2005-05-20

Determining the Optimum Number of Increments in Composite Sampling

John Ellis Hathaway

Brigham Young University - Provo

www.manaraa.com

DETERMINING THE OPTIMUM NUMBER OF INCREMENTS IN COMPOSITE

SAMPLING

by

John E Hathaway II

A Project submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Statistics

Brigham Young University

April 2005

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a Project submitted by

John E Hathaway II

This Project has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

_____	_____
Date	John S. Lawson, Chair
_____	_____
Date	Richard O. Gilbert
_____	_____
Date	G. Bruce Schaalje
_____	_____
Date	David G. Whiting

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the Project of John E Hathaway II in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

_____

Date

_____

John S. Lawson
Chair, Graduate Committee

Accepted for the Department

_____

G. Bruce Schaalje
Graduate Coordinator

Accepted for the College

_____

G. Rex Bryce
Associate Dean, College of Physical and Mathematical Sciences

ABSTRACT

DETERMINING THE OPTIMUM NUMBER OF INCREMENTS IN COMPOSITE

SAMPLING

John E Hathaway II

Department of Statistics

Master of Science

Composite sampling can be more cost effective than simple random sampling. This paper considers how to determine the optimum number of increments to use in composite sampling. Composite sampling terminology and theory are outlined and a model is developed which accounts for different sources of variation in compositing and data analysis. This model is used to define and understand the process of determining the optimum number of increments that should be used in forming a composite. The blending variance is shown to have a smaller range of possible values than previously reported when estimating the number of increments in a composite sample. Accounting for differing levels of the blending variance significantly affects the estimated number of increments.

Acknowledgements

Contents

Chapter

Tables

Table

Figure

xvii

Chapter 1

Introduction

An important issue that arises with data collection is the cost. Environmental and medical data can be very costly and difficult to gather, which motivates the investigation of more cost-effective ways to gather such data. One of these methods is called composite sampling (Johnson and Patil 2001).

The composite sampling process begins by developing the sampling design that will be used to select the increments, the individual samples that will be mixed, to form the composite sample. Next, the increments are gathered and blended into a composite sample that is often treated as a perfect mixture of the increments. Then one or more sub-samples are removed from the composite and sent to the laboratory for analysis. These sub-samples can be further divided and each resulting laboratory sample analyzed (Figure 1.1).

A variety of costs must be accounted for in composite sampling. The researcher must account for the cost to gather each increment, to mix the specified increments into the composites, to analyze the contents of each sample, etc. Often, soil samples are cheap to gather but very expensive to analyze. This fact suggests that some type of compositing technique could reduce costs (Rohlf et al. 1996).
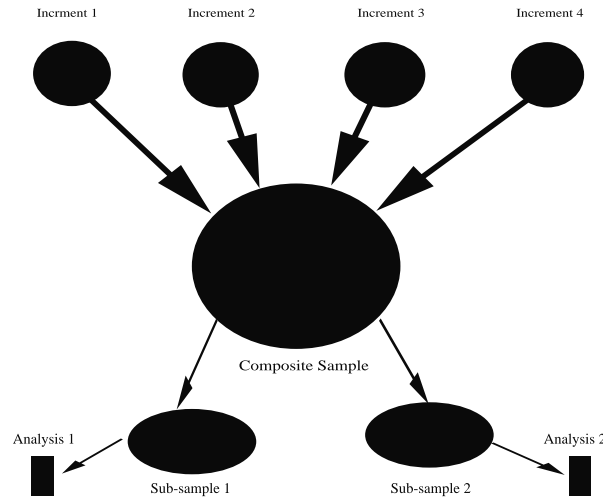
Figure 1.1: Basic Compositing Process

Composite sampling has been used in various fields of application (Boswell et al. 1996). These different applications lead to different purposes for and methods of compositing increments. The general motivation for composite sampling is to reduce the number of measurements that need to be made to estimate some specified parameters. Under this motivation, a plethora of parameters, processes, variations, and objectives reside. This project involves general theory that can be applied to most areas of composite sampling with only slight variations of the underlying model.

This project was funded by Battelle Memorial Institute, which is under contract to the U.S. Department of Energy (Contract DEAC06-76RL01830), to add features to Visual Sample Plan (VSP). VSP is a software program designed to help government agencies, private industry, and individuals develop a sampling procedure that is efficient and statistically sound. One of the options being developed in VSP is a composite sampling plan to assist in prevalence estimates.

Mixing variability must be accounted for in composite sampling. The primary goal of this project is to mathematically define an adequate method for calculating the number of increments to use in a composite sample when there is mixing variation. In the process of developing this methodology, I will propose more stringent bounds on the mixing variation than previously used. Also, I will show that mixing variation highly influences the calculated number of increments to be used in composite sampling.

Chapter 2 is a literature review of the development of the theory of composite sampling. The general field of composite sampling is well represented in statistical, environmental, and other journals (Boswell et al. 1996). With a few exceptions, only the articles that assist in the development of the project required for the VSP program are reviewed. Chapter 3 describes in more detail the features Battelle desired to include in VSP and some proposed ways to account for mixing variation. Finally, chapter 4 is a paper that will be submitted to the *Journal of Environmental & Ecological Statistics*.

Chapter 2

Literature Review

2.1 Introduction

This chapter reviews some of the terminology and concepts that are used in composite sampling. Explanations of some of the methods used to apply composite sampling to various fields are given.

2.2 Terminology

Composite sampling suffers from a lack of common terminology. A researcher who is reading articles on composite sampling will encounter new terminology for similar or even identical aspects of composite sampling in each paper. For example, what Duncan (1962) calls increments have been called samples (Rohde 1976), core samples (Izenman 2001), primary units (Rohlf et al. 1996), and aliquots (Johnson and Patil 2001). The symbols and subscripts that are used to represent the different elements of the calculations appear in an even larger variety. At times, each author's equation can be fundamentally the same, but the definitions of the terms and variables are so different that the equations from each seem starkly different.

For this project, Duncan's (1962) "standardized" terminology will be used.

Lot: The population that is under investigation. A lot could be a truckload of coal, a section of land, or a quantity of blood.

Segment: A portion of the lot to which inference will be made.

Increment: Randomly selected samples that represent the segment and are used to form the composite.

Blending: The mixing or agglomerating of increments to form the composite sample.

Composite sample: A sample formed by blending a certain number of increments from specified segments of the lot.

Sub-sample: A portion of the composite sample that is sent to the laboratory.

Laboratory Sample: A portion of the sub-sample that is measured.

2.3 Fields of Use

Composite sampling can be used in many fields; however, it is often used in areas where the cost of analyzing data is high and gathering increments is low. Composite sampling has one feature that adds complexity in creating a general sampling strategy, namely the mixing process. The physical mixing procedure is different for each area of research. This project uses environmental sampling as an example; however, composite sampling is used in many other fields. For example, the forensic study of illicit drugs has increased the use of composite sampling techniques. Composite sampling is used to test large quantities of illicit drugs at once to estimate the amount of a substance that is possessed by accused

criminals (Izenman 2001). In medical fields there is a great deal of research on composite sampling methods to estimate the prevalence of diseases like HIV (Tu et al. 1994) and Hepatitis (Davis et al. 1973).

Often the goal of these studies is to estimate the overall proportion of infected subjects for a certain disease. This type of estimation requires a detailed sampling plan that takes into account the costs of different steps and the different sizes of composites that will be formed if tests are positive. For example, researchers will form a composite of many individuals' blood samples to check for overall prevalence of HIV. If the test returns positive, a second test is done on a smaller composite sample of the samples used in the previous composite. This procedure is repeated by looking at successively smaller composites of the positive results. A well-defined sampling plan is required to know how much blood is needed from each subject in order to complete the expected divisions (Lancaster and Keller-McNulty 1998).

Medical fields do not have as much difficulty compositing as the environmental fields because they are often analyzing blood. There is a limited number of ways to combine the blood samples into a composite sample. There are some difficulties in how the samples react to dilution; however, compositing samples has been shown to be very effective (Tebbs and Swallow 2003). Composite sampling is effective in medical fields because it is cost-effective when the prevalence rates are low (Stephens et al. 2000).

Environmental researchers also use composite sampling as an efficient cost savings tool. With environmental data a variety of different parameters are of interest. These include the overall mean level of the entire site as well as locations of 'hot spots' (Boswell

and Patil 1987). Much of the composite sampling research for environmental data involves the estimation of contamination levels of soils, air or water (Boswell et al. 1995, 1996). The identification of 'hot spots' from a large sampling area requires complex sampling patterns that mix different combinations of the increments into composite samples (Exner et al. 1985; Gore et al. 2001).

The greatest problems associated with composite sampling in environmental projects arise because of the variety of elements to be mixed. Researchers are faced with the task of mixing many different types of soils in each site. In a quarter acre area, the land could have five or six distinct soil types that require different mixing methods. Research on composite sampling in environmental fields must detail how the samples are prepared and composited because the process of mixing can be controversial (Gilbert and Doctor 1985; Pitard 1993).

2.4 Estimating Variance

The goal of this project is to calculate the appropriate number of increments to include in each composite for a predetermined level of variance. Some researchers have mentioned this issue in their papers, but left the subject with little development (Elder et al. 1980). In Chapter 3, technical articles which define how the different variance components contribute to the overall variance of the composites will be reviewed in more detail. Brown and Fisher (1972) wrote an early paper on the overall variance. This paper was followed by a more complete mathematical evaluation of composite sampling (Rohde 1976).

Elder et al. (1980) outlined in more detail how to estimate the overall variance when accounting for mixing variability; however, they stated that the issue of estimating sample

size while accounting for mixing variability is not very feasible. This lack of feasibility was based on mathematical results of the variance equation they derived and the bounds of the mixing variation. If the $1/\sigma_b$, where σ_b is the deviation of the blending procedure as described in Figure 3.2, is less than the number of increments included then the estimate of the variance will actually increase when more increments are included (Figure 2.1). Chapter 3 discusses in detail how to account for the mixing variation when estimating the overall variation and why it is important.

Relationship of $\sigma_{\bar{x}}^2$ to the number of increments used
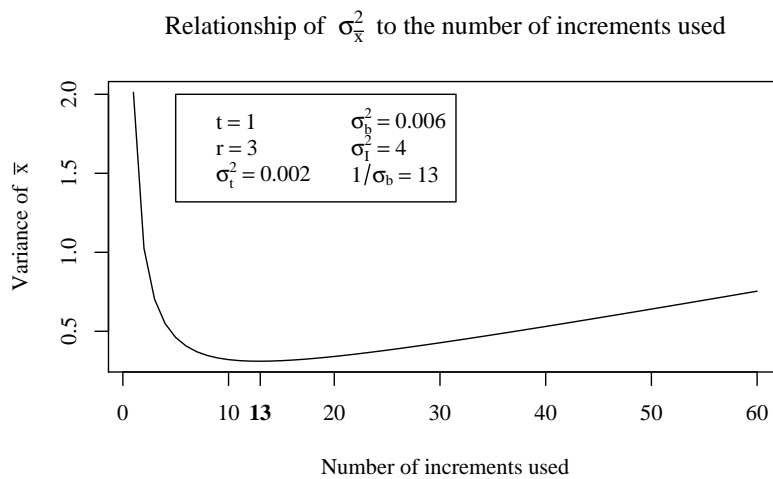
Figure 2.1: Plot of Variance Estimate based on n

Two other papers that have been published in the past ten years attempt to estimate the overall variability, but they do not specifically apply to the scenario of interest in this project. Rohlf, Akcakaya, and Ferraro (1996) derived a basic variance equation that they used along with cost calculations to define the cheapest strategy for a composite sampling project, but they did not include blending variation. Izenman (2001) detailed a specific mathematical strategy for estimating proportions of illicit drug substances for legal issues.

Their paper accounts for mixing variation, but does not deal with a continuous segment to sample from.

Chapter 3

Methods

3.1 Introduction

This project was initiated by Brent Pulsipher of the Pacific Northwest National Laboratory (PNNL). A client contracted to have PNNL design an adequate sampling plan to establish a contamination boundary of a defined site. This project was designed as an enhancement to Visual Sample Plan (VSP) — a software program that facilitates the sampling process based on the Data Quality Objectives (DQO) process developed by the United States Environmental Protection Agency (EPA).

3.2 Visual Sample Plan

The purpose of VSP is to provide simple, defensible tools for defining an optimal sampling scheme for characterizing environmental contamination. VSP is applicable to any two-dimensional sampling plan, including surface soil, building surfaces, water bodies, or other similar applications. VSP is tailored to the environmental professional who values cost effectiveness, simplicity, accuracy, and defensible methods. This professional wants to solve real-world environmental contamination problems using state-of-the-art statistical methods,

www.manaraa.com

but does not have time to master new complex software tools. It is no simple matter to collect and analyze environmental data to reach conclusions that are statistically defensible while minimizing costs. VSP can help scientists and engineers solve these problems (Hassig et al. 2004).

Figure 3.1 shows a screen shot from VSP with a generic map and some of the command options that are available. This specific view has the box on the bottom right that deals with the objective of this project. In that box it details that twenty five increments should be used for the 'multiple-increment sample' or composite sample.

3.3   Proposed Project

This project was designed to fit under the goals of VSP. Specifically, the Navy contracted PNNL to develop the process of using composite samples and sampling segments on the perimeter (boundary) of an enclosed area. The Navy wants to determine the boundary of a specified contaminants spread. If the upper confidence limit (UCL) on the mean of a segment is above the concentration threshold, the boundary is extended outward and new increments are taken on the extended perimeter. This process is repeated until the UCL for all segments is below the defined threshold level. Figure 3.2 shows an image of the sampling process with the components necessary in the estimation process.

When PNNL designed this part of VSP, they did not adequately define how many increments should be included in each of the composites. This was due to the combination of lack of research in this area and the Navy's desire to have an option hurriedly developed in VSP so they could use it in their sampling projects. The research in this project is the first
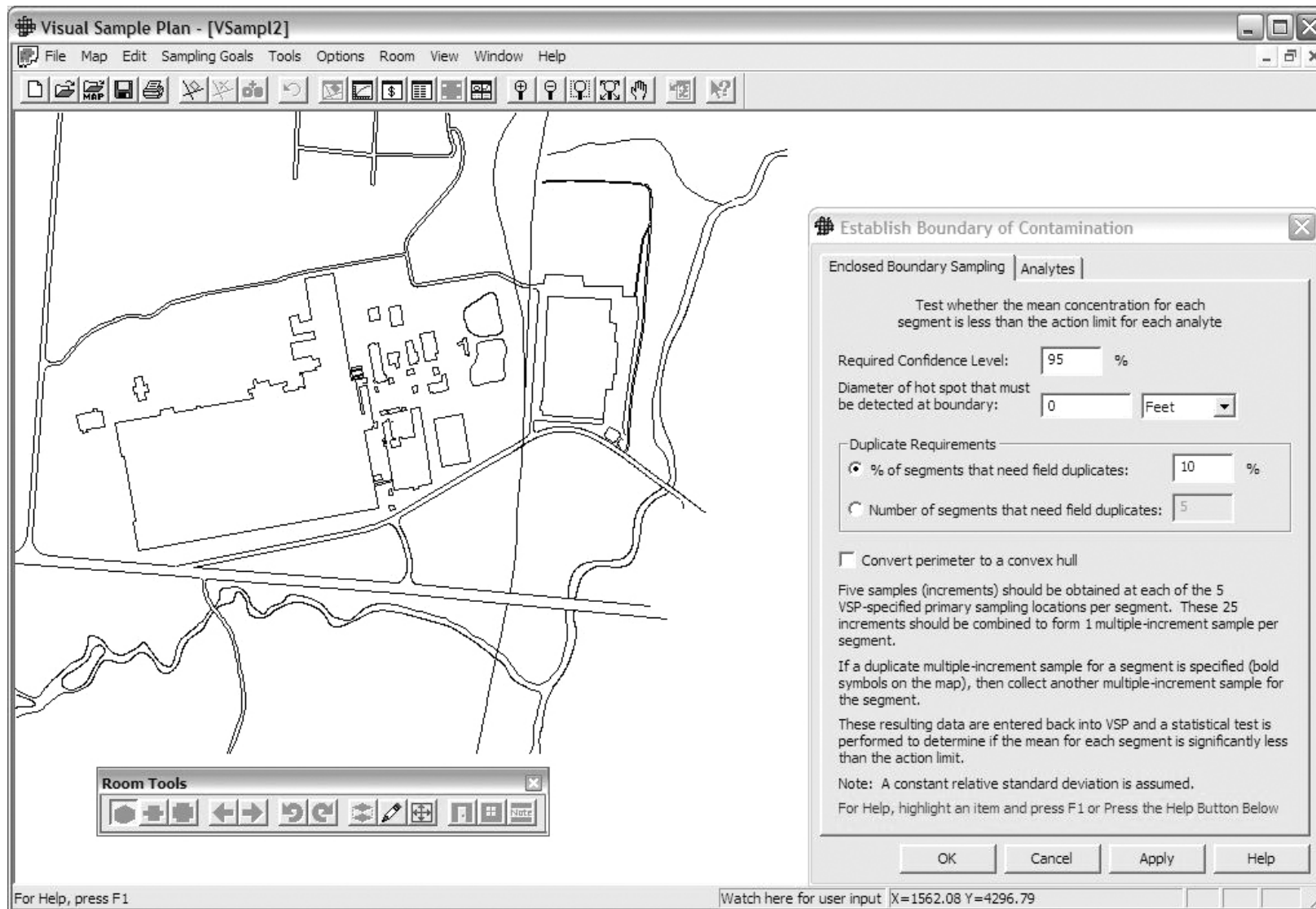
Figure 3.1: Screen Shot from Visual Sample Plan (A windows program for sampling)

step to a more defensible calculation of the number of increments in a composite sample.

3.4   Variance Estimation Equations

Researchers have accounted for different variabilities in estimates of the overall variability of the mean with composite sampling. Some of the early research accounted for between-increment variability and composite-to-composite variability (Elder et al. 1980; Izenman 2001; Rohde 1976; Duncan 1962; Brown and Fisher 1972). Three of the key papers are presented below (Elder et al. 1980; Izenman 2001; Rohde 1976). An in-depth discussion of the variance equation's characteristics as a whole will be further developed in Chapter 4.

3.4.1   Properties of Composite Sampling Procedures

Elder et al. (1980) considered a situation in which each of r composite samples is formed by taking n increments from a segment. Each composite sample consists of S subsamples, but only s of the subsamples are selected. Each of the selected subsamples is measured t times.

They denote the amount of material in subsample j, $j = 1, \ldots, s$ from composite i, $i = 1, \ldots, r$ that was derived from increment l, $l = 1, \ldots, n$ as g_{ijl}. Then the total amount of material in subsample j from composite i is $G_{ijl} = \sum_{l=1}^{n} g_{ijl}$. The proportion of material in subsample j from composite i that was derived from increment l is

$$b_{ijl} = \frac{g_{ijl}}{G_{ij}}. \tag{3.1}$$

Note that $\sum_{i=1}^{n} b_{ijl} = 1$. They assume that the a_{ijl}, $l = 1, \ldots, n$ are random variables with mean $1/n$ and common variance σ_b^2. Because of the constraint, however, they are

13

not independent. Also, the b's in different composites are uncorrelated and the covariance structure for the b's in different subsamples from the same composite is assumed to be the same.

They denote the value of the material in subsample j from composite i that was derived from increment l as z_{ijl}. They assume that the $z_{ijl}, l = 1, \ldots, n$ are random variables and that the $\text{Cov}(b_{ijl}, z_{ijl})$ is zero. When using composite sampling we don't observe each z_{ijl}. We attempt to observe the value associated with the jth subsample from the ith composite which is represented as a linear combination of the b_{ijl} and z_{ijl} as a weighted average

$$y_{ij} = \sum_{l=1}^{n} b_{ijl} z_{ijl}. \tag{3.2}$$

The problem is that y_{ij} is not observed; instead we observe x_{ijk} where

$$x_{ijk} = y_{ij} + t_{ijk} \tag{3.3}$$

where t_{ijk} is the error of measurement for the kth test on the jth subsample from the ith composite. The expected value of t_{ijk} is zero and the variance is σ_t^2.

To estimate the overall mean of the composite samples from a segment the average is taken.

$$\overline{x} = \frac{1}{rst} \sum_{ijk} \left[\sum_{l} (b_{ijl} z_{ijl}) + t_{ijk} \right] \tag{3.4}$$

Elder et al. (1980) show that $\mathbf{E}(\overline{x}) = \mu_x$ where μ_x is the true average level of prevalence. They derive the variance of $\overline{x}$.

$$\text{Var}(\overline{x}) = \frac{\sigma_I^2}{rn} + \frac{1}{rs} \left(\frac{S-s}{S-1} \right) \left[\frac{\sigma_w^2}{n} + n\sigma_b^2(\sigma_I^2 + \sigma_w^2) \right] + \frac{\sigma_t^2}{rst} \tag{3.5}$$

where σ_I^2 is the between-increment variation and σ_w^2 is the within-increment variation.

Often σ_w^2 and σ_b^2 are assumed to be zero. This is most often done because of the difficulty and expense of actually estimating these variation effects. For this project the within-increment variation is assumed to be zero to simplify the development of sample size estimates. When σ_b^2 is assumed to be zero, the calculations reduce to the basic equation for the variance of the sampling distribution of the mean in multi-stage sampling.

If it is assumed that $\sigma_w^2 = 0$ then the Equation (3.5) reduces to

$$\text{Var}(\overline{x}) = \frac{\sigma_I^2}{rn} + \frac{1}{rs} \left(\frac{S-s}{S-1} \right) n\sigma_b^2 \sigma_I^2 + \frac{\sigma_t^2}{rst}. \tag{3.6}$$

If it is further assumed that $\sigma_b^2 = 0$ then Equation (3.6) reduces to

$$Var(\overline{x}) = \frac{\sigma_I^2}{rn} + \frac{\sigma_t^2}{rst}. \tag{3.7}$$

An important implication of the presence of σ_b^2 in the formula for $Var(\overline{x})$ is that there is a relative upper bound on $Var(\overline{x})$ (Elder et al. 1980, lemma 2). This upper bound arises when the composite sample is not mixed and the variation of the "blending" is essentially the between-increment variation. Thus,

$$\frac{\sigma_I^2}{rn} + \frac{\sigma_t^2}{rst} \leq Var(\overline{x}) \leq \frac{\sigma_I^2}{rn} + \frac{1}{rs} \left(\frac{S-s}{S-1} \right) \left[\sigma_w^2 + \left(\frac{n-1}{n} \right) \sigma_I^2 \right] + \frac{\sigma_t^2}{rst} \tag{3.8}$$

3.4.2   Composite Sampling

Rohde (1976) described the variance of the composites in Equation (3.9). Rohde (1976) assumed that σ_t^2 and σ_w^2 as defined above were zero. He also assumed that the number of analytical tests and composites were both 1.

$$Var(\overline{x}) = \frac{\sigma_I^2}{n} + \frac{n\sigma_b^2 \sigma_I^2}{s} \tag{3.9}$$

15

3.4.3 Statistical and Legal Aspects of the Forensic Study of Illicit Drugs

Izenman (2001) developed a similar equation for estimating the variance based on his scenario for estimating purity of segmented containers. He developed his method to identify the purity of the drugs captured for criminal cases. He actually derived his formula from the two previous articles mentioned above. His equation is almost identical to those in the other two articles; however, he added elements unique to his problem. For example, he accounted for the number of segmented containers sampled.

3.5 Summary and Proposal

As stated in Chapter 2, based on the current research there are some theoretical problems in determining the optimal number of increments to use in each composite to obtain a desired level of composite to composite variation (Elder et al. 1980). The original goal was to do a literature review and find journal articles that developed the methods appropriate for use with VSP. The literature reviewed did not contain adequate theory for implementation with VSP. Developers of The VSP software wanted to account for multiple variances in the sampling design. The main source of variability that we wanted to add into the model was mixing variability. While some authors had discussed this mixing variation, there was no development of how to calculate the number of increments to use in each composite sample.

The main goal of this research is the development of a procedure for calculating the optimal number of increments to include in each composite when there is variation in the mixing procedure for use in the VSP software program. Although Elder et al. (1980)

showed how to estimate the overall variance when accounting for mixing variation, they did not propose a solution. They explained how increasing the number of increments can actually increase the composite sample variance instead of decreasing the variation. This is due to the wide bounds that they used for the mixing variation.

Other goals of this project include:

- Use the increment quantity estimation equation to investigate bounds on the range of values for σ_b^2. It turns out that these bounds are more stringent than previously defined by Elder et al. (1980).

- Show how even "small" amounts of mixing variation affect the estimate of the number of increments to use in a composite.

- Show how to estimate the different variance components with actual data from a previous study (Jenkins et al. 1996) of composite sampling, and use these estimates in an example to estimate the number of increments to use in a composite sample.

Segments

n = Number of increments in each composite

σ_I^2 = Between increment Var.

r = Number of composites

$\sigma_{\bar{x}}^2$ = Observed var. of composite mean

s = Number of subsamples taken from composite

rs = Total possible subsamples

t = Number of measurements done per subsample

σ_t^2 = Analytical error

b_{ijl} = The proportion of the ith increment from the jth composite in the lth subsample.

σ_b^2 = The mixing variance. This is bounded.
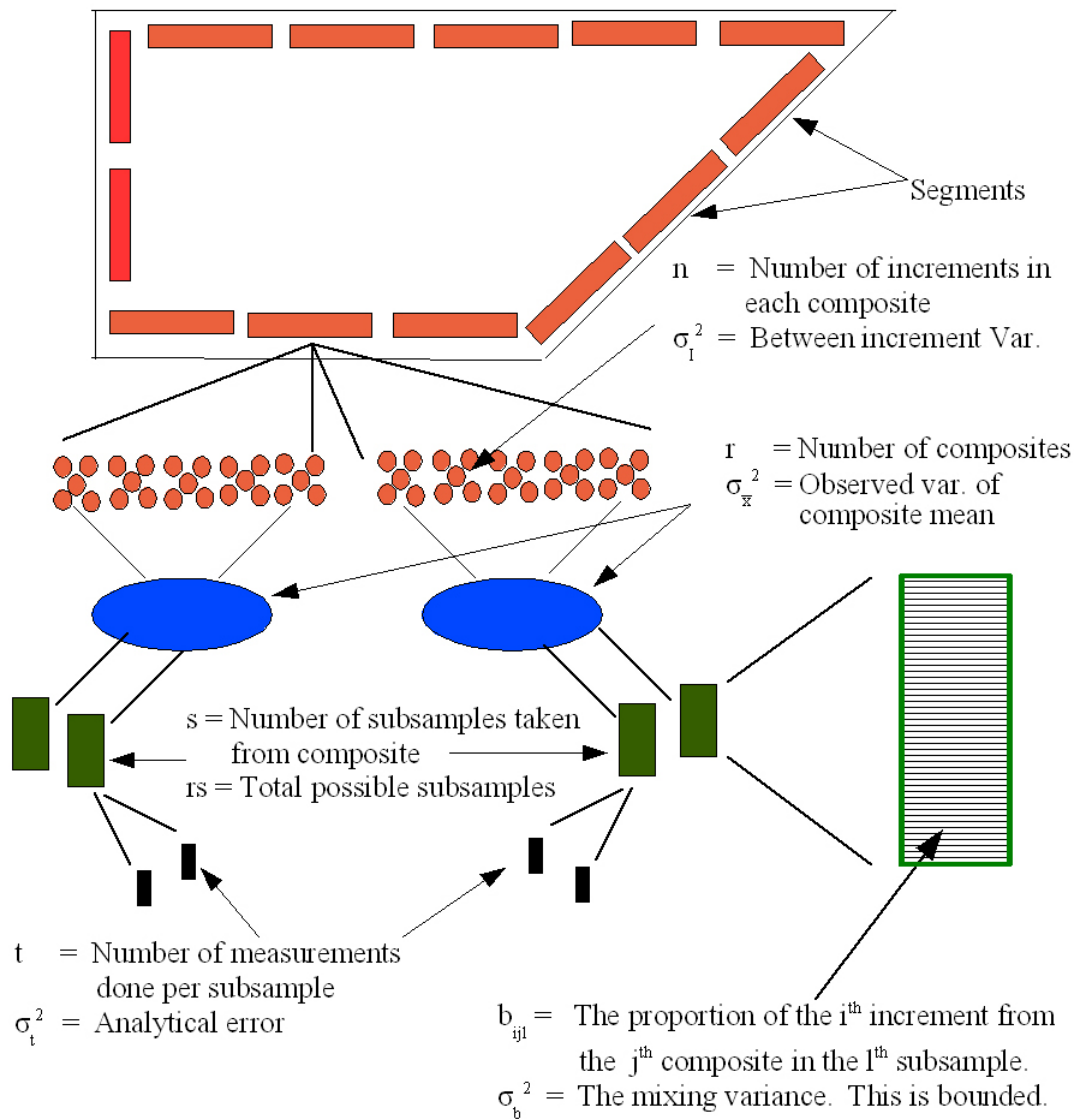
Figure 3.2: The Compositing Process for Perimeter Sampling

Chapter 4

Article

4.1   Introduction

Analyzing for contaminants in soil, sediments, animals, air or water often involve expensive measurement procedures. This expense motivates the use of composite sampling, which is a physical averaging of the samples. If the expense of gathering samples is comparatively low, then the researcher can gather all the needed samples, blend them together, and measure one subsample to get a physical average of all the samples which could be the same as measuring each sample and taking their mathematical average.

Unfortunately, compositing does not always result in a perfectly blended physical sample. Variation in the blending of the composite sample affects the overall variation in unique ways. This uniqueness results from the relationship between sample size and the blending variation; therefore, to estimate the number of samples to use in a composite sampling procedure for a desired level of overall variation requires a different procedure than standard sample size estimation problems.

4.1.1 Definition of Terms

One of the challenges of understanding composite sampling is the terminology that varies across fields of application and even within a specific field. For this article, Duncan's (1962) "standardized" terminology will be used.

Lot: The population that is under investigation. A lot could be a truckload of coal, a section of land, or a quantity of blood.

Segment: A portion of the lot to which inference will be made.

Increment: Randomly selected samples that represent the segment and are used to form the composite.

Blending: The mixing or agglomerating of increments from a segment.

Composite sample: A sample formed by blending a certain number of increments from specified segments of the lot.

Sub-sample: A randomly selected portion of the composite sample that is sent to the laboratory.

Laboratory Sample: A randomly selected portion of the sub-sample that is measured.

4.1.2 Description of Composite Sampling Process

Another difficulty of understanding composite sampling is that compositing is used in a variety of fields of application and materials. This variety leads to different goals and implementations of composite sampling. The general description of composite sampling is the process of combining increments in order to reduce the number of measurements needed

20

for estimating specified parameters (Behets et al. 1990; Izenman 2001; Gilbert and Doctor 1985). Under this overarching concept, a plethora of differing parameters, processes, and objectives reside. The scope of this paper is limited to theory that can be applied to most types of composite sampling with some slight variations of the underlying model.

In general, the compositing process begins by developing the sampling design for selecting the increments to form the composite sample. Next, the increments are gathered and blended into a composite sample. Note that the goal is a perfect blending of the increments, but in practice the mixing may not be perfect. Then one or more sub-samples are removed from the composite and sent to the laboratory for analysis. This sub-sample can be further sampled and each resulting laboratory sample used for chemical analysis or other measurement.

Figure 1 shows the process of forming composite samples and measuring the variables of interest when the lot is the perimeter (boundary) of an enclosed area. This process is of interest when a researcher wants to determine the boundary of a specified contaminant. If the upper confidence limit (UCL) on the mean of a segment is above the concentration threshold, then the boundary is extended outward and new increments are taken on the extended perimeter. This process is repeated until the UCL for all segments is below the defined threshold levels.

4.2   Compositing Theory

Boswell, et al. (1996) compiled an annotated bibliography of the theory and application of composite sampling. Brown and Fisher (1972) developed models and derived

21

Segments

n = Number of increments in each composite

σ_I^2 = Between increment Var.

r = Number of composites

$\sigma_{\bar{x}}^2$ = Observed var. of composite mean

s = Number of subsamples taken from composite

rs = Total possible subsamples

t = Number of measurements done per subsample

σ_t^2 = Analytical error

b_{ijl} = The proportion of the ith increment from the jth composite in the lth subsample.

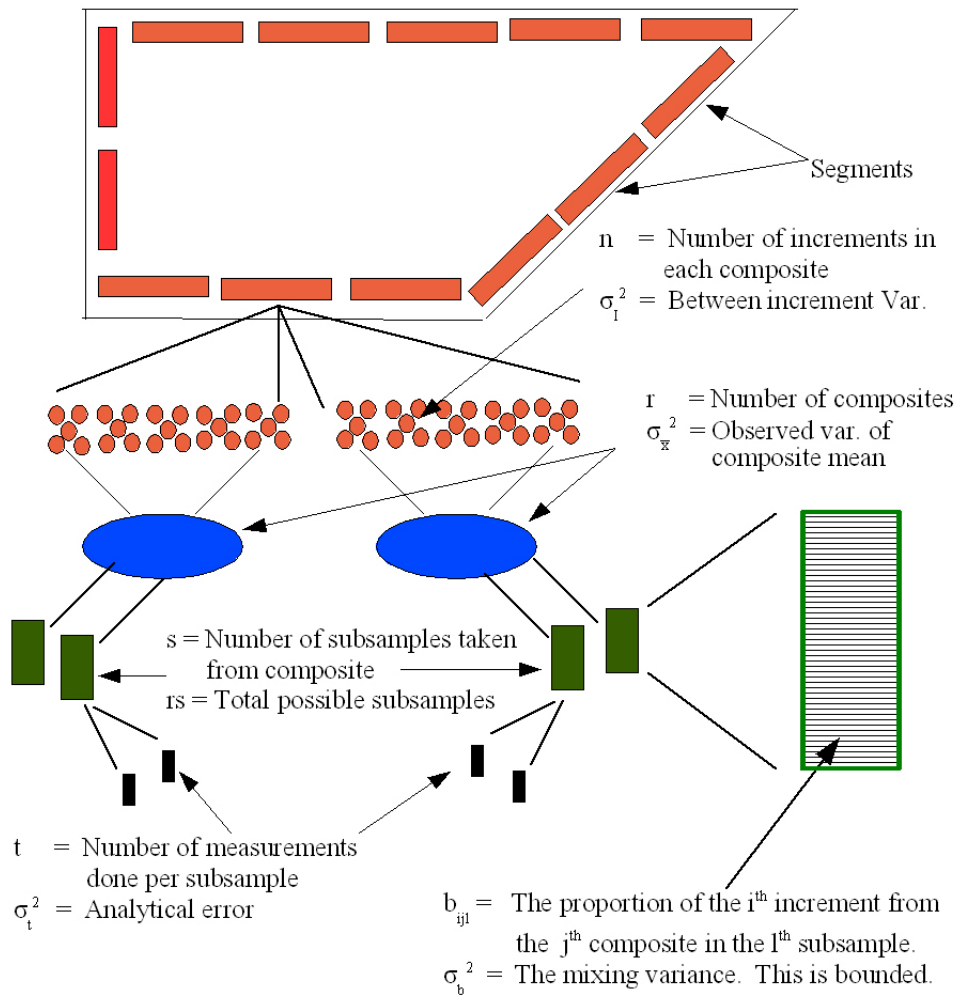σ_b^2 = The mixing variance. This is bounded.

Figure 4.1: Compositing Process

variance formulas for composite sampling of bales of wool. Rohde (1976) extended Brown and Fisher's (1972) work. Finally, Elder, et al. (1980) further extended the work of Rohde (1976) and Brown and Fisher (1972). The Elder et al. (1980) result is used here to develop a method for computing the number of increment samples per composite. They define $\overline{x}$ as the average of the observations from a segment. They then derived the following variance for $\overline{x}$.

$$\sigma_{\overline{x}}^2 = \frac{\sigma_I^2}{rn} + \frac{n\sigma_b^2\sigma_I^2}{rs} + \frac{\sigma_t^2}{rst} \tag{4.1}$$

where $\sigma_{\overline{x}}^2$ is the variance of the composite mean that will be used to compute the confidence interval for the mean,

σ_I^2 is the between-increment variance, which will be greater than or equal to the between composite variance,

σ_b^2 is the variance that results from blending the increments, to be defined in more detail in Section 4.3.1.

σ_t^2 is the between-measurement variance .

n is the number of increments that are taken from a given segment to be blended into a composite sample.

r is the number of replicate composite samples that are obtained in the same segment.

s is the number of sub-samples that are removed from the composite to be analyzed by the laboratory.

Equation (4.1) assumes that the increments, sub-samples, laboratory samples, and

23
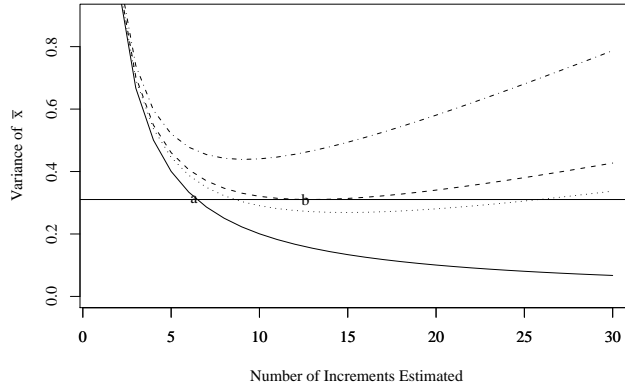
Figure 4.2: Relationship between n, σ_b^2, and $\sigma_{\bar{x}}^2$ when σ_I^2, σ_t^2, r, s, and t are known or fixed. Each line represents a different value of σ_b^2.

composite samples are of equal volumes in their respective stages.

Equation (4.1) reduces to the following expression if $\sigma_b^2 = 0$:

$$\sigma_{\bar{x}}^2 = \frac{\sigma_I^2}{rn} + \frac{\sigma_t^2}{rst} \tag{4.2}$$

where rn equals the total number of increments per segment. This "ideal model" is often used to compute $\sigma_{\bar{x}}^2$ when it is assumed that the blending process is nearly perfect ($\sigma_b^2=0$). This assumption is the cause of debate on the appropriateness of composite sampling techniques (Patil 1995). Johnson et al. (2001) and Rohlf et al. (1996) used costs when determining the optimal numbers of composites, increments, and measurements under the assumption that $\sigma_b^2 = 0$.

4.2.1   Facts about n, σ_b^2, and $\sigma_{\bar{x}}^2$

Elder et al. (1980) mention an interesting relationship between n, σ_b^2, and $\sigma_{\bar{x}}^2$. With their proof they reason that σ_b^2 can be constant over a finite interval of values of n. Over this

24

finite n interval σ_b^2 is a variation that is controlled by the actual blending process. Section 4.3.1 details σ_b^2, its bounds, and how those bounds can affect σ_b^2 when n gets very large. When looking at σ_b^2 as a variation that can be controlled by the blending process alone, Equation (4.1) can be used to show that as n becomes larger than $\sqrt{s}/\sigma_b$ then the estimate for $\sigma_{\bar{x}}^2$ will actually increase with n instead of decreasing.

Figure 4.2 shows the estimates for $\sigma_{\bar{x}}^2$ as a function of differing n and σ_b^2 with $\sigma_I^2 = 4$, $\sigma_t^2 = 0.002$, $r = 2$, $s = 1$, and $t = 2$. The values of σ_I^2, σ_t^2, r, s, and t are arbitrarily chosen for this example and would be known or decided upon in an actual study. The solid line shows the estimates of $\sigma_{\bar{x}}^2$ for differing values of n when $\sigma_b^2 = 0$, the dashed line is for $\sigma_b^2 = 0.006$, the dotted line represents $\sigma_b^2 = 0.0045$, and the dotted-dashed line is for $\sigma_b^2 = 0.012$. The horizontal line represents a desired $\sigma_{\bar{x}}^2$ of 0.31035. The dashed lines in Figure 4.2 are examples of the convex relationship between n and $\sigma_{\bar{x}}^2$ that results from Equation (4.1), when σ_b^2 is considered to be constant in relation to n.

Point 'a' shows the point where an n of 6.3 will result in the desired $\sigma_{\bar{x}}^2$ for $\sigma_b^2 = 0$ and point 'b' gives an n of 12.6 where the line that results from $\sigma_b^2 = .006$ crosses the desired threshold. The dotted-dashed line that results from $\sigma_b^2 = 0.012$ shows that the mixing variation is too large to attain the desired $\sigma_{\bar{x}}^2$, while the dotted line shows that a reduced σ_b^2 will give the desired $\sigma_{\bar{x}}^2$ with a smaller n.

4.3   Expressions for the Number of Increments Per Composite Sample for a Desired $\sigma_{\bar{x}}^2$

Few papers discuss estimating the number of increments to attain a certain level of variation when making inference on a segment. Elder et al. (1980) briefly discuss the problem without a final recommendation of how to estimate n.

If values for σ_I^2 and σ_t^2 are known and values for r, s, t, and $\sigma_{\bar{x}}^2$ can be specified, the reduced model given by Equation (4.2) gives the minimum number of increments (n_o), that would be required if $\sigma_b^2 = 0$. Solving Equation (4.2) for n yields

$$n_o = \frac{\sigma_I^2}{r\sigma_{\bar{x}}^2 - \frac{\sigma_t^2}{st}} \tag{4.3}$$

When σ_b^2 is not zero,

$$n = \frac{\left[r\sigma_{\bar{x}}^2 - \frac{\sigma_t^2}{st}\right] \pm \sqrt{\left[\frac{\sigma_t^2}{st} - r\sigma_{\bar{x}}^2\right]^2 - \frac{4\sigma_b^2(\sigma_I^2)^2}{s}}}{\frac{2\sigma_b^2\sigma_I^2}{s}} \tag{4.4}$$

which is obtained by using the quadratic formula to solve Equation (4.1) for n.

Equation (4.4) gives two solutions since the minimum value is desired, use of the minus sign in Equation (4.4) gives the minimum n value. Figure 4.2 shows examples of having two solution for n of which the solution that results from the minus sign (Equation 4.4 would always be preferable.

$$n = \frac{\left[r\sigma_{\bar{x}}^2 - \frac{\sigma_t^2}{st}\right] - \sqrt{\left[\frac{\sigma_t^2}{st} - r\sigma_{\bar{x}}^2\right]^2 - \frac{4\sigma_b^2(\sigma_I^2)^2}{s}}}{\frac{2\sigma_b^2\sigma_I^2}{s}}. \tag{4.5}$$

4.3.1 σ_b^2-Definition and Bounds

An important part of using Equation (4.5) to estimate the number of increments in a composite sample is to have an accurate understanding of σ_b^2. Three issues are discussed. First, what is σ_b^2 and what are some of its properties? Second, when using Equation (4.5) to estimate n and values for σ_I^2, σ_t^2, r, s, and t are known or specified, what are the permissible values of σ_b^2 in order to attain a desired $\sigma_{\bar{x}}^2$? Finally, how do changes in σ_b^2 within the permissible levels explained in section 4.4 relate to n_o?

Elder et al. (1980) describe the blending variation by defining G as the total amount of material in each subsample and g_{ijl} as the amount of material from the lth increment in the ith composite that appears in the jth subsample from the composite, where $i = 1, \ldots, r$, $j = 1, \ldots, s$, and $l = 1, \ldots, n$. Then the fraction of material from the lth increment in the jth subsample from the ith composite sample is

$$b_{ijl} = \frac{g_{ijl}}{G} \tag{4.6}$$

and thus

$$\sum_{l=1}^{n} b_{ijl} = 1 \tag{4.7}$$

σ_b^2 is the variance of the b_{ijl}, $l = 1, \ldots, n$ and the mean of b_{ijl} is $1/n$. If perfect blending is achieved, there should be $1/n$ of each of the n increments in the subsample taken from the composite. On the other hand, the worst blending would result in n/n of one increment and $0/n$ of the remaining increments in the laboratory subsample. This reasoning leads to the limits of σ_b^2 Elder et al. (1980).

$$0 \le \sigma_b^2 \le \frac{n-1}{n^2} \tag{4.8}$$

27

These bounds show that as the number of increments used in the composite sample are increased toward infinity, the maximum of σ_b^2 will approach zero. However, in actual sampling procedures as n increases the compositing procedure will deteriorate (mix poorly) which will tend to cause σ_b^2 to increase. This creates an interesting dynamic in how σ_b^2 changes with increasing n. It would be beneficial for each composite sampling project to carefully determine how this relationship applies in their procedure.

For small increases n, the mixing procedure should not decrease in efficiency, so that an increase in n may result in some decrease in σ_b^2; however, the amount of decrease in σ_b^2 would be unique to each application and difficult to determine. Rohde (1976) proposed a Dirichlet model for σ_b^2 which results in a precise relationship between n and σ_b^2. Under a Dirichlet model, σ_b^2 when mixing two increments must be larger than when mixing six increments. This forced relationship does not intuitively represent physical blending of soil samples very well over a small interval of n.

In this paper we will assume that σ_b^2 is constant over a reasonable range of values of n. Assuming σ_b^2 is constant with respect to n, as discussed in section 4.2.1, will result in a conservative approach that allows σ_b^2 to be accounted for in the equation to estimate n.

4.4   How σ_b^2 is affected by the other factors.

When all the parts of Equation (4.3) are known or given, then using the discriminant of Equation (4.5) will show the permissible range of σ_b^2. The discriminant from Equation (4.5) is

$$\left[\frac{\sigma_t^2}{st} - r\sigma_{\overline{x}}^2 \right]^2 - \frac{4\sigma_b^2(\sigma_I^2)^2}{s} \tag{4.9}$$

28

which means that

$$\left(\frac{\sigma_t^2}{st} - r\sigma_{\bar{x}}^2\right)^2 \geq \frac{4\sigma_b^2(\sigma_I^2)^2}{s} \tag{4.10}$$

in order that the square root of the discriminant is a real number. Solving for σ_b^2,

$$\sigma_b^2 \leq s\left(\frac{\frac{\sigma_t^2}{st} - r\sigma_{\bar{x}}^2}{2\sigma_I^2}\right)^2. \tag{4.11}$$

Next, taking the square root of both sides of Equation (4.11) yields the only positive solution,

$$\sigma_b \leq \sqrt{s}\frac{r\sigma_{\bar{x}}^2 - \frac{\sigma_t^2}{st}}{2\sigma_I^2} \tag{4.12}$$

which can be expressed as

$$\sigma_b \leq \frac{\sqrt{s}}{2} \times \frac{1}{\frac{\sigma_I^2}{r\sigma_{\bar{x}}^2 - \frac{\sigma_t^2}{st}}}. \tag{4.13}$$

Expression (4.13) is a function of $\sqrt{s}/2$ and n_o in Equation (4.3) which is used to estimate n when there is no blending variation. With this relationship, to estimate n for known values of σ_I^2, σ_t^2, r, s, and t and a desired value for $\sigma_{\bar{x}}^2$, the range of σ_b^2 that will give real solutions for n is

$$0 \leq \sigma_b^2 \leq \frac{s}{2n_o^2}. \tag{4.14}$$

29

4.4.1 Estimation of n Using σ_b^2 and Their Relationship Based on Equation (4.9)

Another important fact that results from Equation (4.5) is that the maximum number of increments (n_m) for a specified set of parameter values is $2n_o$. This result can be seen by using Equation (4.5) and replacing σ_b^2 with its maximum permissible value Expression 4.14). This results in

$$n_m = \frac{\left(r\sigma_{\bar{x}}^2 - \frac{\sigma_t^2}{st}\right) - \sqrt{\left(\frac{\sigma_t^2}{st} - r\sigma_{\bar{x}}^2\right)^2 - \frac{4}{s}\left(\frac{\sqrt{s}}{2}\frac{r\sigma_{\bar{x}}^2 - \frac{\sigma_t^2}{st}}{\sigma_I^2}\right)^2 (\sigma_I^2)^2}}{\frac{2}{s}\left(\frac{\sqrt{s}}{2}\frac{r\sigma_{\bar{x}}^2 - \frac{\sigma_t^2}{st}}{\sigma_I^2}\right)^2 \sigma_I^2} \tag{4.15}$$

which reduces to

$$n_m = 2\frac{\sigma_I^2}{r\sigma_{\bar{x}}^2 - \frac{\sigma_t^2}{st}} = 2n_o. \tag{4.16}$$

Assuming that σ_b^2 is constant with respect to n in this interval will give conservative estimates for the number of of increments (n) needed based on differing levels of σ_b^2. This conservative approach is a result of using the maximum σ_b^2 allowable based on the discriminant of Equation (4.5). Now, σ_b^2 can be examined as a function of the physical mixing process. If the σ_b^2 is to large to obtain an estimate for n, then improvements in σ_b^2 must be attained by improving the physical mixing procedure.

4.5 Estimating the Number of Increments

If σ_I^2, σ_t^2 are known, $\sigma_{\bar{x}}^2$, r, and t are predetermined and $s=1$, then σ_b^2 is bounded by its restraint defined in Expression (4.13). So, σ_b^2 can take one values between 0 and the

value given by Expression (4.13). Now we can look at the range of σ_b and see how changing σ_b affects the estimate of the number of increments.

Figures (4.3) and (4.4) are graphs with specified values of the parameters used in Equation (4.5). The values on the x and y axes will change for different parameter values; however, the shape of the graph will remain the same.

The graph of the standardized values (Figure 4.5) is a unique way that the relationship can be described. This relationship is defined by the discriminant (Equation 4.5). Each axis is standardized using

$$\frac{1}{2\dfrac{\sigma_I^2}{r\sigma_{\bar{x}}^2 - \dfrac{\sigma_t^2}{st}}}. \tag{4.17}$$

Specifically, the y-axis is multiplied by Expression (4.17) and the x-axis is a proportion of the max of σ_b or in other words a proportion of Equation (4.17).

Figure 4.6 has the same curve and x-axis as Figure 4.5; however, the y-axis is now standardized by

$$\frac{1}{\dfrac{\sigma_I^2}{r\sigma_{\bar{x}}^2 - \dfrac{\sigma_t^2}{st}}} \tag{4.18}$$

instead of Expression (4.17). Using this standardization the x-axis starts at n_o, the minimum value of n with no mixing variation. This unique relationship provides a simple method to calculate the number of increments when accounting for mixing variation. Calculate n_o and specify σ_b^2 as a proportion of $\max(\sigma_b^2)$, then use Figure 4.6 to estimate the number of increments to use.
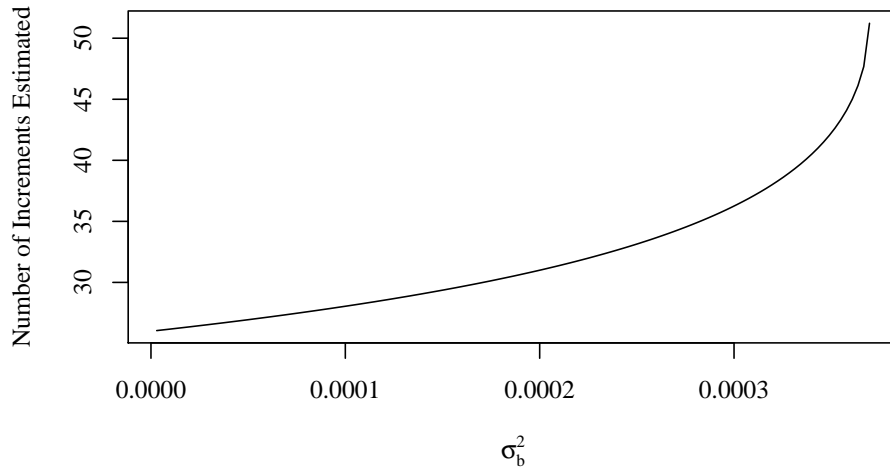
Figure 4.3: Relationship of σ_b^2 to the estimate of the number of increments with $\sigma_t^2 = 0.0001$, $\sigma_{\bar{x}}^2 = 1$, $\sigma_I^2 = 52$, $r = 2$, $s = 1$, $t = 1$

Figure 4.4: Relationship of σ_b^2 to the estimate of the number of increments with $\sigma_t^2 = 0.01$, $\sigma_{\bar{x}}^2 = .5$, $\sigma_I^2 = 19$, $r = 3$, $s = 1$, $t = 1$

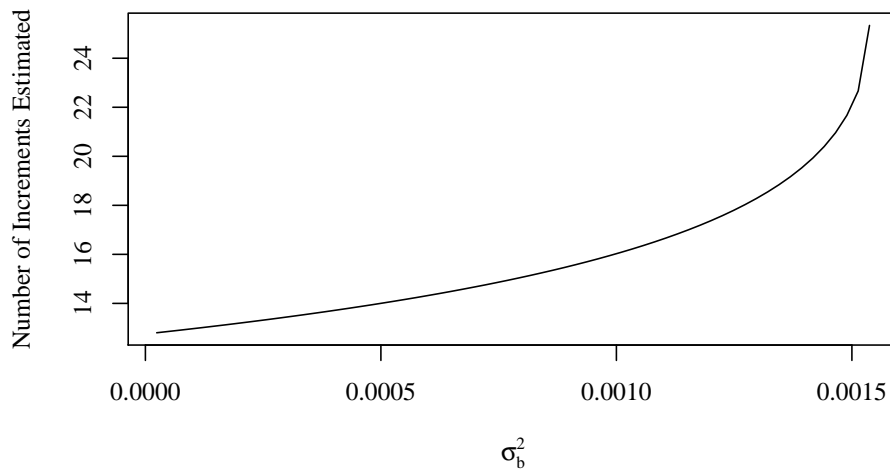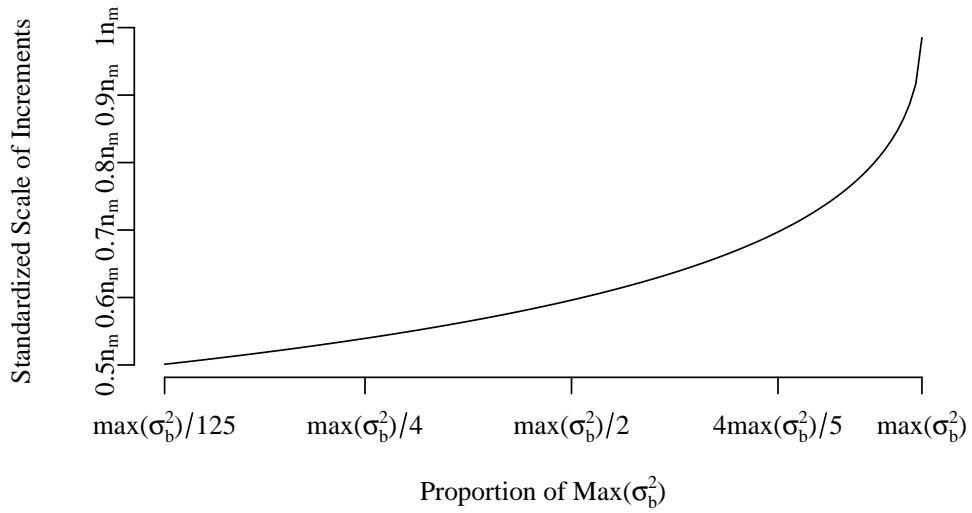Figure 4.5: Relationship of σ_b^2 to the estimate of the number of increments with each axis standardized by Equation (4.17)
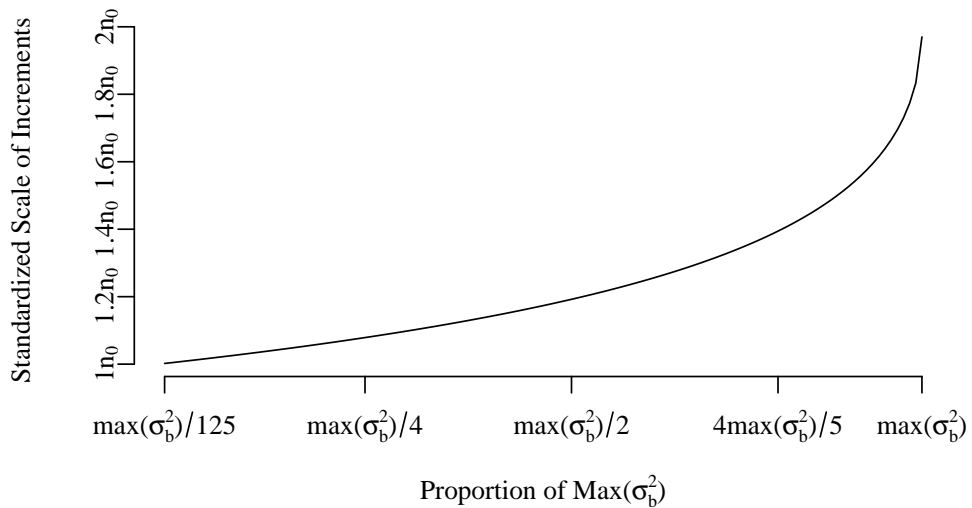

Figure 4.6: Relationship of σ_b^2 to the estimate of the number of increments with the x-axis standardized by Equation (4.17) and the y-axis standardized by Equation (4.18)

33

4.5.1 Example Data Set

Jenkins et al. (1997) describe a composite sampling process that was performed at multiple sites. Two of the 10 data sets, Hawthorne and Volunteer, are used as an example of how to obtain estimates for the desired variance components described previously. Table (4.1) shows the data of the TNT measurements for the increments and for the seven sub-samples taken from the composite. For their study, they took seven samples by randomly selecting the location for sample one and then systematically taking the remaining six samples as illustrated at the top of Figure 4.7. Each of these seven samples was divided into two increments, labeled 1a and 1b through 7a and 7b. The remaining material was placed into the composite sample. After compositing , seven sub-samples were taken from the composite and measured (Each sub-sample measurement is labeled as S1,…,S7).

4.5.2 Analysis

The two data sets in Table (4.1) were analyzed using PROC MIXED in SAS® 9.1. Using the duplicate measurements on each increment, an estimate for the testing variation, σ_t^2, was obtained. The between increment variation, σ_I^2, was also estimated for each dataset. The between subsample variation,σ_s^2, provided a way to find σ_b^2. Elder et al. (1980) showed that

$$\sigma_{\bar{x}}^2 = \frac{\sigma_I^2}{rn} + \frac{\sigma_s^2}{rs} + \frac{\sigma_t^2}{rst} \tag{4.19}$$
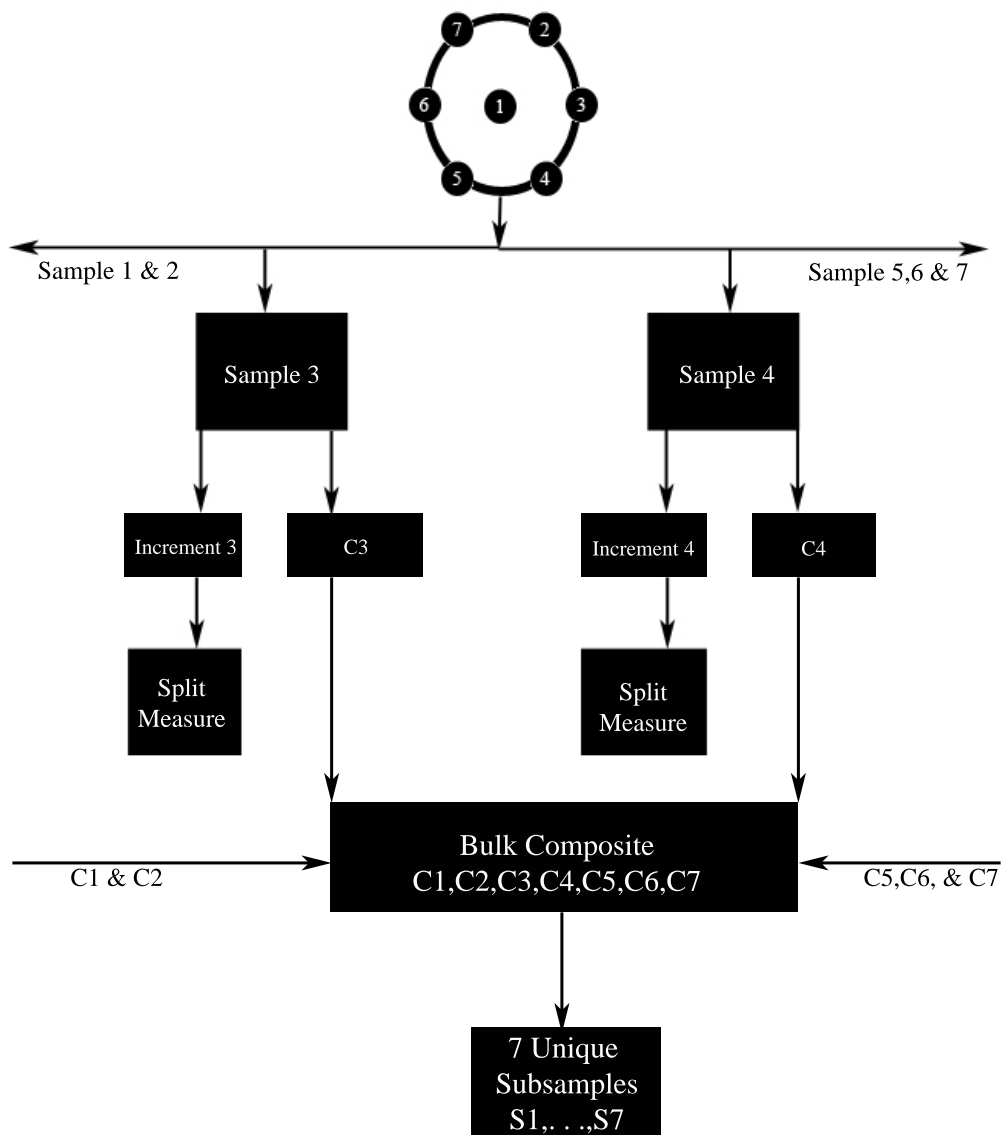
where

$$\sigma_s^2 = n\sigma_b^2\sigma_I^2 \tag{4.20}$$

34

Figure 4.7: Sampling and Measurement Procedure for the data of Jenkins et al. (1997)

Hawthorne		Volunteer	
TNT on-site analysis (mg/g)			
Increments		Increments	
1a	127	1a	4.3
1b	125	1b	4.1
2a	116	2a	6.1
2b	103	2b	5.9
3a	379	3a	17.6
3b	366	3b	19.9
4a	59.1	4a	10.5
4b	56	4b	10.7
5a	12.4	5a	33
5b	13.3	5b	35.8
6a	170	6a	13.9
6b	173	6b	14.4
7a	240	7a	7.8
7b	245	7b	7.3
Subsamples		Subsamples	
S1	129	S1	15.1
S2	137	S2	15.9
S3	116	S3	16.5
S4	138	S4	17.6
S5	139	S5	19.4
S6	147	S6	15.1
S7	170	S7	16.6

Table 4.1: Two data sets from Jenkins et al. (1996). Each sample was split into a part to composite and an increment that was split and measured twice.

	Volunteer (mg/g)	Hawthorne (mg/g)
σ_I^2	14626	108.04
σ_t^2	27.6014	0.9821
σ_s^2	248.68	1.324
σ_b^2	0.00175	0.00243

Table 4.2: The estimates of the variance components from SAS® by their respective data sets.

The data set in Table (4.1) provided estimates for σ_s^2, n, and σ_I^2. With these values an estimate of σ_b^2 can be obtained as

$$\hat{\sigma}_b^2 = \frac{\hat{\sigma}_s^2}{n\hat{\sigma}_I^2} \tag{4.21}$$

4.5.3 Example for estimating n based on the Hawthorne and Volunteer Estimates

For each site, we have chosen values for r, s, and t. In our example, two composites (r) are done and one sub-sample (s) is selected and measured (t) from each composite. The values of r, s, and t can vary by situation. n_o from Equation (4.3) was calculated for the Hawthorne site with desired $\sigma_{\bar{x}}^2$ of 15 and for the Volunteer site with a desired $\sigma_{\bar{x}}^2$ of 750. If $\sigma_b^2 = 0$, the estimate of n_o for the Hawthorne site was 3.73, and the estimate of n_o for the Volunteer site was 9.93. These two values were used in Equation (4.14) to find the highest allowable σ_b^2 to obtain an estimate of n for each site. The maximum σ_b^2 for the Hawthorne site was 0.01797 , and was 0.002534 for the Volunteer site. Table (4.2) shows $\hat{\sigma}_b^2$ for Hawthorne and Volunteer to be 0.00175 and 0.00243, respectively. At each site the blending variation is low enough to attain the desired $\sigma_{\bar{x}}^2$. Figure 4.6 and the value of n_o for each site gave the proposed number of increments to use based on each site's blending variation. These values match the calculations done for each site using Equation (4.5). The calculations for the Hawthorne and Volunteer sites using Equation (4.5) are given in Equations (4.22) and (4.23), respectively.

$$3.825 = \frac{2 \times 15 - \frac{0.9821}{1} - \sqrt{\left(\frac{0.9821}{1} - 2 \times 15\right)^2 - 4 \times 0.00175 \times 108.04^2}}{2 \times 0.00175 \times 108.04} \tag{4.22}$$

37

$$16.52 = \frac{2 \times 750 - \frac{27.6014}{1} - \sqrt{\left(\frac{27.6014}{1} - 2 \times 750\right)^2 - 4 \times 0.00243 \times 14626^2}}{2 \times 0.00243 \times 14626} \qquad (4.23)$$

Thus 4 increments should be composited for Hawthorne and 12 increments should be composited for Volunteer.

4.6 Conclusion

The data of Table 4.1 allowed us to estimate σ_b^2 using Equation (4.21). Another possible method to estimate σ_b^2 is to spike one increment per composite with a substance that performs similarly to the substances of interest and then do the compositing. If the blending process is good, the substance should be $1/n$ of the original spiked amount when the subsample is measured. If a substance could be chosen that did not interact with the elements of interest, this could be used as a possible quality control mechanism on blending.

Three important facts emerge from this study on estimating n when blending variation exists. First, a large blending variance will prevent a solvable solution for n based on certain desired values for $\sigma_{\bar{x}}^2$. Second, when the blending variance is included in the model it can have a large effect on the estimated value of n. Hence, the blending variance should be used in determining the number of increments to use in a composite sample study. Third, creating accurate methods for estimating the number of increments per composite sample will benefit sampling software programs that assist in composite sampling problems.

4.7   Acknowledgments

Chapter 5

Further Research and Contributions to Statistical Practice

There are are a few ideas that could be developed so the theory of estimating the number of increments could be used more in actual research. For example, the data used in this project (Jenkins et al. 1996, 1997) could be used to estimate the number of increments to be used in a future study in the same or similar areas. This would provide an opportunity to see if the estimation maintains the desired variance.

Another crucial step is developing a cost analysis program that uses the costs of different sampling procedures while accounting for the mixing variation. As previously cited some papers have developed a cost analysis program that looks at different combinations of sampling procedures in composite sampling. For example, one could place five increments in each composite sample and gather twenty five composite samples or one could place twenty-five increments in each composite sample and gather five composites. A key element that would affect how many composite samples are gathered is the relationship of mixing variation to the number of increments per composite. Accounting for mixing variation while estimating costs would be a useful extension of this project.

Chapter 6

Appendix: Code

```
varequation1 <- function(iv,bv,tv,r,n,s,t){
mv <- iv/(r*n) + (n*bv*iv)/(r*s)  + tv/(r*s*t)
mv
}
xx <- 1:60

pdf(file=vargraph.pdf",width=6.25, height=4, family="Times",pointsize = 12)
 plot(xx,varequation(4,.006,.002,2,xx,1,2),type="l",xlab=
 "Number of increments used",ylab=quote("Variance of" *bar(x)*""))
 title(main=expression("Relationship of  " *sigma[bar(x)]^2* "  to the number
 of increments used"))
  legend(5,2,c(
  expression(t==1),
  expression(r==3),
  expression(sigma[t]^2 == .002),
  expression(sigma[b]^2 == .006),
  expression(sigma[I]^2 == 4),
  expression(1/sigma[b] == 13)),
  horiz=F,ncol=2)
  axis(1,13,font=2)
dev.off()

pdf("\\figure2.pdf", width = 6.25,height=4, family="Times", pointsize=12)
 plot(stu[,2],stu[,1],xlab=quote(sigma[b]^2),ylab="Number of Increments
    Estimated")
dev.off()

pdf("\\figure3.pdf", width = 6.25, height = 4, family = "Times",pointsize =
    12)
 plot(stu1[,2],stu1[,1],xlab=quote(sigma[b]^2),ylab="Number of Increments
    Estimated")
dev.off()

pdf("\\figure4.pdf", width = 6.25, height=4, family="Times",pointsize=12)
 plot(stu[,2],stu[,4],axes=F,xlab=quote("Proportion of Max("*sigma[b]^2*")"),
  ylab="Standardized Scale of Increments",type="l")
```

41

```r
  axis(1,at=c(0.0000029576,.0001,.0002,.0003,.0003697), labels=F, distn=NULL)
  axis(2, labels=F, distn=NULL)
  a <- expression("max("*sigma[b]^2*")"/125)
  b <- expression("max("*sigma[b]^2*")"/4)
  cc <- expression("max("*sigma[b]^2*")"/2)
  d <- expression("4max("*sigma[b]^2*")"/5)
  e <- expression("max("*sigma[b]^2*")")
  mtext(c(a,b,cc,d,e),side=1,line=1,at=c
      (0.0000029576,.0001,.0002,.0003,.0003697))
  a <- expression(0.5*n[m])
  b <- expression(0.6*n[m])
  cc <- expression(0.7*n[m])
  d <- expression(0.8*n[m])
  e <- expression(0.9*n[m])
  f <- expression(1.0*n[m])
  mtext(c(a,b,cc,d,e,f), side=2, line=.5, outer=F, at=c
      (0.5,0.6,0.7,0.8,0.9,1))
dev.off()

pdf("\\figure5.pdf", width = 6.25, height = 4, family = "Times",pointsize =
    12)
 plot(stu[,2],stu[,3],axes=F,xlab=quote("Proportion of Max("*sigma[b]^2*")"),
  ylab="Standardized Scale of Increments",type="l")
  axis(1,at=c(0.0000029576,.0001,.0002,.0003,.0003697), labels=F, distn=NULL)
  axis(2, at=c(1,1.2,1.4,1.6,1.8,2.0),labels=F, distn=NULL)
  a <- expression("max("*sigma[b]^2*")"/125)
  b <- expression("max("*sigma[b]^2*")"/4)
  cc <- expression("max("*sigma[b]^2*")"/2)
  d <- expression("4max("*sigma[b]^2*")"/5)
  e <- expression("max("*sigma[b]^2*")")
 mtext(c(a,b,cc,d,e), side=1, line=1, outer=F, at=c(0.0000029576,.0001,.0002,
 .0003,.0003697))
  a <- expression(1*n[0])
  b <- expression(1.2*n[0])
  cc <- expression(1.4*n[0])
  d <- expression(1.6*n[0])
  e <- expression(1.8*n[0])
  f <- expression(2.0*n[0])
 mtext(c(a,b,cc,d,e,f),side=2,line=.5,outer=F,at=c(1,1.2,1.4,1.6,1.8,2.0))
dev.off()

#  Code to estimate variance components in SAS  #

data increments;
infile "/increment.dat" firstobs=2;
input location site$ Sample measurement$ TNTOS TNB TNT 24DNT Total;
run;

data composites;
infile "/composite.dat" firstobs=2;
input location site$ composites$ TNTOS TNB TNT 24DNT Total;
run;
```

```
proc mixed data=increments;
class Sample measurement;
model TNTOS=;
random Sample;
by location;
run;

proc mixed data=composites;
class composites;
model TNTOS=;
by location;
run;
```

Bibliography

Behets, F., Bertozzi, S., Kasali, M., Kashamuka, M., A. L., Brown, C., Ryder, R., and Quinn, T. (1990), "Successful Use of Pooled Sera to Determine HIV-1 Seroprevalence in Zaire With Development of Cost-Efficiency Models," *AIDS*, 4, 737–741.

Boswell, M. T., Gore, S. D., Johnson, G. D., Patil, G. P., and Taillie, C. (1995), *EPA Observational Economy Series Vol. 1: Composite Sampling*, United States Environmental Protection Agency.

Boswell, M. T., Gore, S. D., Lovison, G., and Patil, G. P. (1996), "Annotated bibliography of composite sampling. Part A: 1936-92," *Environmental and Ecological Statistics*, 3, 1–50.

Boswell, M. T. and Patil, G. P. (1987), "A perspective of composite sampling," *Communications in Statistics, Part A – Theory and Methods [Split from: @J(CommStat)]*, 16, 3069–3093.

Brown, G. H. and Fisher, N. I. (1972), "Subsampling a mixture of sampled material," *Technometrics*, 14, 663–668.

Davis, C. E., Grizzle, J. E., and Bryan II, J. A. (1973), "Estimation of the Probability of Post-Transfusion Hepatitis in Hemophelia Treatment," *Biometrics*, 29, 386–396.

Duncan, A. J. (1962), "Bulk sampling: Problems and lines of attack," *Technometrics*, 4, 319–344.

Elder, R. S., Thompson, W. O., and Myers, R. H. (1980), "Properties of composite sampling procedures," *Technometrics*, 22, 179–186.

Exner, J. H., Keffer, W., Gilbert, R. O., and Kinnison, R. R. (1985), "A Sampling Strategy for Remedial Action at Hazardous Waste Sites: Clean-up of Soil Contaminated by Tetrachlorodibenzo-p-Dioxin," *Hazardous Waste & Hazardous Materials*, 2, 503–521.

Gilbert, R. O. and Doctor, P. G. (1985), "Determining the Number and Size of Soil Aliquots for Assessing Particulate Contaminant Concentrations," *Journal of Environmnetal Quality*, 14, 286–292.

Gore, S. D., Patil, G. P., and Taillie, C. (2001), "Identifying the largest individual sample value from a two-way composite sampling design," *Environmental and Ecological Statistics*, 8, 151–162.

44

Hassig, N. L., Wilson, J. E., Gilbert, R. O., and Pulsipher, B. A. (2004), *Visual Sample Plan Version 3.0 User's Guide (PNNL-14970)*, Pacific Northwest National Laboratory, Richland, WA.

Izenman, A. J. (2001), "Statistical and legal aspects of the forensic study of illicit drugs," *Statistical Science*, 16, 35–57.

Jenkins, T. F., Grant, C. L., Brar, G., Thorne, P. G., Schumacher, P. W., and Ranney, T. (1996), "Assessment of Sampling Error Associated with Collection and Analysis of Soil Samples at Explosives-Contaminated Sites," Tech. rep., US Army Corps of Engineers: Cold Regions Research and Engineering Laboratory, special Report 96-15.

— (1997), "Sampling Error Associated with collection and Analysis of Soil Samples at TNT-Contaminated Sites," *Field Analytical Chemistry and Technology*, 1, 151–163.

Johnson, G. D. and Patil, G. P. (2001), "Cost analysis of composite sampling for classification," *Environmental and Ecological Statistics*, 8, 91–107.

Lancaster, V. A. and Keller-McNulty, S. (1998), "A review of composite sampling methods," *Journal of the American Statistical Association*, 93, 1216–1230.

Patil, G. P. (1995), "Editorial: Composite sampling," *Environmental and Ecological Statistics*, 2, 169–179.

Pitard, F. F. (1993), *Pierre Gy's Sampling Theory and Sampling Practice Heterogeneity, Sampling Correctness and Statistical Process Control.*, Boca Raton, FL.: CRC Press.

Rohde, C. A. (1976), "Composite sampling," *Biometrics*, 32, 273–282.

Rohlf, F., Akcakaya, H., and Ferraro, S. (1996), "Optimizing Composite Sampling Protocols," *Environmental Science Technology*, 30, 2899–2905.

Stephens, G. M., Raboud, J. M., Karakas, L., and Sherlock, C. H. (2000), "Can Pooling Be Used for Seroprevalence Studies of Hepatitis C?" *Journal of Clinical Microbiology*, 38, 4264–4265.

Tebbs, J. M. and Swallow, W. H. (2003), "More Powerful Likelihood Ratio Tests for Isotonic Binomial Proportions," *Biometric Journal*, 45, 618–630.

Tu, X., Litvak, E., and Pagano, M. (1994), "Studies of AIDS and HIV Surveillance Screening Tests: Can We Get More by Doing Less?" *Statistics in Medicine*, 13, 1905–1919.