A REVIEW OF MINESOIL SAMPLING AND SPATIAL VARIABILITY IN TEXAS[1]

by

J.C. Myers, J.E. Brandt, and W.L. Joseph[2]

Abstract. Accurate characterization of minesoil overburden constituents associated with strip mining is an important part of the pre- and post-mine regulatory process. Characterization of soil material requires sampling of some kind, which implies that 1) the sample material selected must be representative of the area to be characterized and 2) the sample volume (support), size (number of samples), and pattern must be able to support a reasonable decision making process. Therefore, the end use (baseline information, monitoring, or remediation) of this information should dictate the sampling approach; which in turn, is based on the decision to be made and the amount of uncertainty that is allowable. Uncertainties and errors are an integral part of the sampling, laboratory analysis, and spatial characterization processes, arising at each stage. Mathematical approaches such as sampling theory & practice and geostatistics can quantify the amount of error or uncertainty associated with the various stages of sampling, analysis and characterization, as well as distinguishing sampling errors from laboratory errors. These statistical tools can be used to manage errors and uncertainties at each stage of the process, providing confidence to 1) regulatory agencies that compliance has been achieved and 2) mining companies that unnecessary remedial costs will not be incurred. Statistical tools provide a framework for characterizing the wide variety of minesoil constituents and conditions encountered in mine operations. The use of statistically-defined monitoring or remedial decision units of a given area, for example, 5.7-acre grids in Texas, are shown to be consistent with the United States Protection Agency's long-standing guidelines and recommendations for remedial activities. Site-specific variability must always be taken into account when designing a sampling program and caution is recommended in the selection of sampling methods (i.e. compositing versus discrete samples).

Additional Key Words: Data quality objectives, exposure unit, management unit, composite sampling, grid sampling, volume-variance relationship, and error management

Introduction

The establishment of a uniform sampling methodology for minesoils is a difficult task due to the potential variability encountered at a specific site and variability between different sites. Given this, regulatory perspectives in minesoil monitoring and evaluation must account for both the unique conditions and attributes at each site, as well as different conditions between sites. Characteristics of specific sites should be investigated and incorporated into the minesoil monitoring and remedial decision-making processes for those sites. Sites exhibiting more complicated situations can be expected to require greater levels of effort; whereas, sites exhibiting low levels of variability will probably require less scrutiny and sampling.

Analysis of the spatial variability at a particular site provides valuable insights and helps quantify the amount of sampling that will be required to reduce the uncertainty in decision-making to acceptable levels. Geostatistical techniques, such as *variograms* and *kriging*, are valuable tools in assessing both short- and long-range spatial variability. Each minesoil constituent can be evaluated separately, in order to determine the unique nature of its variability within its site-specific context. Figure 1 presents a geostatistically-created graph showing what earth sci-

[1] Paper presented at the 1997 National Meeting of the American Society for Surface Mining and Reclamation Annual National Meeting, Austin, Texas, May 10-15, 1997.

[2] Jeffery C. Myers is GIS/GEM Manager, Morrison Knudsen Corp., P.O. Box 400090, San Antonio, TX 78229; Jon E. Brandt is Soil Scientist, Railroad Commission of Texas, P.O. Box 12967, Austin, TX 78711-2967; and William L. Joseph is Geologist, Office of Surface Mining, 501 Belle Street, Alton, IL 62002.

Proceedings America Society of Mining and Reclamation, 1997 pp 115-127
DOI: 10.21000/JASMR97010115

https://doi.org/10.21000/JASMR98010115

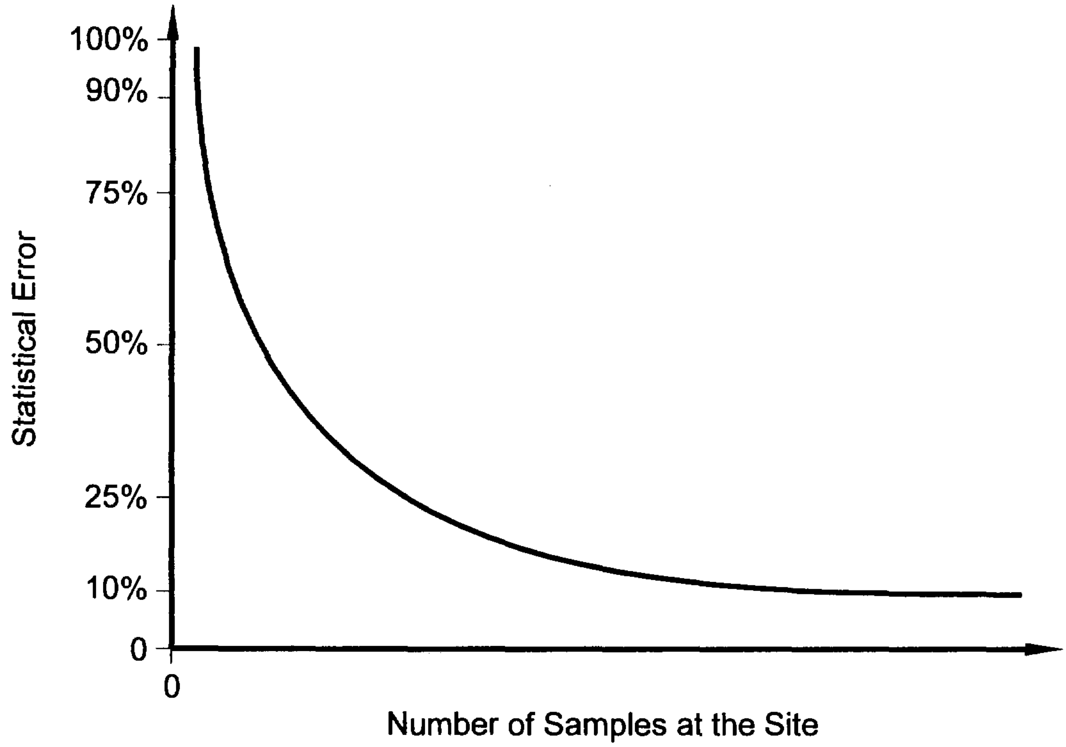

Figure 1. Decline curve showing increased statistical confidence (reduced error) with additional sampling (after Myers, 1997).

entists have known intuitively: with few samples, uncertainties and errors are high, whereas, with many samples, uncertainties and errors are greatly reduced.

Similarly, *sampling theory & practice* (STP), as developed by Pierre Gy, provides value by assessing the magnitude of the errors that are inevitably associated with a sampling or subsampling operation (Pitard 1987). High-quality samples are necessary as poor-quality samples cannot be relied upon for accurate decision making. STP demonstrates the so-called *nugget effect* (short-range variability) is related to the fundamental error (FE). The relationship between geostatistics and STP is not coincidental; both sciences were developed by practitioners working on complex problems in the mining industry.

Both *geostatistical appraisal* (GA) and STP are a function of a *support volume*, that is, the physical size of the sample or subsample. The support volume subsequently influences STP errors, variogram analysis and model parameters, as well as the size of the mapping and/or decision-making unit for monitoring or remediation. Thus, the common theme of support runs through each of the primary activities in minesoil characterization: sampling and subsampling (STP), data analysis (GA), and decision making. This paper will stress the importance of analyzing the variability and uncertainty associated with each support level (sampling, data analysis, decision-making) and their interrelationships. The paper will also present some data that illustrate the need for caution when selecting sampling methods.

The Purpose of Sampling

The objective of any sampling program depends on the questions that need to be answered regarding the target material. In general, there is an almost infinite number of reasons for sampling. For minesoil assessment, three common objectives represent many of the sampling programs initiated: (1) characterization of soil materials (content and variability); (2) as a basis for decision-making; and (3) verification of existing data.

First, it is often necessary to characterize the nature of the soil material. For instance, pre-mine soils will need to be sampled and characterized so that post-mine soils can be compared to them. If post mine soils show unsuitable levels of analytical parameters, remedial treatment of the minesoil may be necessary.

Characterization of the soil can involve many soil parameters, including pH, acid-base account

(ABA), electrical conductivity (EC), heavy metals content, and so forth. In addition to the content of each parameter of interest, it is also important to assess the variability of the parameters. Parameters exhibiting high variability are more difficult to predict. It follows that decisions made at highly variable sites contain more uncertainty than at sites where the parameters exhibit more consistent behavior, assuming equal numbers of samples.

Second, samples provide a basis for decision making. Minesoil characteristics cannot generally be deduced by visual inspection; rather, it is necessary to perform laboratory analysis on the material. Based on the laboratory results, decisions can be made regarding potential remedial actions to the minesoil. The acquisition of "hard" data via sampling is necessary as it provides a defensible basis for subsequent decisions and actions. However; the manner in which samples are taken (based on sampling design) can have a great influence on the analytical data that are produced. This will be discussed later in the paper.

Third, sampling may be used to verify existing data. A sampling program initiated by a regulatory agency or a mining company may produce results that are questioned by the other party. Sampling provides a way in which to obtain additional information and, hopefully, settle disputes.

Difficulties In Sampling

Extrapolation To Larger Volumes

Samples play a key role in minesoil characterization. Samples act as our window on the world, and, by proxy, give us an idea as to the types and proportions of constituents composing a particular volume of soil material. Yet, we impose a great responsibility on and demand a lot of information from a single sample. Typically, a few pounds, or less, of sample material is extracted from the ground. This sample is then sent to the laboratory where a subsampling operation is generally performed. Finally, this tiny bit material is submitted to the analytical procedure. The resulting analytical data are then asked to represent a much larger volume of material, often exceeding multiple acres or tons.

Extrapolating from the specific (i.e. a sample location) to the general (the surrounding area) is an inductive process. Theorems of logic caution that this is a risky endeavor. Therefore, it is prudent to be cautious in extrapolation and interpolation, doing so with an understanding of the risks and errors associated and by using error management techniques to mitigate their impacts.

Separating Laboratory Errors From Sampling Errors

Two areas that are often confused are *laboratory errors*, also known as *analytical errors*, and *sampling errors*. It is important to understand that these errors are distinctly different, arise from different causes, and must be treated independently. They do share the common feature in that both contribute to the error and uncertainty in minesoil characterization.

Laboratory error is familiar to most people who are involved with sampling. Laboratory errors are those that arise from the analytical process itself: the extraction method used, the type of analytical device, and incorrect interpretation or calculation of results. These types of errors have been widely studied and documented in the literature. One well-known example is the U.S. Environmental Protection Agency's (USEPA) SW-846 guidance document.

In contrast, sampling errors are largely unknown or misunderstood. A formalized sampling theory & practice has been developed (Gy 1979, Pitard 1989) based on careful study of the sampling of particulate materials in the mining industry. Important decisions classifying rock into ore versus waste categories are made on a daily basis at a working mine. The implications of incorrect classification are quickly felt in hard, economic terms. Experience has shown that it is often difficult to correctly classify ore and waste, with the mined material commonly differing from that which was predicted or expected.

Sampling errors arise from the inherent *heterogeneities* in the soil material. STP recognizes that two types of heterogeneities exist: (1) constitution heterogeneity and (2) *distribution heterogeneity* (Pitard 1989). Constitution heterogeneity is the variability that is inherent to the composition of each particle or fragment making up the material of interest. Distribution heterogeneity is the manner in which separate and distinct particles are scattered or spread out within the material.

All sampling errors arise from these two heterogeneities. Seven distinct sampling errors result from the heterogeneities in particulate material (Pitard 1989). They are the *fundamental error (FE)*, the *grouping and segregation error (GE)*, the *long-range heterogeneity fluctuation error (CE2)*, the *periodic heterogeneity fluctuation error (CE3)*, the *increment delimitation error (DE)*, the *increment extraction error (EE)*, and the *preparation error (PE)*. Their relationship to one another and to laboratory analytical error is shown in Figure 2. The *discrete model* errors components relate primarily to sample collection and laboratory analysis; the *continuous model* error components relate to spatial and temporal variability. Detailed discussion of heterogeneities and the individual sampling errors is beyond the scope of this paper.

STP provides ways to mitigate or minimize the impacts of sampling errors in a sampling program. Sampling errors, along with laboratory errors and estimation errors, are inevitable. As such, one should accept their existence and use the existing science to manage these errors, constraining them to tolerable levels.

Defining Sampling Objectives

What Question(s) Will Be Answered By Sampling?. Surprisingly, sampling objectives are often not clearly defined. As a result, data may be collected that do not provide the information necessary to support a particular decision. While the lesson is probably self-evident, establishment of sampling objectives is not always easy.

For example, if it is necessary to know whether the average sand content over a 5.7-acre grid unit exceeds a threshold level, a compositing approach may be appropriate to support a decision. If the question is either 1) what areas of the 5.7 acre grid fall above and below 80% sand content; 2) what is the spatial variability within the grid unit; 3) what proportion of the grid area contains greater than 80% sand; or 4) what is the statistical distribution of sand content data in the grid unit, other sampling approaches must be implemented.

Formulation of the appropriate sampling objectives can be achieved in a variety of ways. One model used in environmental applications is the USEPA's *Data Quality Objectives (DQO)* approach. USEPA has been formally committed to data quality for over a decade, first issuing a DQO guidance document in 1987, then updating and streamlining the model in 1994. The DQO process is a seven-step method to assist in assuring that the appropriate type, quantity, and quality of data are collected for decision making purposes. The DQO process also stresses the efficient use of time and financial resources.

The purpose of DQOs is to (1) clarify the study objective; (2) define the most appropriate data to collect; and (3) specify tolerable limits on decision errors, which will be used as the basis for establishing the quantity and quality of data needed to support the decision. It has been used effectively to establish sampling priorities, manage sampling budgets, and reduce conflict between regulatory and industry groups. The seven basic DQO steps are set forth below:

Step 1: State the problem
Step 2: Identify the decision
Step 3: Identify the inputs to the decision
Step 4: Define the study boundaries
Step 5: Develop a decision rule
Step 6: Specify tolerance limits on decision errors
Step 7: Optimize the design

Complete discussion of the DQO process is beyond the scope of this paper; more complete information can be obtained from USEPA (1994).

Support Issues. Geostatistical appraisal (GA) in addition to STP, must also address important support issues. For example, the experimental variogram is sensitive to the support volume of the samples. Variograms calculated on samples having a small support will show more variability and have a higher sill than for samples with a larger support. Also, the *kriging estimation error* for small blocks will be higher than that for large blocks. As such, greater uncertainties surround small blocks as opposed to large blocks, assuming equal numbers of samples and that equivalent variogram parameters are used during estimation (kriging). These examples are the expected result of the *volume-variance relationship* (David 1977).

The DQO model of USEPA also incorporates the idea of support into the characterization and decision making process. DQOs define an *exposure unit* (EU), which can also be considered a *management unit* (MU), that is the fundamental basis for decision making. The size and shape of an MU is dependent upon the sampling objective. It may be a 50 ft by 50 ft rectangle representing a typical residential yard, or it may be a 12 ft. by 1 mile by 6 inch deep parallelpiped along a highway where lead (Pb) contamination in soils (the result of automobile emissions from leaded gasoline) will be assessed.

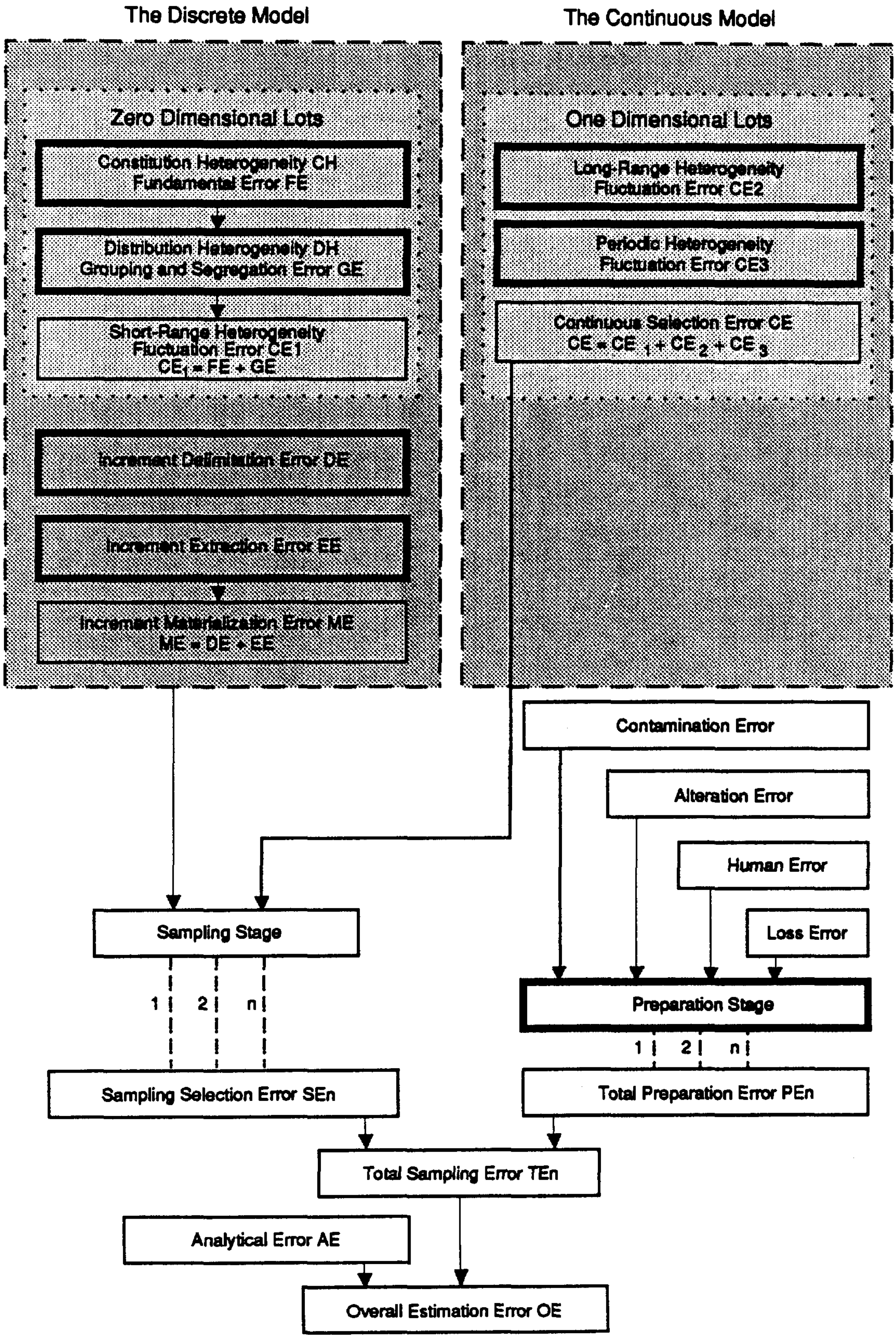

Figure 2. The seven basic sampling errors (after Pitard, 1989).

In the USEPA model, the EU/MU functions as the primary decision unit for remediation. If the average concentration of an EU/MU exceeds a *risk-based threshold*, then the EU/MU must be cleaned up to acceptable levels. If not, no action is necessary. Geostatistical appraisal (GA) can be used to assess the uncertainty on any size of EU/MU, revealing to all parties the probability of misclassification relating to the threshold. Estimated rates of *Type I and Type II errors* (false positives and false negatives) may calculated.

An MU-type approach is used by the state of Texas in pre- and post-mine soil monitoring, where the basic MU can range in size, depending on the site-specific variability. A common MU size is a 5.7-acre grid; however, there are mines that have 20-acre grids when there is less minesoil variability. Based on a composite sample (1 core/acre), the average level of an analytical soil parameter over the grid area is determined. The distribution of post-mine parameter values in 5.7-acre grids is then compared to the pre-mine distribution, to determine if remedial action may be required.

What Is Representative Sampling? As with the heterogeneity terms, the term "*representativeness*" is often used without definition. Pitard (1989) offers a mathematical, and hence, objective, definition of representativeness, where the mean square of the sampling error $r^2(SE)$, the sum of (1) the variance of the *sampling error* $\sigma^2(SE)$ (i.e. the precision), and (2) the square of the *mean sampling error* $m^2(SE)$ (i.e. accuracy), is smaller than a level of *representativeness* $r_o^2(SE)$ defined or agreed upon as acceptable.

Other definitions of representativeness are not exact, an example being the definition in 40 CFR 260.10: *Representative sample* means a sample of a universe or whole (e.g. waste pile, lagoon, ground water) which can be expected to exhibit the average properties of the universe or whole. Please note that the previous definition does not address variability between the sample and the universe or whole, nor does it address tolerable deviations or errors as Pitard (1989) does. Pitard's method, used in conjunction with the DQO process, affords the opportunity to negotiate a definition of representativeness that is acceptable to all parties.

Definition of Appropriate Domains

To the extent possible and practical, it is necessary to respect different geological *domains* during sampling. Native soil conditions may vary due to parent material, climate, vegetation, topography, amount of weathering that has occurred, and the soil horizon. Samples should be confined to domains that are similar in nature; mixing domains invites bias and confusing results. This is equivalent to a stratified sampling approach (Gilbert 1987); which is often difficult to implement as domain differentiation may be difficult, time-consuming, and costly.

Selection of Sampling Approaches

A variety of methods exists to determine the locations and arrangement of soil samples. Accompanying each method are assumptions and implications that should be considered. A few methods are reviewed below.

Non-Statistical Two primary types of non-statistical sampling approaches exist, *haphazard* and *judgmental*. Haphazard sampling takes the attitude that "any sampling location will suffice" (Gilbert 1987). Given this latitude, samples are often taken from locations that are more convenient than representative. This automatically introduces a bias of unknown sign and magnitude into the sampling.

Judgmental sampling yields the power of location selection to one or more individuals. The implication is that representative locations can be determined. There is no way, however, to quantify the degree of accuracy in the resulting sampling. Also, if the credibility of the "expert" is later questioned or shown to be poor, then the resulting samples might also be considered unreliable.

Non-statistical approaches to sampling are generally discouraged because of the biases introduced and the inability to quantify errors.

Statistical Five major statistical approaches to sampling are available for implementation. They are *strict random*, *grids* (and other "patterned" geometries), *randomized grids*, *random stratified grids*, and *probabilistic targeting* (search sampling).

Strict random sampling allows the selection of each sampling location from anywhere within the area or volume of interest. Strict random sampling, while technically unbiased, often results in problems for spatial sampling. Strict random sampling tends to produce clusters of data, meaning that some areas are over-represented and others under-represented by the sampling.

Grid sampling is a common approach to sampling when a spatial area needs to be assessed. Many types or *patterns* of grids are available (Gilbert 1987). The grid helps to eliminate the problem of localized clustering introduced by strict random sampling. However, the selection of the grid origin is often biased, introducing bias into the sampling results.

Randomized grids are an improvement on grid patterns in that the origin is selected by a random process. Once the origin is established, all other grid locations are then known (Myers 1997). While better than non-random grids, a small bias is introduced to the non-origin locations.

Stratified random grids combine the randomization feature of strict random sampling with the spatial coverage feature of grid sampling. Stratified random sampling establishes a grid, then within each grid a sample location is selected at random (Myers 1997). An example is shown in Figure 3. This approach maintains the *equiprobability* conditions necessary for unbiased sampling, that is, every location must be equally available for selection during the sampling process.

Probabilistic targeting (search sampling) focuses on finding areas that contain constituent levels exceeding established thresholds or limits. The sampling program is then designed to yield results such that if no unacceptable areas are discovered, a confidence level (90%, 95%, etc.) will be associated with the results. The process demands that a *target size and shape* be defined, along with a decision as to how much *uncertainty* will be tolerated. Once these parameters and/or assumptions are established, a number of pertinent scenarios can be addressed.

Spatial Variability

Geostatistical Tools. Large amounts of spatial variability are typically encountered at minesites, in both ore reserves and regraded mine soils (David 1977, Myers and Brown 1990). Experience has shown that in addition to spatial variability, *spatial correlation* generally exists between nearby sample data. Since classical statistical methods demand *independence* between samples, techniques capable of dealing with the spatial correlation are necessary.

Geostatistical approaches are able to quantify the spatial correlation components present at a site and incorporate them into estimations and error quantifications. *Variogram analysis* examines and quantifies the degree of spatial correlation for different parameters at a site. An example of a variogram is shown in Figure 4. Plotting distance between samples on the X axis and variance on the Y axis, the variogram shows that, as the distance between samples increase, variability also increases. The rise in the variogram eventually levels off, creating a "*sill*" equal to the population variance. The distance at which the graph reaches the sill is called the range. The range indicates the distance at which samples become independent of each other, that is, nearby samples do not "share" correlation information.

Figure 4 shows that the variogram graph does not always start at the origin. Instead, it starts part-way up the Y axis. This implies that variability exists at zero distance. This is called the *nugget effect*. The nugget effect primarily reflects two phenomena: (1) the short-range variability inherent in the material being sampled, and (2) the fundamental sampling error FE. The nugget effect introduces uncontrolled variability into estimates of spatial areas or volumes, therefore, a high nugget effect is undesirable.

The process of kriging uses the *variogram parameters* (nugget effect, sill, and range) along with sample data to estimate the average constituent level in a defined area or volume. *Kriging* will appropriately weight the sample data points according to their distance from the block. Points closer to the block will receive more weight; points further away will receive smaller weights. Weighting in this manner is consistent with both intuition and with results shown by variographic studies.

In addition to estimating the average of the block, kriging also calculates the estimated error (standard deviation) associated with the estimate. Using the *kriging standard deviation*, the uncertainty on the block estimate can be calculated. Figure 1 showed a decline curve where additional samples contributed to a corresponding reduction in error. Kriging also incorporates this concept into the estimation error: the more samples used, the lower the estimation error.

Resource Limitations

All sampling and characterization efforts are subject to resource availability constraints. The most critical are time and money, both of which are notoriously finite. In addition, regulatory deadlines impose additional constraints on both the mining company to comply and the regulatory agency to respond in a timely manner.

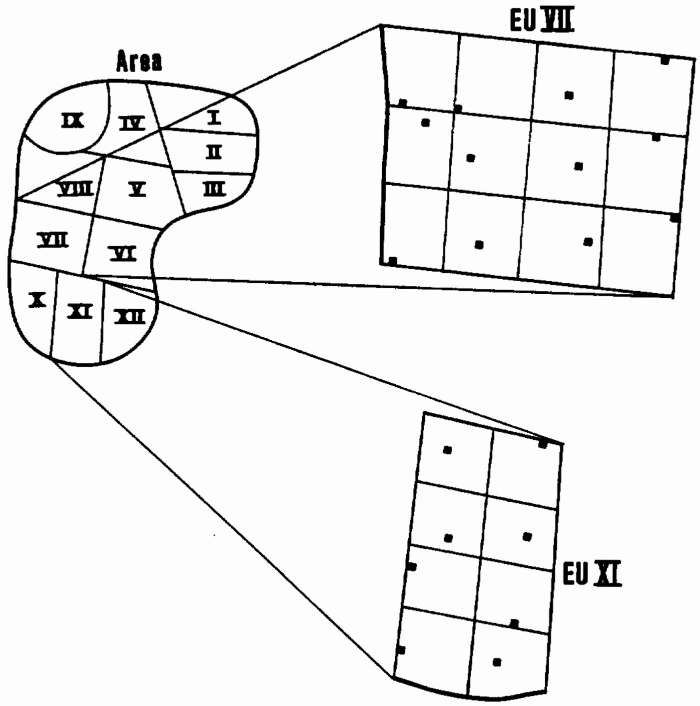

Figure 3. Examples of random stratified grids for sampling in EU/MUs.

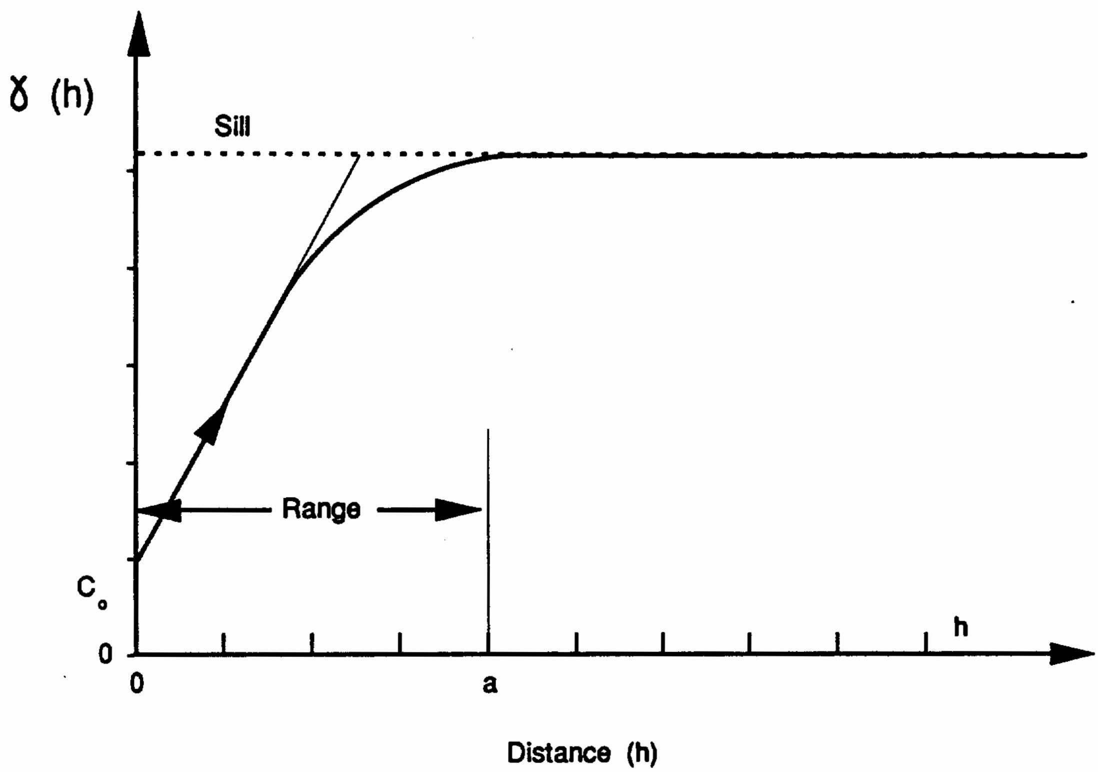

Figure 4. Variogram graph with component features.

Review of Existing Protocols in Texas

The existing system for post-mine soil monitoring in the state of Texas is to base decisions on different-sized MUs (ranging from 5.7- acre to 20-acre grid areas). An MU greater or lesser than 5.7 acres may be instituted based on the site conditions. A sampling density of 1 core per acre is followed. The appropriate number of samples are taken within the grid area and mixed to make one composite sample; with a composite sample being produced for four depth increments: 0-1 ft, 1-2 ft, 2-3 ft, and 3-4 ft. Initial soil samples are to be taken no less than 200 ft from each other. This preceding sampling is a generic example, mine-specific sampling designs are implemented based on the particular minesoils that are found.

A DQO MU decision process is then implemented for the soil constituents. The laboratory result corresponding to the *composite sample*, essentially an average over the 5.7 acre MU, is taken to represent the entire MU. Further decisions pertaining to comparisons of pre- and post-mine conditions, liming requirements, and other relevant parameters are made using the constituent results for the entire MU.

Error Management For Minesoil Constituents

Data Quality Objectives

The use, in Texas, of a 5.7-acre or 20-acre MU follows at least three aspects of the USEPA-approved DQO process for environmental decision making. First, a MU area of concern is defined. Next, an average concentration for the MU is determined by means of composite sample data. Finally, subsequent decisions and actions are based on the average concentration over the MU, as determined by a multiple-sample composite.

It is important to evaluate the assumptions used in the DQO decision rule. The intent of USEPA is to establish the *average concentration* of the entire MU. Under this averaging approach, a wide variety of spatial distributions of the COC may be present at the site, all of which may be tolerated. For example, the distribution of a particular parameter may be quite uniform, with little variation over the MU area. In another case, the distribution of the parameter values may range over two or more orders of magnitude from one side of the MU to the other, showing high variability. Alternatively, the high variability may be distributed somewhat randomly, with high hot and low concentration zones mixed in various (and unknown) arrangements. Different spatial arrangements and levels of spatial variability have been documented and mapped in regraded minesoils for pH and ABA values in Texas mines (Myers and Brown 1990).

Effects of Composite Sampling

By taking a composite sample, certain influences have been exerted on the decision-making process. As described above, any understanding of the spatial variability over the MU has been lost. An assessment of the within-MU variability can only be obtained by analyzing the individual samples that make up the composite. USEPA is comfortable with this loss of information based on three considerations: 1) the risk-based approach, supported by sophisticated health-risk analysis models and a focus on average parameter concentrations; 2) careful definition of an appropriately-sized MU; and 3) a sufficient number of samples taken in each MU to support a defensible level of decision making.

Note that sufficiency of samples is related to the DQOs. The number of samples will be different depending on the question to be answered by sampling. For example, data requirements may vary for situations in which 1) the 95% upper confidence limit (UCL) needs to be below a threshold standard; 2) the number of false positives and false negatives needs to be minimized; 3) the lateral and/or vertical extents of unacceptable concentrations needs to be mapped with confidence; or 4) the pre-mine data distribution needs to be compared with the post-mine data distribution.

The compositing of samples addresses another primary concern of the DQO process, that of efficient utilization of resources. DQOs recognize that money is scarce and sampling should be done cost-effectively. Compositing achieves this objective. The task, then, is to balance the information needs obtained by buying sample information against the limited resources that may be available. All of this must be done within an error management framework.

Compositing of samples can present some problems, however, that should be addressed on a site-specific basis. The compositing of physical samples may not be appropriate for all minesoil parameters. For example, matrix effects (Fundamental error, FE) can significantly alter results for pH values of composited samples Table 1. The calculated sand content and neutralization potential parameters, however, correlated well with the actual analysis values. The data in Table 1 were derived from a bench test that mixed ultra-acidic and alkaline materials in different proportions. These preliminary results indicate that further testing on the parameter value effects of mixing unlike materials is warranted.

The bench test data show that the calculated pH value (based on weight-averaging the samples) of a composite sample may vary greatly from the actual pH obtained when the sample is analyzed. This situation can occur in minesoils where there are both acid and alkaline materials that are composited together in the same sample (i.e. samples from an alkaline material-amended minesoil that may contain a small volume of limestone aggregates). This also suggests that individual samples that are analyzed for pH should not be "composited" using a mathematical average of the values, due to the logarithmic nature of the parameter and variable minesoil chemical and mineralogical composition.

Field data exist that support the need for caution when minesoil samples are composited. The following data (Figure 5) were obtained from minesoils in South Texas that were sampled in 6-acre grids (six individual samples were mixed to produce a composite for each grid), where the predominant field pH of each sample was estimated before the sample was added to the composite. Field pH data were obtained with a universal pH dye indicator kit. The predominant pH of the material was placed in either of four classes: A (pH $\leq$ 4); B (pH > 4.5 and pH $\leq$6); C (pH > 6 and pH $\leq$ 7); and D (pH > 7). There was good correlation between the field pH and the pH measured in the laboratory, as indicated by the two box plots where 100% of the samples had no field pH Class A materials (average lab pH of 7.8) and 100% of the samples were field pH Class A (average pH of 3.8). The median composite pH decreased as the proportion of Class A (field pH $\leq$ 4) samples increased. The data indicate that composite samples from grids, where 4 of 6 samples (approx. 67%) were field pH Class A, could still produce a laboratory pH greater than 5.5 half of the time. The practical implication is that a pH of 5.5 could be assigned (50% of the time) to grids where over two-thirds of the area contained pH values equal to or below 4. This may be acceptable or not; however, this possibility must be addressed when the size of the MU is determined.

<u>Support Considerations</u>

As discussed previously, several support influences must be considered. First, in order to assess the nature and magnitude of the spatial variability, it is necessary to determine which support unit is contributing most to the total variability. As a practical matter, this means determining whether the sampling and/or laboratory error is greater than or equal to the spatial variability.

Pitard (1990) studied duplicate and replicate minesoil samples (neutralization potential, potential acidity, and ABA) in an effort to distinguish nugget-type variability (sampling/subsampling error and laboratory error) versus long-range variability of regional trends (5 acres, 20 acres, etc.). Results indicated that while short-range errors were relatively high, they were not significant in comparison to regional trends, which showed greater variability. However, variability between 5-acre grids was not found to be much different from that shown by 20-acre grids.

Pitard (1990) showed that the greatest amount of spatial variability for the parameters studied appears to exist within a few feet of a given sample, and that long-range trends may be best observed on units of 20 acres or larger. This is consistent with the results obtained by Myers and Brown (1990), where extremely high variability was encountered at distances of 100 ft, while long-range variability (800 to 1200 ft) often produced good spatial correlation structure (variograms). These results indicate the greatest variability in the studied minesoils occurred at a very small scale, less than 100 ft and perhaps as small as one to five feet. Dollhopf and Birkhead (1992), using composited samples, also found small differences between variability in decision units of 5.7 acre versus 20 acre grids for soil pH and ABA. Reductions in variability were 10% and 12% respectively when the MU decision support was increased from 5.7 to 20 acres.

Given the apparent high level of spatial variability at very short distances, two general approaches to improving minesoil characterization arise. The first is to increase the size of the field sample taken. This will reduce variability on the small-scale (nugget type). The second is to increase the number of increments per sample (local compositing).

Table 1. Results from bench test where ultra-acid and alkaline materials were mixed in different proportions (mean values for duplicate samples).

Proportion of Material Mix		pH			Sand Content			Neutralization Potential, meq/100g		
Alk, %	Acid, %	Calc.†	Actual	Rel. Diff., % ‡	Calc.	Actual	Rel. Diff., %	Calc.	Actual	Rel. Diff., %
0.0	100.0		2.7			34			-6	
23.0	77.0	3.8	6.0	37	30	32	6	90	91	1
37.5	62.5	4.6	6.1	25	28	2/	0	150	158	5
54.5	45.5	5.4	6.1	11	25	26	4	221	262	16
71.5	28.5	6.3	6.2	2	23	21	10	291	295	1
91.0	9.0	7.0	6.4	9	19	19	0	373	375	1
100.0	0.0		7.7			18			410	
				17			4			5

† Calc. - calculated values were determined by weight-averaging the corresponding proportions of the alkaline and acid materials.
‡ Relative difference = Actual values - Calculated Values X 100.

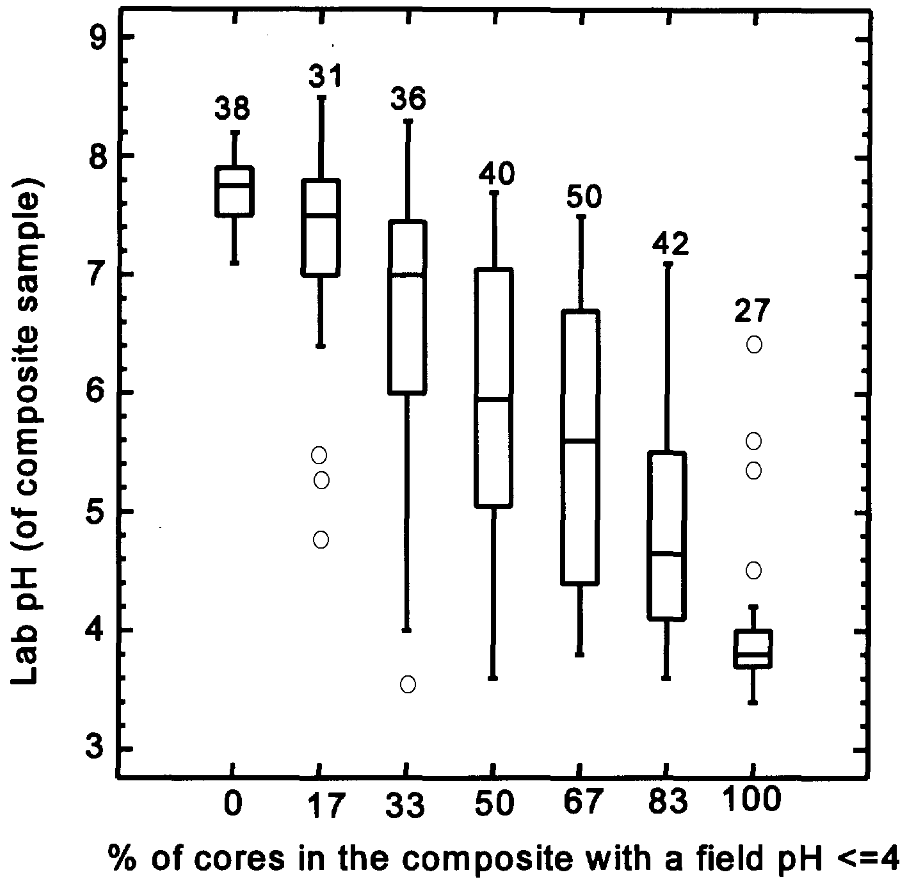

Figure 5. Comparison of the proportion of individual field pH class A (pH $\leq$ 4) subsamples (cores) in a composite sample to the pH values obtained for each composite (numbers within figure correspond to the number of samples used for each box plot; circles represent outliers).

This type of approach will emphasize the large-scale issues of spatial variability, which we can address, versus the micro-scale problems about which little, if anything, can be done economically. Both of these approaches must be considered while determining the appropriate size of the MU.

Summary and Conclusions

Use of a support-based approach to error management and decision making for minesoil constituents offers several advantages. The USEPA-approved DQO/EU or MU model allows for variable-size decision units, allowing regulators to decide on a case-by-case basis whether a 5.7 acre, 20 acre, or other size characterization and decision making unit is appropriate. Studies using STP (small-scale) and GA (large-scale) have shown that both sampling support and grid-size are key issues. STP and GA have also lead to an understanding of the major sources of variability in minesoils, thereby providing insight as to how to reduce and manage variability and decision errors. Site-specific variability must always be taken into account when designing a sampling program and caution is recommended in the selection of sampling methods (i.e. compositing versus discrete samples) for certain minesoil constituents.

Literature Cited

Brandt, J.E. 1993. Baseline Study Bench Test for the Butler-Weddington Area 2A Abandoned Mine Land Project. Unpublished notes.

Dollhopf, D.J. and J.A. Birkhead. 1992. Determination of Minesoil Sampling Grid Size Requirement at the ALCOA Sandow Mine. 37 pages.

David, M. 1977. Geostatistical Ore Reserve Estimation. Elesevier. Amsterdam.

Gilbert, R.O. 1987. Statistical Methods For Environmental Pollution Monitoring. Van Nostrand Reinhold. New York, NY.

Gy, P.M. 1979 Sampling of Particulate Materials. Elsevier. Amsterdam.

Myers, J.C. and T.H. Brown. 1990. Geostatistical Analysis of Regraded Spoil Overburden Materials. Proceedings of the Billings Symposium on Disturbed Land Rehabilitation. Billings, MT.

Myers, J.C. 1997. Geostatistical Error Management (GEM): Quantifying Uncertainty For Environmental Sampling and Mapping. Van Nostrand Reinhold. New York, NY.

Pitard, F.F. 1987. Pierre Gy's Sampling Theory and Sampling Practice: Heterogeneity and Sampling, volume 1. CRC Press. Boca Raton, FL.

Pitard, F.F. 1987. Pierre Gy's Sampling Theory and Sampling Practice: Sampling Correctness and Sampling Practice, volume 2. CRC Press. Boca Raton, FL.

Pitard, F.F. 1990. Study of the Variability of Soil Analyses at the Sandow Mine. Report to The Aluminum Company of America. 38 pages.

U.S. Environmental Protection Agency. 1987a. Data Quality Objectives for Remedial Investigations: Development Process.

U.S. Environmental Protection Agency. 1987b. Data Quality Objectives for Remedial Investigations: Example Scenario.

U.S. Environmental Protection Agency. 1994. Guidance for the Data Quality Objectives Process. EPA QA/G4.