Environmental and Ecological Statistics **8,** 137–149, 2001

Sample support and related scale issues in composite sampling

DOUGLAS E. SPLITSTONE

Principal, Splitstone & Associates, 4530 William Penn Highway, Murrysville, Pennsylvania 15668 USA

Received February 1997; Revised January 2000

Composite sampling offers the great promise of efficiency and economy in environmental decision making. However, if careful attention is not paid to matching the ''support'' of the sample to that required for making the desired decision, the promise is unfulfilled. Obviously this attention must be applied in the design phase of a composite sampling strategy. Less obvious is the potential for alteration of the design sample support during sample collection and assay. The consequences of not paying attention to these aspects of sample design and assay are discussed in this ''issue'' paper illustrated with a series of examples taken from the author's consulting experience.

Keywords: decision unit, heterogeneity, measurement variation, remedial decisions, risk assessment.

Introduction

Evan Englund and Naser Heravi provide the following insight in their 1994 paper, ''Phased sampling for soil remediation'':

> Failure to adequately define [sample] *support* has long been a source of confusion in site characterization and remediation because risk due to long-term exposure may involve areal supports of hundreds or thousands of square meters; removal by backhoe or front-end loader may involve minimum remediation units of 5 or 10 m^2; and sample measurements may be taken on soil cores only a few centimeters in diameter.

The importance of this observation cannot be overstated.

It should be intuitive that a decision regarding the average contaminant concentration over one-half an acre could not be well made from a single kilogram sample of soil taken at a randomly chosen location within the plot. Obviously, a much more sound decision making basis is to average the contaminant concentration results from a number of 1 kg samples taken from the plot. If the formation of a composite sample can be thought of as the ''mechanical averaging'' of concentration, then composite sampling appears to provide for great efficiency in cost-effective decision making. This of course assumes that the formation of the composite sample and its assay truly retain the ''support'' intended by the sampling design. It will be seen in the examples that follow that this may not be the case. First, definitions of a composite sample and sample support are in order.

Webster's (1987) defines composite as (1) made up of distinct parts, and (2) combining the typical or essential characteristics of individuals making up a group. Pitard (1993, p. 10) defines a composite sample as a ''sample made up of the reunion of several distinct subsamples.'' These definitions certainly describe an entity which should retain the ''average'' properties of the whole consonant with the notion of support.

Olea (1991) defines ''support'' as:

> An n-dimensional volume within which linear average values of a regionalized variable may be computed, The complete specification of the support includes the geometrical shape, size, and orientation of the volume. The support can be as small as a point or as large as the entire field. A change in any characteristic of the support defines a new regionalized variable. Changes in the regionalized variable resulting from alterations in the support can sometimes be related analytically.

To this author, it makes sense to define two types of support. The desired ''decision support'' is the sample support required to reach the appropriate decision. Frequently, the desired decision support is that representing a reasonable ''exposure unit'' (for example, see USEPA, 1989, 1996a and 1996b). The desired decision support could also be defined as a unit of soil volume conveniently handled by a backhoe, processed by incineration or containerized for future disposal. In any event, the ''desired support'' refers to that entity meaningful from a decision making point of view. Hopefully, the sampling scheme employed is designed to estimate the concentration of samples having the ''desired support.''

The ''actual support'' refers to the support of the aliquot assayed. Ideally, the decision support and the actual support are the same. However, in this author's experience, the ideal is rarely achieved. This is a very fundamental problem in environmental decision making.

Olea's definition indicates that it is sometimes possible to statistically link the actual support to the decision support when they are not the same. However, in practice the information necessary to do so is rarely generated in environmental studies. While this may seem indeed strange to readers of this article, it should be remembered that most environmental investigations are conducted without the benefit of well thought out statistical design. Generally, if some of the mechanics of design are employed, in practice they are often extracted from various US Environmental Protection Agency (USEPA) guidance documents without a great deal of thought as to whether or not the ''guidance'' really applies to the specific situation.

Because this is a paper discussing the issues associated with environmental decision making and sample support, it addresses the situation as it is, not what one would like it to be. All of the referees correctly made the point that the collection of multiple composite samples from a decision unit would permit estimation of the variation of the average contaminant concentration. Almost all of the environmental engineers and/or managers known to this author think of a composite sample only in terms of the reduction in shipping and analytical cost. Their immediate objective is to minimize the cost of site investigation and remediation. Therefore, the idea of ''why take two when one will do'' will usually win out over assessing the ''goodness'' of estimates of the average concentration.

This is particularly true in the private sector which comprises this author's client base. If there is some potential to influence the design of the study (which is not a frequent occurrence), then it takes a great deal of persuasive power to convince the client to pay for any replicate sampling and/or assay. The statistician's choice, absent the power of design, is to either withdraw, or attempt to guide the decision making process toward the correct interpretation of the results in light of the actual sample support.

If environmental investigators would adhere to the traditional elements of statistical design, the appropriate decisions would be made. These elements are nicely described by the US Environmental Protection Agency's (USEPA's) Data Quality Objectives Process (USEPA, 1994; Neptune, 1990). Flatman and Yfantis (1996) provide a complete discussion of the issues.

The story of the stones

A graphic example of how the actual support of the assay result may be inconsistent with the desired decision support of a composite sample is provided by the story of the stones. In reality, it is an example of how an incomplete sampling design and application of standard sample processing and assay protocols can lead to biased results. This is the story of stone brought onto a site to facilitate the staging of site remediation. The site must remain confidential, however, identification of the site and actual data are not necessary to make the point.

Those who have witnessed the construction of a roadway or parking lot will be able to easily visualize the situation. To provide a base for a roadway and the remediation staging area, 2000 tons of stone classified as No. 1 and No. 24 aggregate by the American Association of State Highway Officials (AASHTO) were brought onto the site. The nominal sizes for No. 1 and No. 24 stone aggregate are $3\frac{1}{2}$ inches to $1\frac{1}{2}$ inches and $2\frac{1}{2}$ inches to $\frac{3}{4}$ inch, respectively. These are rather large stone. Their use at the site was to construct a roadway and remediation support area for trucks and equipment. In addition, 100 tons of AASHTO No. 57 aggregate stone were placed in the access roadway and support area as a top course of stone pavement. No. 57 aggregate has a nominal size of from 1 inch to No. 4 sieve. The opening of a No. 4 sieve is approximately $\frac{3}{16}$ inch.

Upon the completion of the cleanup effort for total DDT, the larger stone was to be removed from the site for use as fill elsewhere. Removal of the stone involves its raking into piles using rear mounted rakes on a backhoe and loading via front-end loader into trucks for transport off-site. In order to remove the stone from the site it had to be demonstrated that the average concentration of total DDT for the stone removed met the Land Disposal Restriction criterion of 87 microgram per kilogram (μg/kg).

The remedial contractor, realizing that the stone was brought on site ''clean,'' and the only potential for contamination was incidental, suggested that two composite samples be taken. Each composite sample was formed in the field by combining stone from five separate randomly chosen locations in the roadway and support area. The total DDT concentrations reported for the two samples were 5.7 μg/kg and 350 μg/kg. Obviously, not a completely satisfactory result from the perspective of one who wants to move the stone off-site.

It is instructive to look at what actually happened to the sample between collection and chemical assay. Because surface contamination was the only concern, the stones comprising each composite were not crushed. Instead several stones, described by the chemical laboratory as having an approximate diameter of 1.5 cm, were selected from each composite until a total aliquot weight of about 30 g was achieved. This is the prescribed weight of an aliquot of a sample submitted for the chemical assay of organic analytes. This resulted in a total of 14 stones in the sample having the 5.7 μg/kg result and 9 stones in the sample showing the 350 μg/kg result.

The stones actually assayed, being less than 0.6 inch (1.5 cm) in size, belong only to the No. 57 aggregate size fraction. They represent less than 5% of the stone placed at the site (100 tons versus 2000 tons). In addition, it represents the fraction most likely to be left on site after raking. Thus, the support of the assayed subsample is totally different than that required for making the desired decision.

In this situation, any contamination of the stone by DDT must be a surface phenomenon. Assuming the density of limestone and a simple cylindrical geometric shape, the 350 μg/kg concentration translates into a surface concentration of 0.15 μg/cm^2. Cylindrical stones of approximately 4 cm in diameter and 4 cm in height with this same surface concentration would have a mass concentration of less than 87 μg/kg. Thus arguably, if the support of the aliquot assayed were the same as the composite sample collected, which is close to describing the stone to be removed by the truck load, the concentration reported would have met the Land Disposal Restriction criterion. Indeed, after the expenditure of additional mobilization, sampling and analytical costs this was shown to be the case.

These expenditures could have been avoided by paying more attention to whether the support of the sample assayed was the same as the support required for making the desired decision. This requires that thoughtful statistical consideration be given all aspects of sampling and subsampling with appropriate modification to ''standard'' protocols made as required.

In the present example, the sampling design should have specified that samples of stone of the size fraction to be removed be collected. Following Gy theory (Gy, 1992; Pitard, 1993), the stone of the collected sample should have been crushed and mixed prior to selection of the aliquot for assay. Alternatively, solvent extraction could have been performed on the entire as collected sample with subsampling of the ''extractate.''

What about soil?

The problems associated with the sampling and assay of the stones are obvious because they are highly visual. Less visual are the similar inferential problems associated with the compositing and assay of all bulk materials. This is particularly true of soil. It is largely a matter of scale. One can easily observe the differences in size and composition of stone chips, but differences in the types and sizes of soil particles are less obvious to the eye of the sample collector.

Yet, because these differences are obvious to the assaying techniques, one must be extremely cautious in assuming the support of any analytical result. Care must be exercised in the sampling design, collection and assay that the sampling assaying processes do not contradict either the needs of the remediator or the dictates of the media and site correlation structure.

In situ soil is likely to exhibit a large degree of heterogeneity. Changes in soil type and moisture content may be extremely important to determinations of bio-availability of import to risk based decisions (for instance, see Miller and Zepp; 1987, Marple *et al.*, 1987; and Umbreit *et al.*, 1987). Consideration of such issues is absolutely essential if appropriate sampling designs are to be employed for making decisions regarding a meaningful observational unit.

Composite samples of soil are usually formed in the field so as to minimize shipping costs. Generally, the subsamples of soil forming the composite sample are mixed in a stainless steel bowl. It will be assumed that these subsamples have been correctly taken from the areal unit of interest and that the field composite does ''reflect'' the average properties of the unit.

A soil sample, composite or otherwise, typically is sent to the analytical laboratory in a container that can be described as a ''quart'' jar. The contents of this container weigh approximately a kilogram depending, of course, on the soil moisture content and density. An aliquot is extracted from this container for assay by the laboratory according to the accepted assay protocol. The weight of the aliquot is 30 g for organics and 5 g for metals. Assuming an organic assay, there are 33 possible aliquots represented in the typical sampling container. Obviously, there are six times as many represented for a metals analysis.

If an organics assay is to be performed, the organics are extracted with a solvent and the ''extractate'' concentrated to a volume of 10 ml. Approximately one-to-five micro liters (about nine drops) are then taken from the 10 ml of ''extractate'' and injected into the gas chromatograph-mass spectrometer for analysis. Thus, there are approximated 2000 possible injection volumes in the 10 ml of ''extractate.'' This means that there are 66,000 possible measurements that can be made from a ''quart'' sample container. While assuming a certain lack of heterogeneity within a 10 ml volume of ''extractate'' may be reasonable, it may be yet another matter to assume a lack of heterogeneity among the 30 g aliqouts from the sample container (see Pitard, 1993).

A properly formed composite sample retains the heterogeneity of the entity sampled although, if throughly mixed, may alter the distributional properties of the in situ material. However, the effects of gravity may well cause particle size segregation during transport. If the laboratory then takes the ''first'' 30 g aliquot from the sample container, without thorough remixing of all the container's contents, the measurement provided by the assay cannot be assumed to be a reasonable estimate of the average concentration of the one kilogram sample.

Assessment of measurement variation

How well a single assay result describes the average concentration desired can only be assessed by investigating the measurement variation. Unfortunately, such an assessment is usually only considered germane to the quality control/quality assurance portion of environmental investigations. Typically there is a requirement to have the analytical laboratory perform a duplicate analysis once every 20 samples. Duplicate analyses involve the selection of a second aliquot from the submitted sample, and the preparation and analysis of it as if it were another sample. The results are usually reported in terms of the relative percent difference (RPD) between the two measurement results. This provides some measure of precision which not only includes the laboratory's ability to perform a measurement, but also the heterogeneity of the sample itself.

The RPD provides some estimate of the ability of an analytical measurement to characterize the material within the sample container. One often wonders what the result would be if a third, and perhaps a fourth aliquot were taken from the sample container and measured. The RPD, while meaningful to chemists, is not adequate to characterize the variation among measures on more than two aliquots from the same sample container. Therefore, more traditional statistical measures of precision are required, such as the variance or standard deviation.

In regard to determining the precision of the measurement, most everyone would agree that the 2000 possible injections to the gas chromatograph/mass spectrometer from the 10 ml extractate would be expected to show a lack of heterogeneity. However, everyone might not agree that the 33 possible 30 g aliqouts within a sample container would also be lacking in heterogeneity.

Extending the sampling frame to ''small'' increments of time or space, introduces into the measurement system sources of possible heterogeneity which include the act of composite sample collection as well as those inherent to the media sampled. Gy (1992), Liggett (1995a, 1995b, 1995c) and Pitard (1993) provide excellent discussions of the statistical issues.

Having an adequate characterization of the measurement system variation may well assist in defining appropriate sampling designs for estimation of the desired average characteristic for the decision unit. Consider this example extracted from data contained in the site Remedial Investigation/Feasibility Study (RI/FS) reports for a confidential client. Similar data may be extracted from the RI/FS reports for almost any site.

Fig. 1 presents the results of duplicate measurements of 2,3,7,8-TCDD in soil samples taken at a particular site. These results are those reported in the quality assurance section of the site characterization report and are plotted against their respective means. The ''Prediction Limits'' shown in this figure will, with 95% confidence, contain an additional single measurement (Hahn, 1970a, 1970b). If one considers all the measurements of 2,3,7,8-TCDD made at the site and plots them versus their mean, the result is shown in Fig. 2. Note that all of these measurements lie within the prediction limits constructed from the measurement system characterization. This reflects the results of an Analysis of Variance indicting that the variation in log-concentration among sample locations at the site is not significantly different than the variation among repeated measurements made on the same sample.

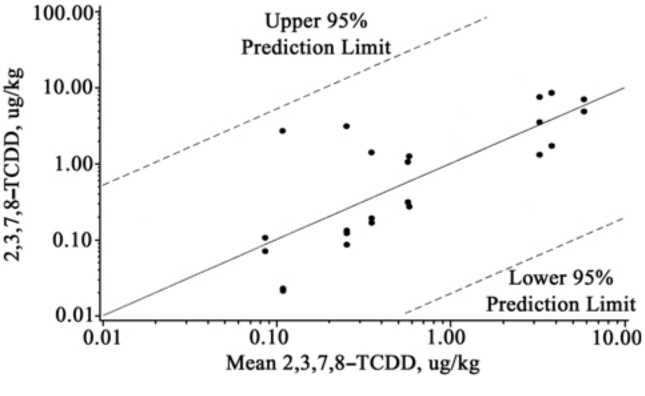

Figure 1. Example site, 2,3,7,8-TCDD. Sample repeated analyses versus mean.

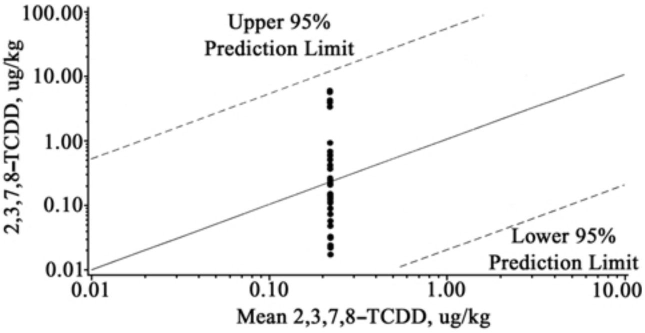

Figure 2. Example site, 2,3,7,8-TCDD. Samples in close proximity versus their mean. Neighborhood: Total site.

Two conclusions come to mind. One is that the total variation of 2,3,7,8-TCDD concentrations across the site is the same as that describing the ability to make such measurement. The second is that had a composite sample been formed from the soil at this site, a measurement of 2,3,7,8-TCDD concentration made on the composite sample would be no closer to the site average concentration than one made on any single sample. This is because the inherent heterogeneity of 2,3,7,8-TCDD in the soil matrix is a major component of its concentration variation at the site. Thus, the composited sample will also have this heterogeneity.

The statistically inclined are likely to find the above conclusion counterintuitive. Upon reflection however, one must realize that regardless of the size of the sample sent to the laboratory, the assay is performed on only a small fractional aliquot. The support of the resulting measurement extends only to the assayed aliquot. In order to achieve support equivalent to the size of the sample sent, it is necessary to either increase the physical size of the aliquot assayed, or increase the number of aliquots assayed per sample and average their results. Alternatively, one could grind and homogenize the entire sample sent before taking the aliquot for assay. In light of this, one wonders what is really implied in basing a risk assessment for 2,3,7,8-TCDD on the upper 95% confidence limit for the mean concentration of 30 g aliquots of soil.

In other words, more thought should be given to the support associated with an analytical result during sampling design. Unfortunately, historically the ‘‘relevant guidance’’ on site sampling contained in many publications of the USEPA does not adequately address the issue. Therefore, designing sampling protocols to achieve a desired decision support is largely ignored in practice.

Useful compositing—mixing oil and water

The assay procedure for determining the quantity of total oil and grease (O&G) in groundwater via hexane extraction requires that an entire 1 liter sample be extracted. This also includes the rinsate from the sample container. Certainly, the measurement of O&G via the hexane extraction method characterizes a sample volume of 1 liter. Therefore, the actual ‘‘support’’ is a 1 liter volume of groundwater. Rarely, if ever, are decisions required for volumes this small.

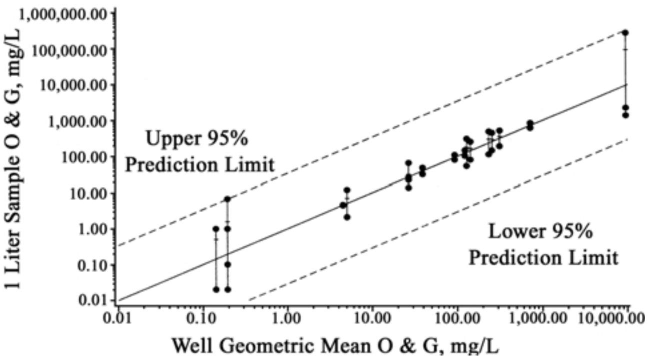

Figure 3. Groundwater oil and grease concentration—hexane extraction. Individual 1 liter sample analyses by source well geometric mean.

A local municipal water treatment plant will take 2400 gallons (9085 liters) per day of water, if the average O&G concentration is less than 50 milligrams per liter (mg/l). To avoid fines and penalties, water averaging greater than 50 mg/l O&G must be treated before release. Some wells monitoring groundwater at a former industrial complex are believed to monitor uncontaminated groundwater. Other wells are thought to monitor groundwater along with sinking free product. The task is to develop a means of monitoring groundwater to be sent to the local municipal treatment plant.

Fig. 3 presents the results of a sampling program designed to estimate the variation of O&G measurements with 1 liter support. This program involved the repeated collection of 1 liter grab samples of groundwater from the various monitoring wells at the site over a period of several hours. Obviously, a single grab sample measurement for O&G does not provide adequate support for decisions regarding the average O&G concentration of 2400 gallons of groundwater. However, being able to estimate the within well mean square assists the development of an appropriate sampling design for monitoring discharged groundwater.

Confidence limits for the true mean O&G concentration, as would be estimated from composite samples having 24-hour support, are presented in Fig. 4. This certainly suggests that an assay of a flow-weighted composite sample would provide a reasonable estimate of

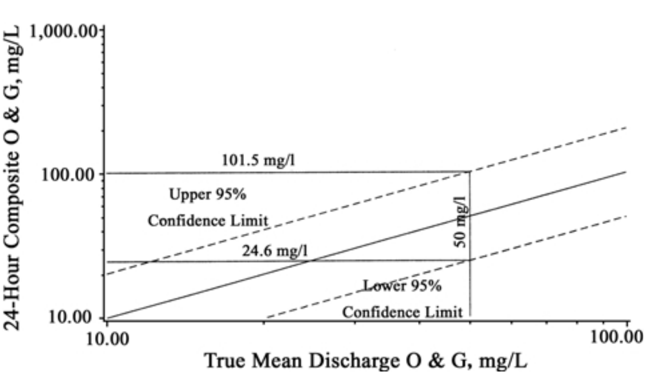

Figure 4. Site discharge oil and grease concentration. Proposed compliance monitoring design based upon 24-hour composite sample.

the true mean O&G concentration during some interesting time span. The exercise also provides material to begin drafting discharge permit conditions based upon a composite over a 24 hour period. These might be stated as follows: (1) If the assay of the composite sample is less than 24 mg/l O&G then the discharge criteria is met; (2) If this assay result is greater than 102 mg/l then the discharge criteria has not been met. While this example may seem intuitively obvious to statisticians, it is this author's experience that the concept is totally foreign to many engineers and environmental managers.

Useful compositing—the dirty floor

Another example of the potential for compositing to provide adequate support for decision making is given by determination of surface contamination by polychlorinated biphenyls (PCBs). Consider the case of a floor contaminated with PCBs during an electrical transformer fire. The floor is located remotely from the transformer room, but may have been contaminated by airborne PCBs via the building duct work. The criteria for reuse of PCB contaminated material is that the PCB concentration must be less than 10 micrograms per 100 square centimeters ($\mu g/100\,cm^2$). That is the entire surface must have a surface concentration of less than $10\,\mu g/100\,cm^2$.

The determination of surface contamination is usually via ''wipe'' sampling. Here a treated filter type material is used to wipe the surface using a template which restricts the amount of surface wiped to $100\,cm^2$. The ''wipes'' are packaged individually and sent to the laboratory for extraction and assay. The final chemical measurement is preformed on an aliquot of the ''extractate.''

Suppose that the floor has been appropriately sampled (Ubinger, 1987). A determination regarding the ''cleanliness'' of the floor may be made from an assay of composited extractate if the following conditions are satisfied. One, the detection limit of the analytical method must be at least the same fraction of the criteria as the number of samples composited. In other words, if the extractate from four wipe samples is to be composited, the method detection limit must be $2.5\,\mu g/100\,cm^2$ or less. Two, it must be assumed that the aliquot taken from the sample extractate for composite formation is ''representative'' of the entity from which it was taken. This assumes that the wipe sample extractate lacks heterogeneity when the subsample aliquot is selected.

If the assay results are less than $2.5\,\mu g/100\,cm^2$, then the floor will be declared clean and appropriate for reuse. If, on the other hand, the result is greater than $2.5\,\mu g/100\,cm^2$, the remaining extractate from each individual sample may be assayed to determine if the floor remains uniformly contaminated, or only a portion of it.

Comments on stuff blowing in the wind

Air quality measurements are inherently made on composite samples. Most are weighted by the air flow rate through the sampling device. The only air quality measure that comes to mind as not being a flow weighted composite is a particulate deposition measurement. It appears to this writer that it is the usual interpretation that air quality measurements made by a specific monitor represent the quality of ambient air in the general region of the monitor. It also appears to this writer that it is legitimate to ask how large an ambient air region is described by such a measurement.

Fig. 5 illustrates the differences in hourly particulate (PM_{10}) concentrations between co-located monitors. Fig. 6 illustrates the differences in hourly PM_{10} between two monitors separated by approximately 10 feet. All of these monitors are located at the Lincoln Monitoring site in Allegheny County, Pennsylvania. This is an industrial area with a multiplicity of potential sources of PM_{10}. These potential sources can be described as both major and minor. The inlets for the co-located monitors are at essentially the same location.

The observed differences in hourly PM_{10} measurements for the monitors with 10 foot separation is interesting for several reasons. The large magnitude of some of these differences certainly will affect the difference in the 24 hour average concentrations. This magnitude is as much as 7–8 μg/cubic meter on June 17 and 19. During periods when the measured concentration is near the 150 μg/cubic meter standard, such a difference could effect the determination of attainment. Because the standard is health based and presumes a 24-hour average exposure, the support of the ambient air quality measurement takes on increased importance.

If the support of an ambient air quality measurement is only in regard to inferences regarding a rather small volume of air, say within a 10-foot semi-sphere around the monitor, it is unlikely to describe the exposure of anyone not at the monitor site. Certainly, there is no support from this composite sample measurement for the making of inferences regarding air quality within a large region unless it can be demonstrated that there is no heterogeneity within the region. This requires a study of the measurement system variation utilizing monitors placed at varying distances apart. In truth, any ambient air quality monitor can only composite a sample of air precisely impinging on the monitor's inlet. It cannot form an adequate composite sample of air in any reasonable spatial region surrounding that monitor.

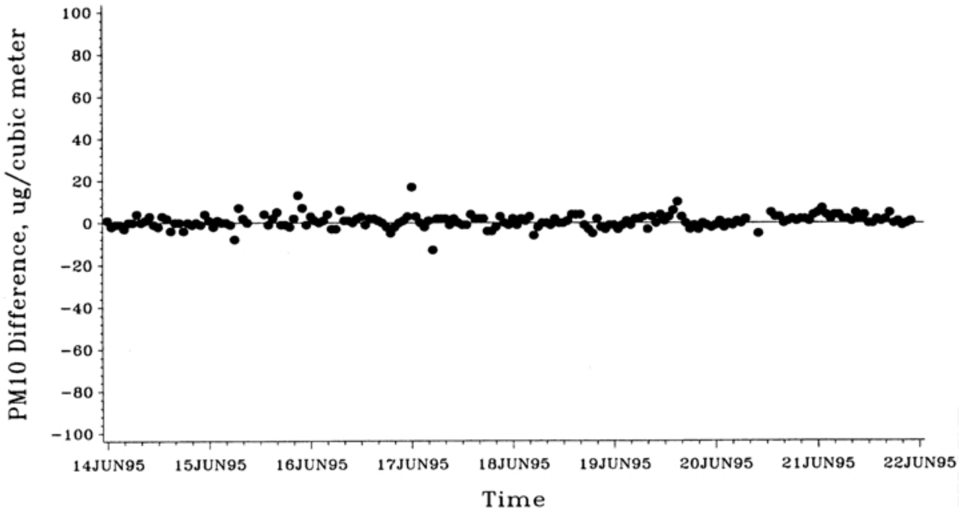

Figure 5. Hourly particulate (PM10) monitoring results. Single monitoring site, June 14–21, 1995. Differences between co-located monitoring devices.

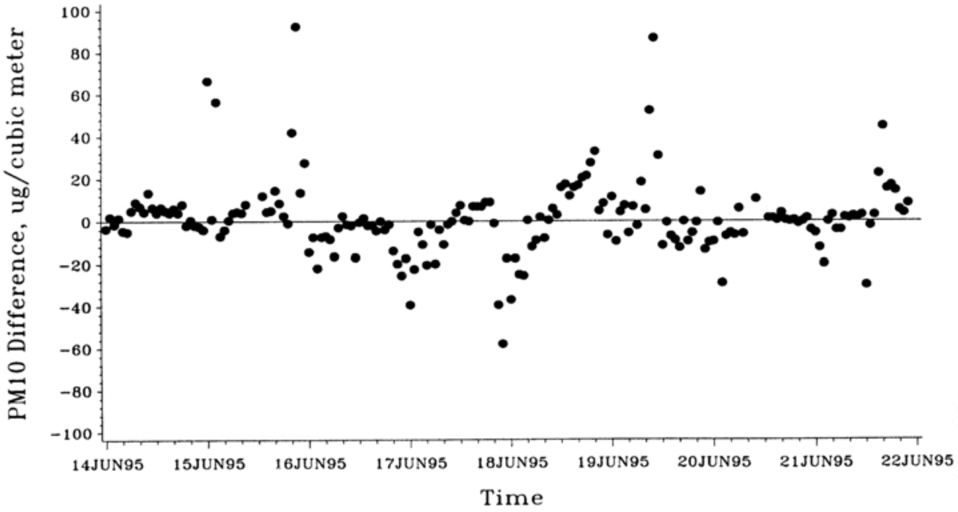

Figure 6. Hourly particulate (PM10) monitoring results. Single monitoring site, June 14–21, 1995. Differences between monitoring devices 10 feet apart.

Summary

On the surface composite sampling has a great deal of appeal. In practice this appeal is largely economic in that there is a promise of decreased sample processing, shipping and assay cost. However, if one is not very careful this economy may come at a large cost due to incorrect decision making. While the desired support may be carefully built into the formation of a composite soil sample, it may be poorly reflected in the final assay result.

This is certainly a problem that can be corrected by appropriate design. However, the statistician frequently is consulted only as a last resort. In such instances, we find ourselves practicing statistics in retrospection. Here the statistician needs to be particularly attuned to precisely defining the support of the measurement made before assisting with any inference. Failure to do so would just exacerbate the confusion as discussed by Englund and Heravi.

The confusion is also frequently exacerbated by regulation. Current regulations regarding the characterization of hazardous waste contains the following definition:

> (c) For purposes of this subpart, the Administrator will consider a sample obtained using any of the applicable sampling methods in appendix I to be a representative sample within the meaning of part 260 of this chapter. (40CFR §261.20).

Part 260 contains the following definition:

> *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. (40CFR §260.10).

If the phrase ''a sample'' is taken to mean any single physical entity upon which a measurement is made, there is a potential conflict between these two statements. Unless ‘‘a sample’’ refers to a composite sample with sufficient support to describe the average properties of the whole, the two statements are not compatible. If the measurement made on ‘‘a sample’’ does not truly support inferences regarding the average properties as intended, then clearly bad decision making will be the result. In addition, the measurement made on ‘‘a sample’’ provides no means for determining the ‘‘goodness’’ of the measurement in terms of its variance as required by USEPA regulations.

Acknowledgments

The publishing of articles is not an integral part of my consulting practice. I therefore asked several colleagues to review and comment on various drafts of this manuscript prior to submittal for publication. I gratefully acknowledge the efforts and suggestions of George Flatman, Mike Givevan, Walter Liggett and John Warren. Without their encouragement this article would probably never been finished. I also wish to thank the referees for their thoughtful comments and suggestions. The referees are quite correct that this is the type of article that, today, is not typically found in a refereed journal. However, maybe refereed journals ought to entertain more such articles on the problems of statistical consulting practice as a means of decreasing the communication gap between those interested in solving statistical problems and those whose interest is in solving problems using statistics.

References

Englund, E.J. and Heravi, N. (1994) Phased sampling for soil remediation. *Environmental and Ecological Statistics*, **1**, 247–63.

Flatman, G.T. and Yfantis, A.A. (1996) Geostatistical sampling designs for hazardous waste site, *Principles of Environmental Sampling*, ed. L. Keith, American Chemical Society, pp. 779–801.

Gy, P.M. (1992) *Sampling of Heterogeneous and Dynamic Material Systems: Theories of Heterogeneity, Sampling, and Homogenizing*, Elsevier, Amsterdam.

Hahn, G.J. (1970a) Statistical intervals for a normal population, Part I. Tables, examples and applications, *Journal of Quality Technology*, **2**, 115–25.

Hahn, G.J. (1970b) Statistical intervals for a normal population, Part II. Formulas, assumptions, some derivations, *Journal of Quality Technology*, **2**, 195–206.

Liggett, W.S. and Inn, K.G.W. (1995a) Pilot studies for improving sampling protocols, *Principles of Environmental Sampling*, L. Keith, (ed.) American Chemical Society, Washington, D.C.

Liggett, W.S. (1995b) Functional errors-in-variables models in measurement optimization experiments, *1994 Proceedings of the Section on Physical and Engineering Sciences*, American Statistical Association, Alexandria, VA.

Liggett, W.S. (1995c) Right measurement tools in the reinvention of EPA, *Corporate Environmental Strategy*, **3**, 75–8.

Marple, L., Brunck, R., Berridge, B., and Throop, L. (1987) Experimental and calculated physical constants for 2,3,7,8-tetrachlorodibenzo-*p*-dioxin, *Solving Hazardous Waste Problems Learning from Dioxins*, J. Exner, (ed.) American Chemical Society, Washington D.C., pp. 105–13.

Miller, G.C. and Zepp, R.G. (1987) 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin: Environmental chemistry, *Solving Hazardous Waste Problems Learning from Dioxins*, J. Exner, (ed.) American Chemical Society, Washington D.C., pp. 82–93.

Neptune, D., Brantly, E.P., Messner, M.J., and Michael, D.I. (1990) Quantitative decision making in superfund: A data quality objectives case study, *Hazardous Material Control*, May/June.

Olea, R. (1991) *Geostatistical Glossary and Multilingual Dictionary*, Oxford University Press, New York, NY.

Pitard, F.F. (1993) *Pierre Gy's Sampling Theory and Sampling Practice*, Second Edition, CRC Press, Boca Raton, FL.

Ubinger, E.B. (1987) Statistically valid sampling stratergies for PCB contamination, *Presented at the EPRI Seminar on PCB Contamination*, Kansas City, Missouri, October 6–9.

Umbreit, T.H., Hesse, E.J., and Gallo, M.A. (1987) Differential bioavailability of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin from contaminated soils, *Solving Hazardous Waste Problems Learning from Dioxins*, J. Exner, (ed.) American Chemical Society, Washington D.C., pp. 131–39.

USEPA. (1989) *Risk Assessment Guidance for Superfund: Human Health Evaluation Manual Part A*, EPA/540/1-89/002.

USEPA. (1994) *Guidance for the Data Quality Objectives Process, EPA QA/G-4*.

USEPA. (1996a) *Soil Screening Guidance: Technical Background Document*, EPA/540/R95/128.

USEPA, (1996b) *Soil Screening Guidance: User's Guide*, Pub. 9355.4–23.

Webster's, (1987) *Webster's Ninth New Collegiate Dictionary*, Merriam-Webster Inc. Springfield, MA.

Biographical sketch

Douglas E. Splitstone, Principal of Splitstone & Associates, has more than 30 years of experience in the application of statistical tools to the solution of environmental problems. He has served as a member of the Task Group on Epidemiology and Statistical Methodology for the USEPA's Center for Environmental Epidemiology at the University of Pittsburgh's Graduate School of Public Health; a member of the Peer Review Group, Salt Host-Rock Portion of the Department of Energy's Civilian Radioactive Waste Management Program for Argonne National Laboratory; a faculty member of the 1992 Health Physics Society's Annual Summer School; and a member of the adjunct faculty at The Pennsylvania State University and Indiana University of Pennsylvania. He is currently a consultant to the USEPA's Science Advisory Board serving on the Environmental Engineering Committee's Quality Management, Secondary Data Use and Statistical Consultation Subcommittees. His efforts in the application of statistical techniques to environmental problem solving were recognized by the American Statistical Association in 1993 with the award of the Distinguished Achievement Medal by the Section on Statistics and the Environment. He holds the degree of Master of Science in Mathematical Statistics awarded by Iowa State University in 1967.