Survey Methodology. December 1993
Vol. 19, No. 2, pp. 199-204
Statistics Canada

An Alternative View of Forest Sampling

FRANCIS A. ROESCH, JR., EDWIN J. GREEN and CHARLES T. SCOTT[1]

ABSTRACT

A generalized concept is presented for all of the commonly used methods of forest sampling. The concept views the forest as a two-dimensional picture which is cut up into pieces like a jigsaw puzzle, with the pieces defined by the individual selection probabilities of the trees in the forest. This concept results in a finite number of independently selected sample units, in contrast to every other generalized conceptualization of forest sampling presented to date.

KEY WORDS: Forest sampling; PPS sampling.

1. INTRODUCTION

The sampling of forests is often accomplished as a two part process: first a random point is located in the forest and then a cluster of trees in the vicinity of the point is selected for the sample by some rule. The two most common rules are known as (circular, fixed-area) plot sampling and (horizontal) point sampling. In the former, all trees for which the center of the cross-section of the bole at 4.5 feet above the ground is within a constant horizontal distance (d) of the random point are included in the sample. In the latter, tree i is selected for the sample if this center is within a horizontal distance αr_i of the random point, where r_i is the radius of the cross-section and α is a constant, chosen appropriately to obtain a desired sampling intensity. Tree i would be selected with probability proportional to πd^2 in plot sampling (the probability is the same for all trees) and with probability proportional to πr_i^2 (basal area of tree i) in point sampling (larger trees have a higher probability of selection).

There has been much discussion in the forestry literature about what the sample unit actually is in the various methods of forest sampling. The tree is considered the sample unit from one point of view (e.g. Oderwald 1981), while from other points of view, the cluster of trees associated with the point (e.g. Palley and Horwitz 1961; Schreuder 1970), the circular plot (e.g. Cunia 1965), and the point (e.g. Husch 1955) are considered the sample units. These various viewpoints are supported by different statistical tools. For example, treating the tree as the sample unit requires the use of finite population sampling theory, while considering the point as sample unit requires the use of the somewhat more advanced theory of infinite population sampling. In addition, plot sampling has traditionally been presented from the viewpoint of the plot as the sample unit, whereas point sampling has usually been presented from the viewpoints of the tree or the point as the sample unit. Therefore, these very common and quite similar sampling mechanisms artificially appear disparate.

We will show a conceptualization of the primary sample unit that is applicable to every type of forest sampling scheme which selects trees based on the location of a random point. We will also show that this conceptualization is simple and that it provides a finite number of mutually exclusive and independently selected sample units. This is in contrast to the view of the tree or the cluster of trees as the sample unit, because trees are not selected independently and clusters of trees are not mutually exclusive. It also differs from the views of the randomly placed point or the plot as the sample unit, because there are an infinite number of units in these cases. We will also suggest that this alternative conceptualization is often more appropriate.

2. THE JIGSAW PUZZLE VIEW

Suppose that there are Ntrees in the forest with labels 1, 2, . ..) N. Associated with the N frees are values of interest $\tilde{y} = \{\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_N\}$, K-circles $\boldsymbol{K} = \{K_1, K_2, \ldots, K_N\}$, and selection areas of sizes $\tilde{A} = \{\tilde{A}_1, \tilde{A}_2, \ldots, \tilde{A}_N\}$. Grosenbaugh and Stover (1957) first defined the K-circle in the context of point sampling. For our purposes the K-circle of tree i, K_i, is an imaginary circle, centered at tree center, with radius d in plot sampling and radius αr_i in point sampling. The selection area for tree i, of size A_i (in acres), is the portion of tree i's K-circle which is within the forest, and is the area from within which a random point will select the tree for the sample.

When discussing point sampling, Palley and Horwitz (1961) contend that "... the primary sampling unit is a cluster of trees associated with a locus of origin. The locus

[1] Francis A. Roesch. Jr., Mathematical Statistician, Institute for Quantitative Studies, Southern Forest Experiment Station, USDA Forest Service, 701 Loyola Avenue, New Orleans, LA 70113: Edwin J. Green, Professor of Forestry, Cook College-Rutgers University, P.O. Box 231, New Brunswick, NJ 08903; and Charles T. Scott, Project Leader, Forest Ecosystem Modeling Unit, Northeastern Forest Experiment Station, USDA Forest Service, 359 Main Road, Delaware, Ohio 43015.

of origin is a point in the case of point sampling ...". Actually the locus of origin is not a point because the cluster of trees is not selected only from that point but rather from an infinite set of points within a specific area.

We offer the alternative view of the sample units being the mutually exclusive sections of ground resulting from the overlapping selection areas of the individual trees in the forest.

The treatment of the ground broken up into primary sampling units is clearly shown in Figure 1, for example. The correspondence between the population, sampling frame and sample unit as given in say Cochran (1977, p. 6) is apparent: the population (or *the puzzle picture*) is divided up into mutually exclusive, exhaustive sample units (the *puzzle pieces*) which together comprise the sample frame. Each ground segment has a definite probability of selection and the total of these probabilities over all segments is 1. We will call this the jigsaw puzzle view.

Associated with each ground segment are attributes of interest, the measurement of which will result in identical values from any point in that segment of ground. The crux of the matter is that individual points are equivalent within any particular segment. The ground segments, of course, are selected with probability proportional to size. In the case of point sampling, the segment size is determined by the basal areas and spatial distribution of the trees and the constant α chosen. Once α is chosen, the sample frame at a particular point in time is fixed. In the case of plot sampling, the size of the segment is determined by d and the spatial distribution of the trees. Thus, regardless of the method used to determine the sample trees (e.g., plot sampling or point sampling), all schemes can be thought of as cutting the puzzle up in some way, selecting the pieces with probability proportional to their size, and then turning each piece over to read the attributes associated with it.

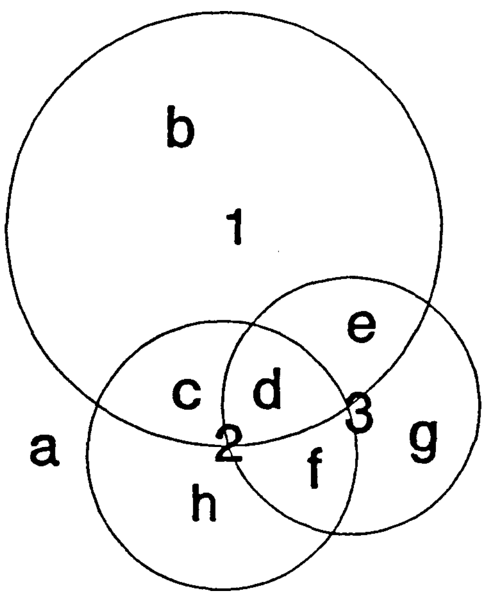

Figure 1 The Puzzle Pieces. Trees 1, 2 and 3 are centered at their respective numbers. The surrounding circles represent the selection areas of the trees. Each of the lettered segments represents a sample unit.

Returning to our proposition that this view is often more appropriate, we note that the purpose of most forest surveys is to describe the *forest*, not the individual trees. Our aggregations are usually made on a per acre or hectare basis, ***i.e.*** units of the forest land, ***not*** units of the tree. From the same place we may measure many other things besides the trees such as topographic and site characteristics. It is therefore usually more appropriate to view pieces of the forest as the sample units rather than individual trees in the forest.

Although we will be working mostly in the context of forest sampling in general, our discussion is easily applied to any specific type of forest sampling which relies on the selection of trees by some function of randomly placed points. The only difference is the definition of the ground segments, or how we dissect the picture into puzzle pieces. For example, in plot sampling the ground is divided into pieces defined by overlapping circles of equal size, while in point sampling the definition is by overlapping circles of sizes proportional to each corresponding tree's basal area.

To examine this further, suppose that we randomly drop a point on the surface of a forest and use any function to select sample trees. Suppose also that within our forest are three trees (1, 2, and 3) whose selection areas overlap. In Figure 1, trees 1, 2 and 3 are centered at their respective numbers with their selection areas shown as circles. Each lettered segment represents a different sample unit. If the point falls in segment a, the empty cluster is chosen, in segment b, the cluster containing only tree 1, in segment d, the cluster of all three trees, efc. Tree 1 would therefore be selected from segments b, c, d or e. This results in a situation somewhat analogous to that described in Kish (1965, sec. 11.2), if we were to consider the tree to be the primary sample unit, in which a list to be sampled from contains duplicate listings of the same unit. In this case, the list would be one of clusters of trees, in which ***most*** trees are associated with more than one cluster. The clusters are selected with probability proportional to the size of the ground segment. The standard technique of weighting duplicate elements of a list, discussed by Kish, considers rather the selection of primary units with equal probability.

The jigsaw puzzle view reduces the complexity of the sampling mechanism in one sense by first mapping the tree population into the ground segment population and thereby reducing the sample list from a list of clusters of trees in which trees belong to more than one cluster to a list of unique ground segments. Our claim below that forest sampling simulations can be simplified by the jigsaw puzzle view is supported wholly by the tradeoff between the one time cost of this reduction in the complexity of the sample list and the need to select from that list many times.

To man the tree population into the segment population, an observation for a segment would preferably be the sum of weighted tree values, the weight for each tree being proportional to its probability of being observed from that particular segment. The probability that sampled tree i was selected from the particular ground segment j is:

$$p_{ij} = \left(\frac{A_j}{\tilde{A}_i}\right) Z_{ij},$$

where:

A_j = the area of segment j in acres, and

$$Z_{ij} = \begin{cases} 1 \text{ if segment j is part of the k-circle of tree } i \\ 0 \text{ otherwise.} \end{cases}$$

The sum over j of p_{ij} is 1. We can now write the observation for each segment as a sum of weighted tree values:

$$y_j = \sum_{i=1}^{N} p_{ij}\,\tilde{y}_i. \tag{1}$$

Now suppose that we randomly drop m points on the surface of a forest with the same assumptions as above (our sampling is with replacement). An unbiased estimator of the total value of interest for a sample selected with probability proportional to size is:

$$\begin{aligned} \hat{Y} &= \frac{A_T}{m} \sum_{j=1}^{m} \frac{y_j}{A_j} \\ &= \frac{A_T}{m} \sum_{j=1}^{M} \frac{y_j}{A_j} W_j, \end{aligned} \tag{2}$$

where:

$A_T = \sum_{j=1}^{M} A_j$; the total area of the forest in acres,

m = the number of sample points,

M = the number of ground segments, and

W_j = the number of times the jth unit appears in the sample.

Note that W_j is an integer between 0 and m, inclusive. A_j and y_j are fixed and W_j is random. In addition, we will define:

$Y = \sum_{i=1}^{N} \tilde{y}_i$; the total value of interest across all trees, and

$Y^* = \sum_{j=1}^{M} y_j$; the total value of interest across all segments.

To show that $\hat{Y}$ is unbiased for Y, we will first show $\hat{Y}$ to be unbiased for Y^* and then show that Y^* equals Y. Following Cochran (1977, p. 252-255), we can show $\hat{Y}$ to be unbiased for Y^*:

$$\begin{aligned} E[\hat{Y}] &= E\left[\frac{A_T}{m} \sum_{j=1}^{M} \frac{y_j}{A_j} W_j\right] \\ &= \frac{A_T}{m} \sum_{j=1}^{M} \frac{y_j}{A_j} E[W_j]. \end{aligned} \tag{3}$$

W_j is a multinomial random variable and its expected value is equal to $m(A_j/A_T)$. Therefore

$$E[\hat{Y}] = \sum_{j=1}^{M} y_j = Y^*. \tag{4}$$

We can now show that $\hat{Y}$ is unbiased for Y by showing that Y* = Y. Substituting the right hand side of equation (1) for y_j in the definition of Y^*, we get:

$$Y^* = \sum_{j=1}^{M} \sum_{i=1}^{N} p_{ij}\,\tilde{y}_i. \tag{5}$$

After substituting in the definition Of p_{ij} and rearranging the order of summation:

$$Y^* = \sum_{i=1}^{N} \tilde{y}_i \left[\frac{1}{\tilde{A}_i} \sum_{j=1}^{M} A_j Z_{ij}\right]. \tag{6}$$

Because

$$\tilde{A}_i = \sum_{j=1}^{M} A_j\, Z_{ij},$$

the term within the brackets on the right hand side of (6) equals 1, and

$$Y^* = \sum_{i=1}^{N} \tilde{y}_i = Y. \quad \text{Q.E.D.} \tag{7}$$

By definition, the variance of $\hat{Y}$ is

$$V(\hat{Y}) = \left(\frac{1}{mA_T}\right) \sum_{j=1}^{M} A_j \left(\frac{A_T y_j}{A_j} - Y\right)^2. \tag{8}$$

The sample estimate of the variance is then (Cochran 1977):

$$v(P) = \frac{1}{m(m-1)} \sum_{j=1}^{m} \left(\frac{A_T y_j}{A_j} - \hat{Y} \right)^2. \qquad (9)$$

The general development in equations (1) through (9) can be used for any specific type of forest sampling which follows the two part process of selecting trees from randomly placed points.

As a further example of the use of the jigsaw puzzle view, we will illustrate the sample frame when point samples are used to measure forest growth. For the greatest efficiency, measurements are taken at two points in time and the same random points are used both times. This type of sampling for forest growth is known as remeasured point sampling and has been discussed at length in the literature, most recently by Van Deusen *et al.* (1986) and Roesch *et al.* (1989, 1991, 1993). If a remeasured point sample had been taken, and Figure 1 represented time 1, the puzzle for the overall sample might be cut up into pieces like those in Figure 2. Trees 1, 2 and 3 are the same as those in Figure 1 and tree 4 is a tree which grew into the stand between times 1 and 2. The inner circles represent the trees' point sample areas of selection at time 1 (say αr_{i1}, including a subscript for time) and the outer circles represent the point sample areas of selection at time 2 (αr_{i2} is larger due to an increase in basal area). Tree 4 only has an outer circle since it did not exist at time 1 and tree 2 only has an inner circle since it died prior to time 2. The dotted circle represents the selection area tree 2 would have had at time 2 if time 2 had occurred just prior to the tree's demise. Therefore, the dotted circle does not contribute to the definition of the segments.

If the random point lands in segment *a*, trees 1 and 3 would be measured at both times and tree 2 would be measured only at time 1; in segment *b*, tree 1 would be measured at both times and tree 3 would only be measured at time 2. This exemplifies the fact that even though another dimension was added to the sample (the time dimension), the forest sample concept remains the same, since the time dimension can be collapsed down onto the puzzle picture. So, in addition to the conditions mentioned above, the definition of the segments depends upon the exact times of each measurement. This concept of the sample unit is helpful in understanding the estimators of the components of change from time 1 to time 2 given in Van Deusen *et al.* (1986) and Roesch *et al.* (1989 and 1991).

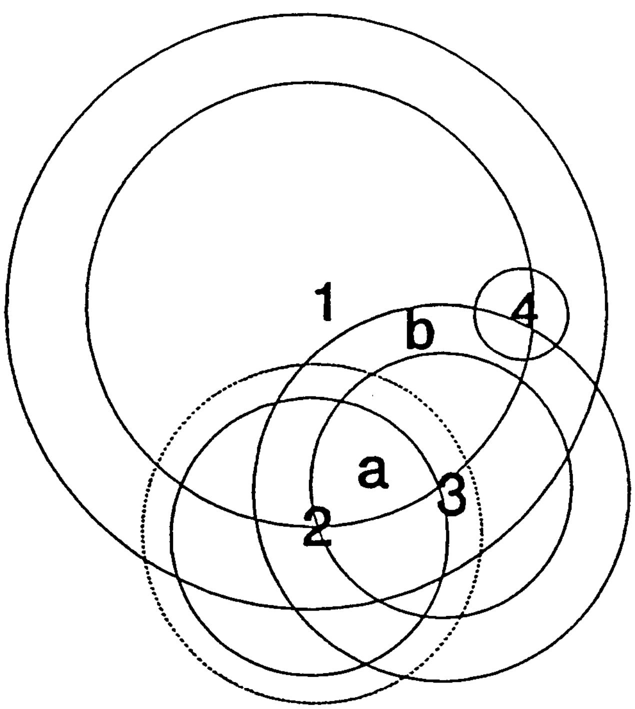

Figure 2. Puzzle pieces defined by location, size, and time. An example of sample units in a remeasured point sample. Trees 1 and 3 have grown and survived, tree 2 grew somewhat before dying and tree 4 is ingrowth.

3. DISCUSSION

Given the simplicity of the jigsaw puzzle concept, one might wonder why this view of forest sampling has not been proposed before. The most compelling reason is probably that the above estimators cannot be calculated when the A_j's are unknown. Since a particular tree's area of selection might be divided between many of the puzzle pieces and the size of a particular puzzle piece may be limited by trees not sampled by that piece, the selection areas of both sample and non-sample trees must be known to calculate the A_j's of the selected segments. For example, referring to Figure 1, if our point landed in section c, we would sample trees 1 and 2 and the area of c + *d* would be readily calculable. However, to calculate $\hat{Y}$ and v ($\hat{Y}$), we need the area of c alone, for which we do not have adequate sample information. We will show that this apparent deficiency is unimportant by showing that $\hat{Y}$ can be reexpressed in terms which are calculable. This will, in fact, always be the case no matter which sampling method is described by the jigsaw puzzle view.

The jigsaw puzzle view of point sampling is actually a mapping of the tree population into the associated ground segment population. We can reexpress $\hat{Y}$ to show that it is equivalent to the usual point sampling estimator which is based upon the tree population. Expanding equation (2) to include the definition of y_j and subsequent rearrangement gives:

$$\hat{Y} = \frac{A_T}{m} \sum_{j=1}^{M} \frac{y_j}{A_j} W_j$$

$$= \frac{A_T}{m} \sum_{j=1}^{M} \frac{\sum_{i=1}^{N} p_{ij} \bar{y}_i}{A_j} W_j$$

$$= \frac{A_T}{m} \sum_{j=1}^{M} \sum_{i=1}^{N} \frac{\bar{y}_i Z_{ij} W_j}{\bar{A}_i} \qquad (10)$$

$$= \frac{A_T}{m} \sum_{i=1}^{N} \frac{\bar{y}_i}{\bar{A}_i} \sum_{j=1}^{M} Z_{ij} W_j$$

$$= \frac{A_T}{m} \sum_{i=1}^{N} \frac{\bar{y}_i}{\bar{A}_i} w_i,$$

where w_i equals the number of rime\ tree i is selected for the sample. The final expression in (10) is the usual point sample estimator.

The purpose of this paper, therefore, is not to introduce a new set of estimators for sampling systems which already have reasonably good estimators, but rather to show how sampling schemes of quite disparate justifications in the literature are related in general. This alternative avenue of understanding may be useful in many ways. For one, we believe that some abstract forest sampling systems may be easier to understand if put into the framework described above. Our experience is that students, for instance, readily grasp the idea of point sampling when taught as merely a method of dividing the forest up into non-overlapping jigsaw puzzle pieces which are then sampled with probability proportional to size. Researchers who are interested in developing new forest sampling schemes or new estimators for existing schemes may benefit from this view because it provides another path for understanding new sampling schemes and for programming the forest sampling simulations used to test the new methods. The simulation discussed in Roesch (1993), for example, was simplified by using the jigsaw puzzle view rather than the other conceptualizations of the forest sampling frame which had been suggested up to that time. The simplification stemmed from the fact that the bulk of the simulation could be used for many different sampling schemes with only minor modifications to the subroutine which dissected the puzzle.

Because forest sampling simulations often start with a mapped forest, the A_j's are readily obtainable. Once the puzzle is dissected, y_j can be calculated for each piece. The simulator then simply selects these pieces from a list in proportion to their size. Contrast this with the simulation resulting from the view of the point as the sample unit. In this latter simulation, a random point would be dropped and the tree list searched for all of the trees close enough to that point to be selected for the sample. Then the attributes of interest would be calculated. Since the probability of selecting a point from an infinite population twice is zero, this list search and calculation would have to be repeated for each random point, possibly resulting in repeated calculation of the attributes from the same cluster of trees. For simulation purposes, the optimal approach to programming will depend upon the length of the tree list to be searched, the degree of clustering in the tree population, and the number of random points.

4. CONCLUSION

We've presented a generalized forest sampling concept which utilizes a finite number of ground segments as the sample units existing within a land-area based sample frame. We have also given estimators based on this concept. The jigsaw puzzle view should be of help in understanding the similarities and differences between different methods of forest sampling by putting all of the methods into the same framework. Although we would not normally utilize the associated estimators in their given form in an actual forest survey, we can always find an equivalent calculable form. The additional benefit of an alternative route for sampling simulations is not only one of academics but also economics. Given the amount of time and money it takes to acquire data in forestry studies, the ability to easily test the properties of different sampling methods before they are applied in the field is of paramount importance. We would not endeavor to undermine the importance of a thorough theoretical development of proposed forest sampling schemes as the crucial first step, but simulation of these schemes before implementation may help uncover overlooked problems. This alternative conceptualization will, in general, facilitate comparisons within any group of forest sampling schemes.

ACKNOWLEDGEMENTS

The authors would like to thank the Associate Editor, two anonymous referees and Hans Schreuder for their helpful comments.

REFERENCES

COCHRAN, W.G. (1977). *Sampling Techniques,* (3rd Ed.). New York: John Wiley.

CUNIA, T. (1965). Continuous forest inventory, partial replacement of samples and multiple regression. *Forest Science,* 11, 480-502.

GROSENBAUGH, L.R., and STOVER, W.S. (1957). Point-sampling compared with plot-sampling in southeast Texas. *Forest Science, 3,* 2- 14.

HUSCH, B. (1955). Results of an investigation of the variable plot method of cruising. *Journal of Forestry, 53, 570-574.*

KISH, L. (1965). Survey *Sampling. New* York: John Wiley.

ODERWALD, R.G. (1981). Point and plot sampling - the relationship. *Journal of Forestry, 79, 377-378.*

PALLEY, M.N., and HORWITZ, L.G. (1961). Properties of some random and systematic point sampling estimators. *Forest Science, 7, 52-65.*

ROESCH, F.A. Jr. (1993). Adaptive cluster sampling for forest inventories. *Forest Science, 39.* In press.

ROESCH, F.A. Jr., GREEN,E.J., and SCOTT, C.T. (1989). New compatible estimators for survivor growth and ingrowth from remeasured horizontal point samples. *Forest Science, 35, 281-293.*

ROESCH, F.A. Jr., GREEN, E.J., and SCOTT C.T. (1991). Compatible basal area and number of trees estimators from remeasured horizontal point samples. *Forest* Science, 37, 136-145.

ROESCH, F.A. Jr., GREEN, E.J., and SCOTT, C.T. (1993). A test of alternative estimators for volume at time 1 from remeasured point samples. *Canadian Journal of Forest Research, 23, 598-604.*

SCHREUDER, H.T. (1970). Point sampling theory in the framework of equal-probability cluster sampling. *Forest Science, 16, 240-246.*

VAN DEUSEN, P.C., DELL, T.R., and THOMAS, C.E. (1986). Volume growth estimation from permanent horizontal points. *Forest* Science, 32, 415-422.