&EPA

Project Summary

Preparation of Soil Sampling Protocol: Techniques and Strategies

Benjamin J. Mason

This report sets out a system for developing soil sampling protocols. The body of the report discusses the factors that influence the selection of a particular sampling design and the use of a particular sampling method. Statistical designs are discussed along with the appropriate analysis of the data. Three appendices are included. One consists of a discussion of the steps that must be taken to arrive at the desired protocol. The remaining appendices present two examples of protocols--one for a shallow spill situation and the second for a deep contamination plume. A technique called kriging is presented as an approach to the analysis of data collected during a soil sampling program. This technique allows the researcher to develop maps of the pollution levels and to assign a statistical precision to the data at each point on the map. The data presented on these two maps can be evaluated to identify areas where additional samples are needed to reach a desired level of precision for the area covered by the maps.

This Project Summary was developed by EPA's Environmental Monitoring Systems Laboratory, Las Vegas, NV, to announce key findings of the research project that is fully documented in a separate report of the same title (see Project Report ordering information at back).

Introduction

Considerable effort has been expended in developing protocols for monitoring airborne and waterborne pollutants. The complexities of sampling the soil system hinder the development of appropriate field procedures.

This report provides a means for developing a protocol to satisfy the soil sampling needs of the U.S. Environmental Protection Agency (EPA) and other agencies with similar needs. The methods outlined are adequate tools for producing reliable estimates of the spatial distribution of soilborne pollutants. The investigator should already be somewhat familiar with the general properties of soils and the behavior of pollutants in the soils environment. A soil chemist and a statistician should be available from the outset to consult about the selection processes used in developing the protocol. The investigator must be able to modify the protocol to meet unusual conditions that frequently arise in the field and were not specifically covered during the design or planning phase of the program.

EPA scientists need to follow rigorous chain-of-custody and quality assurance procedures in conducting their investigations. A brief discussion of those requirements is included in this report, as is a treatment of the application of a relatively new geostatistical technique (kriging) for estimating the distribution of pollutants in soils.

The Soil System

Because factors such as clay and organic matter content, texture, permeability, pH, and cation exchange capacity influence the rate and route of migration, these parameters and their variability must be considered in the sampling design. One statistical measure of variability is the coefficient of variation (CV). Coefficients of variation for soil parameters have been

reported ranging from as low as one to two percent to as high as 850 percent, with the variation slightly higher for physical than for chemical properties.

$$CV = s/\bar{y} \times 100$$

where CV = coefficient of variation in percent
s = standard deviation of sample
$\bar{y}$ = mean of sample

The inherent variation found in data collected from any preliminary or previous soil sampling is of particular value in designing a sampling plan for the same area.

Initiating the Soil Sampling Study

The soil sampling protocol should begin with a clear statement of the objectives of the study followed by a series of steps or decision elements leading to the proper sampling design. The goals of the study *must* be elucidated and *must* be agreed upon by all parties involved prior to designing and implementing the field sampling program. The objective statement should also include data reliability desired and the resources committed to the study.

The scientist must establish the confidence level desired and the allowable margin of error to be met by the results. Laboratory systems have been developed to the point where considerable confidence can be placed in the results. The sources of uncontrolled variation are too great, however, for a field study to meet the precision found in a laboratory. The only alternative is to select a level of confidence that is acceptable and attainable within the limits of the resources available for the study.

Types of Soil Pollution Situations

Major types of sampling situations that the environmental scientist is likely to encounter include large or localized areas where pollutants are either in the surface layers, moved down into the soil profile, or contained within well-defined plumes.

Factors such as length of time over which the contamination has occurred, physical and chemical properties of the pollutant(s), type of soil and past use of the area must all be considered in determining the appropriate study design.

Review of Background Data

All available documents dealing with the study area, including pertinent newspaper accounts, should be collected and eval-

uated. In general, the information of value should deal with the historical use of the area, current and old drainage patterns, groundwater flow and use, and associated environmental and health problems.

Geological characteristics are important not only for determining routes of pollutant migration but also for stratifying an area by homogenous soil types. Parent materials and bedrock often play an important part in determining how the pollutants will react in the soil. Information on the nature of the bedrock, the groundwater elevations, the direction of groundwater flow and the sources of recharge to the aquifer should be acquired before finalizing the monitoring plan.

Statistical Designs

Four basic statistical sampling designs can be used in soil studies -- simple random, stratified random, systematic and judgmental sampling.

Simple random sampling is the basis for all probability sampling techniques used in soil sampling and serves as a reference point from which modifications to increase the efficiency of sampling are evaluated. Simple random sampling in itself may not give the desired precision because of the large statistical variations encountered in soil sampling. Therefore, one of the other designs may be more useful. Where little information exists about the area to be studied or the pollution distribution, the simple random sampling design is the only design other than the systematic grid that can be used effectively.

The procedures for determining the number of samples required to meet a predetermined precision in a simple random sampling design are the basis for allocating samples to a strata in the stratified random design. They can also be used to determine the number of sample points needed in the systematic sample design.

If an estimate of the variance can be obtained from either a preliminary experiment, a pilot study, or from the literature, the number of samples required to obtain a given precision with a specific confidence level can be obtained from the following equation:

$$n = t^2_\alpha s^2/D^2$$

where D is the precision given in the specifications of the study; s^2 is the sample variance and t is the two-tailed t-value obtained from the standard statistical tables at the α level of significance and $(n-1)$ degrees of freedom. D is usually expressed as $\pm$ a specified number of concentration units (i.e., $\pm$ 5.00 ppm).

The equation can also be written in terms of the coefficient of variation (CV) as:

$$n = (CV)^2 t^2 \alpha/p^2$$

where CV is the coefficient of variation; p is the allowable margin of error expressed as a percentage $(D/\bar{y})$ and $\bar{y}$ is the mean of the samples.

The margin of error is needed in determining the number of samples required to meet the precision specified. This is often expressed as the percentage error that the scientist is willing to accept or it may be a difference that he hopes to detect via the study. The margin of error chosen is combined with the confidence level to derive an estimate of the number of samples required. The smaller the margin of error, the larger the number of samples required.

As the t-value is dependent upon the number of degrees of freedom, it is necessary to use an iterative approach to calculate the number of samples required. Curves can be prepared to plot the number of samples against the coefficient of variation, thus avoiding iterations.

The total cost of soil studies often follows a linear form of equation which sums the overhead, sampling and analytical costs. That equation is used along with the equations designed to provide the number of samples required to ultimately arrive at a number that will satisfy the budget and yield the desired precision.

Once the number of samples has been determined, their locations can be planned. A map of the study area is overlaid with an appropriately scaled grid. The starting point of the grid should be randomly selected rather than located for convenience. This can be accomplished by selecting four random numbers from a random number table. The first two numbers locate a specific grid square on the overlay. The second two identify a point within that grid square. This point is fixed on the map and the entire grid shifted so that the lower right corner of the original grid square lies on the point chosen. This technique eliminates some of the questions often raised about bias in grid sampling. A second alternative is to select a pair of coordinates at random, which becomes the starting point for the grid.

Prior knowledge of the sampling area and information obtained from the background data can be combined with information on pollutant behavior to reduce the number of samples necessary to attain a specified precision. The statistical technique used to produce this savings is called stratification. Basically it operates on the fact that environmental factors play

a major role in leaching and concentrating pollutants in certain locations. Stratified random sampling should lead to increased precision if the strata are selected in such a manner that the units within each stratum are more homogenous than the total population. Each stratum is handled as a separate simple random sampling effort.

At least two samples must be taken from each stratum in order to obtain an estimate of the sampling error. The number of sampling units is usually allocated according to a proportion based on the land area covered by each stratum, i.e., if the area of soil in one stratum is 25 percent of the total study area, then 25 percent of the samples would be taken from that stratum. Proportional allocation is used in soil sampling work primarily because the variance within a general area tends to be constant over a number of soil types. A pilot study would confirm if this is, in fact, the case. If the variances are materially different, the basis of allocation must be changed or stratification may be inappropriate.

The systematic sampling plan is an attempt to provide better coverage of the soil study area than could be provided with the simple random sample plan, and may result in better precision and less bias. The method is easy to apply and, therefore, has been frequently used. Samples are collected in a regular pattern (usually a grid or line transects) over the area under investigation. The starting point is located by some random process; then all other samples are collected at regular intervals in one or more directions. The orientation of the grid lines should also be randomly selected except when a pollution plume is the subject of the investigation. The orientation of the grid should be such that the lines are parallel to the general trace of the plume.

The spacing on the grid is particularly important if regionalized variable theory is used to design the study. This theory is the basis of a sampling program termed "kriging," which has the advantage of developing estimates of concentrations over geographic areas and also provide a measure of the point-to-point confidence limits. The theory is based upon the spacing of data points along the grid lines. The samples must be close enough to provide a measure of the continuity of the location to location variation within a soil study area. If, however, a measure of the mean and variance of the population is the focus of the grid sampling array, the samples must be placed outside of the "range" of the variance of each point. This allows the collection of samples that are not influenced

by regionalized variables. If a significant percentage of the variation occurs within the first few meters of a point, the range beyond which kriging is not effective probably lies at a distance of ten meters or greater. This lies within the range within which the mean and variance of the population are the only parameters that can be determined.

The systematic sampling plan is ideal when a map is the final product. It provides for a uniform coverage of the area and allows the scientist to have points to use in generating a plot of isoconcentrations.

Two factors may limit the use of this design. First, the estimation of the sampling error is difficult to obtain from the sample itself unless replicate sampling is used at a number of sites. The variance cannot be calculated unless a mean successive difference test is used to evaluate the data. Secondly, the presence of trends and periodicity in the data create problems when the direction of the grid aligns with the pattern in the data. Soil sampling efforts seldom encounter a cyclic pattern or periodicity to a degree that creates a problem. Trends, however, are common in soil pollution work and frequently are the whole purpose for the sampling effort.

The final statistical sampling design considered is termed judgmental sampling. It is typically used in concert with one of the other methods in order to cover areas of unusual pollution levels or where effects have been observed. The major problem with this approach is that it is subject to bias and, therefore, faulty conclusions. If judgmental sampling is used, duplicate or triplicate samples should be taken to increase the level of precision.

When historical data are not available for use in planning a study, the use of a phased approach is required. The first phase of the study might incorporate a simple random study design with a 68 percent confidence level, the results of which would be used to design a more definitive study with a 95 percent confidence level. A stratified random design or a systematic sampling grid approach could be used to obtain data with a higher confidence limit. The grid design would allow the researcher to analyze the data using kriging and thus find where additional samples are needed to further refine the sampling design so that the entire area is covered at the 95 percent confidence level.

Control sites are used quite often in major soil studies, especially if the study is designed to determine the extent and presence of local pollution. Sites for

controls must be as representative as possible of the study area and subject to the same type pollution sources except for the specific pollutant under investigation. In most cases it is desirable to spend as much time searching out data on the control as on the study area.

Sample Collection

There are two layers within the soil column of primary importance. The surface layer (0-15 cm) includes recently deposited pollutants. Pollutants deposited as a result of liquid spills or long-term seepage of water soluble materials may be found at depths ranging up to several meters. Plumes emanating from hazardous waste dumps or leaking storage tanks may be found at considerable depths. The methods of sampling these layers are only slightly different, however. Samples can either be collected with some form of core sampling or auger device or through the use of excavations or trenches. In the latter case, samples are cut from the soil mass with spades or short punches.

Several surface soil studies have made use of a punch or thin walled steel tube (15 to 20 cm long) to extract short cores from the soil. The soil punch method is fast and can be adapted to a number of analytical schemes provided precautions are taken to avoid cross-contamination during shipping and in the laboratory.

Perhaps the most undesirable sample collection device for the surface layer is the shovel or scoop. This device is often used in agriculture, but where samples are being taken for chemical pollutants, the inconsistencies are too great.

Percolation or precipitation will move surface pollutants into the lower soil horizons where the use of devices, that will extract a longer core is required. Soil probes collect a 30 to 45 cm column of relatively undisturbed soil, while augers collect a "disturbed" sample in approximately the same or slightly longer column. The detail often desired in research studies or in cases where pollutants are suspected cannot be met effectively through the use of augers. In these cases some form of core sampling or trenching must be used. For measuring the vertical distribution of a pollutant, that augers and probes are not recommended as they tend to contaminate the lower samples with materials from above.

Trenching is used to carefully remove sections of soil when a detailed examination of pollutant migration patterns and detailed soil structure is required. It should be used only in those cases where detailed information is desired because of the

relatively high costs associated with excavating. Typically a trench, approximately 1 meter wide, is dug to a depth of about 30 cm below the desired sampling depth. The samples are taken from the sides of the pit using a soil punch or a trowel. The maximum effective depth for this method is about 2 meters unless done in some stepwise fashion.

Sampling for underground plumes is perhaps the most difficult of all soil sampling requirements. Often it is conducted along with groundwater and hydrological sampling. The equipment required usually consists of large, vehicle mounted power auger and coring devices although some portable tripod-mounted coring units are available.

Frequently, several soil samples collected from a given location are composited to minimize analytical costs. The key to any statistical sampling plan is the use of the variation within the sample set to test hypotheses about the population and to determine the precision or reliability of the data. The composite sample provides an excellent estimate of the mean but does not give any information about the variation within the sampling area. A compromise is possible, however, by collecting and analyzing duplicates or triplicates at a percentage of the locations. A similar requirement will be imposed by the quality control program.

A vital component of the protocol is an adequate definition of the records required during the study. Good records are essential should litigation be a potential end point for the study. Each result may be questioned in an attempt to either discredit or verify the data presented.

Data Analysis

Finally, the data analyst must keep in mind the purpose(s) for which the samples were collected. These purposes can usually be characterized as: An estimate of the mean level of pollutant in a geographic area; a determination as to whether the pollution measured is above some standard or is higher than the ambient levels found in the control areas; or to quantitatively document the areal extent and depth of the pollution and the confidence with which it is known.

Protocol Development

A scheme for preparing a soil sampling protocol via a decision tree process is included as an appendix to the report. It is not possible to address in a single protocol all of the variables that may be encountered in a field program. The procedure outlined permits, therefore, the development of a protocol specifically tailored to the requirements of a study. An example protocol for a described scenario is presented.

Benjamin J. Mason is with ETHURA, McLean, VA 22101.
***Robert D. Schonbrod** is the EPA Project Officer (see below).*
The complete report, entitled "Preparation of Soil Sampling Protocol: Techniques and Strategies," (Order No. PB 83-206 979; Cost: $13.00, subject to change)
will   be available only from:
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone: 703-487-4650
The EPA Project Officer can be contacted at:
Environmental Monitoring Systems Laboratory
U.S. Environmental Protection Agency
P.O. Box 15027
Las Vegas, NV 89114

United States
Environmental Protection
Agency

Center for Environmental Research
Information
Cincinnati OH 45268

Postage and
Fees Paid
Environmental
Protection
Agency
EPA 335