

Kansas State University Libraries  
**New Prairie Press**

---

Conference on Applied Statistics in Agriculture

1990 - 2nd Annual Conference Proceedings

---

## COMPOSITE SAMPLING TECHNIQUES FOR DETERMINING PESTICIDE CONCENTRATIONS

R. S. Parrish

G. O. Ware

C. N. Smith

P. A. Banks

*See next page for additional authors*

Follow this and additional works at: <https://newprairiepress.org/agstatconference>

 Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

---

### Recommended Citation

Parrish, R. S.; Ware, G. O.; Smith, C. N.; and Banks, P. A. (1990). "COMPOSITE SAMPLING TECHNIQUES FOR DETERMINING PESTICIDE CONCENTRATIONS," *Conference on Applied Statistics in Agriculture*.  
<https://doi.org/10.4148/2475-7772.1429>

---

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact [cads@k-state.edu](mailto:cads@k-state.edu).

---

**Author Information**

R. S. Parrish, G. O. Ware, C. N. Smith, and P. A. Banks

## COMPOSITE SAMPLING TECHNIQUES FOR DETERMINING PESTICIDE CONCENTRATIONS

R. S. Parrish, Computer Sciences Corporation

G. O. Ware, University of Georgia

C. N. Smith, U. S. Environmental Protection Agency

P. A. Banks, University of Georgia

### ABSTRACT

Composite sampling techniques are compared with random sampling methods for determining pesticide concentrations in agricultural fields. Estimates of mean pesticide concentrations and associated standard errors are presented for different experimental conditions. Variance components defined in extended forms of the Brown-Fisher model are estimated. The method of nonlinear least squares was employed to obtain numerical estimates of variance components by equating observed mean squares to expected mean squares for appropriate sampling designs.

Key Words: Composite Sampling, Nonlinear Models

### 1. INTRODUCTION

For experimental plots on which pesticides are applied, the usual method for determining pesticide concentrations involves taking individual random soil samples, determining precise chemical or biological characteristics of each sample, and computing the mean concentrations with associated bounds on the error of estimation. The expense of processing a large number of random samples to determine precise chemical or biological characteristics of the soil, however, encourages the use of composite sampling strategies that, by definition, involve physical averaging of many samples and result in laboratory analysis of relatively few samples. In addition, information on point-to-point variability generally is lost when composite sampling approaches are used.

From 1984 through 1987, the U. S. Environmental Protection Agency (USEPA) and the U. S. Geological Survey (USGS) were involved in a joint research project in southwest Georgia (Lee County). The major objective of the field study was to develop an extensive database for studying pesticide movement through the soil profile and testing models that predict such movement. Through extensive site characterization, random sampling, and monitoring activities, data were gathered to estimate statistical characteristics of the field site and to estimate mean concentrations, point-to-point variability, and standard errors of applied pesticides at various points in space and time. These estimates were required for performing uncertainty analyses. Due to the extensive number of soil samples that were being collected each year and the expense involved in the laboratory analysis of these samples, it was decided that at the beginning of the fourth year of the project a pilot study would be undertaken to investigate composite sampling techniques for estimating these

statistical parameters and to compare them with those obtained by random sampling methods for determining pesticide concentrations.

At about the same time, a weed science group in the Department of Agronomy at the University of Georgia (UGA) had several research projects that involved the monitoring of pesticide residues in agricultural fields. Likewise, the project staff was very interested in exploring composite sampling strategies that might provide the same sampling efficiency but at a reduced cost. In 1988, a small pilot project was initiated on the Plant Sciences Farm at the University of Georgia to compare composite sampling with random sampling.

Through collaboration among statisticians and soil scientists involved in these independent research studies, strategies were developed for establishing composite sampling designs that would enable the estimation of mean pesticide concentrations with associated standard errors and still provide the capability for estimating other variance components. The collaboration permitted the investigation of composite sampling techniques and the comparison of composite sampling with random sampling as methods for estimating pesticide concentrations. The joint investigation focused on the composite sampling models proposed by Brown and Fisher (1972), Rhode (1976), and Elder *et al.* (1980).

## 2. COMPOSITE SAMPLING METHODOLOGY

In one form of composite sampling, groups of individual soil samples are physically mixed to form composite samples. Subsamples are drawn from these composite samples and analyzed to determine chemical characteristics. This process results in subsamples that are comprised of differing proportions from the individual samples. Because the relative contributions are random, this sampling technique is referred to as composite sampling with random proportions. The proper estimation of population means and variance components for this type of experimental situation requires specific statistical methods. Elder *et al.* (1980) discussed this type of problem; their approach can be described as follows.

Suppose  $m$  equal-sized portions of soil material (increments) are drawn randomly from the field. From these,  $r$  composites are formed by randomly partitioning the set of increments into  $r$  subsets of  $n$  increments each and physically mixing the increments in each subset. Then,  $s$  subsamples are drawn randomly from each composite, and  $t$  analyses are run on each subsample. Each composite is assumed to contain enough material to provide a maximum of  $S$  subsamples ( $s \leq S$ ). If the subsample size is the same as the increment size, then  $S = n$ . This composite sampling approach is displayed graphically in Figures 1 and 2.

To mathematically represent the variables involved in the composite sampling model, the following notation is defined:

$a_{ijl}$  = the proportion of the  $j$ th subsample from the  $i$ th composite that comes from the  $l$ th increment in that composite,

$X_{ijl}$  = the concentration value associated with the portion of the  $l$ th increment that appears in the  $j$ th subsample from the  $i$ th composite.

Figure 2 shows the case of three increments and corresponding subsamples. Each variable  $a_{ijl}$  has mean  $\mu_a$  and variance  $\sigma_a^2$ . The variables  $X_{ijl}$  have equal means  $\mu_x$  and variances  $\sigma_x^2 + \sigma_w^2$ , where  $\sigma_x^2$  is the between-increment variance and  $\sigma_w^2$  is the within-increment variance subsequent to compositing and subsampling.

In different composites, the variables  $a_{ijl}$  are uncorrelated, but they are correlated within the same increment, in the same subsample, and in different increments and subsamples within a composite. Similarly, the variables  $X_{ijl}$  are correlated within the same increment, in the same subsample, and in different increments and subsamples within a composite. Elder *et al.* (1980) provided explicit expressions for these covariances.

The concentration value associated with the  $j$ th subsample from the  $i$ th composite is based on contributions from all increments in that composite. It can be expressed as the weighted mean

$$y_{ij} = \sum_{l=1}^n a_{ijl} X_{ijl} . \quad (1)$$

The value observed in the laboratory from the  $k$ th analysis, however, is

$$z_{ijk} = y_{ij} + e_{ijk} \quad (2)$$

where  $e_{ijk}$  is the measurement error associated with the  $k$ th test on the  $j$ th subsample from the  $i$ th composite. The variance of  $e_{ijk}$  is defined as  $\sigma_e^2$ . If the analysis procedure requires that additional subsampling be done, then an additional variance component should be incorporated.

### 3. ESTIMATION OF THE MEAN AND STANDARD ERROR

The estimator of  $\mu_x$  for composite sampling is

$$\bar{z} = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^t z_{ijk} / rst . \quad (3)$$

As a result of the covariance structures for the  $a_{ijl}$  and  $X_{ijl}$  variables, the variance for this estimator involves four individual variance components. Elder *et al.* (1980) provided

$$Var(\bar{z}) = \sigma_x^2/(rn) + [(S-s)/(S-1)][\sigma_w^2/n + n\sigma_a^2(\sigma_x^2 + \sigma_w^2)]/(rs) + \sigma_e^2/(rst). \quad (4)$$

Also, they established that the expected mean squares (between groups, within groups, and for error) associated with the analysis of variance for the composite sample groups are:

$$EMS_B = (st/n)\sigma_x^2 + [t(S-s)/(S-1)][\sigma_w^2/n + n\sigma_a^2(\sigma_x^2 + \sigma_w^2)] + \sigma_e^2 \quad (5)$$

$$EMS_W = [tS/(S-1)][\sigma_w^2/n + n\sigma_a^2(\sigma_x^2 + \sigma_w^2)] + \sigma_e^2 \quad (6)$$

$$EMS_E = \sigma_e^2 \quad (7)$$

Clearly, the variance of the estimator  $\bar{z}$  is equivalent to the value  $EMS_B/(rst)$ . Thus, the standard error of the estimate of  $\mu_x$  can be obtained from a standard analysis of variance computation.

#### 4. ESTIMATION OF VARIANCE COMPONENTS

Estimates of the individual variance components are necessary for some statistical studies. For example, it may be desirable to know the among-increment variance  $\sigma_x^2$ . Because of the nonlinearity involved in the EMS expressions, the difference between  $EMS_B$  and  $EMS_W$  does not result in a simple function of  $\sigma_x^2$  alone. In fact,

$$\frac{n}{st} [EMS_B - EMS_W] = \sigma_x^2 - \left( \frac{s}{S-1} \right) [\sigma_w^2 + n^2\sigma_a^2(\sigma_x^2 + \sigma_w^2)]. \quad (8)$$

Thus, an ANOVA-based estimate of the form  $(n/st)(MS_B - MS_W)$  generally will underestimate  $\sigma_x^2$ . If appropriate sampling designs are used, however, it is possible to utilize a nonlinear algorithm to obtain estimates for the four variance components.

One strategy that can be used to estimate the variance components involves the formation of several composite samples that are subsampled more than one time each. ANOVA computations can be performed to produce mean square values that then are equated to expected values as given above. A nonlinear least squares solution can be used to produce the estimates for the variance components.

For example, suppose  $r=15$  composite samples are constructed on the basis of  $n=20$  increments and  $s=3$  subsamples per composite, as described in Section 2. Mean squares between samples and among samples can be computed. Suppose also that an additional  $r=1$  composite sample is constructed separately on the basis of, say,  $n=10$  increments and  $s=10$  subsamples. These could be used to compute a sum of squares analogous to within-composites variation. Additional composite samples of  $n=20$  and  $n=30$  increments and  $s=15$  subsamples could be constructed separately in the same way. Such an approach provides several mean square values that can be equated to the EMS expressions above. A nonlinear minimization routine such as that provided by the NLIN procedure of the SAS System (1985) can be used to solve the resulting system of equations.

Figure 3 displays a sample SAS implementation of this algorithm. The input data consists of the mean square values (MS), number of subsamples (S), number of increments (N), and weights (WT). It is reasonable to assume that the weights should reflect the number of increments involved in the mean square computation; as such, it is suggested that the product of degrees of freedom (DF) from the ANOVA table and the number of increments (N) provides an appropriate weight for a given mean square (i.e.,  $WT = DF \cdot N$ ).

## 5. DESCRIPTION OF STUDIES

### USEPA Study

The USEPA study involved the measurement of pesticide concentrations applied on a 3.9-ha field in which peanuts were grown under modified conventional tillage practices. The pesticides aldicarb and metolachlor were applied to the site at the beginning of each of four cropping seasons. Aldicarb is a granular formulated insecticide and nematicide incorporated within the top 5.1 cm of the soil in a banded application at the time of planting. Metolachlor is a liquid formulated preemergence herbicide uniformly sprayed on the soil surface at the time of planting using a tractor-mounted broadcast spray rig. The target rate of application was 3.36 kg/ha for both pesticides. Individual soil samples (ca. 500g) were collected from the surface zone immediately after application, packed on ice, and transported to the laboratory where they were stored at 0°C. Compositing of the samples was accomplished at a later date using two different methods: hand mixing and machine mixing. Hand mixing was accomplished by using a large spoon to mix soil in a large stainless-steel pail. Machine mixing involved using a small concrete mixer for a period of 10 minutes for each composite sample. The composited samples were stored again at 0°C until laboratory analysis. Measurements for both compounds were made in units of kg/ha. For metolachlor, a  $\log_{10}$  normalizing transformation was applied.

Several composite samples were formed by collecting random samples in groups of 10, 20, and 30 increments. As indicated in Table 1, 15 composite samples were formed using 20 increments each, one composite sample was formed using 10 increments, one using 20 increments, and one using 30 increments. The numbers of subsamples taken from each composite also are given in the table. In addition, 80 random samples were collected for comparison purposes. The corresponding mean squares are given in Figure 4.

Table 1. Composite sampling designs.

| Number of Composites<br>(r) | Number of Increments /Composite<br>(n) | Number of Subsamples /Composite<br>(s) | Mean Square Between <sup>+</sup><br><i>MS<sub>B</sub></i> | Mean Square Within <sup>+</sup><br><i>MS<sub>W</sub></i> |
|-----------------------------|--|--|---|--|
| 15                          | 20                                     | 3                                      | x   | x  |
| 1                           | 10                                     | 10                                     |   | x  |
| 1                           | 20                                     | 15                                     |   | x  |
| 1                           | 30                                     | 15                                     |   | x  |

<sup>+</sup>x denotes that the mean square component can be formed for this specific design.

Table 2 summarizes the results corresponding to the 15 composite samples for both pesticides and both methods of compositing samples. Values for the estimated mean, its variance and standard error are given. The machine-mixing method resulted in larger mean values and larger standard errors when compared to the hand-mixing method for both compounds. This is perhaps due to the fact that the machine mixing tended to stratify the soil particles rather than to mix them uniformly. In contrast, the hand mixing technique provided a more homogeneous mixture.

Table 2. Estimated means, variances, and standard errors from the analysis of the composite samples where  $r=15$ ,  $n=20$ , and  $s=30$ .

|                           | Aldicarb |         | Metolachlor |         |
|---------------------------|----------|---------|-------------|---------|
|                           | Hand     | Machine | Hand        | Machine |
| $\bar{z}$                 | 2.07541  | 2.21040 | 0.38915     | 0.41343 |
| $\hat{V}(\bar{z})$        | 0.01098  | 0.01861 | 0.00026     | 0.00031 |
| $\sqrt{\hat{V}(\bar{z})}$ | 0.10478  | 0.13642 | 0.01598     | 0.01763 |

The results of the random sampling, in which individual increments were analyzed, are presented in Table 3. In the case of aldicarb, the estimate of the mean concentration from random sampling was larger than the values observed from either of the composite samples. To compare the standard errors of the mean for the composite versus random sampling methods, a standard error can be projected based on an equivalent number of

laboratory observations. On the basis of the estimated variance, the standard error projected for a random sample of size 45 is 0.19725 for aldicarb and 0.03389 for metolachlor. The projected standard errors for random sampling are approximately twice those for composite sampling. The smaller standard errors for composite sampling presumably are due to the physical averaging of 20 increments in each composite.

Table 3. Estimated means and variances from the analysis of the random samples.

|                    | Aldicarb | Metolachlor |
|--------------------|----------|-------------|
| $\bar{x}$          | 2.63314  | 0.40271     |
| $\hat{\sigma}_x^2$ | 1.75077  | 0.05167     |

The variance components were estimated using PROC NLIN as described above. These results are summarized in Table 4. The estimate of the among-increments component  $\sigma_x^2$  was larger for machine mixing than for hand mixing for both compounds. The corresponding value obtained from random sampling was smaller in both cases (1.75077 and 0.05167 for aldicarb and metolachlor, respectively). The within-increments component similarly was larger for machine mixing in both cases. There were no obvious differences between the other variance components.

Table 4. Estimates of variance components using PROC NLIN to equate mean squares with expected values.

|                    | Aldicarb |         | Metolachlor |         |
|--------------------|----------|---------|-------------|---------|
|                    | Hand     | Machine | Hand        | Machine |
| $\hat{\sigma}_x^2$ | 2.40815  | 3.08282 | 0.06523     | 0.08142 |
| $\hat{\sigma}_a^2$ | 0.00317  | 0.00353 | 0.00063     | 0.00023 |
| $\hat{\sigma}_w^2$ | 0.00000  | 0.33692 | 0.01298     | 0.02946 |
| $\hat{\sigma}_e^2$ | 0.00000  | 0.00000 | 0.00000     | 0.00000 |

## UGA Study

The UGA study involved a relatively small area (30 x 30 m), but the same number of soil samples as in the USEPA study were taken for compositing. The area was divided into 100 3-m by 3-m plots and random numbers were generated for determining the location of the soil samples to be collected. Trifluralin, a herbicide, is commonly used to control annual grasses in row crops such as cotton, soybeans, and peanuts. It was applied using a backpack sprayer at a rate of 1.12 kg/ha over the area and incorporated with a roto tiller to a depth of approximately 11.4 cm. Individual soil samples (ca. 100 g) were collected from the top 11.4 cm with a core sampler 24 hours after application, composited in the field, and stored at 0°C until laboratory analysis. Subsamples of approximately 20 g each were taken from the composite for analysis. The herbicide concentrations were measured in parts per billion and, for statistical analysis, a  $\log_{10}$  normalizing transformation was applied.

The same design as given in Table 1 for the USEPA study was used in the UGA study. Because the subsample size (20 g) differed from the increment size (100 g), the total number of subsamples that could be formed from a single composite was five times the number of increments. Consequently, the NLIN procedure as given in Figure 3 was modified to reflect the value of  $S = 5n$  (i.e., SCAP =  $5^*N;$ ). The first mean square in the data step of the SAS listing corresponds to  $MS_B$ , and the remaining four mean squares are  $MS_W$  values corresponding to Table 1.

Results of the analysis of variance of composite samples where  $r=15$ ,  $n=20$ , and  $s=3$  were:

$$\bar{z} = 2.75590 \quad \hat{\sigma}_x^2 = 0.25019$$

$$\hat{V}(\bar{z}) = 0.00157 \quad \sqrt{\hat{V}(\bar{z})} = 0.03958$$

The estimate for  $\sigma_x^2$  given above is an approximation based on the traditional approach of using the difference between the mean square between groups and the mean square within groups ( $(n/st)(MS_B - MS_W)$ ).

From 49 random samples, the mean and variance were obtained as:

$$\bar{x} = 2.39497 \quad \hat{\sigma}_x^2 = 0.22391$$

On the basis of these values, the standard error projected for a sample size equivalent to the 45 laboratory analyses is 0.07054, which is approximately twice that observed for composite sampling.

The variance components estimated using PROC NLIN were:

$$\hat{\sigma}_x^2 = 0.25345 \quad \hat{\sigma}_w^2 = 0.00000$$

$$\hat{\sigma}_a^2 = 0.00292 \quad \hat{\sigma}_e^2 = 0.01696$$

The estimates of the variance component among increments ( $\sigma_x^2$ ) compared very favorably in all cases. This is probably due to the fact that the study area was very homogeneous in terms of soil properties.

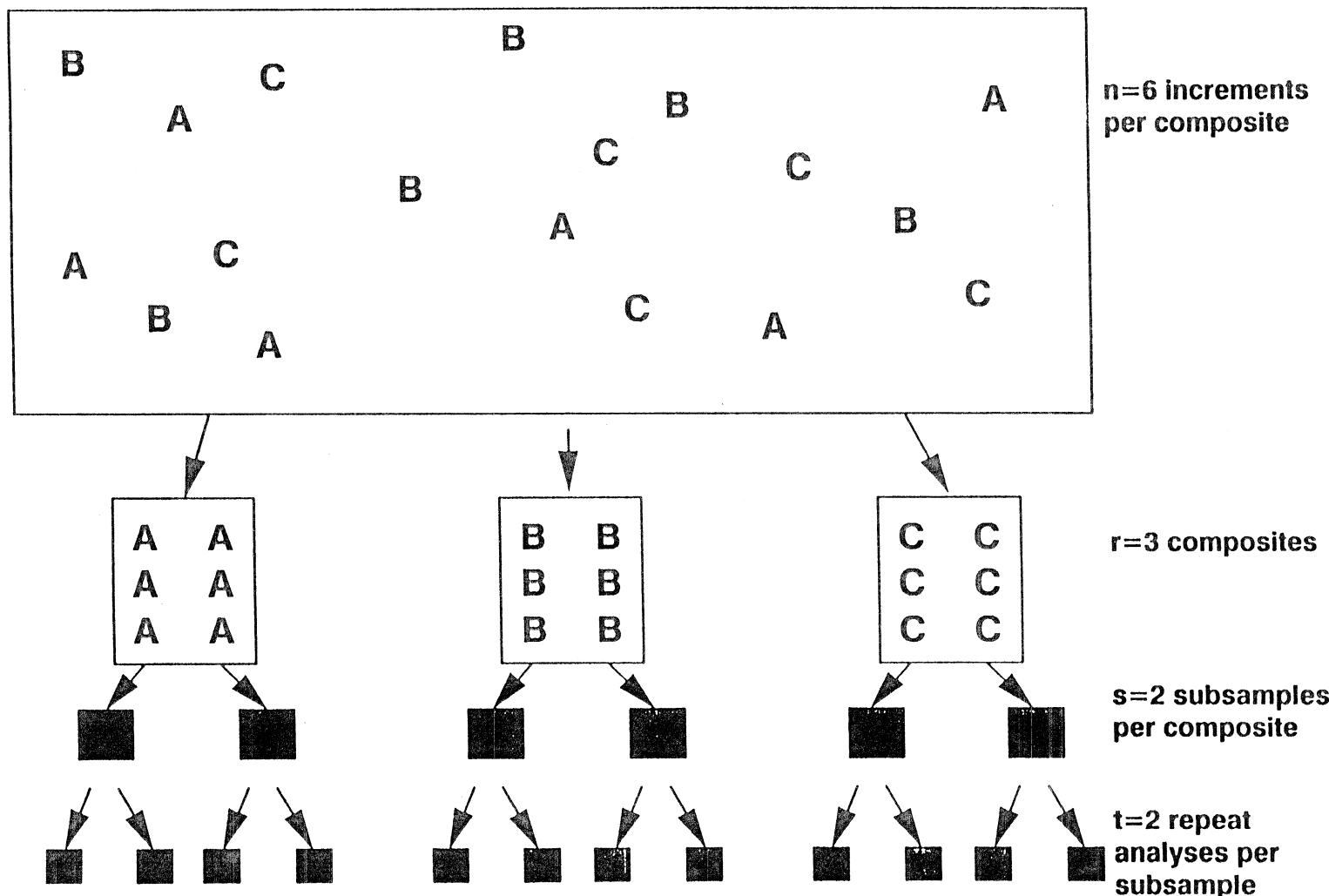
## 6. SUMMARY

Composite sampling techniques are being used very effectively in situations that involve mainly monitoring efforts where only means are of interest. In cases where standard errors are needed, several composite samples can be collected. In the research environment, however, additional considerations must be given to the importance of point-to-point variability. By use of appropriate design strategies for composite sampling, the associated variance components can be estimated using nonlinear least squares methods. The current efforts have shown that composite sampling significantly reduces the standard error of the mean, that the variance components can be affected by the compositing procedure, and that elements of bias can be introduced by the compositing process. In addition, composite sampling can be a very labor intensive activity.

When implemented appropriately, substantial gains in precision and savings in analytical efforts can be realized. The added complexity of implementation and the increased level of field-based labor, however, are disadvantages that often may negate other advantages of composite sampling. All of these factors should be considered carefully when choosing between composite and random sampling approaches. It is recommended that composite sampling procedures be used only in cases where it has been established that bias is not being introduced and variance components are not being adversely affected for the specific methods employed.

## 7. REFERENCES

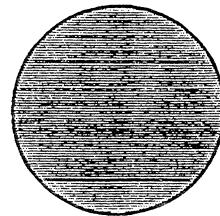
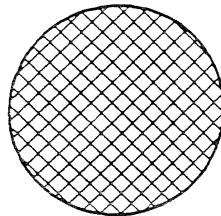
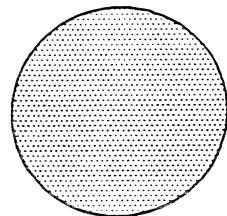
- Brown, G. H. and N. I. Fisher. 1972. Subsampling a mixture of sampled material. *Technometrics* 14(3): 663-668.
- Elder, Robert S., William O. Thompson, and Raymond H. Myers. 1980. Properties of composite sampling procedures. *Technometrics* 22(2): 179-186.
- Rhode, Charles A. 1976. Composite sampling. *Biometrics* 32, 273-282.
- SAS Institute, Inc. 1985. *SAS User's Guide: Statistics, Version 5 Edition*. SAS Institute Inc., Cary, NC.



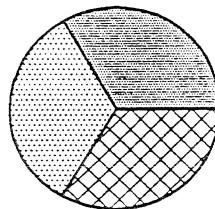
**Figure 1. Formation of Composite Samples**

Figure 2. Composite Sampling (Random Proportions)

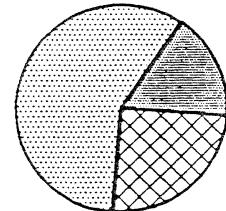
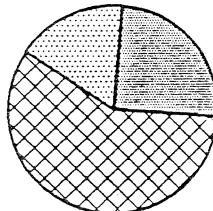
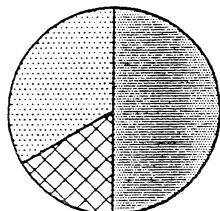
Increments (n)



Composite Sample



Subsamples (s) from same composite



$\sigma_x^2$  = Variability among increments

$\sigma_a^2$  = Variability among volume proportions for l-th increment

$\sigma_w^2$  = Variability among concentrations due to l-th increment

$\sigma_e^2$  = Variability associated with measurement error

**Figure 3. PROC NLIN code for estimating variance components.**

```

DATA;
  INPUT MS DF S N;
  WT = DF*N;
  CARDS;
0.07050606 14 3 20          /* UGA Study */
0.03297726 30 0 20
0.02052736 9 0 10
0.03610572 14 0 20
0.03584159 14 0 30

PROC NLIN METHOD=MARQUARDT;
  SCAP = N;
  SS = (SCAP-S)/(SCAP-1);
  MODEL MS = (S/N)*SIGx + SS*(SIGw/N + N*SIGa*(SIGw + SIGx)) + SIGe;
  DER.SIGe = 1;
  DER.SIGx = (S/N) + SS*N*SIGa;
  DER.SIGw = SS*(1/N + N*SIGa);
  DER.SIGa = SS*N*(SIGw + SIGx);
  PARMS SIGx = 0.2, SIGa = 0.01, SIGw = 0.011, SIGe = 0.01;
  BOUNDS SIGx>0, SIGa>0, SIGw>0, SIGe>0;
  WEIGHT = WT;
  OUTPUT P=MSHAT R=MSRES;

```

**Figure 4. Data Sets for EPA Study.**

| <u>Aldicarb - Hand</u>    |    |   |    | <u>Aldicarb - Machine</u>    |    |   |    |
|---------------------------|----|---|----|------------------------------|----|---|----|
| 0.48307                   | 14 | 3 | 20 | 0.83747                      | 14 | 3 | 20 |
| 0.10739                   | 29 | 0 | 20 | 0.16337                      | 30 | 0 | 20 |
| 0.07596                   | 9  | 0 | 10 | 0.17321                      | 9  | 0 | 10 |
| 0.18180                   | 13 | 0 | 20 | 0.19045                      | 14 | 0 | 20 |
| 0.27325                   | 14 | 0 | 30 | 0.51773                      | 12 | 0 | 30 |
| <u>Metolachlor - Hand</u> |    |   |    | <u>Metolachlor - Machine</u> |    |   |    |
| 0.01124194                | 14 | 3 | 20 | 0.01398475                   | 14 | 3 | 20 |
| 0.00143009                | 29 | 0 | 20 | 0.00213379                   | 30 | 0 | 20 |
| 0.0024059                 | 9  | 0 | 10 | 0.004251                     | 8  | 0 | 10 |
| 0.0018507                 | 14 | 0 | 20 | 0.001306                     | 14 | 0 | 20 |
| 0.0021427                 | 14 | 0 | 30 | 0.002079                     | 13 | 0 | 30 |