
Dimensionality and the Sample Unit

Francis A. Roesch¹

Abstract.—The sample unit and its implications for the Forest Service, U.S. Department of Agriculture's Forest Inventory and Analysis program are discussed in light of a generalized three-dimensional concept of continuous forest inventories. The concept views the sampled population as a spatial-temporal cube and the sample as a finite partitioning of the cube. The sample serves to cut the volume of the cube into a finite number of pieces like a three-dimensional jigsaw puzzle. Each puzzle piece is defined by the spatial-temporal selection volumes of observation sets on the individual trees existing in the forest during the period of interest. The concept is developed as a temporal extension of the alternative view of forest sampling offered in Roesch *et al.* (1993).

Introduction

Roesch *et al.* (1993) gave an alternative view of forest sampling that could be applied to all forest sampling schemes that select trees based on the location of a random point. They explain the idea as a jigsaw puzzle view of forest sampling. In this view, the sample units are the mutually exclusive sections of ground resulting from the overlapping selection areas of the individual trees in the forest. The population is the puzzle picture and it is partitioned into puzzle pieces that are mutually exclusive, exhaustive sample units that together define the sample frame. Each ground segment had a defined probability of selection proportional to its size, and the total of these probabilities over all segments is 1. In the case of point sampling, the size of each segment is determined by the basal areas and spatial distribution of the trees and the sampling angle chosen. In the case of plot sampling, the size of each segment is determined by the plot radius and the spatial distribution of the trees. Thus, all schemes can be thought of in a common

fashion, differing only in the method used to cut the puzzle into pieces. After the puzzle is segmented, the segments are selected with probability proportional to their size, and each segment contains a set of attributes. Roesch *et al.* (1993) also showed the theory applicable to remeasured samples for two specific points in time. Since the paper's publication, the Forest Service Forest Inventory and Analysis program has adopted new sampling methods that require the addition of a temporal dimension to the jigsaw puzzle view. A formal description of that extension is the focus of current research by the author. That description involves the extension of the visualization to include time as a third dimension of the puzzle. The addition of time (to the puzzle) is required when the set of observation times is randomly determined. This addition results in a sampled population and sampling frame that are three dimensional.

Theory of the Three-Dimensional Sample

Assume a three-part process: first, a random point is located in two-dimensional space; second, a set of observation times is determined; and third, at each selected time, a cluster of trees near the point is selected for measurement by some rule. The two most common temporally specific rules for selecting the clusters of trees (i.e., two dimensional) are known as (circular, fixed-area) plot sampling and (horizontal) point sampling. In this article, we will concentrate on point sampling, in which a tree i , with radius of the cross-section r_i , is selected with probability proportional to tree basal area, πr_i^2 . In considering continuous forest inventories, if a tree is selected at sample time t_i from a permanent point and it lives until sample time t_i+k , most forest sampling rules would result in it also being measured at time t_i+k . Therefore, the probability of selection of the previously measured tree at time t_i+k is actually equal to 1 and not independent of the time t_i sample. Successively applying the two-dimensional view is erroneous because new *independent* samples are not taken each time. One could argue that the population is an infinite set of points, with a point

¹ Mathematical Statistician, U.S. Department of Agriculture, Forest Service, Southern Research Station, 200 W.T. Weaver Blvd., Asheville, NC 28804-3454. E-mail: FRoesch@fs.fed.us.

being the sample unit, and therefore the same sample is merely reobserved. Alternatively, one could argue that because the set of trees occurring over the area of interest throughout the period of interest is the biological population of interest, our sampled population should be directly associated with it. This position requires knowledge of the probability of inclusion for the realized set of observations on each tree in the sample over the course of the period of interest. Because, when we also randomly sample time, potentially many sets of observations are realizable for each tree in the population, we must reconsider the sampled population and the sample unit. For example, table 1 shows the nine potential sets (excluding the null set) of observations on a single live tree over two cycles of a three-panel rotating panel design.

The sample unit is a three-dimensional puzzle piece created by partitioning the population volume with a solid of revolution for each tree, created by integrating πr_i^2 over time. The individual tree spatial-temporal volumes may be truncated on the sides when adjacent to an areal edge of the population and on the tops and bottoms by time limits of the population. We would divide the volume of a sample unit (in area x time units)

by the volume of the population to determine the probability of selection for each unit. Although time is a continuous variable, we will treat it as discrete.

The three-dimensional view removes potential confusion related to inclusion probabilities. Note that inclusion probabilities can only differ between samples, not within samples. The population in this case is three-dimensional, with time being one of the dimensions; ergo, it is obvious that the inclusion probability cannot change over time unless a new sample is taken. Any change that would appear to change an inclusion probability in the two-dimensional view can be seen to actually define a separate subpopulation or sample unit by the three-dimensional definition. Subpopulations are defined by land area and/or by time.

If time is treated as discrete with a partition length of 1 year, the sample unit appears as a set of puzzle pieces created by partitioning the volume by overlapping sets of discs, one set for each tree in each panel. The disc size varies over time with tree size, enclosing an area known as the K-circle (Grosenbaugh and Stover 1957). The K-circle of tree i at time j , K_{ij} , is an

Table 1.—The marginal and conditional probability expressions for the nine possible sets of observations of a single live tree over two cycles (6 years) of an annual rotating panel design consisting of three consecutive panels. A nondecreasing probability occurs in live round trees that do not shrink.

Year	1	2	3	4	5	6	Case
$p(Y_i)$	p_1	p_2	p_3	p_4	p_5	p_6	
Panel	1	2	3	1	2	3	
$p(P_j)$	1/3	1/3	1/3	1/3	1/3	1/3	
Set							Unrestricted probability
1	X						$p_1(1-p(Y_4 Y_1))/3$
2	X			X			$p_1(p(Y_4 Y_1))/3$
3				X			$((1-p_1)(p(Y_4 Y_1)))/3$
4		X					$p_2(1-p(Y_5 Y_2))/3$
5		X			X		$p_2(p(Y_5 Y_2))/3$
6					X		$((1-p_2)(p(Y_5 Y_2)))/3$
7			X				$p_3(1-p(Y_6 Y_3))/3$
8			X			X	$p_3(p(Y_6 Y_3))/3$
9						X	$((1-p_3)(p(Y_6 Y_3)))/3$
	$(p_1)/3$	$(p_2)/3$	$(p_3)/3$	$(p_4)/3$	$(p_5)/3$	$(p_6)/3$	
	$(p_1 + p_2 + p_3) / 3$			$(p_4 + p_5 + p_6) / 3$			
							Nondecreasing Probability
							0
							$(p_1)/3$
							$(p_4-p_1)/3$
							0
							$(p_2)/3$
							$(p_5-p_2)/3$
							0
							$(p_6)/3$
							$(p_6-p_3)/3$

imaginary circle, centered at tree center, with radius αr_{ij} . Assume that N trees are present, with labels $1, 2, \dots, N$, associated with the population of temporal length P starting in year 1. The selection area for tree i at time j , of size $a_{i,j}$ (in acres), is the portion of tree i 's K-circle that is within the population at time j and is the area from within which a random point will select the associated observations of tree attributes for the sample. Index $a_{i,j}$ annually, assuming $a_{i,j}$ remains constant for an entire year. Collect the $a_{i,j}$ into the matrix:

$$\mathbf{A} = \begin{bmatrix} a_{1,P} & a_{1,P-1} & \cdots & a_{1,1} \\ a_{2,P} & a_{2,P-1} & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{N,P} & & \vdots & a_{N,1} \end{bmatrix}$$

We wish to estimate change over a defined area (\mathbf{A}) and temporal period, and we'll assume a continuous forest inventory using a rotating panel design and leave the simplification to a single panel (or nonpaneled design) to the reader. An example of a rotating panel design is in use by the Forest Service's Forest Inventory and Analysis units (e.g., see Roesch and Reams 1999). Designs of this type consist of g mutually exclusive temporal panels. One panel per year is measured for g consecutive years, after which the panel measurement sequence reinitiates. Assume that the continuous inventory consists of n_c cycles. A cycle is a complete set of measurements on all panels, so $P = n_c g$ years. That is, if panel 1 is measured in year t , it will also be measured in years $t+g$, $t+2g$, etc., until year $t + (n_c - 1)g$. Panel 2 would then be measured in years $t+1$, $t+1+g$, $t+1+2g$, etc. Set t to 1 and reindex \mathbf{A} as follows:

$$\mathbf{A} = \begin{bmatrix} a_{1,n_c g} & a_{1,(g-1)+(n_c-1)g} & \cdots & a_{1,1} \\ a_{2,n_c g} & a_{2,(g-1)+(n_c-1)g} & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{N,n_c g} & & \vdots & a_{N,1} \end{bmatrix}$$

Without loss of generality, assume that a tree's assignment to temporal panel p is random, with probability equal to $1/g$.

Under this assumption, the unconditional probabilities of observation owing to a random point in three-dimensional spatial-temporal volume can be represented by the following matrix:

$$\mathbf{V} = \mathbf{g}^{-1} \mathbf{A}^{-1} \mathbf{A} = \begin{bmatrix} v_{1,n_c g} & v_{1,(g-1)+(n_c-1)g} & \cdots & v_{1,1} \\ v_{2,n_c g} & v_{2,(g-1)+(n_c-1)g} & & \vdots \\ \vdots & & \ddots & \vdots \\ v_{N,n_c g} & & \vdots & v_{N,1} \end{bmatrix}$$

Potentially n_c sets of observations are present on each tree for each panel. The probability of selection, by a random point in three-dimensional space, for a specific set, s , of observations (indexed by time of observation o) on tree i (π_{is}) is equal to the intersection of the selection volumes (for each time in the set) divided by the population (areal-temporal) volume, or

$$\pi_{is} = \frac{\bigcap_{o=1}^{n_s} v_{i,o \in s}}{V^T}, \text{ where } n_s \text{ is the number of observations in the}$$

set. The joint probability of selection, by a random point in three-dimensional space, for a set, s_1 , of observations (indexed by time of observation o) on tree i , and a set s_2 on tree j in panel p is

$$\pi_{is_1 j s_2} = \frac{\bigcap_{i,j} \left(\bigcap_{o=1}^{n_{s_1}} v_{i,o \in s_1}, \bigcap_{o=1}^{n_{s_2}} v_{j,o \in s_2} \right)}{V^T}.$$

When we consider all the possible intersections of all observation sets for all trees associated with the population, we have fully defined all sample units and the sample frame. That is, we have carved the spatial-temporal volume (or statue) into chunks that are selected with probability proportional to their size. Each chunk is associated with a unique set of observations on trees in the forest over time.

Assuming that tree centers do not move over time, then the intersection is equal to the smallest volume in the set,

$$\pi_{is} = \frac{\min(v_{i,o \in s})}{V^T}.$$

If we further assume the selection areas do not shrink, then the intersection is equal to the first volume in the set,

$$\pi_{is} = \frac{V_{i, \min(o) \in S}}{V^T}$$

The probability of having made any observation from panel p

on tree i is $\pi_{ip} = \frac{\sum_{s=1}^{n_s} v_{ips}}{V^T} = \max_{s=1}^{n_s} (\pi_{ips})$, where v_{ips} is the selection

volume for observation set s , of panel p for tree i , and n_s is the number of unique sets of observations on tree i in panel p .

Suppose that we randomly drop a point on the surface of a forest and use any rule to observe tree attributes over time. If the function is temporally dependent, then one must integrate area over time to determine a probability of inclusion. The probability of observing the attributes associated with tree i in year o is

$$\pi_{o(i)} = \sum_{j \in S} \pi_{o(i)j} Z_{o(i)j}, \text{ where}$$

$$Z_{o(i)j} = \begin{cases} 1 & \text{if segment } s \text{ partitions the selection volume intersecting} \\ & \text{observation year } o \text{ for tree } i \text{ and} \\ 0 & \text{otherwise} \end{cases}$$

and

$\pi_{o(i)j}$ is the probability that observation in year o for sample tree i was selected from the particular population segment s , corresponding to observation set S .

Therefore, $\pi_{o(i)j} = \left(\frac{\pi_{is}}{\pi_{o(i)}} \right)$. The sum over s of $\pi_{o(i)j}$ is 1.

Let y_{io} be the value of an attribute of interest for tree i at time o

and $Y_o = \sum_{i=1}^N y_{io}$ be the total value of interest at time o across all trees.

We can now write a temporally specific observation for each segment as a sum of weighted tree values:

$$\bar{y}_{o,s} = \sum_{i=1}^N \pi_{o(i)s} y_{io} \quad (1)$$

Now suppose that we randomly drop m points into the population volume with the same assumptions as those mentioned previously in the text and we are sampling with replacement. An unbiased estimator of the temporally specific total value of interest for a sample selected with probability proportional to size, V_s , is

$$\hat{Y}_o = \frac{V^T}{m} \sum_{s=1}^M \frac{\bar{y}_{o,s}}{V_s} = \frac{V^T}{m} \sum_{s=1}^M \frac{\bar{y}_{o,s}}{V_s} W_s \quad (2)$$

where

$V^T = \sum_{s=1}^M V_s$, the total areal-temporal volume of the population,

m = the number of sample points,

M = the number of volume segments, and

W_s = the number of times the unit s appears in the sample.

Note that W_s is an integer between 0 and m , inclusive. V_s and $\bar{y}_{o,s}$ are fixed and W_s is random. From the development above, recall that o ranges from 1 to n_g years, so it is of interest to estimate the vector $\mathbf{y} = Y_1, Y_2, \dots, Y_{n_g-1}, Y_{n_g}$ with a vector. $\hat{\mathbf{y}} = \hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_{n_g-1}, \hat{Y}_{n_g}$. Focusing on the elements of the vectors, we define

$Y_o^* = \sum_{s=1}^M \bar{y}_{o,s}$ as the total time o value of interest across all segments.

Similar to Roesch *et al.* (1993) \hat{Y}_o is easily shown to be unbiased for Y_o . By definition, the variance of \hat{Y}_o is

$$\text{Var}(\hat{Y}_o) = \left(\frac{1}{m V^T} \right) \sum_{s=1}^M V_j \left(\frac{V^T \bar{y}_{o,s}}{V_s} - Y_o \right)^2 \quad (3)$$

The sample estimate of the variance is then represented by the following equation (Cochran 1977):

$$\text{var}(\hat{Y}_o) = \frac{1}{m(m-1)} \sum_{s=1}^M \left(\frac{V^T \bar{y}_{o,s}}{V_s} - \hat{Y}_o \right)^2 \quad (4)$$

Expanding equation 2 to include the definition of $\bar{y}_{o,s}$ and subsequent rearrangement gives

$$\begin{aligned}
 \hat{Y}_o &= \frac{V^T}{m} \sum_{s=1}^M \frac{\bar{y}_{o,s}}{V_s} W_s \\
 &= \frac{V^T}{m} \sum_{s=1}^M \frac{\sum_{i=1}^N \pi_{o(i)s} y_{io}}{V_s} W_s \\
 &= \frac{V^T}{m} \sum_{s=1}^M \sum_{i=1}^N \frac{y_{io} Z_{o(i)s}}{v_{o(i)}} W_s \quad (5) \\
 &= \frac{V^T}{m} \sum_{i=1}^N \frac{y_{io}}{v_{o(i)}} \sum_{s=1}^M Z_{o(i)s} W_s \\
 &= \frac{V^T}{m} \sum_{i=1}^N \frac{y_{io}}{v_{o(i)}} w_{o(i)}
 \end{aligned}$$

where $w_{o(i)}$ equals the number of times tree i is selected for observation at time o . The final expression in equation 5 is the three-dimensional point sample estimator.

Conclusion

This article discusses a unified discrete, three-dimensional population sampling theory for continuous forest inventories. It shows that, regardless of the method used to determine

the observation points in space or time (e.g., remeasured plot sampling or point sampling, with or without temporal panels), all schemes can be thought of as carving the population volume into pieces and selecting the pieces with probability proportional to their size. The general development in equations 1 through 5 can be used for any specific type of areal-temporal forest sampling. This alternative view of continuous forest inventories may be useful in many ways. Most notably, it can contribute to database design by highlighting the minimally necessary information unit.

Literature Cited

- Cochran, W.G. 1977. Sampling techniques. 3rd ed. New York: John Wiley. 444 p.
- Grosenbaugh, L.R.; Stover, W.S. 1957. Point-sampling compared with plot-sampling in southeast Texas. Forest Science. 3: 2-14.
- Roesch, F.A.; Reams, G.A. 1999. Analytical alternatives for an annual inventory system. Journal of Forestry. 97(12): 33-37.
- Roesch, F.A., Jr.; Green, E.J.; Scott, C.T. 1993. An alternative view of forest sampling. Survey Methodology. 19(2): 199-204.