.2-22

Notes

- **Worthwhile as long as:** The simple equation in this slide can be used to determine whether adaptive compositing will save analytical costs. As an example, if there were four samples contributing to the composite, then as long as less than 75% of the resulting composites failed the decision criteria (or more than 25% passed), analytical money would be saved. Of course, adaptive compositing also introduces some other costs into the sampling program that have to be considered as well. These include the costs of splitting and archiving samples, and the costs of homogenizing the composite before analysis.

How Many Samples to Composite?

Addressing the Unknown



- A function of the probability of contamination being present
- The less likely contamination is present, the larger the number of samples to composite
- Graph at left shows the case when one has 20 sampled locations

6.2-23

Notes

- The number of samples to composite is a function of the probability of contamination being present. The less likely that contamination is present, the larger the number of samples to composite. The graph above shows the case when there are 20 sampled locations. Along the x-axis is the number of samples contributing to a composite. The y-axis is the relative expected analytical cost (the cost of the adaptive compositing program relative to simply analyzing each individual sample). Note that in a worst case scenario, an adaptive compositing program can cost more than simply analyzing each of the original samples. The probability associated with each line refers to the fraction of the site that might contain contamination above the action level (not the composite decision criteria, which will be something less than the action level). A smaller "hit probability" indicates that it is less likely that contamination is present above the action level. Each curve plots out the expected relative cost as a function of the number of samples contributing to the composite for various "hit probability" scenarios. A couple of things to note from this graph. First, as the "hit probability" drops, the potential cost savings improves (simply another way of saying that the spottier the contamination, the better the performance from a cost perspective of adaptive compositing). The second is that for each hit probability scenario, there is a "best" number of samples to composite (the points on the graph circled in red) that minimize analytical costs. As hit probability falls, this optimal number of samples to composite increases. Finally, if the probability of encountering contamination above the action level is greater than 20% (i.e., more than 20% of the site is likely contaminated at levels of concern), there is no benefit cost-wise from adaptive compositing. Adaptive compositing is a very effective way of clearing large areas of concerns about contaminants such as PCBs that are not naturally occurring. It is not effective for contaminants such as arsenic where the action level is typically close to background. It does lend itself to "real time" techniques such as the XRF where the project team can quickly determine whether the splits of samples contributing to the composite need analysis.

Adaptive Sampling Approaches

Addressing the Unknown

- Comparing means/medians to standards:
  - Allocating XRF numbers dynamically
  - Binomial Sequential Probability Ratio Test
- Delineating contamination
  - Simple decision rules
  - Adaptive cluster sampling
  - GeoBayesian approaches

6.2-24

Notes

- **Comparing means/medians to standards:** The next portion of this module will discuss comparing means/medians to standards or action levels using an XRF.
- **Delineating contamination:** This module will conclude with a discussion of methods for delineating contamination, including the use of simple decision rule, adaptive cluster sampling, and GeoBayesian approaches.

Adaptive Strategies for Mean Estimation with XRF

- Use XRF as the primary data source
- Goal is to determine whether average concentration is above or below standard
- Will be estimating mean and 95%LCL/UCL based on bagged sample XRF data:
  - How many sampling locations?
  - How many shots per bag?

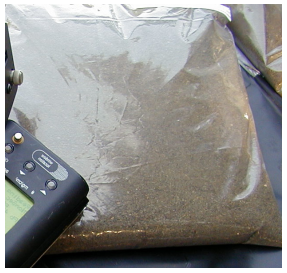
6.2-25

Notes

- The project team can use dynamic strategies that employ the XRF to estimate mean values for decision units (e.g., an exposure unit, or a yard). The assumptions are that the XRF data are quantitative and can be used as the primary data source when estimating mean values. It is also assumed that the project team will be pulling soil samples, bagging them after minimal preparation (e.g., stone and stick removal, aggregate crushing within the bag), and then estimating the average concentration within each bag by multiple XRF shots across the bag's surface. Finally, assume that the project team will calculate the average for the yard or exposure unit using all of the bag results, and that the comparison of XRF results to our action level will consist of comparing the 95%UCL (upper confidence level) and 95%LCL (lower confidence level) to the action level…in other words, a form of statistical test. The question is how many samples are needed from each yard, and how many measurements should be taken for each bag when doing this comparison.

Example: Lead in Residential Yards

- 500 ppm goal, averaged over a yard
- Minimum data collected (e.g., 5 locations per yard, each bag shot 4 times)
- Sample, measure, and calculate 95%LCL and 95%UCL for mean for yard
- If 95% UCL < 500, clean!
- If 95% LCL > 500, contaminated!
- If neither, evaluate sources of uncertainty and either:
  - » Re-measure bagged samples more times (e.g., 6 more times), or
  - » Add an additional 5 sampling locations to the decision unit, or
  - » Do both (i.e., increase sample numbers and measurements per bag)
- Recalculate LCL and UCL



6.2-26

Notes

- Here is an example of how this might be approached. The example is based on reality…a lead-in-residential-yard problem. In this particular case the action level is 500 ppm averaged over a yard. The approach consists of the following. First, the project team determines what detection limit they would like the XRF to have for any particular bag, and consequently what the total measurement time should be for the bag. In this particular case the project team will use 120 seconds. Each yard has five samples collected and bagged initially. Each bag is measured four times (2 times on each side w/ 30-sec acquisitions giving a total of 120 seconds per bag). The data for the yard are downloaded from the XRF to an Excel spreadsheet. The project team calculates the following using simple spreadsheet formulas: the average for all measurements from the yard, the standard deviation for this pooled set of measurements, the averages for each individual bag, the standard deviations observed among data for each of the individual bags, and finally the standard deviation observed among the averages for each of the bags.

  With the information, the project team calculates the 95%UCL and 95%LCL for the yard. The team can use ProUCL for this, or do it in Excel. Remember that the 95%LCL and the 95%UCL put brackets around the mean value, and basically are an indication of how good the mean value is.

  The decision criteria is this: if the 95%UCL (our mean upper bound) is less than 500 ppm, the yard is in compliance. If the 95%LCL is greater than 500 ppm, there is a lead problem. If 500 ppm falls between our 95%LCL and 95%UCL, then the data is inconclusive and more data is needed (note that collecting more data will tighten the spread between the 95%LCL and 95%UCL).

The question now is…are more samples needed, or should more measurements have been taken on the bags, or both? The answer is derived by comparing the standard deviations observed among the bag measurements with the standard deviation seen among the averages of the bag measurements. If the former is greater, then the project team should be taking more measurements per bag (i.e., re-measure our bags adding an additional four measurements). If the latter, then the project team needs more samples from the yard. If they are approximately the same, the project team will need to do both. The project team can estimate the number of additional measurements needed per bag, or the number of additional samples required by simply remembering that, in general, one needs to increase the amount of data by a factor of four to cut the size of a 95%LCL/95%UCL spread in half. Once the project team has the new data, they pool it with the old and redo the calculations.

It's important to note that one outcome of this may be a "true" average that is so close to 500 ppm that we will never be able to show definitively that is above or below…so there needs to be a reality check that says STOP!...collecting more samples is no longer justified, we will just consider the yard contaminated.

Compare this to the performance of the standard lead yard sampling protocols which typically call for a 5-increment MI sample for each yard and direct comparison of the resulting lead value with the action level. If it fails or passes based on this comparison we have no idea how confident we are that we are right. If it does fail, we have no idea where contamination might be in the yard.

A slightly more complex version of the above approach was recently used in Region 3 with great success on a set of problematic properties where historical sampling had been contradictory and inconclusive.

Binomial Sequential Probability Ratio Test (SPRT)

Addressing the Unknown

- A sequential non-parametric statistical test
- GOAL: If a portion of a decision unit is contaminated above an AL, determine whether the amount of contaminated material is less than some limit.
  - » The actual concentration of contamination is not important, only the amount of contaminated material.
- For example, suppose the decision unit is ¼-acre. Suppose also that no more than 1/5th of the ¼-acre is allowed to exceed the A/L.
  - » This test statistically determines whether the contaminated area (within the ¼-acre under investigation) exceeds 1/20th-acre

6.2-27

Notes

- **A sequential non-parametric statistical test:** The sequential probability ratio test is the adaptive or sequential alternative to the non-parametric Sign test. It is common in industrial applications where it is used for industrial quality control purposes. It is used to determine if the fraction of a particular decision unit that is above the cleanup standard is above or below some requirement (typically 50%). Like the Sign test, it assumes that sampling the decision unit follows a binomial distribution.

- **Goal:** This method assumes a goal will be set for that fraction of a decision unit which must be below some action level. If the true fraction contaminated above the cleanup standard is less that the goal, the decision unit is deemed to meet the criteria. The actual concentration of contamination is not important, only the amount of contaminated material is evaluated. A real-time method that can confidently determine whether sample concentrations are above or below a standard is required. A methodology for selecting sampling locations within the unit as sampling proceeds, such as adaptive fill methodology, is also necessary. Using this approach, a sampling location is selected at random, analyzed, and the result is compared to a standard. Based on the comparison, it must be decided if another sample is required or if a conclusion can be drawn that the fraction of the unit above the standard is either greater than the requirement (e.g., 50%) or below the lower bound of the grey region at the desired error levels. If no decision can be made about the fraction, then another sampling location is selected and the process is repeated.

VSP has a version of this test that requires the underlying distribution to be normally distributed and requires the user to know the underlying standard deviation of the contaminant. The more general version can be implemented using Excel.

- **For example:** In the example, the decision unit is ¼-acre and the decision criteria is that no more than 1/5th of the ¼-acre is allowed to exceed the action level. The test statistically determines whether the contaminated area (within the ¼-acre decision unit under investigation) exceeds 1/20th-acre.

Binomial SPRT Assumptions

Addressing the Unknown

- ◆ No underlying distributional assumptions
- ◆ Specify how small the contaminated portion (which exceeds the AL) must be in order to be confidently declared "in compliance." (sets the lower bound of the gray region)
- ◆ Specify acceptable decision error rates
- ◆ A real-time analytical method is available that can confidently determine whether a particular sample contaminant concentration is above or below the AL
- ◆ A method is available for selecting the next sampling location in the sequence (e.g., adaptive fill)

(continued)

6.2-28

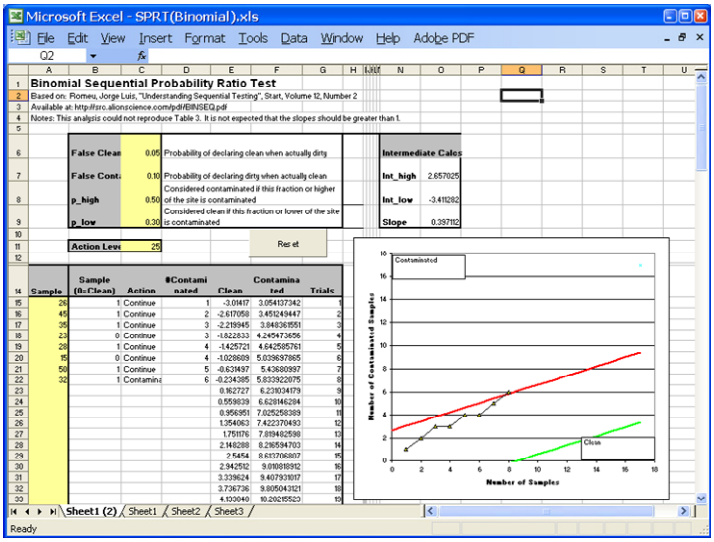
Notes

- ◆ **No underlying distributional assumptions:** The binomial SPRT assumes there are no underlying distributional assumptions for the decision unit or contamination.
- ◆ **Specify how small the contamination portion (which exceed the AL) must be in order to be confidently declared "in compliance" (sets the lower bound of the gray region):** This approach requires decisions to be made about how small the contaminated portion of the decision unit must be in order for the decision unit as a whole to be considered "clean" or "in compliance." The confidence associated with this decision sets the lower bound of the gray region.
- ◆ **Specify acceptable decision error rates:** This approach also assumes that acceptable decision error rates have been specified.
- ◆ **A real-time analytical method is available that can confidently determine whether a particular sample contaminant concentration is above or below the AL:** This approach requires that a real-time measurement technology be available that can accurately determine sample concentration levels so that there is confidence in the determination that the sample concentration is above or below the action level. The XRF is a real-time instrument that can be used with this approach.
- ◆ **A method is available for selecting the next sampling location in the sequence (e.g., adaptive fill);** This approach also requires the use of a method, such as adaptive fill, for determining the location of the next sample in the sequence.

Binomial SPRT Assumptions

Sampling continues until it can be concluded that the decision unit is contaminated, clean, or the approach is abandoned

Addressing the Unknown

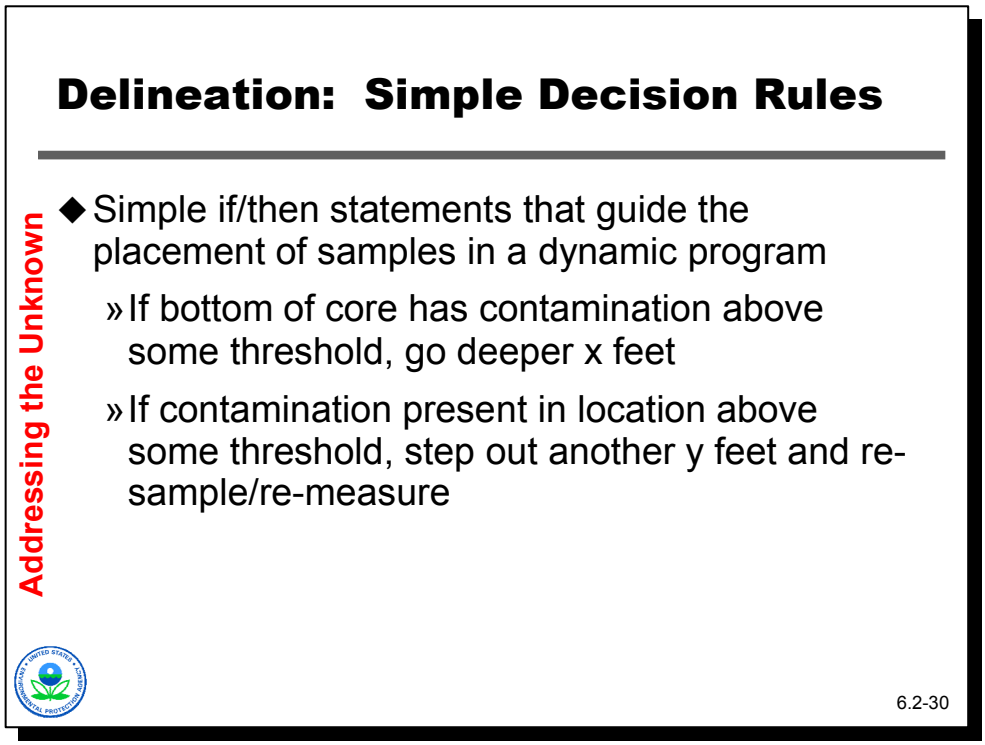


6.2-29

Notes

- This graphic is a screen shot of the Excel format for a binomial SPRT approach. The decision error rates are identified in the box in the upper left of the screen. Sampling continues until it can be concluded that the decision unit is contaminated, clean, or the approach is abandoned. For the example displayed, we are trying to show that less than 50% of the site is contaminated above our action level ("p_high" in the spreadsheet). "False Clean" sets the acceptable error when deciding that less than 50% is contaminated above the action level. "p_low" helps us set our second error rate which tries to control for the other mistake we might make, i.e., deciding that more than 50% of the site is contaminated when in fact that is not the case. With "p_low" set to 0.3 and "False Cont." set to 0.1, we are saying that when 30% or less of the site is contaminated, we want to make the mistake of deciding that more than 50% is contaminated 10% of the time or less. With these parameters set, we can start sampling, as shown in the bottom left portion of the spreadsheet. Our action level has been set to 25 ppm. Here we have sequentially sampled eight times. Six of those samples yielded a result above the detection limit and two did not. By the time we encountered our 6th "hit", we could confidently conclude more than 50% of our site was contaminated. The progress of this sampling program from a decision-making perspective is shown on the right graph. The red and green lines divide our graph into three regions…above the red line we are confident we are "dirty", below the green line we are confident we are "clean", and the area between we can't decided. Along the x-axis we have our sample numbers. The stair-step black line traces our sampling program progress…if it walks above the red line we stop and conclude our site is contaminated. If it walks below the green line we stop and conclude our site is clean. For our example it approaches the red line, and then finally crosses after the 8th sample has been collected and analyzed. Using the "50% of the site" as the decision

rule, depending on how the other 3 parameters are set, one could have potentially decided that the site was "contaminated" in as few as five samples (if all had been above the action level), or "clean" with as few as seven samples (if all had been below the action level).



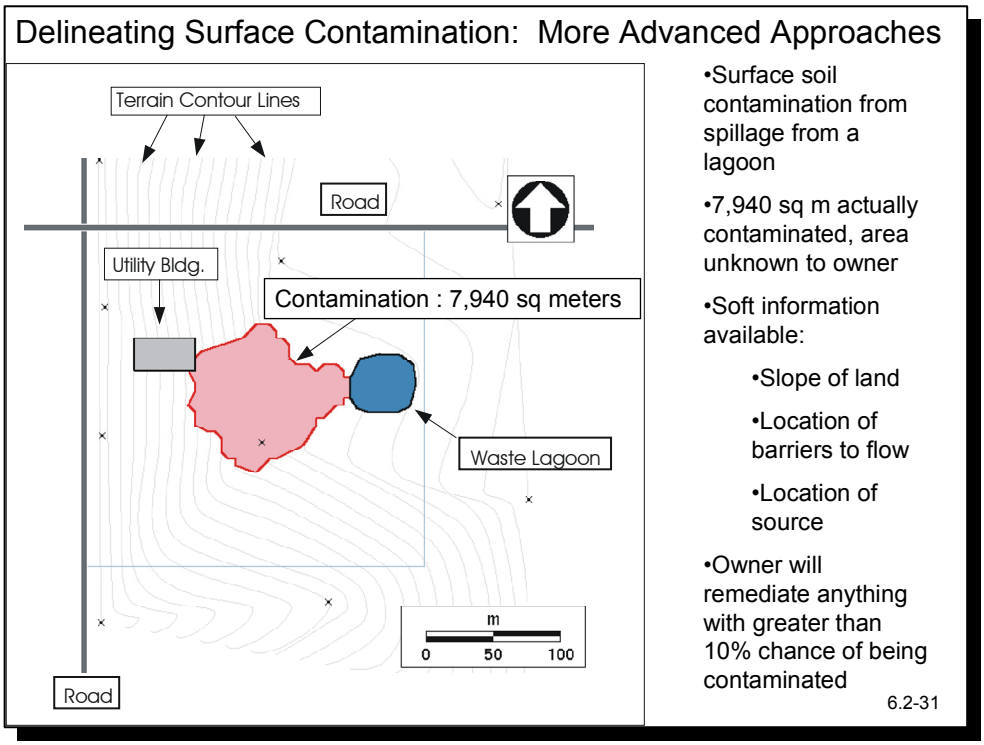
Notes

This module ends with a discussion of several different dynamic methods for delineating contamination using an XRF. There are three different approaches presented: simple decision rules, adaptive cluster sampling, and GeoBayesian adaptive sampling.

The most basic approach is simple if/then statements that guide the progress of a dynamic field effort.

- **Simple if/then statements that guide the placement of samples in a dynamic program:** The delineation of contamination can employ simple if/then statements to guide the selection of additional sample locations in a dynamic work strategy. For example, a simple if/then statement could be: If the bottom of the core has contamination above a specific threshold, go 2 feet deeper. Another example could be: If contamination is present in a location above a specific threshold, step out another 3 feet and re-sample/re-measure.

  In each of these cases, the rules are simply defined and easy to follow.



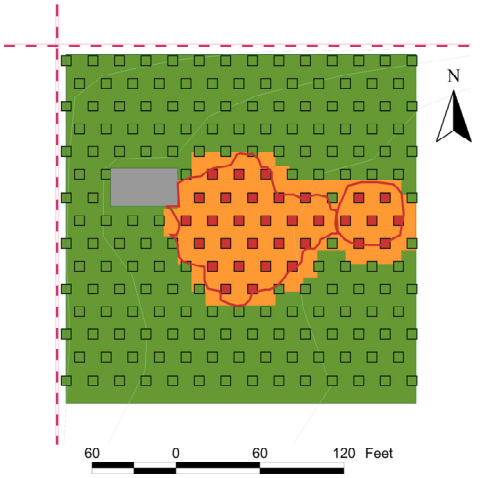
Notes

◆ The next several slides will present two different techniques for delineating contamination footprints. The example is a site where there has been a surface release from a waste lagoon that resulted in 7,940 square meters of contaminated surface soils. The owner knows there is a problem, but is unsure of the footprint. The actual footprint of contamination is outlined in red (but the owner doesn't know exactly where this is). "Soft" information for the site includes the slope of terrain (presumably a predictor of transport), the locations of barriers to flow, and the location of the source. The owner will remediate any soil area with greater than a 10% chance of being contaminated. The challenge is to design a sampling program that will delineate this contamination for him is a cost-effective but accurate way.

Traditional Approach Results

Addressing the Unknown

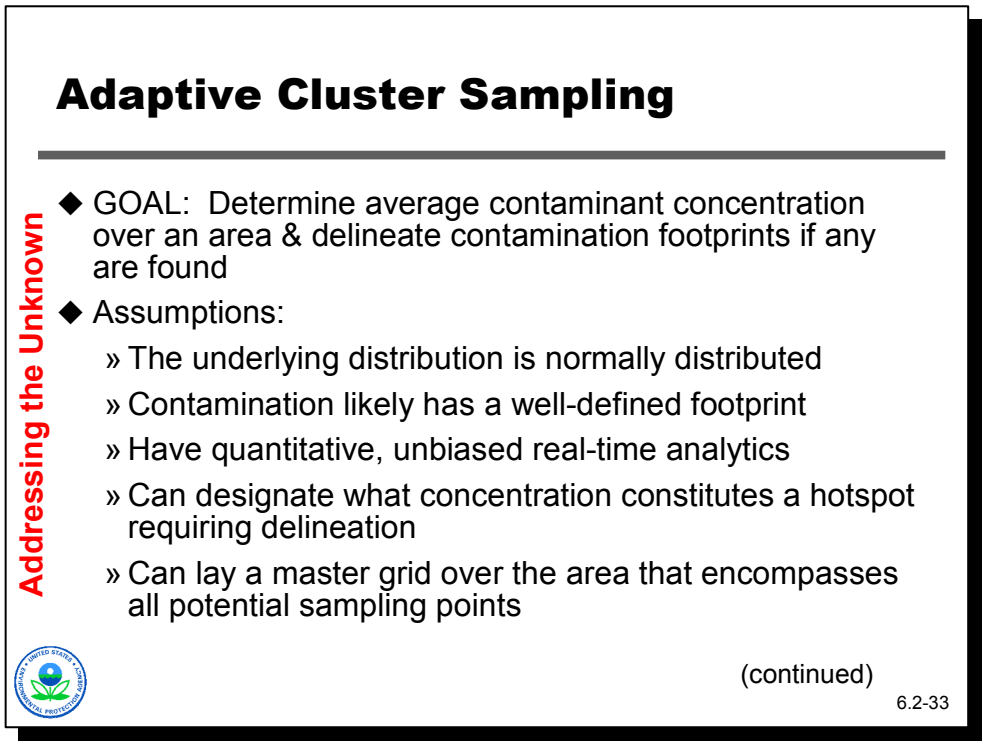
- Triangular gridded program
- 203 samples allocated
- Observed error rates:
  - » Missed contamination: 0 ft2
  - » Incorrectly excavated clean: 3,500 ft2 (35% over-excavation)



6.2-32

Notes

- A traditional approach to this site would involve laying down a regular grid of sampling points, sampling them all in one round, sending the samples off-site for analysis (e.g., ICP), and then estimating contamination volumes and determining excavation footprints based on the results.
- This graphic shows the results from such a program based on 203 samples. Sample results are color-coded either green or red, depending on whether the laboratory results were below or above the action level. The orange footprint indicates the portion of the site that has greater than a 10% probability of having contamination above action levels based on these data. With this number of samples we would have gotten all of the contamination for this site, but over-excavated by about 35%.



Notes

One alternative to a traditional approach is something call adaptive cluster sampling. Adaptive cluster sampling design and data analysis is an option in Visual Sample Plan (VSP). VSP is free software available at http://vsp.pnl.gov/.

- **Goal:** Adaptive cluster sampling can be used to both determine the average level of contamination over an area and delineate contamination footprints if they are found.

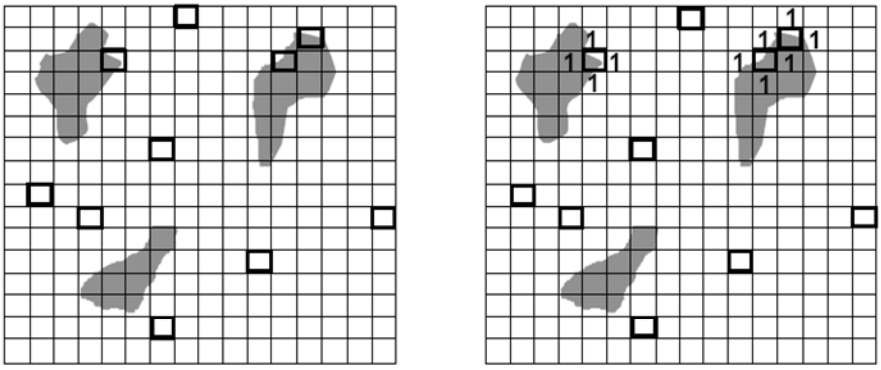
- **Assumptions:** Adaptive cluster sampling assumes the underlying contamination distribution is normal, that contamination has a well-defined footprint, that there is a quantitative, unbiased real-time analytical technique, that an action level requiring delineation can be established; and that a master grid can be laid over the area that encompasses all potential sampling points (not all of which will be sampled). The latter is important because it determines the spatial resolution that must be attained when delineating footprints.

Adaptive Cluster Sampling

Addressing the Unknown

Initial grid sampling (10 random samples)

1st batch of adjacent units



From EPA QA/G5S (Dec 2002) page 107, Figure 9-1

(continued) 6.2-34

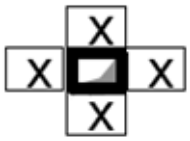
Notes

- Sampling begins by first estimating the number of samples required to get a simple estimate of the mean and randomly or systematically assigning those samples to the grid. For the samples that hit contamination, all neighboring grid nodes are then sampled and analyzed as depicted in the slide. In this example graphic, the grey areas are the contamination footprints, and the cells are the locations where samples could potentially be taken. The map on the left show the initial set of sampling locations. The map on the right shows additional samples collected in response to what was initially observed (indicated by the number 1 in cells). Sampling continues until the last round of samples are all "clean" (results below action levels).

Adaptive Cluster Sampling

Addressing the Unknown

- Requires initial grid - number of grid nodes to be sampled in the 1st round determined based on the number needed to estimate a simple mean
- Any contamination found is surrounded by samples from adjacent nodes
- Estimating the mean concentration
- Available in VSP

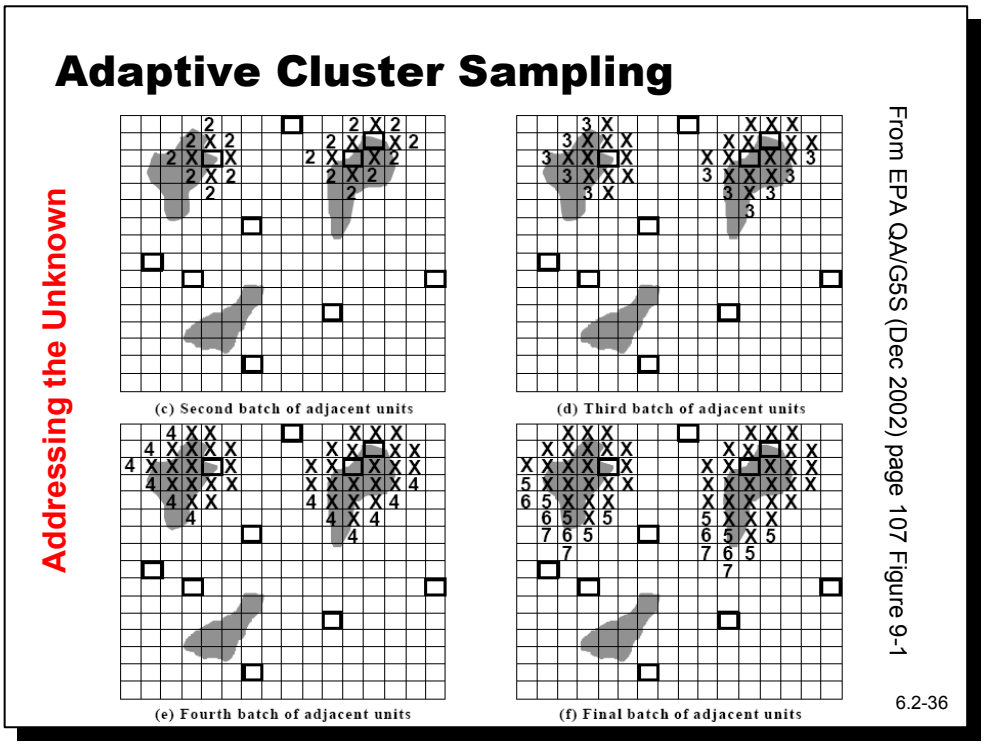


(continued)

6.2-35

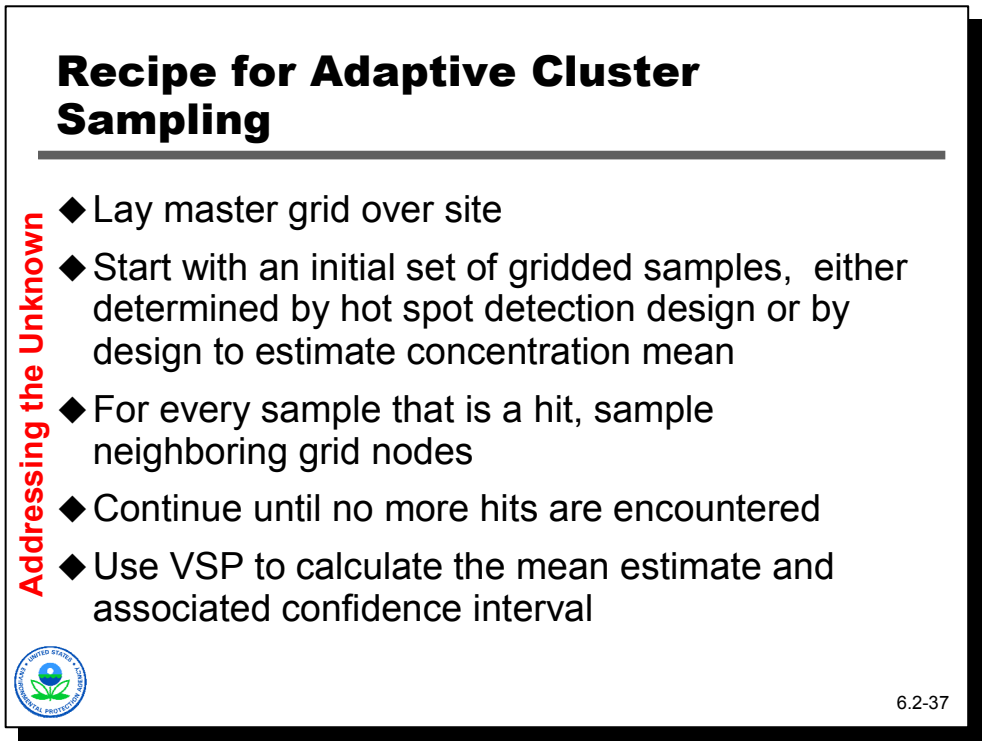
Notes

- **Requires initial grid:** Once the initial grid is established, the number of grid nodes to be sampled in the first round is determined based on the number needed to estimate a simple mean.
- **Any contamination found is surrounded by samples from adjacent nodes:** Sampling continues until every contaminated sample is surrounded by samples below the level of concern.
- **Estimating the mean concentration:** After sampling is complete, estimating the mean concentration is a more complicated computation because of the biased nature of the sampling.
- **Available in VSP:** Virtual Sample Plan (VSP) can perform the calculation.



Notes

- This graphic shows the second, third, fourth, and final batch of sampling of adjacent units using the adaptive cluster sampling approach.



Notes

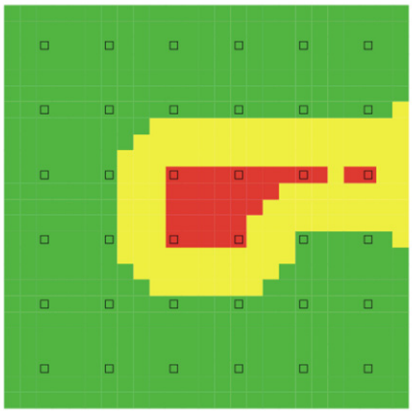
Recipe for adaptive cluster sampling: The steps for this approach are as follows:

- Lay master grid over site
- Start with an initial set of gridded samples, either determined by hot spot detection design or by design to estimate the concentration mean
- For every sample that is above the action level, sample neighboring grid nodes
- Continue until no more hits are encountered
- Use VSP to calculate the mean estimate and associated confidence interval

Adaptive Cluster Results

Addressing the Unknown

- 36 samples selected initially, looking for circular hot spot with radius = 50 ft.
- Results shown on map



(continued)

6.2-38

Notes

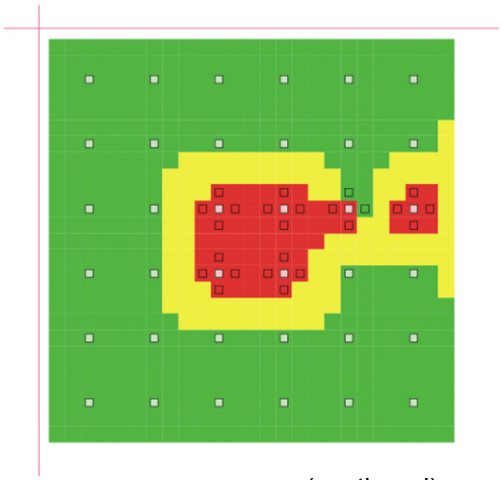
- The next series of slides shows how apply adaptive cluster sampling is applied to the example.

  The project team selects 36 initial sampling locations. These 36 locations were chosen because the project team wanted a high degree of confidence that they could identify a 50-ft radius contamination location if it were present. The results from these 36 locations are shown on the map, color-coded by whether contamination was encountered above or below the action level, and then the results interpolated so that the project team could identify what area had a greater than 10% chance of being contaminated (the area identified by red and yellow).

Adaptive Cluster Results

Addressing the Unknown

- 2nd round of sampling includes 24 additional locations
- Results shown on map



(continued)

6.2-39

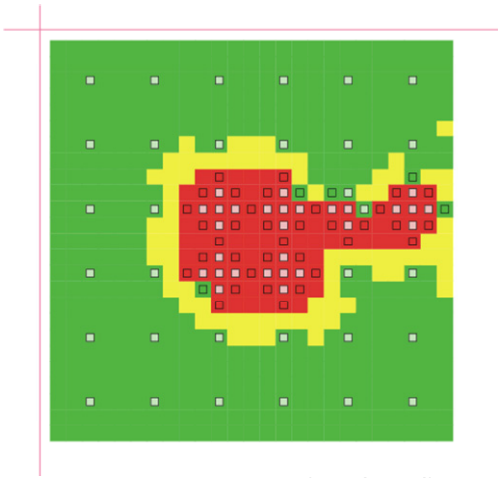
Notes

- The second round of sampling targeted those locations that encountered contamination in the first round, resulting in an additional 24 sampling locations. The map shows the results from these samples.

Adaptive Cluster Results

Addressing the Unknown

- 3rd round of sampling includes 40 additional locations
- Results shown on map



(continued)

6.2-40

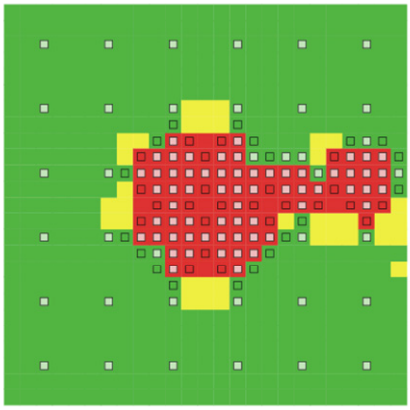
Notes

- The third round required another 40 additional locations.

Adaptive Cluster Results

Addressing the Unknown

- 4th round of sampling includes 43 additional locations
- Results shown on map



(continued)

6.2-41

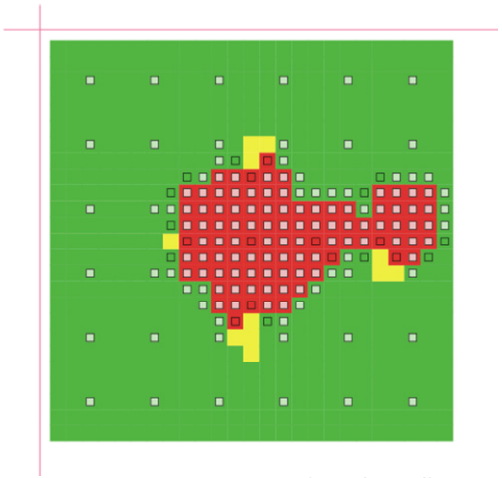
Notes

- The 4th round needed another 43 locations.

Adaptive Cluster Results

Addressing the Unknown

- 5th round of sampling includes 21 additional locations
- Results shown on map



(continued)

6.2-42

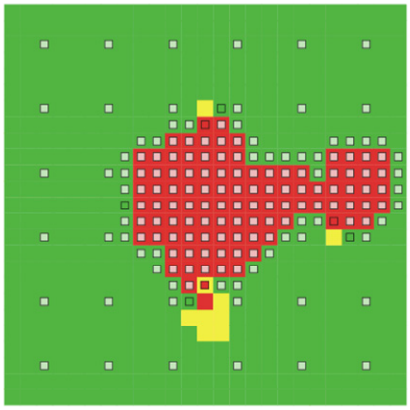
Notes

- The 5th round required another 21 locations.

Adaptive Cluster Results

Addressing the Unknown

- 6th round of sampling includes 7 additional locations
- Results shown on map



(continued)

6.2-43

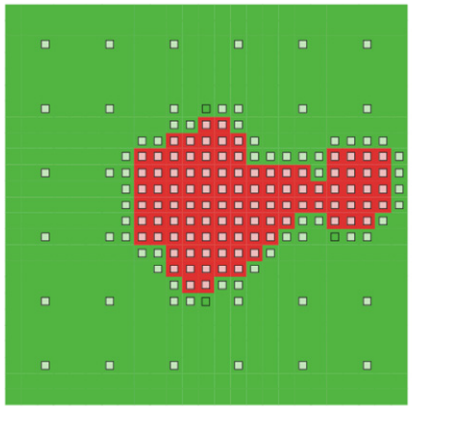
Notes

- The 6th round took only seven locations.

Adaptive Cluster Results

Addressing the Unknown

- 7 rounds of sequential sampling: 36, 24, 40, 43, 21, 7, and 3 samples
- Sampling complete and footprint delineated after a total of 174 samples
- Works when contamination is "blocky"



6.2-44

Notes

By the 7th round, sampling was complete (the 7th round took a final 3 additional locations).

Applying adaptive cluster sampling to example site: To summarize:

- 7 rounds of sequential sampling: 36, 24, 40, 43, 21, 7, and 3
- Sampling complete and footprint delineated after a total of 174 samples (about a 15% savings as compared to a strictly gridded approach)
- Works when contamination is "blocky"

Adaptive GeoBayesian Approaches

Addressing the Unknown

- Goal: Hot spot and boundary delineation
- Assumptions:
  - » Appropriate real-time technique is available
  - » Yes/no sample results are adequate
  - » Spatial autocorrelation is significant
  - » Desire to leverage collaborative information

(continued)

6.2-45

Notes

The final approach for delineation that we will discuss is adaptive GeoBayesian approaches.

- **Goal:** Adaptive GeoBayesian approaches are used for searching and boundary delineation. They focus on indicator results, i.e., are the results of sampling above or below requirements? As such they can be used with a wide variety of analytical techniques. They assume that real-time analytical methods are available. These approaches can be used with multi-increment sampling to improve sample representativeness. They can also be used with collaborative data sets.

- **Assumptions:** Use of GeoBayesian techniques assume the following:
  - » There are appropriate real-time techniques available
  - » Yes/no sample results are adequate for the investigation
  - » Spatial autocorrelation is significant
  - » There is a desire to leverage collaborative information

Adaptive GeoBayesian Approaches

Addressing the Unknown

- Method uses geostatistics and Bayesian analysis of lines of evidence and collaborative data to guide sampling program and estimate the probability of contamination at any location
- Design requirements include appropriate investigation levels for the real-time technique and estimate of the spatial autocorrelation range. Estimates of the concentration mean can be obtained using block kriging.
- Available in BAASS software; download from http://web.ead.anl.gov/baass/register2/

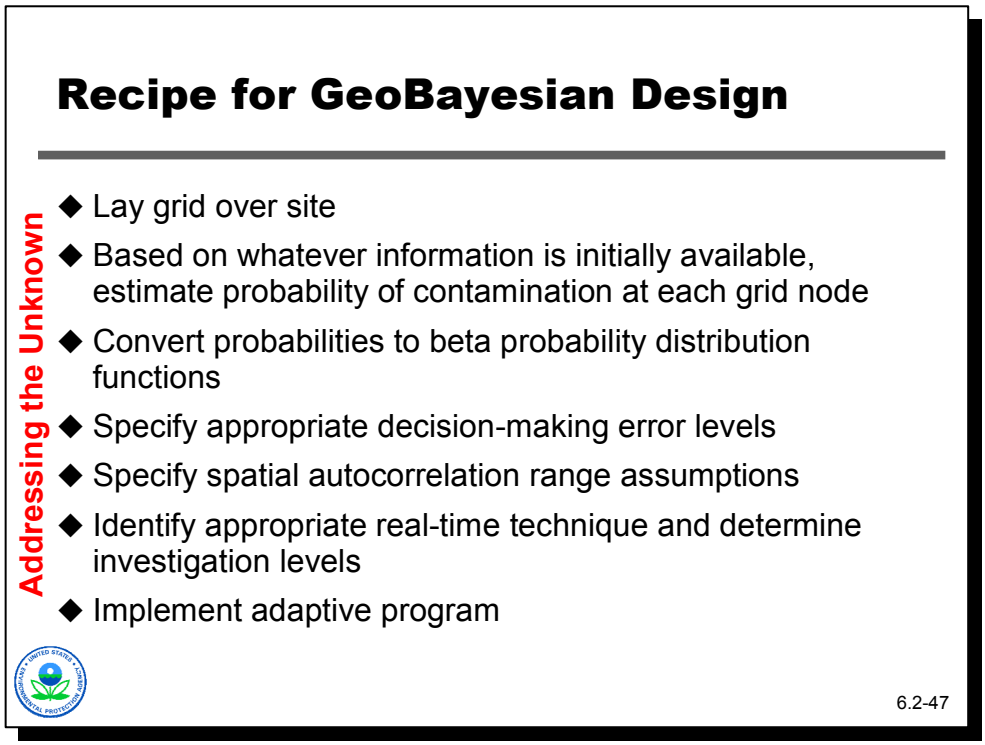
6.2-46

Notes

- **Method uses geostatistics and Bayesian analysis of lines of evidence and collaborative data to guide sampling program and estimate the probability of contamination at any location:** GeoBayesian techniques explicitly address spatial autocorrelation in the analysis, and in particular the indicator variogram range that would be expected from a site under investigation. Bayesian techniques allow the use of "soft" information (e.g., anecdotal information, aerial photos, stressed vegetation, modeling results, etc.) when designing the program. Soft information forms the initial Conceptual Site Model (CSM) for the site, which in turn is captured by probability density functions assigned to a grid laid over the site. Bayesian techniques also provide a way of "updating" the initial CSM as sample results become available. The approach provides best estimates of the probability of contamination exceeding requirements at each grid node, as well as confidence levels associated with the probabilities. Grid node spacing determines resolution, but does not affect sample number determinations.

- **Design requirements include appropriate investigation levels for the real-time technique and estimate of the spatial autocorrelation range. Estimates of the concentration mean can be obtained using block kriging:** These techniques are particularly effective for identifying hot spots and delineating contamination boundaries when a significant autocorrelation is expected to be present (i.e., the dump site, or sites where contaminant transport has or is occurring). These techniques do not provide any information about the average level of contamination in a decision unit but this can be obtained using block kriging.

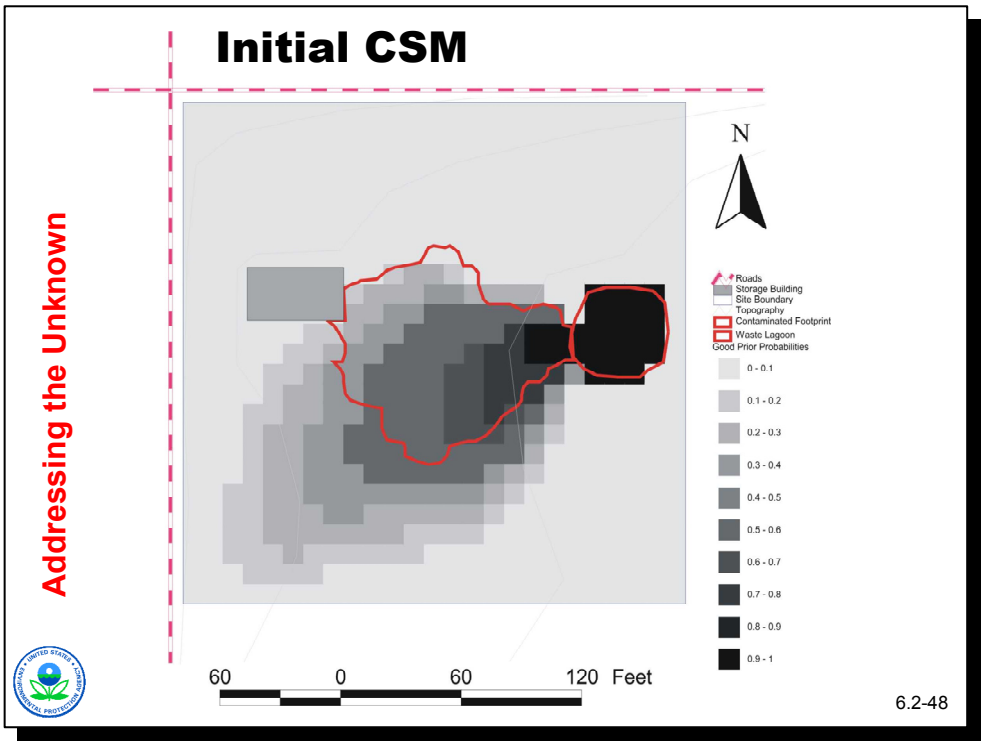
- **Available in BAASS software:** Download from http://web.ead.anl.gov/baass/register2/



Notes

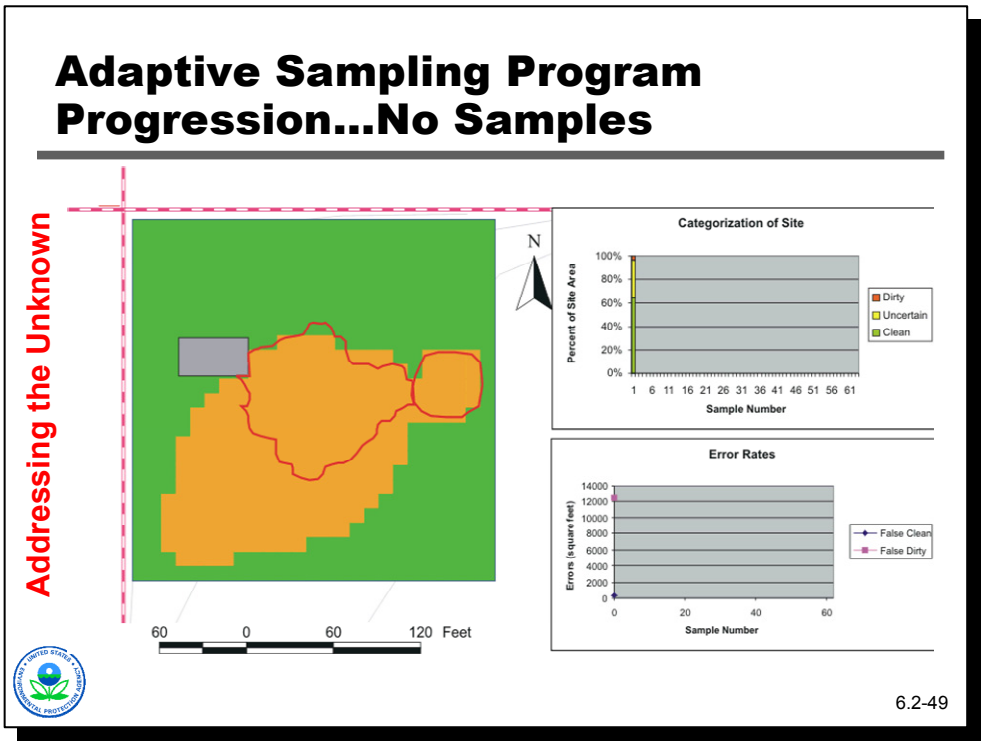
Recipe for GeoBayesian design: The steps for this approach are as follows:

- Lay grid over site
- Based on whatever information is initially available, estimate probability of contamination at each grid node
- Convert probabilities to beta probability distribution functions
- Specify appropriate decision-making error levels
- Specify spatial autocorrelation range assumptions
- Identify appropriate real-time technique and determine investigation levels
- Implement adaptive program



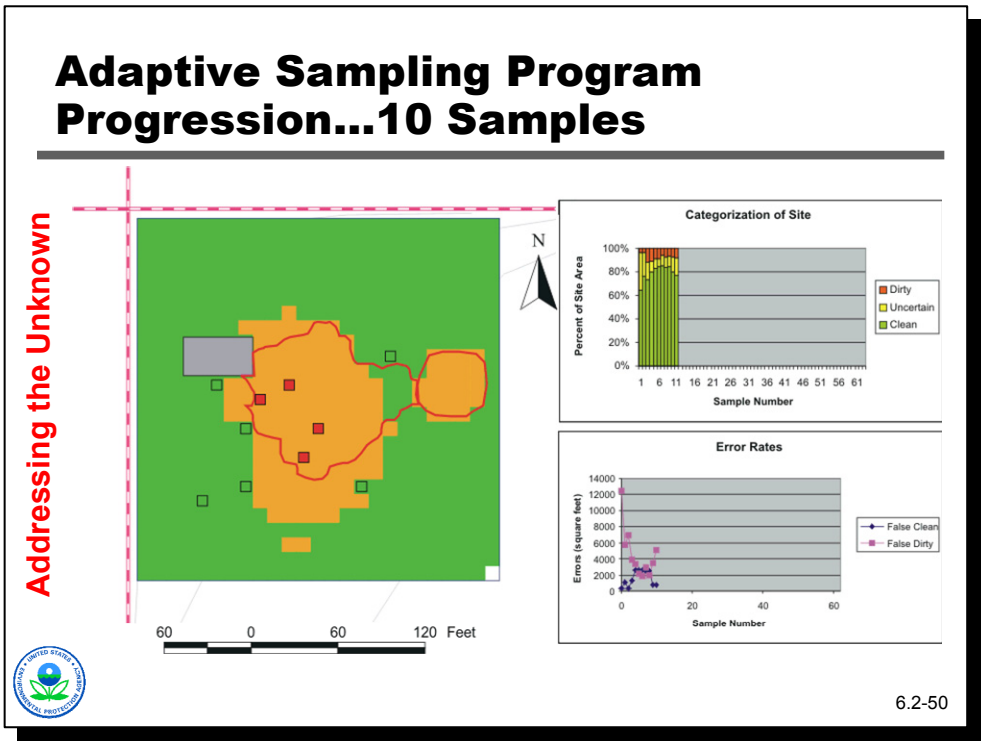
Notes

- Based on soft information, assign a probability of contamination being present.
- The map shows this CSM pictorially, it is color-coded based on contamination probability.
- This CSM drives subsequent sampling decisions and becomes an important point of concurrence for stakeholders.
- In a GeoBayesian approach, the first step is to gather all of the soft information available for the site, and build a quantitative initial Conceptual Site Model (CSM) that reflects the confidence that contamination is or is not present for particular areas. The map shows the results of this CSM effort, gray-scale coded by probability. The red line shows the actual footprint of contamination.
- This initial CSM will drive subsequent sample collection, whose purpose will be to delineate the contamination footprint. Note that this CSM becomes an important point of concurrence for stakeholders. Areas where the initial CSM indicate less than a 10% probability of contamination being present will not be sampled, unless subsequent sampling ends up pursuing contamination into these areas. That is because the available soft information is sufficient to support decision-making (i.e., no action) for these areas.



Notes

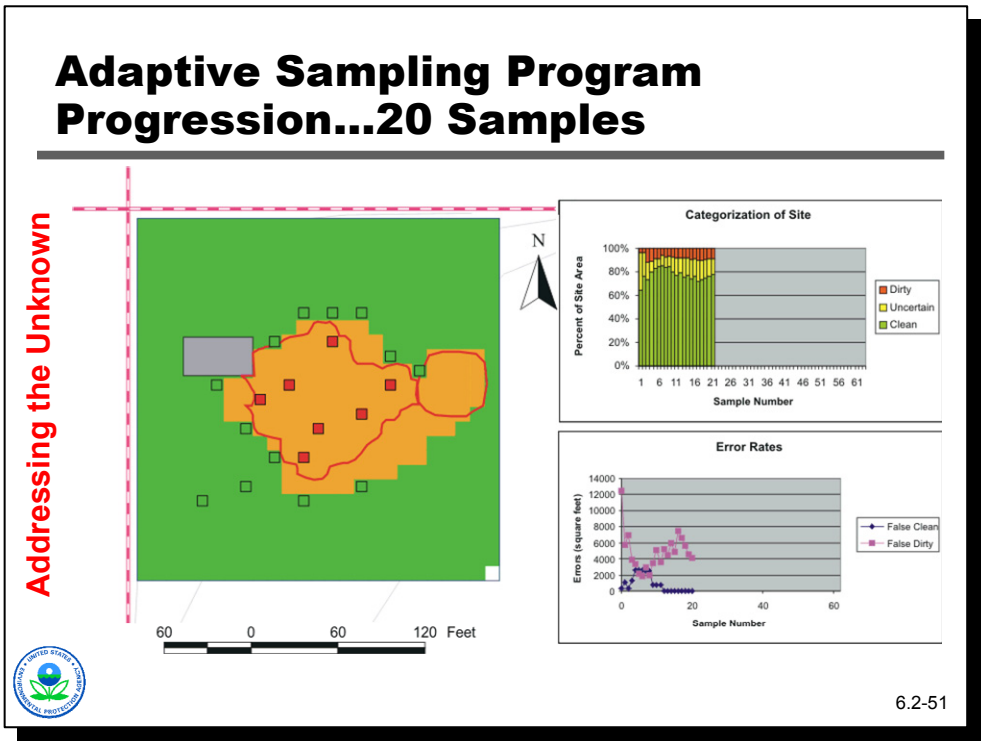
- Based on the initial CSM, this map shows what the owner would do if no samples were collected. The area in orange indicates what would need to be removed. Not surprisingly it would be a conservative excavation, and would remove a lot more soil than necessary, while at the same time missing some soil that should be removed.



Notes

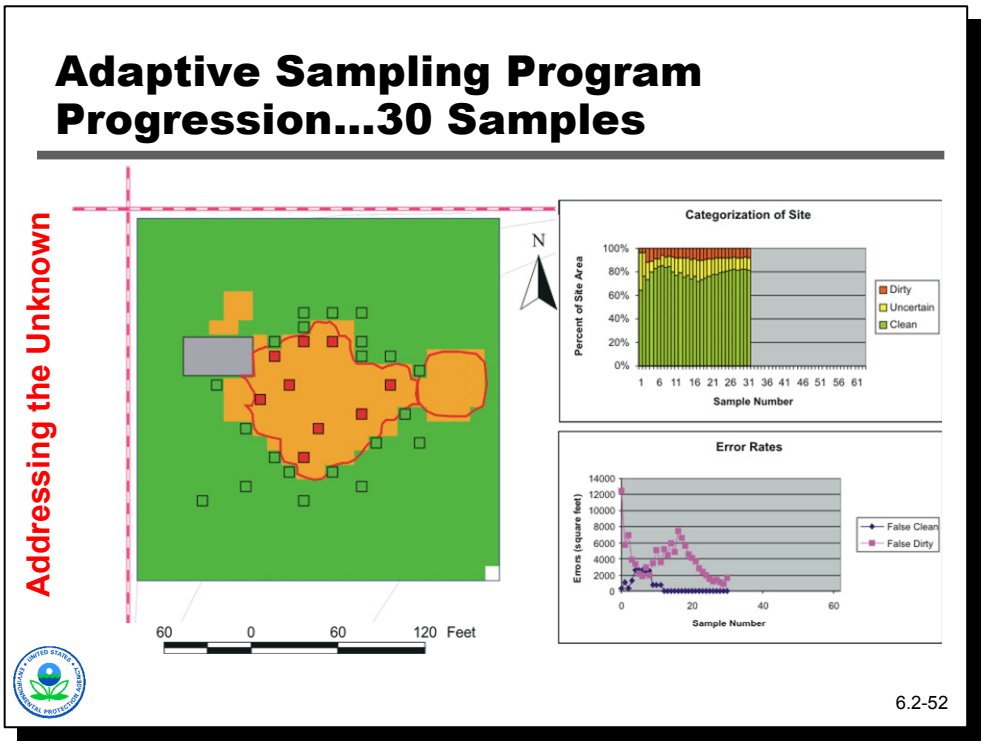
- With this type of adaptive program, sampling can be done perfectly sequentially (i.e., pick a location, sample, analyze, and then determine the next location), or new sample locations can be selected in sets. For this example, the sampling was done perfectly sequentially. This slide shows the status of the program after 10 samples had been collected and analyzed. After each new sample is collected and analyzed, the results are fed into the software, which then determines where the next sample should be. The orange area identifies the portion of the site that has been determined to have greater than a 10% chance of being contaminated based on the information collected up to that point in time.

  The two graphs to the right track sampling program progress based on two different metrics. The top graphic indicates how much of the site would be classified as clean (green, contamination probability <10%), contaminated (red, contamination probability >90%), or uncertain (yellow, contamination probability between 10% and 90%). As sampling continues, the yellow should be "squeezed out". The bottom graphic tracks decision errors (i.e., how much area is being called clean when in fact it is contaminated, and how much is being called contaminated when in fact it is clean). This is a metric invisible to the owner, but useful to see how the sample program is performing.



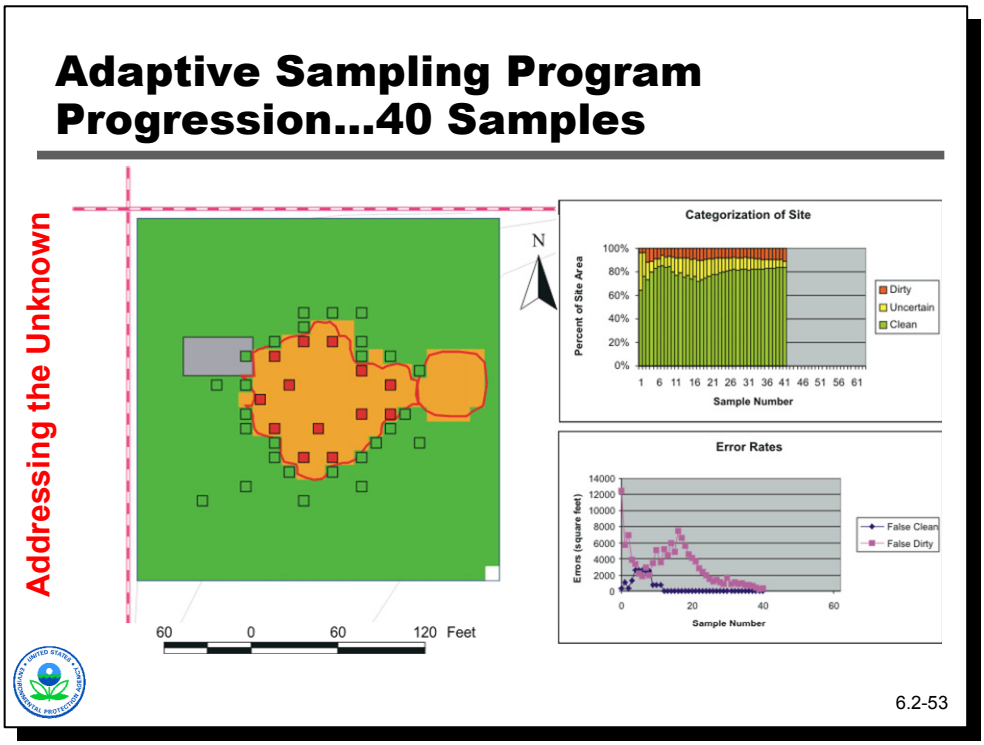
Notes

- This graphic shows the sampling status after 20 samples.



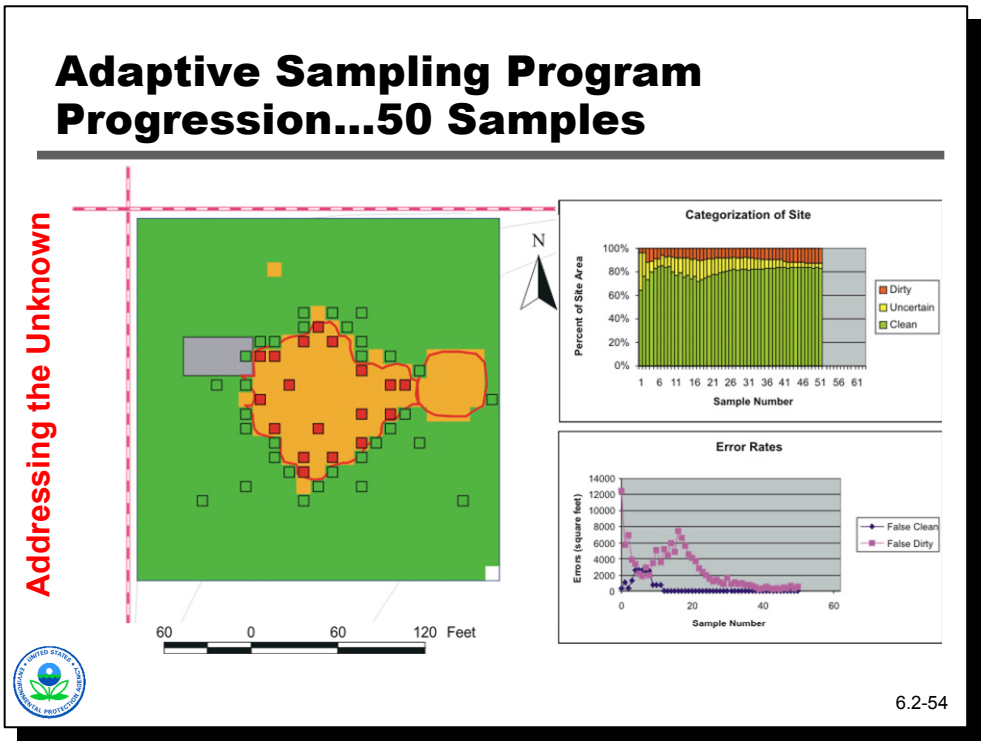
Notes

- This graphic shows the sampling status after 30 samples.



Notes

- This graphic shows the sampling status after 40 samples.



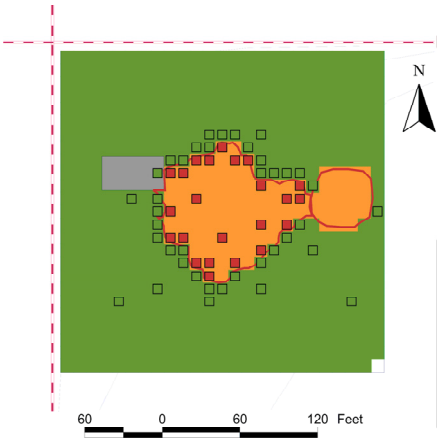
Notes

- This graphic shows the sampling status after 50 samples.

Adaptive Program Performance

Addressing the Unknown

- Completely done with 62 samples
- After only 22 samples, outperformed traditional 203 gridded program from an error rate perspective
- Works best when contamination is "blocky"



6.2-55

Notes

- **Adaptive program performance:** After 62 samples the sampling is complete. The contamination footprint is completely delineated. As the top graphic to the right shows, the yellow uncertain area has been squeezed out. As the bottom graphic shows, decision errors have been forced to zero. The resulting sampling program looks neither like the traditional gridded sampling program nor like the adaptive cluster program. The number of samples required is less than 30% of what the traditional program required, and less than 40% of what the adaptive cluster approach required. This type of approach works best when there is "soft" information available to leverage and/or contamination is expected to be "blocky".

When are adaptive delineation methods cost-effective?

Addressing the Unknown

- In general, will always outperform traditional methods
- Level of out-performance depends on:
  - How "wrong" traditional method design assumptions were
  - How "spotty" contamination is (the more spotty, the poorer the performance)

6.2-56

Notes

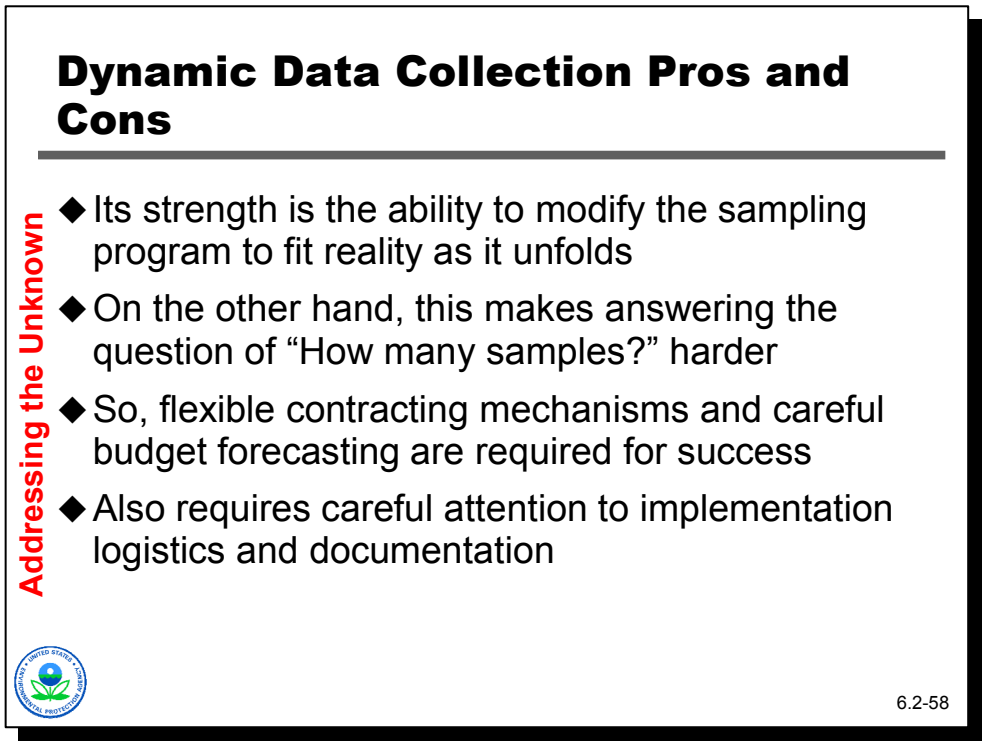
- **In general, will always outperform traditional methods:** Adaptive sampling methods will always outperform traditional methods by providing a more dense characterization picture at a lower cost and in less time.
- **Level of out-performance depends on:** The level of performance depends on how "wrong" traditional method design assumptions were and how "spotty" contamination is (the more spotty, the poorer the performance of the traditional method).

The Biggest Bang Comes from Combining…

Addressing the Unknown

- CSM knowledge, with…
- Multi-increment sampling, with…
- Collaborative data sets, with…
- Adaptive analytics, with…
- Adaptive compositing, with…
- Adaptive sample location selection.

6.2-57



Notes

To summarize adaptive sampling technique strengths and weaknesses:

- **Its strength is the ability to modify the sampling program to fit reality as it unfolds:** The strength of adaptive sampling techniques lies in their ability to modify sampling program progress (sample numbers and/or locations and/or analytics) to fit reality as it unfolds. The resulting program is as efficient as possible.

- **On the other hand, this makes answering the question of "How many samples?" harder:** This strength is also the basis for their primary weakness from a program planning perspective. It can be difficult to accurately evaluate sample number needs before work begins. However, "what if" scenarios can be constructed to estimate the most likely sample numbers and the worst case scenario.

- **So, flexible contracting mechanisms and careful budget forecasting are required for success:** To implement these types of programs requires careful budget forecasting and flexible contracting mechanisms to be successful.

- **Also requires careful attention to implementation logistics and documentation:** Also careful attention must be paid to implementation logistics and all activities must be thoroughly documented.

Q&A – If Time Allows

6.2-59