

Composite sampling

G.P. Patil

Volume 1, pp 387–391

in

Encyclopedia of Environmetrics
(ISBN 0471 899976)

Edited by

Abdel H. El-Shaarawi and Walter W. Piegorsch

© John Wiley & Sons, Ltd, Chichester, 2002

Composite sampling

The high costs of laboratory analytical procedures frequently strain environmental and public health budgets. Whether soil, water or biological tissue is being analyzed, the cost of testing for chemical and pathogenic contaminants can be quite prohibitive.

Composite sampling can substantially reduce analytical costs because the number of required analyses is reduced by compositing several samples into one and analyzing the composited sample. By appropriate selection of the composite sample size and retesting of select individual samples, composite sampling may reveal the same information as would otherwise require many more analyses.

Many of the limitations of composite sampling have been overcome by recent research, thus providing more widespread potential for the use of composite sampling to reduce costs of environmental and public health assessments, while maintaining and often increasing the precision of sample-based inference.

When the objective is to estimate the population mean or total, compositing will always reduce analytical cost; however, a sufficient number of composite samples must still be obtained for estimating the variance.

When the objective is to classify each individual sample, with say subsequent estimation of the prevalence of a binary trait or proportion of non-compliant measurements (*see* **Binary data**), testing composite samples with selective retesting becomes cost-effective when the prevalence or proportion is low. Examples where composite sampling can be very cost-effective for **classification** include estimating the prevalence of a rare disease or verifying whether a hazardous waste site has been sufficiently remediated (*see* **Restoration, environmental**).

Conventional statistical techniques allow for the reduction of either cost or uncertainty. However, the reduction of one of these factors is typically at the expense of an increase in the other. Composite sampling can maintain cost or uncertainty at a specified level while decreasing the other component.

Compositing simply refers to physically mixing individual samples to form a composite sample, as visualized in Figure 1. A single analysis is performed on the composite, which is used to represent each of the original individual samples.

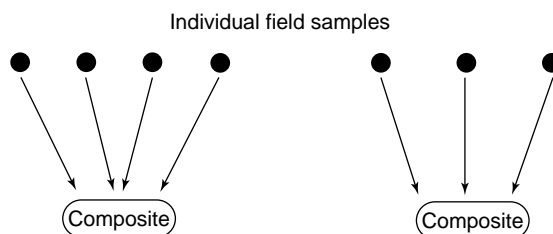


Figure 1 Forming composite samples from individual samples

When analytical costs dominate over sampling costs, the savings potential is obviously high; however, the immediate question is ‘How is it possible to compensate for information that is lost due to compositing?’ More specifically, if we are testing whether or not a substance is present or existing at a concentration above some threshold, then we do not want to dilute individual ‘contaminated’ samples with clean samples so that the analysis does not detect any contamination. Furthermore, if the measurements are of a variable such as a chemical concentration, it may be necessary to know the actual values of those individual samples with the highest concentrations. For example, ‘hot spots’ need to be identified at hazardous waste sites.

Through judicious choice of a strategy for retesting some of the original individual samples based on composite sample measurements, many limitations of composite sampling can be overcome. Furthermore, other innovative applications of composite sampling are emerging, such as combining with **ranked set sampling**.

Composite Sampling Method

By way of clarification, a ‘sample’ in this entry refers to a physical object to be measured, whether an individual or a composite, and not a collection of observations in the statistical sense. Individual sample units are what is obtained in the field, such as soil cores or fish fillets; or obtained from subjects, such as blood samples. Meanwhile, a composite sample may be a physical mix of individual sample units or a batch of unblended individual sample units that are tested as a group. Most compositing for environmental assessment and monitoring consists of physically mixing individual units to make a composite sample that is as homogeneous as possible.

2 Composite sampling

With classical sampling, no distinction is made between the process of sampling (i.e. selection or inclusion) and that of observation or measurement. It is assumed, with classical sampling, that any unit selected for inclusion in a statistical sample is measured and hence its value becomes known (*see Sampling, environmental*). In composite sampling, however, there is a clear distinction between the sampling and measurement stages. Compositing takes place between these two stages, and therefore achieves two otherwise conflicting goals. While a large number of samples can be selected to satisfy sample size requirements, the number of analytical measurements is kept affordable.

If a variable of concern is a measurement that is continuous in nature such as a chemical concentration, then the mean (arithmetic average) of composite samples provides an unbiased estimate of the true but unknown population mean. Also, if measurement error is known, the population variance based on the scale of the individual samples can be estimated by a simple weighting of the measured composite sample variance.

With selective retesting of individual sample units, based on initial composite sample results, all of the individual sample units can be classified according to the presence or absence of a trait, or exceedance (vs. compliance) of a numerical standard (*see Exceedance over threshold*). The prevalence of a trait or proportion of noncompliance can subsequently be estimated. If a composite measurement does not reveal a trait in question or is in compliance, then all individual samples comprising that composite are classified as 'negative'. When a composite tests positive, then retesting is performed on the individual samples or subsamples (aliquots) in order to locate the source of 'contamination'.

Retesting, as visualized in a general sense in Figure 2, may simply be exhaustive retesting of all individuals comprising a composite or may entail more specialized protocols. Generally, as the retesting protocol becomes more sophisticated, the expected number of analyses decreases. Therefore, one must consider any increased logistical costs along with the expected decrease in analytical cost when evaluating the overall cost of a compositing/retesting protocol.

Owing to recent research [22], the individual samples with the highest value, along with those individual samples comprising an upper percentile, can

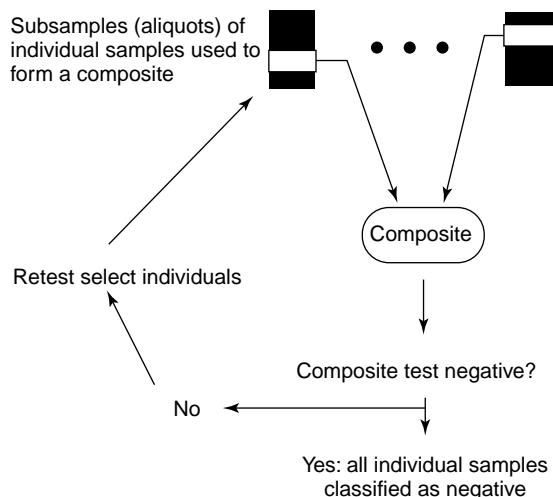


Figure 2 Composite sampling with retesting

be identified with minimal retesting. This ability is extremely important when 'hot spots' need to be identified such as with soil monitoring at a hazardous waste site.

Whether the data being dealt with are from binary (presence/absence) measurements or from measurements on a continuum, composite sampling can result in classifying each individual sample without having to analyze each one separately. While composite sampling may not be feasible when the prevalence of contamination is high, the analytical costs can be drastically reduced as the number of contaminated samples decreases.

Circumstances that may presently disqualify composite sampling from being applied may change with further advances in technology. Long turn-around time for laboratory results and large labor costs may eliminate optimal retesting designs from consideration. However, retesting designs in the future may be automated and guided by an expert system (Rajagopal, 1990, personal communication). Also, advances in statistical methodology may further extend the utility of composite sampling.

For other reviews of composite sampling, see [2], [8]–[10], [25] and [26]. For an overview, see [23].

Applications

Composite sampling has its roots in what is known as *group testing*. An early application of group testing

was to estimate the prevalence of plant virus transmission by insects [31]. In this application, insect vectors were allowed to feed upon host plants, thus allowing the disease transmission rate to be estimated from the number of plants that subsequently become diseased.

Apparently, the next important application of group testing occurred during the Second World War when US servicemen were tested for syphilis by detecting the presence or absence of a specific antigen of the syphilis-causing bacterium in samples of their blood [7].

Composite sampling is increasingly becoming an acceptable practice for sampling soils (*see Soil surveys*), biota, and bulk materials when the goal is estimation of some population value under restrictions of a desired standard error and/or limits on the cost of sampling. Additional applications include:

- Establishing and verifying attainment of remedial cleanup standards in soils using sample compositing and **bootstrap resampling** techniques.
- Use of compositing to obtain adequate support in geostatistical sampling.
- Optimal compositing strategies for screening material for **hazardous agents**.
- A soil sample design utilizing techniques of compositing, binary search, and confidence limits on proportions.
- Composite sampling for analyzing foliage and other biological materials.

Examples that are particularly relevant to environmental and public health studies include, soil sampling for characterization of soil polychlorinated biphenyl (PCB) contamination at gas pipeline compressor stations [11, 23] and characterization of soil polycyclic aromatic hydrocarbon (PAH) contamination at a **superfund site** [16]; **groundwater monitoring** [24, 28]; indoor air monitoring for allergens [15] (*see Personal exposure monitoring*); measuring **bioaccumulation** in human adipose tissue [19]; assessing contamination in fish [20]; assessing contaminants in mollusks [17]; and measuring average fat content in bulk milk [5, 32]. Also see the special issue of *Environmental and Ecological Statistics* on composite sampling [1, 3, 4, 6, 12–14, 18, 21, 27, 29, 30].

References

- [1] Blomqvist, P. (2001). A proposed standard method for composite sampling of water chemistry and plankton in small lakes, *Environmental and Ecological Statistics* **8**, 121–134.
- [2] Boswell, M.T. & Patil, G.P. (1987). A perspective of composite sampling, *Communications in Statistics – Theory and Methods* **16**, 3069–3093.
- [3] Carson, J.H. (2001). Analysis of composite sampling data using the principle of maximum entropy, *Environmental and Ecological Statistics* **8**, 201–211.
- [4] Colón, S., Patil, G.P. & Taillie, C. (2001). Estimating prevalence using composites, *Environmental and Ecological Statistics* **8**, 213–236.
- [5] Connolly, J. & O’Connor, F. (1981). Comparison of random and composite sampling methods for the estimation of fat content of bulk milk supplies, *Irish Journal of Agricultural Research* **20**, 35–51.
- [6] Correll, R.L. (2001). The use of composite sampling in contaminated sites – a case study, *Environmental and Ecological Statistics* **8**, 185–200.
- [7] Dorfman, R. (1943). The detection of defective members of large populations, *Annals of Mathematical Statistics* **14**, 436–440.
- [8] Elder, R.S. (1977). Properties of composite sampling procedures. Ph.D. Dissertation. Virginia Polytechnic Institute and State University, Blacksburg.
- [9] Elder, R.S., Thompson, W.O. & Myers, R.H. (1980). Properties of composite sampling procedures, *Technometrics* **22**, 179–186.
- [10] Garner, F.C., Stapanian, M.A. & Williams, L.R. (1988). Composite sampling for environmental monitoring, in *Principles of Environmental Sampling*, L.H. Keith, ed., American Chemical Society, Washington, pp. 363–374.
- [11] Gore, S.D., Patil, G.P. & Taillie, C. (1992). Studies on the applications of composite sample techniques in hazardous waste site characterization and evaluation: II. Onsite surface soil sampling for PCB at the Armagh Site, Technical Report Number 92-0305, Center for Statistical Ecology and Environmental Statistics, Penn State University, University Park.
- [12] Gore, S.D., Patil, G.P. & Taillie, C. (2001). Identifying the largest individual sample value from a two-way composite sampling design, *Environmental and Ecological Statistics* **8**, 151–162.
- [13] Johnson, G.D. & Patil, G.P. (2001). Cost analysis of composite sampling for classification, *Environmental and Ecological Statistics* **8**, 91–107.
- [14] Kosmelj, K., Cedilnik, A. & Kalan, P. (2001). Comparison of a two stage sampling design and its composite sample alternative: An application to soil studies, *Environmental and Ecological Statistics* **8**, 109–119.
- [15] Lintner, T.J., Maki, C.L., Brame, K.A. & Boswell, M.T. (1992). Sampling dust from human dwellings to estimate the prevalence of indoor allergens, Technical Report Number 92-0805, Center for Statistical Ecology

4 Composite sampling

- and Environmental Statistics, Penn State University, University Park.
- [16] Messner, M.J., Clayton, C.A., Michael, D.I., Neptune, M.D. & Brantly, E.P. (1990). Retrospective design solutions for a remedial investigation. Supplement to quantitative decision making in superfund: a data quality objectives case study, *Hazardous Materials Control* **3**.
- [17] National Oceanic and Atmospheric Administration. (1989). A Summary of Data on Tissue Contamination from the First Three Years (1986–1988) of the Mussel Watch Project. NOAA Technical Memorandum, NOS OMA 49.
- [18] Nussbaum, B.D. & Gilbert, R.O. (2001). Editorial: Special issue on composite sampling, *Environmental and Ecological Statistics* **8**, 89–90.
- [19] Orban, J.E., Lordo, R. & Schwemberger, J. (1990). Statistical methods for analyzing composite sample data applied to EPA's human monitoring program, manuscript.
- [20] Paasivirta, J. & Paukku, R. (1989). Use of composited samples to optimize the monitoring of environmental toxins, *Chemosphere* **19**, 1551–1562.
- [21] Patil, G.P. & Taillie, C. (2001). Use of best linear unbiased prediction for hot spot identification in two-way compositing, *Environmental and Ecological Statistics* **8**, 163–169.
- [22] Patil, G.P., Gore, S.D. & Sinha, A.K. (1994). Environmental chemistry, statistical modeling, and observational economy, in *Environmental Statistics, Assessment and Forecasting*, C.R. Cothorn & N.P. Ross, eds, Lewis, Boca Raton, pp. 57–97.
- [23] Patil, G.P., Gore, S.D. & Taillie, C. (1994). Design and analysis with composite samples: a novel method to accomplish observational economy in environmental studies. Technical Report Number 94–0410, Center for Statistidcal Ecology and Environmental Statistics, Penn State University, University Park.
- [24] Rajagopal, R. & Williams, L.R. (1989). Economics of sample compositing as a screening tool in ground water quality monitoring, *Ground Water Monitoring Review* **9**, 186–192.
- [25] Rohde, C.A. (1976). Composite sampling, *Biometrics* **32**, 273–282.
- [26] Rohde, C.A. (1979). Batch, bulk and composite sampling, in *Sampling Biological Populations*, R.M. Cormack, G.P. Patil & D.S. Robson, eds, International Cooperative Publishing House, Fairland, pp. 365–377.
- [27] Schaeffer, D.J. (2001). Discussion on 'A proposed standard method for composite sampling of water chemistry and plankton in small lakes' by P. Blomqvist, *Environmental and Ecological Statistics* **8**, 135–136.
- [28] Schaeffer, D., Kerster, H.W. & Janardan, K.G. (1982). Monitoring toxics by group testing, *Environmental Management* **6**, 467–469.
- [29] Splitstone, D.E. (2001). Sample support and related scale issues in composite sampling, *Environmental and Ecological Statistics* **8**, 137–149.
- [30] van Belle, G., Griffith, W.C. & Edland, S.D. (2001). Contributions to composite sampling, *Environmental and Ecological Statistics* **8**, 171–180.
- [31] Watson, M.A. (1936). Factors affecting the amount of infection obtained by aphid transmission of the virus Hy. III, *Philosophical Transactions of the Royal Society of London, Series B* **226**, 457–489.
- [32] Williams, C.J. & Peterson, R.G. (1978). Variation in estimates of milk fat, protein and lactose content associated with various bulk milk sampling programs, *Journal of Dairy Science* **61**, 1093.

G.P. PATIL