ERDC TN-DOER-D12
December 2011

Sampling Strategies for Confined Disposal Facilities (CDF) Characterization

by *Joan U. Clarke, Christian J. McGrath, and Trudy J. Estes*

PURPOSE: Upland disposal capacity for dredged material is diminishing in many Corps Districts, resulting in deferral of navigation dredging in some areas. There are many site-specific limitations to construction or expansion of confined disposal facilities (CDFs), including high waterfront land values, cost of construction, siting conflicts, and foundation strength limitations. Removal of dredged material from CDFs for beneficial use appears to offer the best potential for sustaining existing CDF capacity. There has been no baseline established, however, regarding the intensity of sampling necessary to adequately characterize materials in a CDF for this purpose, nor guidance developed regarding parameters that should be evaluated to assess feasibility and benefit of material recovery. A previous Technical Note (Estes and Clarke 2011) examined the Chicago Area CDF as a case study to illustrate various aspects of CDF characterization for beneficial reuse of dredged material, including physical and chemical characterization with associated uncertainty analysis.

The purpose of this technical note is to describe several sampling strategies that may be appropriate for CDF characterization. Sampling considerations and the use of interpolation tools are described for the quantification of CDF material of interest for removal, including ways to estimate volume of material and visualize its distribution within the CDF. For the goal of estimating mean concentrations of contaminants of concern, recommendations are given for determining sample size (number of samples) and selecting a sampling strategy. A metric based on sampling and analysis cost, and on evaluation of precision and bias in the resulting parameter estimates, was developed to facilitate selection of an optimum sampling and compositing strategy for a given site.

BACKGROUND: Measurement **error** encompasses both lack of accuracy and lack of precision. **Accuracy** refers to the degree of closeness of measurements to the true value of a quantity, while **precision** is the closeness of repeated measurements to each other. Thus, measurements can be precise but not accurate or vice versa. Lack of precision, i.e. variability, is measured statistically by the variance (or standard deviation, standard error, coefficient of variation, etc.). **Bias** is a measure of inaccuracy, i.e. the difference between the estimate and the true value. The objective of sampling is the accurate (unbiased) and precise estimation of parameters of interest.

Many sources of error in the process of obtaining an estimate potentially affect the precision and accuracy of that estimate, including sampling, sample pre-treatment and preparation, instrument measurement, and data analyses. Of these, sampling is decidedly the "weak link," contributing the largest proportion of variability (Keith et al. 1996). The relative amounts of error attributed by

Rasemann (2000) to the various aspects of sample collection and analysis are given in Table 1, along with typical per-sample costs[1] that might be expected in CDF characterization work.

Table 1. Relative amounts of error and typical costs attributable to the process of sample collection and analysis in CDF characterization.

Sample Processing	Relative Error (% of true value)[a]	Cost per Sample
Sample collection	1000	\$7,000 - \$11,000[b] \$80 - \$1400[c]
Sample pre-treatment	100	
Sample preparation	100 - 300	\$750[d]
Instrument analysis	2 - 20	\$600 - \$900[e]
Data analysis (use of incorrect model)	up to 50	

a From Rasemann (2000)
b High cost sample collection (machine coring); includes mobilization and demobilization costs
c Low cost sample collection (hand coring); includes travel and supplies
d Includes sample preparation; analysis; quality control; and data reporting for grain size, total organic carbon, metals, PCBs and PAHs
e Additional cost for dioxins and furans

The error resulting from sampling can be up to three orders of magnitude greater than that of subsequent sample analyses, yet the most attention is often focused on reducing variability in the laboratory, with the hope that whatever samples were collected are somehow adequately representative of the area to be characterized. Unfortunately, the main way to reduce variability associated with sampling is by taking more samples, which increases costs for both collection and analysis. Cost considerations may constrain sampling efforts such that the desired optimum number of samples may not be affordable (and this is usually the case), but reduction in variability may still be achieved by employing improved sampling design (Flatman et al. 1988).

The desired result of sampling may be a point estimate such as an average, or it may be a two- or more dimensional surface in space and/or time. In the CDF, point estimates may be sufficient for characterizing contaminant concentrations for comparison with applicable soil or sediment quality criteria. On the other hand, two- and three-dimensional surfaces such as contour and isopach maps may be needed to determine the volume and distribution of material suitable for removal and beneficial reuse.

TYPES OF SAMPLING: A **discrete sample** is collected, prepared, and analyzed as a unique entity, i.e. not combined with other samples. As an example, 12 discrete core samples were collected from the Chicago Area CDF, of which all were analyzed for grain size, nine were analyzed for organic carbon and seven for contaminants. After collection, discrete samples from a specified area may be combined and homogenized to form a **composite sample**. Composite samples are routinely used in dredged sediment disposal evaluations (U.S. Environmental Protection Agency/U.S. Army Corps of Engineers (USEPA/USACE) 1991, 1998). A special type of composite, the ***multi-increment* sample**[2], combines 30 or more increments (i.e. small discrete samples) collected within a pre-defined boundary. Multi-increment sampling has been

[1] Cited costs are based on reported sampling and analytical costs from various sources; confirmation of site-specific costs is recommended as part of the planning process.
[2] *MULTI-INCREMENT*® is a registered trademark of Envirostat, Inc.

recommended for collecting soil samples of munitions residues at military instillations (Jenkins et al. 2005), evaluating contaminants in soil stockpiles (Lamé et al. 2005), and environmental site assessments in Hawaii (Hawai'i Office of Hazard Evaluation and Emergency Response (HEER) 2008).

Discrete sampling. When the cost of sample collection is high relative to the cost of subsequent sample processing and laboratory analysis, cost constraints may restrict sampling to a limited number of discrete samples. Location of samples can be random, systematic, or judgmental. **Random** samples are located by overlaying the sampling area with a grid having more cells than the desired number of samples, and then randomly selecting cells from which samples will be collected. Stratified random sampling imposes restrictions on sample location by allocating a specified number of random samples to each of several predetermined strata. In a CDF, strata might consist of areas of sand versus silt, or inundated areas versus exposed areas. When strata are known to affect the distribution of parameters of interest, the use of stratified random sampling can be more efficient than simple random sampling and will reduce variability. **Systematic** sampling locates samples along a linear transect or at the nodes of a geometric grid. This type of sampling is generally necessary for geostatistical analyses such as Kriging and evaluation of spatial variability, which often require many more samples than might typically be collected in a random design.

Random and systematic sampling are examples of probabilistic sampling designs that enable calculation of uncertainty and allow statistical inferences. A general disadvantage of both random and systematic sampling is that some of the selected locations may be difficult or impossible to sample. **Judgmental** sampling, which is non-probabilistic, locates samples based on criteria such as convenient access or desire to sample specific locations based on historical information about the site. Judgmental sampling is inherently biased and does not produce statistically valid measurements. Use of judgmental sampling in CDF characterization would mean, for example, that comparisons of contaminant concentrations with remediation criteria would not be valid statistically; however, such samples are amenable to interpolation and geostatistical analysis and could be used to map the distribution of the remediation material. Random, systematic, judgmental, and other sampling designs are covered in considerable detail in USEPA (2002).

Random sampling is acceptable when a detailed characterization of the sampling area is not needed, delineation of specific subregions is unnecessary, and the spatial location of any data point is not of interest. It is generally more efficient than systematic sampling (i.e. produces a lower standard error) (Rasemann 2000).

Composite sampling. Several discrete samples can be combined to form a composite that is more representative of the mean concentrations of analytes of interest over the area sampled, and will reduce error due to spatial variability (Jenkins et al. 1999). (Note that vertical compositing is also often used to estimate mean properties of materials in the vertical plane; however, compositing over selected vertical intervals is usually desirable where a more representative picture of the vertical distribution of materials is desired.) Combining many discrete samples into a few composites can substantially reduce sample analysis cost. Allocating samples from a defined area to more than one composite allows estimation of the composite mean and variability for that area, but the relationships among multiple analytes in individual discrete samples will be

lost, as will the ability to map spatial distributions of analytes. Compositing can increase the amount of material available for sample analysis, in turn lowering method detection limits and reducing the rate of false negatives (Garner et al. 1996). However, compositing can also lead to dilution of the analyte below the limit of detection; this is a particularly important consideration when samples are taken for the purposes of criteria comparisons, and when the criterion is near the detection limit of the analyte. Thus, the use of compositing over either the horizontal or vertical planes must be considered carefully in view of the data objectives.

***Multi-increment* (MI) Sampling.** While the typical composite sample is composed of a few discrete samples, many more discrete samples (increments) are needed to make up the MI sample. In a simulation study using munitions residue data, Jenkins et al. (2005) found that composite samples of n = 5 increments had a range (maximum observation divided by minimum observation) greater than a factor of eight. However, with n = 30 increments or more, the range was less than a factor of three, the mean and median agreed closely, and the frequency distribution of repeated mathematically generated composites was much more normal; hence the recommendation to use 30 or more increments in the MI sample. The increments are collected randomly or systematically within a defined area ("decision unit"), combined and thoroughly mixed, then spread out and systematically subsampled to obtain the material that will be analyzed (Hawaii HEER 2008). MI sampling is appropriate for easily collected surface samples when the average analyte concentration in the decision unit is needed. To estimate variability, three field replicate MI samples per decision unit are recommended (Hawaii HEER 2008). MI sampling is not suitable for determining maximum or minimum concentrations, or spatial distribution of analytes within a decision unit.

SAMPLING CONSIDERATIONS AND INTERPOLATION TOOLS FOR QUANTIFICATION OF DREDGED MATERIAL IN CDFs: In determining the amount and distribution of material (e.g., sand) available for recovery from the CDF, it will be necessary to collect discrete samples for sediment grain size analysis. To illustrate the sampling concepts and interpolation tools, sediment characterization data from 12 core samples from the Chicago Area CDF at the mouth of the Calumet River in Lake Michigan, Illinois (Figure 1) are used to demonstrate multiple approaches to assess sand distribution and estimate volume. This CDF is a lake-margin facility serving as a disposal area for sediment from the Chicago River and Harbor and the Calumet River and Harbor. The total capacity of this 43-acre (208,120-yd^2 or 174,015-m^2) facility is 1,300,000 yd^3 (993,921 m^3), which would suggest an average total sediment thickness of approximately 19.2 ft (6.4 yd or 5.9 m). Core positions are indicated by the red symbols within the interior of the contoured areas in Figure 2.

All available data and information regarding a site should be considered in deriving a reasonable estimate of resource content including, but not limited to, the following:

- Bathymetry or topography prior to sediment disposal and at the time of characterization will define the multidimensional extent of the materials.
- Sediment placement history – including the location, amount, and nature of sediment being disposed – will assist in interpreting observed data and in guiding the visualization tools toward consistent realizations.

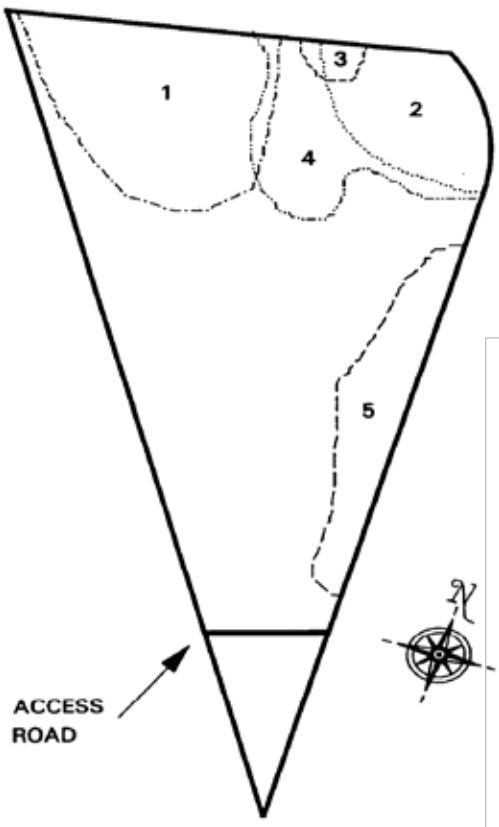

Figure 1. Chicago Area CDF aerial photo (above; from 2008 fact sheet) and the locations of five disposal events (#1-4 described in Dorkin et al. 1988; #5 described in Lutz et al. 1991).

Key: For each numbered area on the map view (left; from Lutz et al. (1991)), the sediment volume (yd^3 1,300,000 capacity), source, and year of placement (if known) are as follows:
1-100, Calumet River, 1984-85
2-100, Calumet River, 1984-85
3-5, Demonstration project
4-70, Chicago River/Harbor, 1986
5-N/A, Debris (pilings, concrete, rock)

- Physical characterization of the sediment, e.g., particle size distribution (sand, silt, clay), bulk density, porosity.
- Chemical characterization of the sediment, e.g., contaminant concentrations, organic content, or redox condition may be of concern.

If available data do not support a satisfactory analysis of CDF content, the preliminary analysis may still be useful in guiding collection of new data, e.g., to fill data gaps or better delineate sand-rich areas.

There are no standard sampling protocols for CDF material quantification, and limited examples of CDF sampling to provide insight. However, the following are some relevant considerations:

- Site history, including placement method and discharge points
- Historical character and variability of channel sediments
- Settling behavior of various sediment fractions; expected degree of size separation during placement
- Characterization objectives
- Sampling and analytical budget

For the sake of this analysis, an awareness of the CDF construction and discharge history is assumed. This awareness affords at least a qualitative concept of total sediment thicknesses across the CDF. Assuming that discharged sediments are heterogeneous in grain size, the

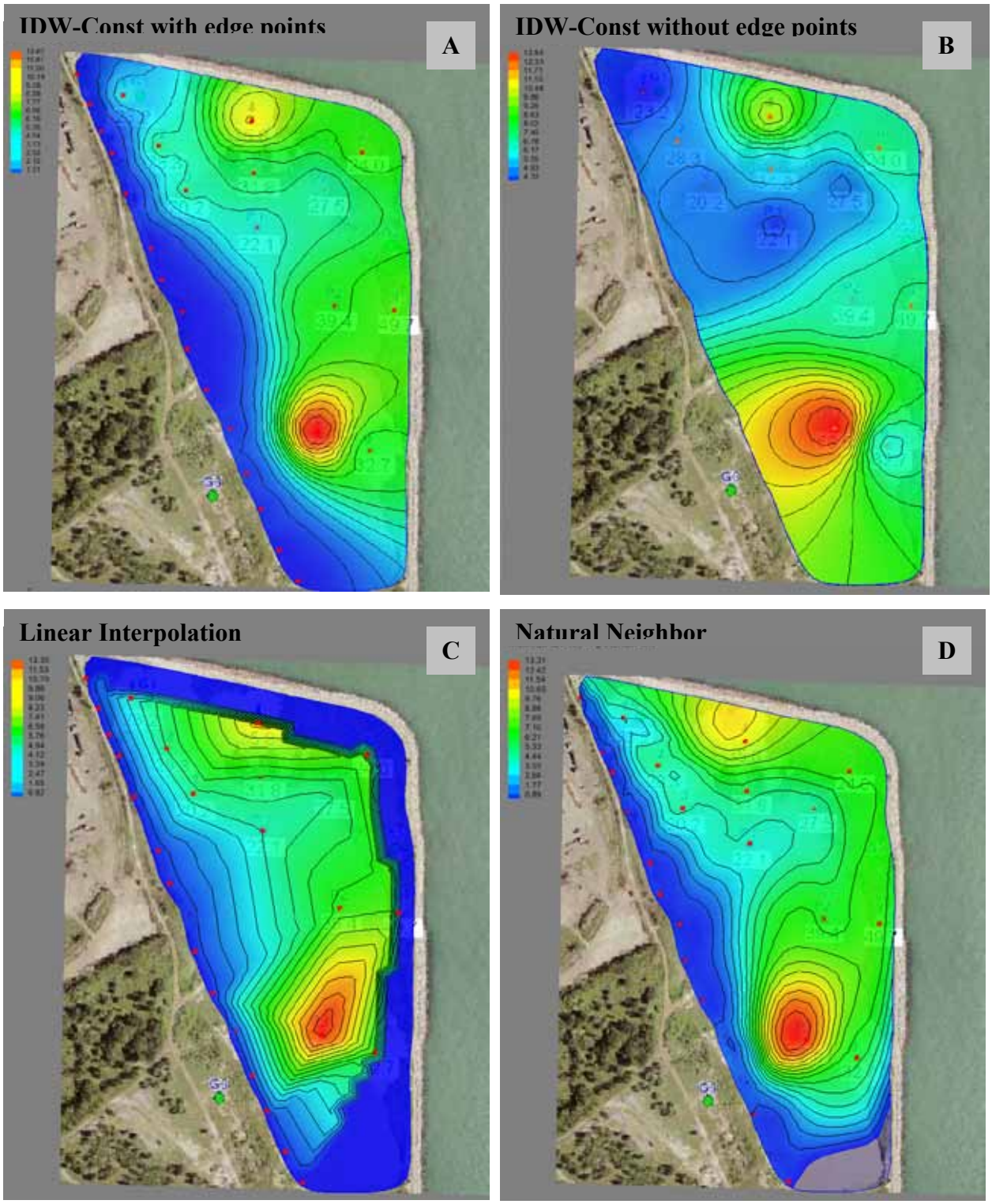

Figure 2. Comparison of GMS (version 7.1) interpolation schemes for the Calumet CDF sand isopach data. (A) Inverse distance weighted. (B) Inverse distance weighted, as in 'A', but without synthetic data points on the west edge to constrain extrapolation; all other realizations include the synthetic data. (C) Linear interpolation. (D) Natural neighbor interpolation. Interior points indicate core locations; marginal data points represent synthetic data.

sediment settling proximal to discharge points is likely to be relatively coarse-grained, while finer sediments would tend to settle more slowly at greater distances. Some level of subjectivity is unavoidable in developing a sampling plan, starting with the level of accuracy required (directly proportional to the sample density and number), as well as in prioritizing sample locations. Highest priority, and approximately 50 percent of the samples, should be assigned to locations with the greatest potential return of desired information. In quantification of CDF sand content, highest priority would be given to the positions believed to be the sandiest and/or thickest. In the Chicago Area CDF, this would be weighted toward the north and east margins. Medium priority (~30 percent of the sites) would be given to points that will help delineate the extent of sand-rich areas. Lowest priority would be given to positions expected to be least sandy and/or thinner, which would tend to be the western and southern regions in the Chicago Area CDF (~20 percent of the sites). Samples would likely be collected during a single field excursion, particularly if mobilization of sampling equipment is expensive. The sampling plan for the Chicago Area CDF was quite good in meeting the priorities described above. However, the absence of data in the central area and along the westernmost and southernmost margins requires more extrapolation, which contributes to increased uncertainty.

Interpolation tools. Given sufficient data, computer interpolation tools can be used to estimate volume of material and develop visualizations of material distribution or even contaminant concentration contours within the CDF. This discussion will illustrate interpolation tools provided in the Department of Defense (DoD) Groundwater Modeling System (GMS) and Surface-water Modeling System (SMS). Similar contouring tools are available in other commercial, groundwater modeling suites (e.g., Groundwater Vistas, Visual Modflow, Argus ONE, 3D Master) and more generic suites (e.g., Surfer, ArcGIS), but the GMS and SMS suites are available to DoD employees at no charge (also to contractors for DoD project applications).

Six interpolation methods are available in GMS that serve to project two- or three-dimensional point data onto a network of points suitable for visualization, analysis, or model preparation. These interpolation schemes include: (1) Linear; (2) Inverse distance weighted, IDW; (3) Clough-Tocher; (4) Natural neighbor; (5) Kriging; and (6) Gaussian field generator. All but the linear method require the selection or development of options and parameters too numerous to discuss here. GMS includes default options and values where appropriate, as well as tools with which to generate the required characterization (e.g., variograms for Kriging). Brief discussions of each method and options are presented in the online documentation wiki (Aquaveo 2010). All of these methods are derived from the GSLIB geostatistics library, with more rigorous descriptions of methods presented in its user's guide (Deutsch and Journel 1998).

The accuracy of any interpolated realization is generally greatest within the physical space defined by the extent of the data points (assuming high quality data). Most often, however, the area of interest extends beyond the range of data. The trends defined by the interpolation method, conditioned on the available data, are extended by extrapolation beyond the data range to the margins of the user-defined domain. Such extrapolation can generate counterintuitive or nonsensical realizations. Users can impose reasonable and justifiable constraints to control such extrapolations. Constraining limits are often intuitively obvious, such as near-zero thickness values at the margins of a CDF, which can be implemented by supplementing the measured data with synthetic data at the margins.

Data from the Chicago Area CDF will serve to demonstrate three interpolation methods. The original coring dataset includes total thickness (ΔZ) and the vertically averaged weight-percent sand (%S) at 12 locations within the CDF. The point symbols in the interior of the CDF (Figure 2) represent locations of the cores. The symbols at the west margin of the CDF represent the synthetic data points used to constrain interpolation with minimal distortions (e.g., kinks) in natural-looking contours. Local total sand thickness (T_s) is calculated ($T_s = \Delta Z \cdot \%S$), neglecting any dependence of bulk density on sediment grain size distribution. The areal distribution of T_s constitutes a sand isopach map as shown in Figures 2 and 3.

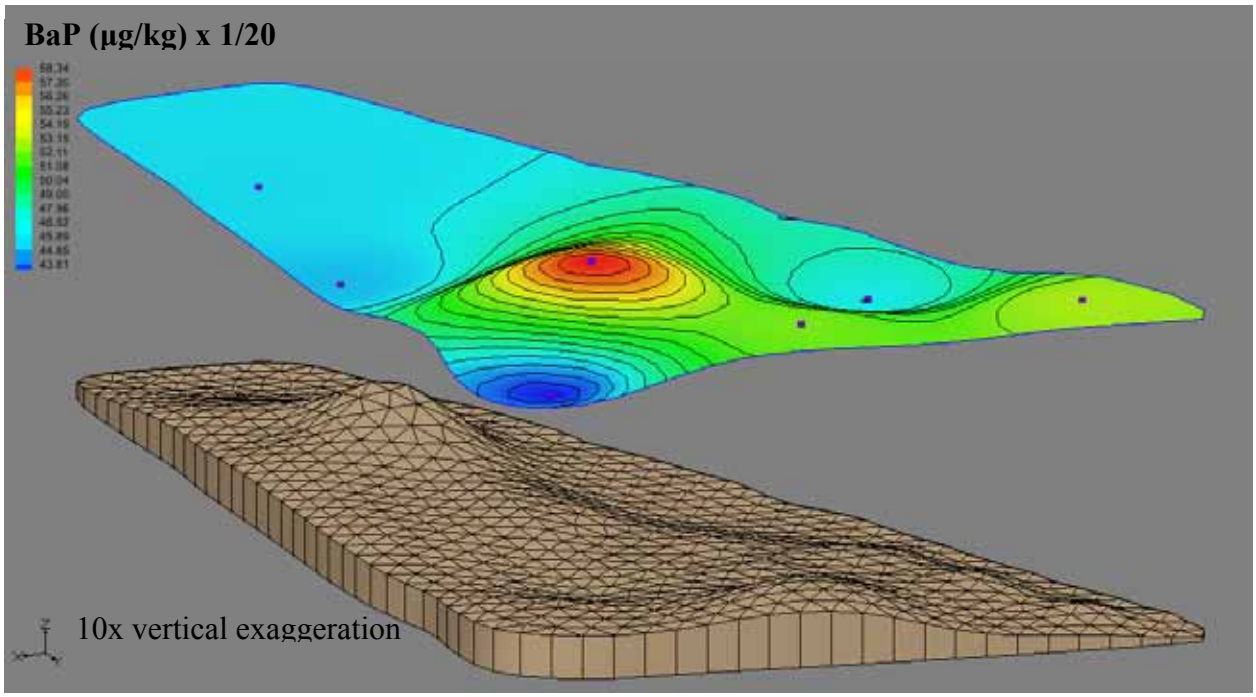

Figure 3. 3D visualization of sand isopach (solid below) and benzo(a)pyrene distribution (contoured surface), looking toward land. An estimate of over 254,000 yd^3 (194,197 m^3) of sand is conditioned on the IDW interpolation scheme (as in Figure 2A). Note: 10x vertical exaggeration; BaP concentrations (C_{BaP}, µg/kg) are scaled as (0.01 C_{BaP} + 40) for visualization purposes.

The inverse distance weighted (IDW) method applied in Figures 2A and 2B is one of the more commonly used interpolation tools. The IDW employs a constant nodal function (Shepard's method), interpolated using up to 16 nearest data points. These default options for IDW generally yield a reasonable interpolated realization. Figure 2B uses the same IDW options as in A, but without the synthetic data points on the west edge to constrain extrapolation. Note that the effect of unconstrained extrapolation is to greatly expand the already suspect peak in the southern part of the CDF to impossible levels, given that the sediments are thin at the west margin.

Linear interpolation yields very angular contours (Figure 2C) because this method cannot extend beyond inter-point areas, i.e., cannot extrapolate. Synthetic data points would be required along the domain margins, as done along the western edge, to achieve complete coverage. There are no secondary options with the GMS linear interpolation method. Although the realization looks unnatural, there is no extrapolation, and it does provide a stark visual reminder that this is an approximation, not reality, which "prettier" interpolations tend to disguise.

Natural neighbor interpolation (Figure 2D) is similar to IDW, but with subtle differences in the weighting of surrounding data points in the interpolation. This particular realization used a gradient nodal function conditioned on all scatter points, with a bounding limit of 10 percent beyond the convex hull (default settings). These settings generate negative values, which are counterintuitive here. In the contoured image, the negative values were truncated by limiting the contour range to positive values, making the resemblance to the first IDW image more evident.

Which contouring approach is best can only be tested objectively by: (a) quantification of the sand content during recovery for beneficial use, or (b) testing realizations with a second set of infilling sand-content data. The second option is more useful because it could be attempted prior to excavation. A set of training data would be used to develop alternative realizations. A separate, smaller set of testing data would be used to quantify how well alternative interpolations predicted the test set. This general approach of leaving out one or more data points to test interpolation scheme alternatives is referred to as cross validation or jackknife methods (Deutsch and Journel (1998) and most textbooks on geostatistics). If all samples are collected in a single field sampling, two or three of the medium to low priority data points might serve as the test set. However, if one has the luxury of collecting new data in a second field sampling, that event could also serve to resolve gaps or suspect data in the initial data and realizations.

Estimation of the total sand content within a CDF may be achieved in GMS by several approaches. The simplest method would be to define one TIN (triangulated irregular network) using the isopach values, another identical TIN at a zero reference horizon, and then calculate the volume between the two TINs. This is readily achieved in GMS by defining a 3D domain (referred to as a *solid*) between the two TINs. The volume of the solid will be among the characterization data calculated by GMS. Figure 3 shows the resulting isopach map in 3D (the tan solid), representing only the sand content as if it were placed on a flat surface. Note that the sand thickness decreases toward shore (the long distal edge in the isopach image). The network of triangles on the top surface is the sand isopach TIN; an identical flat TIN is on the underside of the solid; both are interpolated using the IDW method. GMS estimates 6,859,456 ft^3 (or 254,053 yd^3 or 194,237 m^3) of sand based on the available data set. The accuracy of such estimates is limited by the quality and extent of the characterization data. The color-contoured surface in Figure 3, representing the areal distribution of benzo(a)pyrene (BaP, μg/kg) in bulk sediment, is included only as an example of how GMS can be used to visually map the spatial correlation of contaminant and sediment type.

SAMPLING STRATEGIES FOR ESTIMATING MEAN CONTAMINANT CONCENTRATIONS

Approach. Sediment contaminant concentration data from the Chicago Area CDF, and from Grassy Island CDF in the Detroit River, Michigan, were used to evaluate several possible CDF sediment sampling strategies with the ultimate goal of estimating mean contaminant concentrations for comparison with remediation criteria. The original data sets consisted of multiple contaminant measurements in seven discrete samples from the Chicago Area CDF and 41 discrete samples from Grassy Island CDF. Composite, multi-increment, and discrete sampling strategies were evaluated using bootstrap resampling, a technique in which a data set is randomly

resampled a large number of times (e.g., 1,000) with replacement[1] in order to calculate statistics of interest. The goal of this exercise was to determine which sampling strategy would optimize the combination of low cost, low bias, and high precision in estimating mean contaminant concentrations within the CDF.

Since the original data sets were analyzed only as discrete samples, composites were simulated by randomly assigning a specified number of sampling locations from the CDF to each composite and averaging the data from those locations to obtain the concentration for the composite. For the Chicago Area CDF, 2- and 3-composite strategies were tested. In the 2-composite strategy, for example, one of the two composites would average data from three of the seven sampling locations, while the other composite would average data from the remaining four locations. The Grassy Island CDF with its much larger number of sampling locations allowed evaluation of 2-, 3-, 4-, 7-, 11-, and 20-composite sampling strategies. MI sampling, for which at least 30 increments per composite have been recommended, could only be tested with the Grassy Island data. In each test, the averaged composite concentrations were used to calculate a mean concentration and variance for the CDF. The MI compositing process was repeated three times to obtain three MI replicates for calculation of the CDF mean and variance.

Since the true value of any analyte concentration in the CDF is unknown, the original discrete sample mean was taken as the best estimate of the true value and bias was therefore defined as the absolute value of the percent difference between the composite mean and the original discrete sample mean relative to the discrete mean:

$$\text{Bias} = | (\text{Mean}_{\text{composite}} - \text{Mean}_{\text{discrete}})100 / \text{Mean}_{\text{discrete}}| \quad (1)$$

By definition, bias of the original discrete sample mean = 0. The process of randomly assigning sampling locations to composites and calculating the composite mean, variance, and bias was repeated 1,000 times for each analyte to obtain an average mean, variance, and comparative bias for each sampling strategy. Precision was evaluated by dividing the square root of the average variance by the square root of the sample size N (where N = number of composites or MI replicates) to obtain the standard error of the mean. Coefficients of variation were also calculated by dividing the average standard deviation (square root of the average variance) by the mean.

Estimated costs for the various sampling strategies are shown in Table 2, based on an average per-sample collection cost of $9000 for high-cost sampling (machine coring); average total sample collection costs for low-cost sampling (hand coring) of $4100 (1 to 30 samples), $5700 (31 to 60 samples), or $7300 (61 to 90 samples); and average per-sample analysis cost of $750 for a typical suite of analytes including grain size, organic carbon, metals, PAHs and PCBs. Note that the cost differences between discrete and composite samples shown in Table 2 reflect the savings in analytical costs achieved by compositing.

[1] Each data point may be resampled multiple times.

Table 2. Estimated total cost for sample collection and analysis.

Sampling Strategy	Cost
Chicago Area CDF (high-cost coring)	
7 discrete samples	$68,250
2 composites[a]	$64,500
3 composites[a]	$65,250
Grassy Island CDF (low-cost coring)	
41 discrete samples	$36,450
2 composites[b]	$7,200
3 composites[b]	$7,950
4 composites[b]	$8,700
7 composites[b]	$10,950
11 composites[b]	$13,950
20 composites[b]	$20,700
Multi-Increment[c]	$9,550

[a] 7 discrete samples randomly assigned to composites
[b] 41 discrete samples randomly assigned to composites
[c] 30 discrete samples randomly assigned to each of 3 MI replicates

Finally, to compare the quality of the parameter estimates obtained with the different sampling strategies, an "Inefficiency Function" (IEF) was created, incorporating bias, cost, and precision (estimated by standard error of the mean) as follows:

$$IEF = Cost * StdErr * (1 + Bias) \quad (2)$$

For convenience of presentation and subsequent analyses, the IEF was scaled down by 10^{-6} for high-cost sampling and 10^{-5} for low-cost sampling.

In addition to the composite and MI sampling strategies, the Grassy Island CDF data were also used to compare the performance of smaller numbers of discrete samples to that of the 41 discrete samples that were originally collected. This was done by randomly selecting 3, 4, 5, 6, 8, 10, 12, 15, 20, 25, 30, 35, or 40 of the 41 sample locations and repeating this process 1,000 times to calculate statistics and determine bias, precision, and IEF.

RESULTS: Average bias and standard error for the composite, discrete, and MI sampling strategies are shown in Figure 4 using Grassy Island CDF arsenic data as an example. The bootstrap means and standard errors for the composite sampling schemes are consistent with each other, and are also similar to the original 41 discrete samples. Bootstrap means for reduced discrete sample sizes[1] show some inconsistency and the standard errors increase as sample size decreases. MI sampling results in the smallest standard error of all the sampling strategies. Average bias is lowest for the composite sampling strategies, slightly higher for the MI samples, and increasingly problematic as discrete sample size gets smaller.

Figure 5 displays the estimated sampling and analysis cost and the inefficiency function using the same sampling strategies and Grassy Island arsenic data. The Grassy Island samples were obtained using a low-cost sampling method (hand-held coring). Clearly, cost increases primarily with the number of samples analyzed. The IEF, which integrates cost, precision, and bias, was

[1] Number of samples.

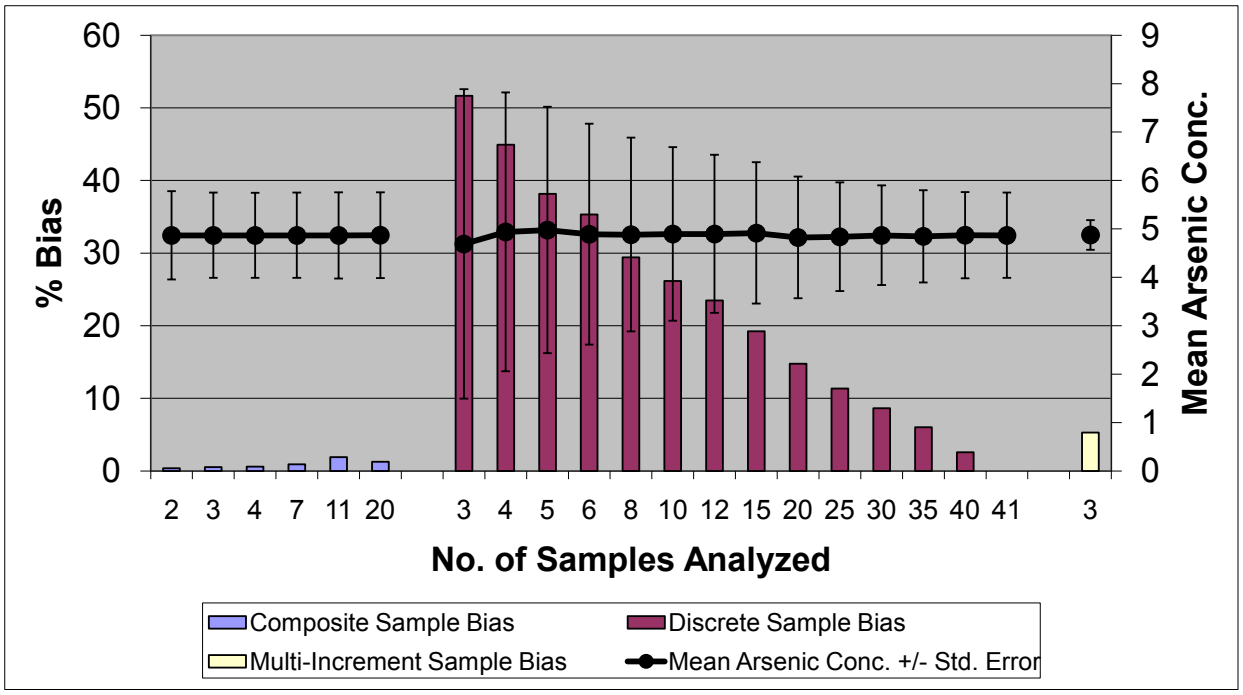

Figure 4. Bias and precision of sampling strategies (composite, discrete, and MI) using arsenic data from Grassy Island CDF. Means, standard errors and percent bias were calculated using bootstrap resampling for all sampling strategies except the original 41 discrete sample strategy (which was the basis for calculation of bias in parameter estimates obtained using the other sampling strategies). Histograms along base of figure use % Bias scale (left); black dots with standard error bars use scale to the right.

lowest for the 2-, 3-, 4- and 7-composite, and MI sampling strategies. Even though the lowest number discrete sampling strategies (3 and 4 samples) have a slight cost advantage over the cheapest compositing strategy, that advantage is more than offset by the huge increase in bias and loss of precision.

The patterns observed in Figures 4 and 5 were consistent for all Grassy Island sediments contaminants that were examined. In the Chicago Area CDF, seven discrete sediment samples were collected for contaminant analysis, using a high-cost sampling technique (track-mounted hollow-stem auger drill rig). This method enables collection of core samples up to several meters in length so that sub-surface sediment characterization is possible. The limited number of discrete samples allowed evaluation of only the 2- and 3-composite sampling strategies using bootstrap re-sampling. The IEF and percent bias are illustrated for lead in Figure 6. Although seven discrete samples resulted in a slightly higher estimated cost than either the two-or three-composite sampling strategies (Table 2), compositing produced greater bias and higher IEFs. Had a low-cost sampling method been used, the comparative results would have been the same (Figure 6). The same patterns hold for other individual contaminants and for the average of all detected contaminants in the Chicago Area CDF samples.

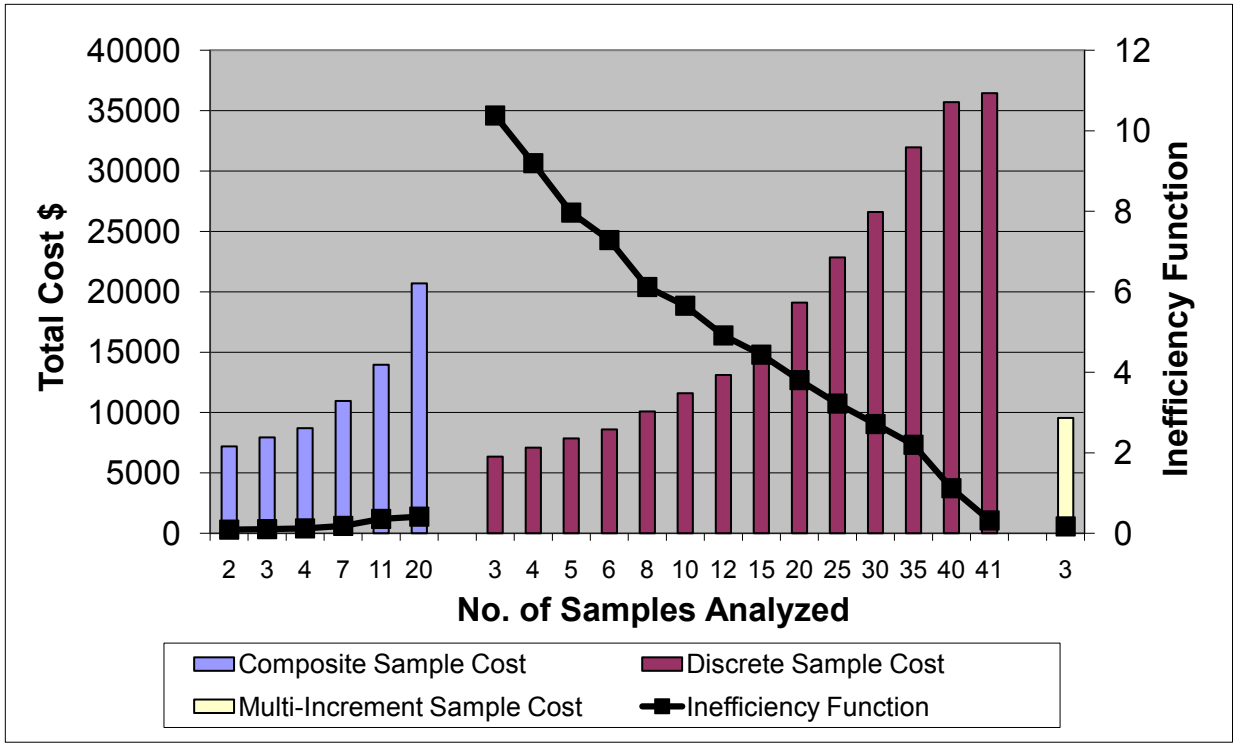

Figure 5. Cost and Inefficiency Function of sampling strategies (composite, discrete, and MI) using arsenic data from Grassy Island CDF. Histograms along base of figure use cost scale (left); black squares use scale to the right.

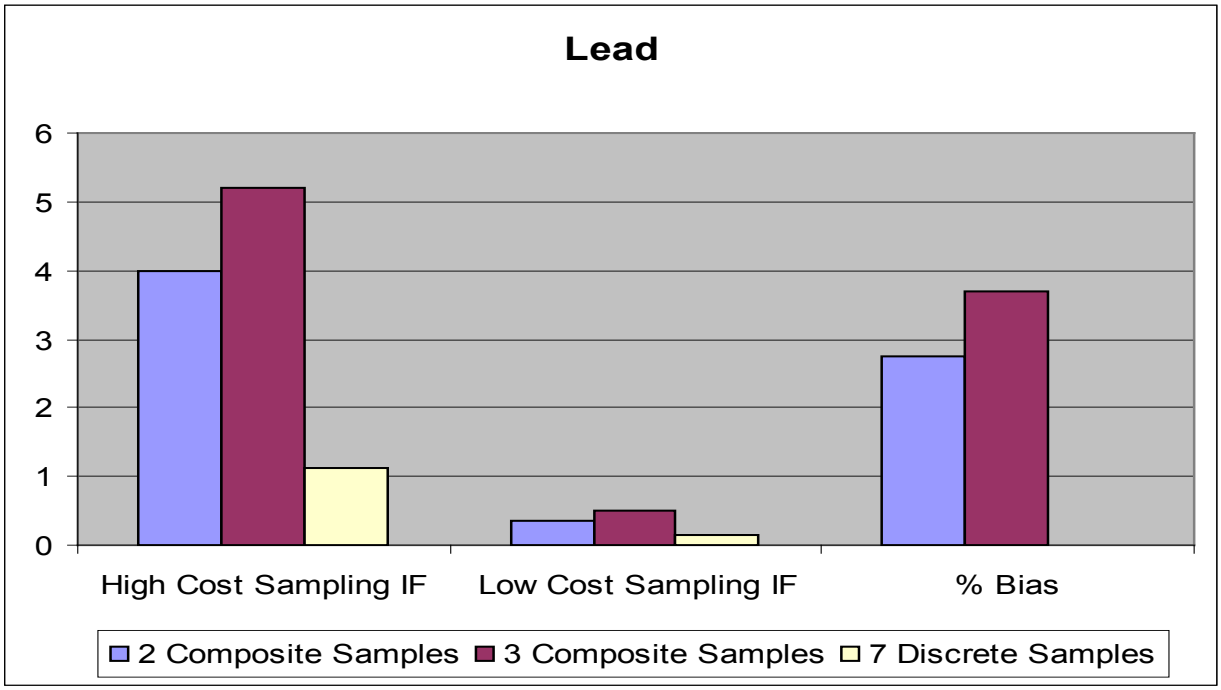

Figure 6. Inefficiency Function (IEF) and percent bias of sampling strategies using lead data from the Chicago Area CDF.

Effective sample size[1] for compositing. Bootstrap analyses of the Grassy Island CDF data demonstrated that compositing of samples into two to seven composites, or into three MI replicates, was certainly more cost-effective and slightly more efficient than analysis of 41 discrete samples for estimation of mean analyte concentrations, based on low-cost sample collection methods. On the other hand, similar analyses of the Chicago Area CDF data, where high-cost sampling had been employed, showed that analysis of seven discrete samples was more efficient than either the 2- or 3-composite sampling strategy. Compositing results in savings primarily in the cost of sample analysis, rather than collection. Therefore, with high-cost sampling methods, little or no benefit is likely to be realized by compositing, unless many samples are combined into few composites, or sample analysis costs are very high, as in the case of dioxin analyses. Otherwise, the slight cost advantage of compositing is likely to be offset by increased bias and possibly reduced precision in the estimation of endpoints. Bootstrap analyses of 33 analytes from the Chicago Area CDF showed that bias in the estimation of the mean ranged from 0.8 to 13.8 percent for two composites, and 1.1 to 15.6 percent for three composites as compared to means estimated from the seven discrete samples (where bias was assumed = 0). Bootstrap standard errors of estimated mean analyte concentrations were also greater as a result of compositing as compared with the seven discrete samples, by as much as 5.4 percent for two composites and 4.7 percent for three composites.

2-composite sampling strategy. Contaminant concentration data from the Grassy Island and Chicago Area CDFs were analyzed using bootstrap resampling, together with cost figures for low-cost sample collection, to determine how many samples (i.e. *x* number of discrete samples combined into two composites) would need to be collected for the 2-composite sampling strategy to become more efficient than analysis of the discrete samples. To determine this, the average bootstrap IEF for the 2-composite strategy was calculated at a given sample size and compared with the IEF for the same total number of discrete samples. The process was then repeated while increasing the sample size[1] until the composite sample IEF was lower than its respective discrete sample IEF, indicating the minimum sample size needed for compositing to be more efficient than analysis of discrete samples. Minimum sample sizes were determined for individual analytes and plotted against the corresponding coefficient of variation (CV) for that analyte (Figure 7). Figure 7 illustrates that as analyte variability increased, there was also an increase in the minimum sample size required before compositing became more efficient than discrete sample analysis. (The CV was utilized in Figure 7 rather than the associated IEF for each analyte, which varied by orders of magnitude.) For all but the most highly variable analytes, compositing became more efficient with a minimum of 11 to 15 samples total.

Three-composite sampling strategy. Using the 3-composite sampling strategy, somewhat more samples would be required to achieve greater efficiency than analysis of discrete samples. For most analytes, a total of 13 to 19 samples would be needed for compositing to be more efficient (Figure 8).

[1] Number of samples.

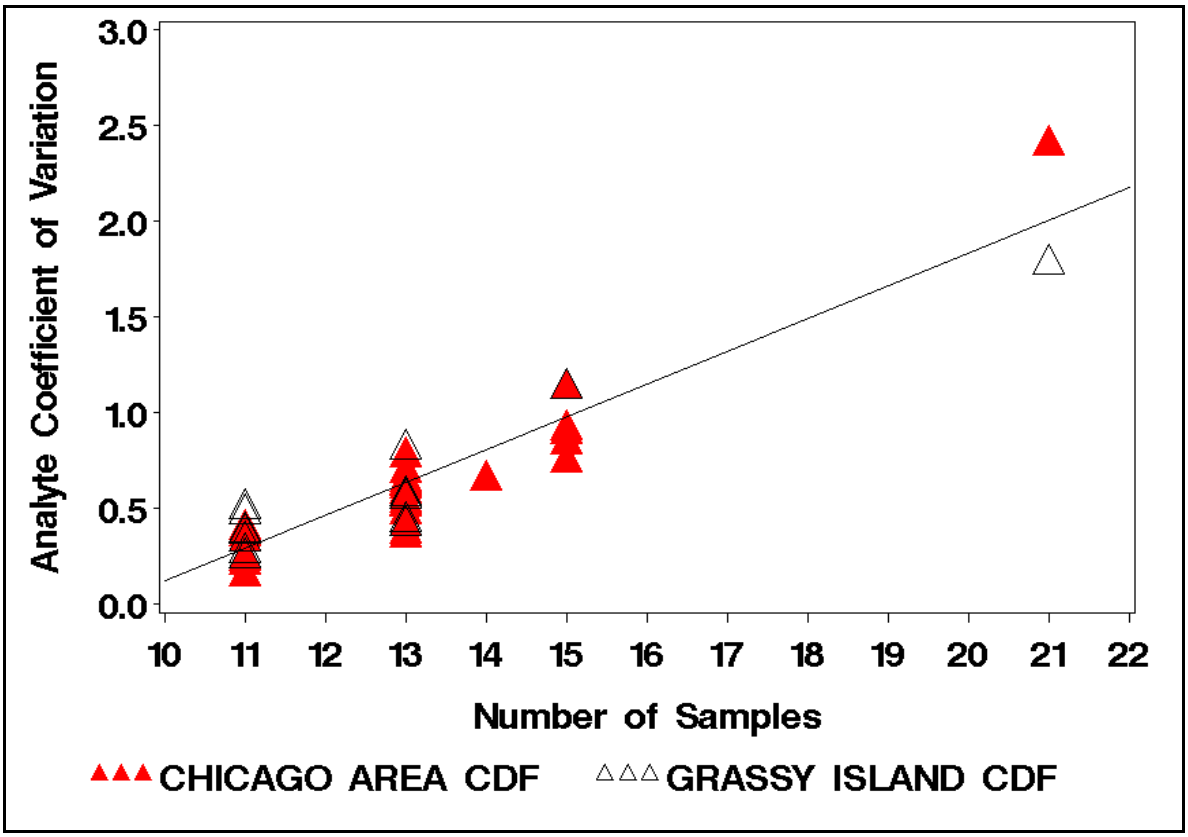

Figure 7. Two-composite sampling strategy - minimum number of samples at which the 2-composite sampling strategy becomes more efficient than analysis of discrete samples, shown as a function of analyte coefficient of variation. The line is the linear regression for all points on the graph; the regression equation is given in Table 3.

Table 3. Predictive equations for determining minimum number of samples (N) for compositing to be more efficient than discrete sampling given analyte coefficient of variation (CV) (n is number of analytes used to develop each regression).

Number of Composites	Equation	R^2	n
2	N = 9.77 + 5.09 * CV	0.87	48
3	N = 10.06 + 7.96 * CV	0.85	46
4	N = 11.53 + 8.63 * CV	0.90	15
7	N = 16.95 + 9.27 * CV	0.93	15
11	N = 19.34 + 21.58 * CV	0.68	13

Figure 9 compares various metrics for the 2-, 3-, 4-, 7-, and 11-composite sampling strategies using bootstrapped cadmium data from the Grassy Island CDF. The CV for cadmium from the complete suite of 41 discrete samples was 0.6. N is the minimum number of samples required for the composite sampling strategy to be more efficient than analysis of discrete samples, as determined using IEF and assuming a low-cost sample collection method. N increases from 13 samples total for two composites to 34 samples total for 11 composites. The standard error for both the discrete and composite samples decreases slightly as the number of samples increases, but differs little between the discrete and composite sampling strategies for the sample numbers

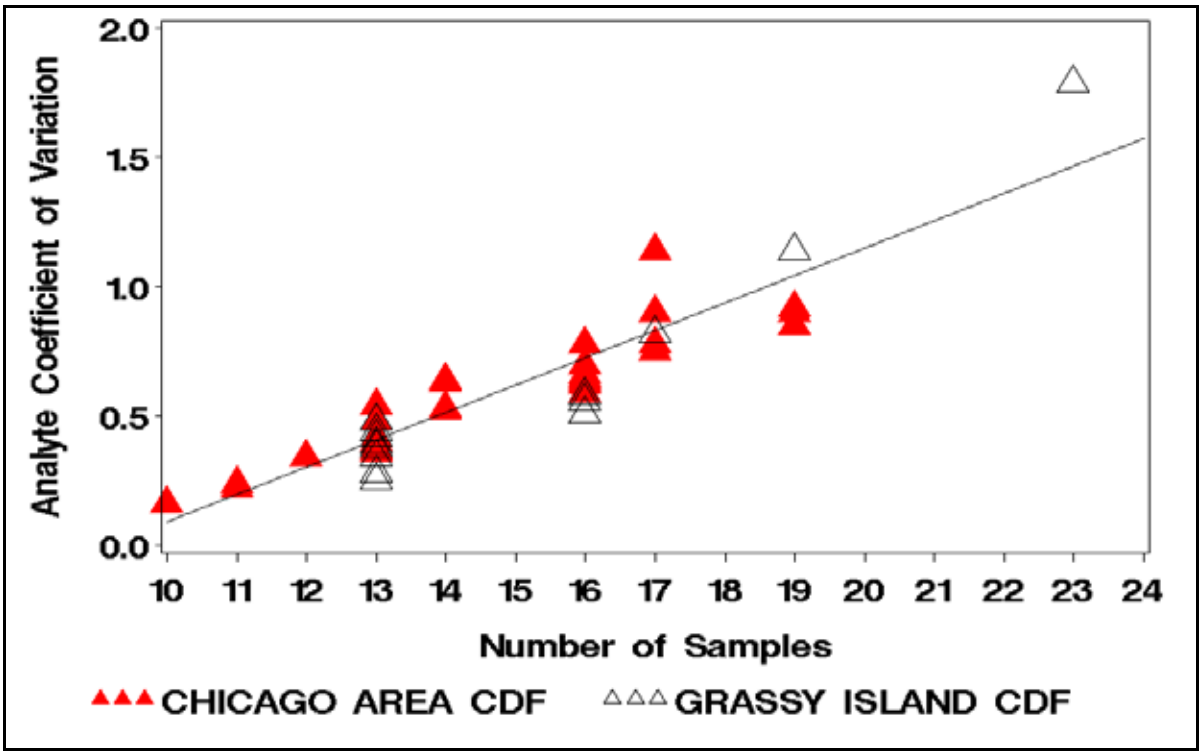

Figure 8. Three-composite sampling strategy - minimum number of samples at which the 3-composite sampling strategy becomes more efficient than analysis of discrete samples, shown as a function of analyte coefficient of variation. The line is the linear regression for all points on the graph; the regression equation is given in Table 3.

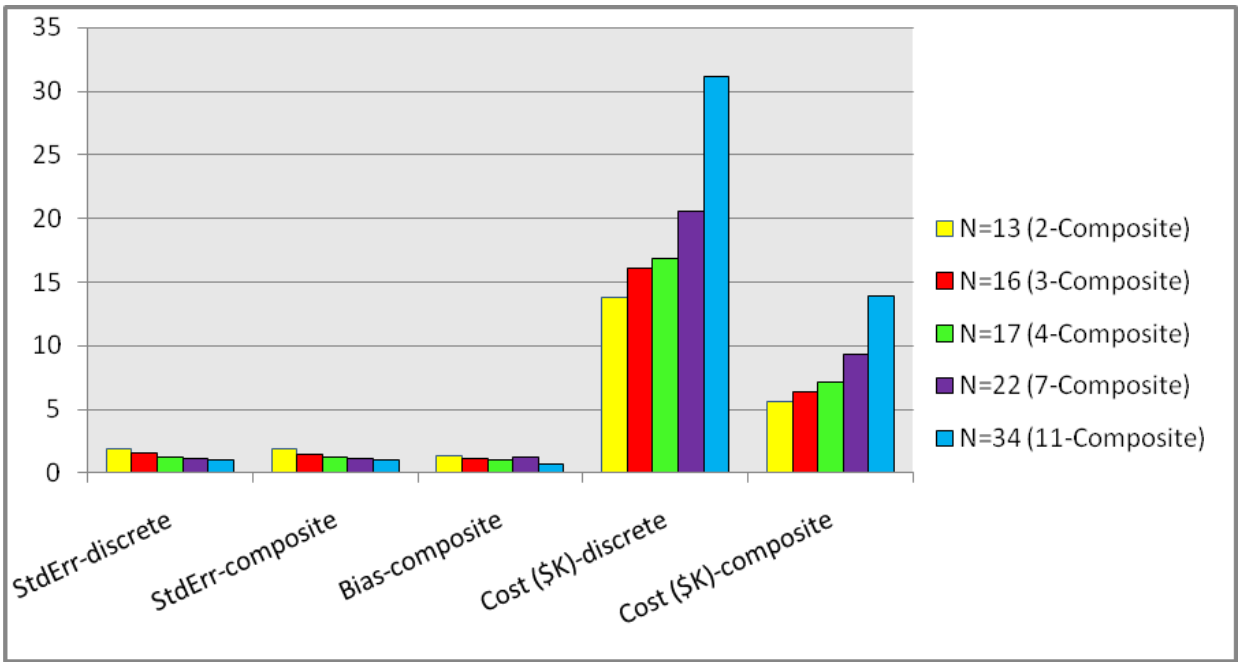

Figure 9. Comparison of bootstrap metrics for discrete and several composite sampling strategies using cadmium data from Grassy Island CDF.

evaluated. Bias introduced by compositing (as compared to the discrete sample means) also tends to decrease somewhat as the number of composites increases (recall that bias is assumed to be zero for the discrete sampling strategy). Cost, of course, increases with increasing sample numbers, but is decidedly lower for the composites than for discrete sampling.

Using the combined bootstrap results for the Grassy Island and Chicago Area CDF data, linear regression equations were developed for predicting the required minimum number of samples needed for a composite sampling strategy to be more efficient than discrete sampling, given an estimate of analyte CV (Table 3). The required minimum number of samples is shown in Table 4 for estimated CVs of 0.25, 0.5, 1 and 2, which represent a typical range from very low to very high analyte variability. The average CV of all analytes from the Grassy Island and Chicago Area CDFs combined was 0.65, which would require a minimum of 13 samples for a 2-composite sampling strategy to be more efficient, or 15 samples for a 3-composite sampling strategy to be more efficient, assuming low-cost sampling.

Table 4. Minimum number of samples required for compositing to be more efficient than discrete sampling.

Number of Composites	Analyte Coefficient of Variation			
	0.25	0.5	1.0	2.0
2	11	12	15	20
3	12	14	18	26
4	14	16	20	29
7	19	22	26	35
11	25	30	41	63

EFFECT OF SAMPLING STRATEGY ON COMPARISONS WITH CRITERIA: In CDF sediment characterizations, it may be necessary to compare estimated contaminant concentrations with applicable screening or remediation criteria. Contaminant levels in the CDF could be judged too high for beneficial reuse of material if a specified parameter such as the mean, or even the upper 95-percent confidence limit (UCL95) of contaminant concentrations, exceed given thresholds. Since compositing and MI sampling strategies may introduce bias into the estimation of the mean and standard error compared with randomized discrete sampling, the outcome of comparisons with criteria could also be different for these sampling strategies than for discrete sampling. To evaluate compositing effects, bootstrap re-sampling was used with the Chicago Area CDF and Grassy Island CDF data to generate 1,000 estimates of the UCL95 for each contaminant using each of the sampling strategies described above. The bootstrap UCL95's and the original discrete sample UCL95 were then compared with one or more freshwater sediment criteria. Criteria were obtained from several sources (Manny 1999; *SQuiRT@NOAA.gov*[1]; Estes and Clarke 2011). Criteria were chosen to be as close as possible to the discrete sample UCL95, because it was observed that in this region small differences in UCL95 estimates could result in different outcomes in the criteria comparison (i.e. different

[1] *http://response.restoration.noaa.gov/topic_subtopic_entry.php?RECORD_KEY(entry_subtopic_topic)=entry_id,subtopic_id,topic_id&entry_id(entry_subtopic_topic)=783&subtopic_id(entry_subtopic_topic)=5&topic_id(entry_subtopic_topic)=2*

pass/fail determinations). For a given contaminant, the percent disagreement between the outcome of the original discrete sample UCL95 comparison with a criterion and that of another sampling strategy was recorded. For example, if the discrete sample UCL95 was below the criterion but 235 out of 1,000 bootstrap UCL95s for a composite sampling strategy exceeded the criterion, the percent disagreement for the composite sampling strategy would be 23.5. In the same example, if the discrete sample UCL95 was instead above the criterion, the percent disagreement with the bootstrapped composites would be 76.5 (100 – 23.5). Percent disagreement between the 2-composite sampling strategy and the original discrete samples is shown in Figure 10.

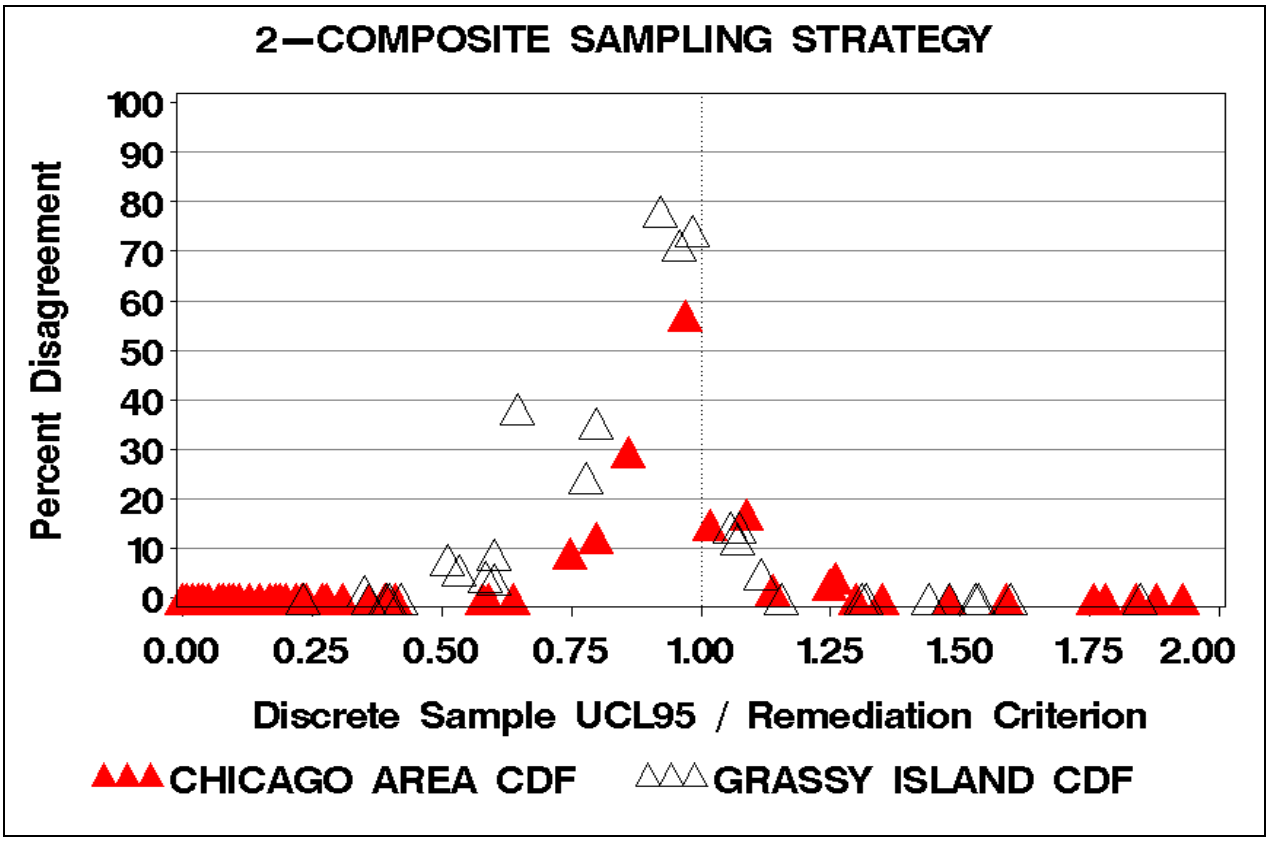

Figure 10. Percent disagreement between the 2-composite sampling strategy and the original discrete samples in comparisons of various analyte UCL95s with contaminant remediation criteria. The horizontal axis is the ratio of the discrete sample UCL95 to the remediation criterion, i.e. data points to the left of 1.00 on the x-axis represent values of the discrete sample UCL95 that were below the selected criteria, while data points to the right represent values of the discrete sample UCL95 that were above the criteria.

When the ratio of the discrete sample UCL95 to the remediation criterion is less than 0.5 or greater than 1.3, the pass/fail conclusions resulting from the criteria comparisons would be the same for the 2-composite sampling strategy as for the discrete sampling strategy. However, as the ratio approaches one, the percent disagreement between the two sampling strategies increases greatly and the 2-composite sampling strategy is much more likely to produce a UCL95 that exceeds the criterion, for the datasets evaluated. When the ratio of the discrete sample UCL95 to the criterion slightly exceeds one (resulting in a "fail"), there is about a 15-percent likelihood that

the 2-composite sampling strategy will produce a UCL95 lower than the criterion (i.e., a "pass"). As the number of composites in a sampling strategy increases, the amount of disagreement with the discrete sample UCL95 to criterion comparisons diminishes; see, for example, the 7-composite sampling strategy results in Figure 11. For all compositing strategies, the greatest disagreement occurs where the ratio of the UCL95 and the selected criterion is near 1.00.

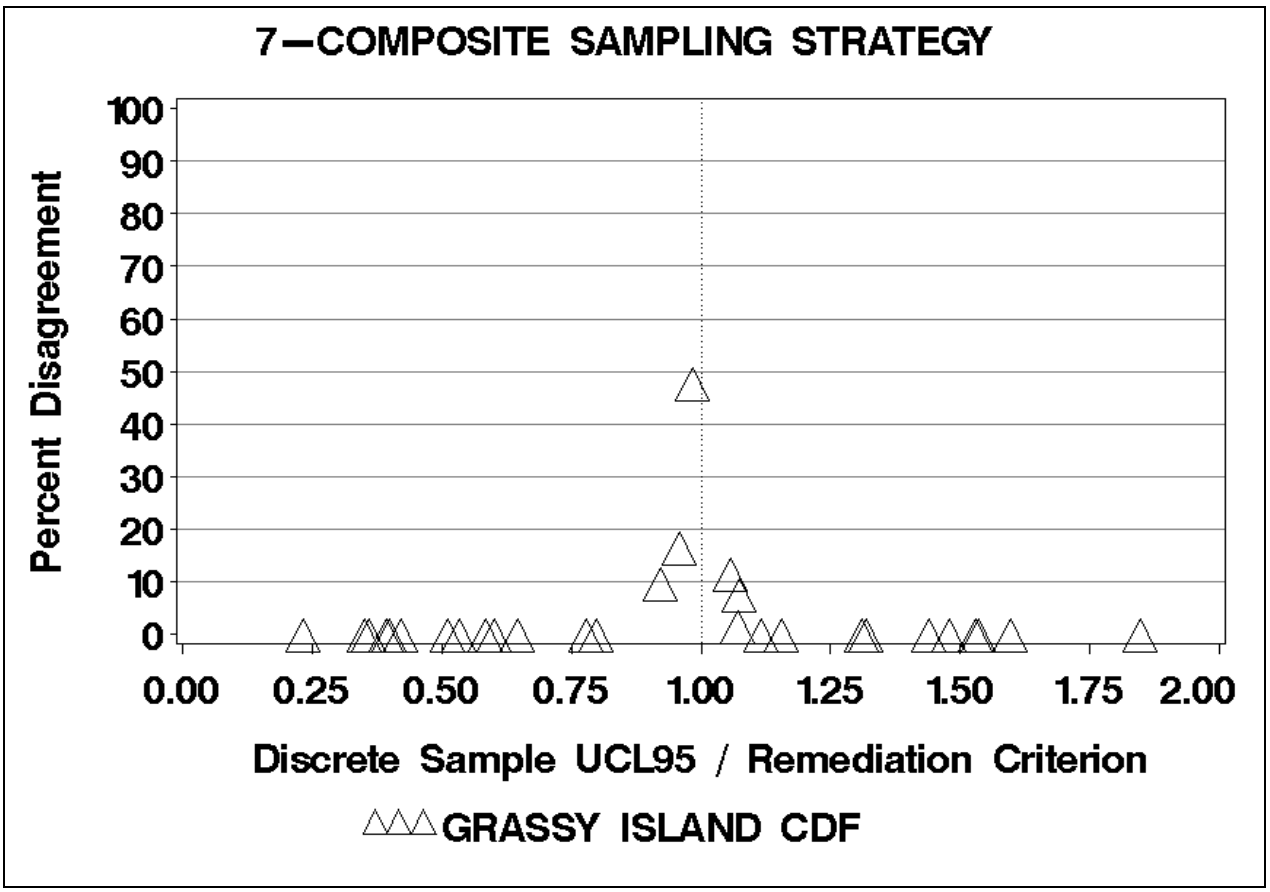

Figure 11. Percent disagreement between the 7-composite sampling strategy and the original discrete samples in comparisons of various analyte UCL95s with contaminant remediation criteria. The horizontal axis is the ratio of the discrete sample UCL95 to the remediation criterion.

There is a large disparity between MI and discrete sample UCL95 comparisons with criteria, but only when the ratio of the discrete sample UCL95 to the criterion is slightly greater than one (Figure 12). Because MI sampling increases precision by reducing the standard error, the UCL95 will also be reduced compared with the discrete sample UCL95. Thus, MI UCL95s are less likely to exceed criteria than discrete sample UCL95s. From a practical perspective, discrete sample UCL95's appear to be more conservative than MI UCL95s. Similarly, discrete sample UCL95s appear to be generally less conservative than composite UCL95s. This would not be problematic in cases where all samples "pass," but might require further re-sampling and re-assessment based on discrete or MI samples where samples "fail." This analysis could thus inform subsequent field sampling efforts, compositing, and data analysis.

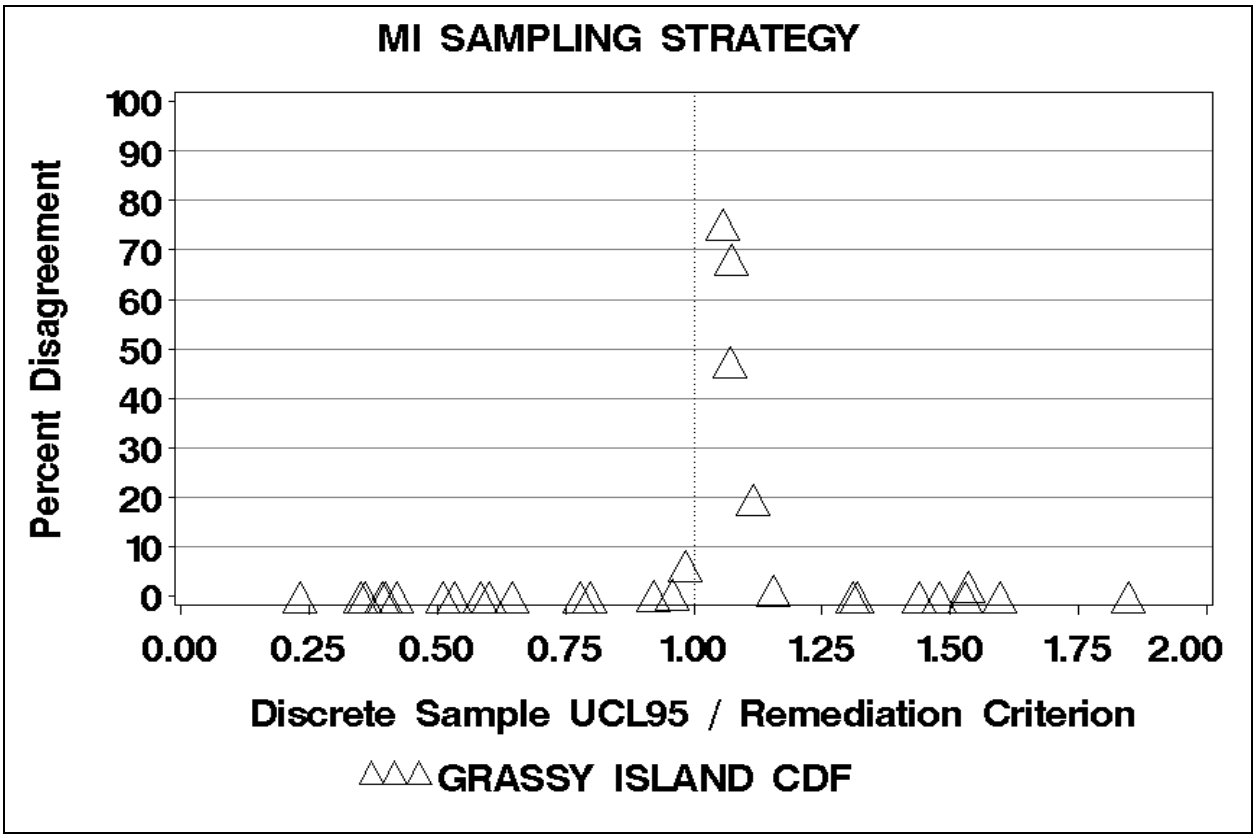

Figure 12. Percent disagreement between the MI sampling strategy and the original discrete samples in comparisons of various analyte UCL95s with contaminant remediation criteria. The horizontal axis is the ratio of the discrete sample UCL95 to the remediation criterion.

Discrete sample size[1]. Bootstrap analyses of contaminant data from the Grassy Island CDF using reduced numbers of discrete samples compared with the original 41 discrete samples clearly showed profound negative consequences of smaller sample sizes (Figures 4 and 5). As sample size decreased, precision decreased (i.e. standard error increased) and bias increased rapidly. Despite lower cost, discrete sample sizes less than 41, with higher standard errors and bias, would be much less likely to provide an accurate or precise estimate of average contaminant concentration in the CDF than the original 41 samples. But is analysis of all 41 discrete samples necessary? Using Table 4, and considering that some analytes had high CVs (>1), the 41 samples could efficiently be assigned to seven or fewer composites, for an analytical cost savings of at least $25,500. Using fewer composites reduces cost, but increases the likelihood that the outcome (pass/fail) of contaminant comparisons with criteria will differ from discrete sample comparisons with those criteria.

DISCUSSION: Sample size[1] formulas were given in a previous Technical Note (Estes and Clarke 2011) for determining the total sample size needed for comparison with a criterion given a specified difference from the criterion, or for estimating a mean within a specified margin of error. CDF characterizations may require obtaining information about a large number of sediment constituents, including grain size, organic carbon, and various contaminants. Sampling

[1] Number of samples.

for grain size is intended to provide information for estimating volume of material of interest and visualization of its distribution within the CDF, and possibly for correlation with organic carbon content and contaminant concentration data. Although sample size formulas are less pertinent for these applications, most likely the same samples collected for grain size analysis will also be analyzed for other parameters such as mean contaminant concentrations where the formulas are applicable.

Sample size can be calculated based on either the most important or the most variable of the constituents, using historical data to obtain the necessary estimates of variability. In the absence of historical data, sample size can be approximated from Table 5, based on an assumed CV and the desired maximum acceptable relative error in estimating the mean for any given constituent. Although most constituents will have CVs less than one, there likely will be a few with CVs approaching or even exceeding 1.5 or 2. As seen in Table 5, the need to accurately estimate means for highly variable constituents will require large sample numbers. The sample sizes in Table 5 were calculated using the following equation, derived from Equation 2 in Estes and Clarke (2011):

$$n \geq \frac{z_{1-a/2}^2 CV^2}{d^2} + 0.5 z_{1-a/2}^2 \quad (3)$$

where:

z = standard normal deviate or z-score corresponding to a selected confidence probability 1-α

CV = decimal coefficient of variation of the analyte of interest

d = maximum acceptable error in the estimate, expressed as a decimal (e.g., 0.1 = 10-percent error)

$0.5 z_{1-\alpha}^2$ = correction factor for small sample size and the use of the sample standard deviation in place of the unknown population standard deviation in the calculation of CV

Table 5. Minimum total sample size required so that acceptable relative error in the estimation of mean analyte is not exceeded, given analyte coefficient of variation and 95-percent confidence level.

Acceptable Relative Error[a]	Analyte Coefficient of Variation				
	0.25	**0.5**	**1.0**	**1.5**	**2.0**
0.1	26	98	387	867	1539
0.25	6	18	64	141	248
0.5	3	6	18	37	64
1.0	3	3	6	11	18

[a] Expressed as fraction of analyte mean, i.e. 0.1 indicates an acceptable relative error of +/- 10% of the mean, 1.0 indicates an acceptable relative error of +/- 100% of the mean.

Once the optimum total sample size has been determined, selection of a sampling strategy for CDF characterization requires careful consideration of the relative advantages and disadvantages of the

various strategies as described in this Technical Note. The primary considerations are the cost of sampling and analysis, and the accuracy and precision of estimated endpoints. Certainly cost is an overriding concern, but cutting costs by taking too few samples is a false economy that can virtually guarantee inaccurate results and wide variability. In this Technical Note, the Inefficiency Function (IEF) is presented as a simple metric to simultaneously evaluate cost, inaccuracy, and imprecision for the various sampling strategies. Increasing values of IEF represent increasing disadvantage in terms of higher cost, bias, and variability. Extensive data sets from two CDFs are employed to demonstrate IEF and the comparative performance of various sampling strategies. Selection of a sampling strategy will depend in large measure on whether the CDF characterization objectives require high-cost sampling methods such as a track-mounted auger drill rig for vertical sediment profiling, or whether shallow cores or surface samples are sufficient and relatively low-cost sample collection methods such as hand augers can be used.

High-cost sampling methods (deep cores) likely will be needed in most CDFs, where the sediment strata to be characterized are relatively thick. When the per-sample cost of sample collection greatly exceeds the per-sample cost of subsequent analyses, there will generally be no benefit to compositing unless: (1) a large number of samples are combined into few composites, and (2) the bias in estimates of mean values introduced by compositing is negligible, or (3) standard errors are greatly reduced by compositing. Bootstrap analyses of Chicago Area CDF data and Grassy Island CDF data gave no indication that either of the latter conditions was true for these datasets. MI sampling does reduce standard errors but so many samples (increments) are required that cost likely would be prohibitive for deep cores.

Low-cost sampling methods (shallow cores or surface samples) may be sufficient in CDFs where the sediment strata of interest are relatively thin and/or relatively uniform, such that the vertical profile is well represented by these shallow samples. When the per-sample cost of sample collection is more on a par with that of subsequent sample analyses, compositing can be efficient, providing total sample size meets a certain threshold. Those thresholds are given in Table 4 based on analyte CV, and on representative estimated costs for hand coring and for chemical analytical laboratory analyses of a typical suite of sediment analytes. (Site-specific costs should be applied where available.) Compositing produces relatively low bias, and consistent means and standard errors regardless of the number of composites. Cost increases about 10 percent with each additional composite, but the IEF increases only slightly with increasing number of composites. The analyses presented in this Technical Note would suggest the use of two or three composites when compositing is indicated. A disadvantage of compositing is that information about individual sample locations is lost. For example, compositing would not differentiate a localized contaminant hot spot that would potentially need to be managed separately from the remainder of the CDF. However, neither would discrete sampling unless sampling intensity were sufficient to detect and delineate the hot spot, or to allow reasonably accurate modeling of contaminant spatial distribution.

When total sample size is low, location of hot spots is important, or individual sample sites need to be characterized, discrete samples must be analyzed. As total sample size decreases, cost obviously decreases, but there are severe trade-offs: the mean eventually becomes unstable and the IEF actually increases because the variance increases and bias (inaccuracy) increases exponentially.

MI sampling is limited to situations where a large number of samples can easily be collected in a systematic fashion from a defined decision unit, combined and thoroughly mixed, and subsampled. This process results in low variability, but the analyses summarized in this Technical Note showed higher bias than typical compositing. While the IEF was low, total cost was higher than the 2-, 3- or 4-composite sampling strategies.

Note that if a compositing or MI strategy is to be employed to improve sampling efficiency and reduce the cost of subsequent chemical analyses, it would be advantageous to remove subsamples for grain size analysis from the original discrete samples before they are combined to form the composites. Discrete samples with their spatial coordinates are needed for interpolation and visualizations of the material of interest. While there is no minimum sample size for these exercises, increasing numbers of samples will provide increased reliability and confidence in the accuracy of visualizations and volume estimates. Twelve core samples from the Chicago Area CDF were sufficient to estimate sand volume and construct sand isopach maps, but the reliability of at least one sand peak in the map was suspect and additional data would have been desirable to test the realizations constructed by the interpolation software.

Nonnormal data and nondetects. The analyses described in this Technical Note generally assume uncensored data (i.e., without nondetects) from normally distributed populations. However, contaminant concentrations often are skewed and frequently are censored. When the CV > 1, nonnormality should be suspected. Data can be checked for nonnormality by visual inspection, using graphics such as box plots and Q-Q plots, or by testing, using procedures such as Shapiro-Wilk's test. Sources of skewness can include an underlying lognormal population, the random occurrence of a few high outliers, or the mixture of more than one population (e.g., samples from a low level background population mixed with samples from a high concentration contaminated area). Although the median is a better descriptor of central tendency than the mean in nonnormal populations, comparisons with criteria are generally based on the mean and upper confidence limit. With skewed data, nonparametric methods can be used to obtain more reliable estimates of the mean and UCL than standard arithmetic computation. Singh et al. (1997) describe several such methods, including the H-statistic, jackknife, bootstrap, Central Limit Theorem, and Chebychev Theorem methods. The presence of nondetects (censored data) greatly complicates statistical analyses because traditional, easy methods of handling nondetects such as substituting half the detection limit, have statistically undesirable ramifications. A program is available for free download from the USEPA ("Scout 2008," USEPA (2008), *http://www.epa.gov/esd/databases/scout/abstract.htm*), which offers a wide variety of classical and robust statistical methods for environmental analyses. Scout also includes ProUCL (Singh et al. 2007), which contains an array of parametric and nonparametric UCL computation methods for uncensored or censored data sets.

SUMMARY: Adequate and considered sampling is essential in CDF characterization, whether for the determination of general sediment characteristics such as grain size and organic carbon content, the delineation of the specific material to be removed from the CDF for subsequent beneficial reuse, or the estimation of mean concentrations of contaminants of concern. Careful consideration should be given to the number of samples required for adequate characterization, and to the type of sampling strategy that will most efficiently achieve the objectives of the characterization, based on the sample collection method to be employed and what is known

about the variability of the constituents to be characterized. Sampling efficiency can be defined in terms of cost of collection and analysis, as well as accuracy and precision in the estimation of desired endpoints. This Technical Note presents a simple function, the Inefficiency Function (IEF), that incorporates cost, bias (inaccuracy), and standard error (variability, or imprecision). The IEF can be used for comparison of competing sampling strategies. The sampling strategies evaluated herein include analysis as discrete samples; combining samples into two or more composites; and a special type of composite known as a multi-increment (MI) sample, consisting of 30 or more discrete subsamples combined according to a specific protocol.

For a given sampling and analysis budget, when the per-sample cost of the sample collection method is high, as when machine coring is used for vertical profiling of the sediment, the most efficient sampling strategy likely will be analysis of discrete samples. When the cost of the sample collection method is low relative to the cost of subsequent sample analysis, as when sampling the top meter of sediment using hand augers, compositing can be a useful way to improve sampling efficiency by allowing a greater total number of samples to be taken without a corresponding increase in analytical costs. Two or three composites are likely sufficient unless the area to be characterized is large or complex. Generally, compositing will only be more efficient than discrete sample analysis when a minimum of 13 to 19 total samples is planned, with the minimum increasing as the variability of the data and the number of composites increase. MI sampling, though highly recommended for certain venues like military firing ranges where surficial samples can easily be collected, may be considered but likely is impractical for CDF characterization because of the need to characterize the vertical profile in the CDF, the high cost of collecting sufficiently large numbers of cores to satisfy the MI procedure requirements, and the logistical difficulty in physically combining large volume samples.

Discrete samples and their spatial coordinates are needed for volume estimation and spatial distribution visualization of the material of interest in the CDF, using geostatistical interpolation tools such as GMS and SMS.

ACKNOWLEDGEMENTS: The Chicago Area CDF data set was supplied by the U.S. Army Engineer District, Chicago, and the Grassy Island CDF data were obtained from Dr. Stephanie Millsap, U.S. Fish and Wildlife Service, Grosse Ile, MI.

POINTS OF CONTACT: For additional information, contact Dr. Trudy J. Estes, research civil engineer (601-634-2125, *Trudy.J.Estes@usace.army.mil*); Joan U. Clarke, statistician (601-634-2954, *Joan.Clarke@usace.army.mil*); Christian J. McGrath (601-634-3798, *Chris.McGrath@usace.army.mil*) or the program manager of the Dredging Operations and Environmental Research Program, Dr. Todd S. Bridges (601-634-3626, *Todd.S.Bridges@usace.army.mil*). This technical note should be cited as follows:

> Clarke, J. U., C. McGrath, and T. J. Estes. (2011). *Sampling strategies for Confined Disposal Facilities (CDF) characterization.* DOER Technical Notes Collection. ERDC TN DOER-D12. Vicksburg, MS: U.S. Army Engineer Research and Development Center.

REFERENCES

Aquaveo. 2010. *GMS: Interpolation*. On-line documentation available at the following URL: http://www.xmswiki.com/xms/GMS:Interpolation [9/14/10]

Deutsch, C. V., and A. G. Journel. 1998. *GSLIB: Geostatistical Software Library and User's Guide, 2nd ed.,* Oxford: Oxford University Press.

Dorkin, J., P. Ross, M. S. Henebry, J. Miller, and M. Wetzel. 1988. Biological and toxicological investigations of Chicago area navigation projects. *Illinois Natural History Survey Contract Report*, 1-150. [digital version available at http://www.ideals.illinois.edu/handle/2142/15217]

Estes, T. J., and J. U. Clarke. 2011. *CDF characterization for beneficial reuse of dredged material—establishing precedents based on a case study*. DOER Technical Notes Collection. ERDC-TN-DOER-D11. Vicksburg, MS: U.S. Army Engineer Research and Development Center.

Flatman, G. T., E. J. Englund, and A. A. Yfantis. 1988. Geostatistical approaches to the design of sampling regimes. In *Principles of environmental sampling, 1st ed*., L. H. Keith, 73-83. Washington, DC: American Chemical Society.

Garner, F. C., M. A. Stapanian, and L. R. Williams. 1996. Composite sampling for environmental monitoring. In *Principles of Environmental Sampling, 2nd ed.*, ed. L. H. Keith, 679-690. ACS Professional Reference Book. Washington, DC: American Chemical Society.

Hawai'i Office of Hazard Evaluation and Emergency Response (HEER). 2008. Section 4. Soil Sample collection approaches. In *Technical Guidance Manual for the implementation of the Hawai'i State Contingency Plan*. http://www.hawaiidoh.org/

Jenkins, T. F., A. D. Hewitt, M. E. Walsh, T. A. Ranney, C. A. Ramsey, C. L. Grant, and K. L. Bjella. 2005. Representative sampling for energetic compounds at military training ranges. *Environmental Forensics* 6:45-55.

Jenkins, T. F., C. L. Grant, M. E. Walsh, P. G. Thorne, S. Thiboutot, G. Ampleman, and T. A. Ranney. 1999. Coping with spatial heterogeneity effects on sampling and analysis at an HMX-contaminated antitank firing range. *Field Analytical Chemistry and Technology* 3(1):19-28.

Keith, L. H., G. L. Patton, D. L. Lewis, P. G. Edwards. 1996. Determining what kinds of samples and how many samples to analyze. In *Principles of environmental sampling, 2nd ed.*, ed. L. H. Keith, 3-40. ACS Professional Reference Book. Washington, DC: American Chemical Society.

Lamé, F., T. Honders, G. Derksen, and M. Gadella. 2005. Validated sampling strategy for assessing contaminants in soil stockpiles. *Environmental Pollution* 134(2005):5-11.

Lutz, C. H., J. M. Brannon, and F. J. Reilly, Jr. 1991. *Influence of sediment properties on bioaccumulation potential of PCBs: Field studies at the Calumet confined disposal facility*. Environmental Effects of Dredging Technical Notes Collection. EEDP-02-16. Vicksburg, MS: U.S. Army Engineer Waterways Experiment Station. [http://el.erdc.usace.army.mil/elpubs/pdf/eedp02-16.pdf]

Manny, B. A. 1999. *Contamination of surface soils and wildcelery tubers at Grassy Island in the Wyandotte National Wildlife Refuge in the Detroit River, Michigan*. Completion Report, Research Partnership Program – FY 1997. Kearneysville, WV: U.S. Geological Survey, Biological Resources Division (Eastern Region).

Rasemann, W. 2000. Industrial waste dumps, sampling and analysis. In *Encyclopedia of analytical chemistry*, ed. R. A. Myers, 2692-2719. Chichester: John Wiley & Sons Ltd.

Singh, A., R. Maichle, A. K. Singh, S. E. Lee, and N. Armbya. 2007. *ProUCL Version 4.00.02 User Guide*. EPA/600/R-07/038. Las Vegas, NV: U.S. Environmental Protection Agency, Office of Research and Development, Las Vegas, NV. www.epa.gov

Singh, A. K., A. Singh, and M. Engelhardt. 1997. *The lognormal distribution in environmental applications*. EPA/600/R-97/006. Las Vegas, NV: U.S. Environmental Protection Agency, Office of Research and Development.

U.S. Environmental Protection Agency (USEPA). 2002. *Guidance on choosing a sampling design for environmental data collection for use in developing a quality assurance project plan*; EPA QA/G-5S. EPA/240/R-02/005. Washington, DC: Office of Environmental Information.

U.S. Environmental Protection Agency (USEPA). 2008. *Quick overview, Scout 2008 Version 1.0; A statistical software package for the classical and robust analysis of univariate and multivariate data sets with and without non-detect observations*. EPA/600/F-08/016. Office of Research and Development. www.epa.gov

U.S. Environmental Protection Agency/U.S. Army Corps of Engineers (USEPA/USACE). 1991. *Evaluation of dredged material proposed for ocean disposal—Testing Manual*. EPA-503/8-91/001. Washington, DC.

U.S. Environmental Protection Agency/U.S. Army Corps of Engineers (USEPA/USACE). 1998. *Evaluation of dredged material proposed for discharge in waters of the U.S.—Testing Manual*. EPA-823-B-98-004. Washington, DC.