

DESCRIPTION OF SAMPLING DESIGNS USING A COMPREHENSIVE DATA STRUCTURE

John C. Byrne and Albert R. Stage¹

ABSTRACT. Maintaining permanent plot data with different sampling designs over long periods within an organization, as well as sharing such information between organizations, requires that common standards be used. A data structure for the description of the sampling design within a stand is proposed. It is based on the definition of subpopulations of trees sampled, the rules used to sample each subpopulation, and linking variables that tie these two elements together. In addition, the data structure includes the spatial relationship of plots within the stand as well as the precision with which tree characteristics are measured. These elements provide all the information necessary for the compilation of stand attributes. Changes in design through the life of a set of permanent plots are common, and the data structure is able to incorporate such changes. The data structure adequately describes an actual, complex subsampling design.

INTRODUCTION

Sharing of data on tree growth between organizations offers many opportunities for improving knowledge of forest yield capabilities. However, common definitions and standards are needed when data from diverse sources are merged. Although standards for measuring trees and describing stand conditions have been established (U.S. Department of Agriculture, 1933; Curtis, 1983), standards for describing underlying sampling designs have been ignored.

Differences in sampling designs may be due to changing objectives and personnel within and between organizations, to changing forest conditions, or to innovation in sampling techniques. Permanent plot catalogs (i.e., COSMADS--Committee on Standards of Measure and Data Sharing; Western Forestry and Conservation Association, 1977) and permanent plot data base systems, including PDMS (Plot Data Management System)², have only recently added the capability for describing designs more complex than separate fixed-area plots, such as variable-radius plots or simple cluster and concentric plot designs. Many of the permanent plots that we maintain at the Intermountain Research Station have more complex designs or have had design changes at some time in their life. To incorporate these permanent plots into a shared

¹Forester and Principal Mensurationist, USDA Forest Service, Intermountain Research Station, 1221 South Main Street, Moscow, ID 83843.

²Curtis, R. D., and G. W. Clendenen. 1981. Plot data management system (PDMS) Study Plan. Olympia, WA: U.S. Department of Agriculture, Forest Service, Forestry Sciences Laboratory.

Presented at the IUFRO Forest Growth Modelling and Prediction Conference, Minneapolis, MN, August 24-28, 1987.

data base, as well as for ease in compilation of stand attributes, we propose a data structure that can describe a wide variety of designs and changes in design.

ELEMENTS OF DESIGN DESCRIPTION

The objective of this data structure is to facilitate creation of a data set in which each "record" represents an estimate of stand parameters accompanied by treatment, site, and environmental attributes of the stand. The stand represented by the "record" is defined to be an area of ground with relatively uniform conditions that has received a unique treatment combination. In research studies, the experimental unit would correspond to our "stand." This paper will be concerned only with the design characteristics within the stand.

Whereas conventional specification of sampling designs focused on description of "the plot" or "plots" as pieces of ground, the procedure we propose focuses on description of subpopulations of trees and their spatial distribution. The key change in thinking--and in describing the sampling design--is the change from defining the sampling unit as the "plot" drawn from a population of all possible plots in the stand being sampled, to defining the sampling unit as the tree drawn from the population of all trees in the stand being sampled. This concept allows both fixed-area plots and variable probability points to be adequately and similarly described, without resorting to the use of such confusing concepts as "ongrowth" or "offgrowth," that are caused by forcing variable probability points to be described using the "plot" concept.

Designs are characterized by three elements. One element is the spatial layout. This spatial information determines how measurements taken at one place are to be used with measurements taken at a nearby place. The second element describes how different members of the population are to be sampled. For example, the subpopulation of large trees might be sampled on larger plots while small trees might be sampled on one or several smaller plots. The third element describes how tree characteristics are measured for each sampled tree.

SPATIAL RELATIONS

The spatial layout in a stand is described by the numbering scheme of the samples. For finding numbered trees during remeasurement, and for variance calculations, it is essential to know which sample units are nested within sample units used to describe another subpopulation. The fundamental unit in the numbering scheme within the stand is the "plot." A "plot" is defined as the collection of trees included within a single sampling unit. Each plot is given a number. When sampling units are nested, subordinate levels of numbering are needed, i.e., subplots, sub-subplots, etc. Our work with a variety of complex designs suggests that one level subordinate to plot (subplot) would be sufficient for most designs. All sampling units within a nest have the same plot number but different subplot numbers. Once plot and subplot numbers are created, they must never change.

Providing a unique tree identification is an important part of the structure. Each tree is uniquely identified by its plot/subplot/tree number. In repeated sampling of nested plots, trees will often grow from one subpopulation into another subpopulation. Proper compilation of per-acre values requires that a tree growing into a different subpopulation be treated as a new individual in that subpopulation, although identified by its original plot/subplot/tree numbers. A tree can never occur in two or more subpopulations at the same time and can occur only once in each subpopulation.

SAMPLING SUBPOPULATIONS

The totality of trees to be included in stand statistics may be divided into subpopulations. In our data structure each subpopulation must be mutually exclusive and, in the aggregate, encompass the whole population. Each subpopulation is to be sampled with one or more simple sampling procedures. A sampling procedure can be described by the number of samples and the rule that establishes the sampling probability for each individual tree (i.e., plot area for fixed-area plots, basal area factor (BAF) for variable-radii points, etc.). In some cases the stand and plot are synonymous so that all trees in the plot are tallied (and therefore have sampling probability of unity). But in other cases, the stand is too large for complete enumeration and is then sampled with some combination of fixed-area plots, variable-radii points, etc.

In defining each subpopulation, one must record the tree characteristics that are used in delimiting the subpopulation. The most common way of defining tree subpopulations in permanent sample plot work has been by diameter at breast height (DBH) limits. In some studies, the presence or absence of a certain tree characteristic--i.e., whether the tree is alive or dead, the presence of a particular disease organism, etc.--may define the subpopulations. Once the delimiting variable is recorded, the delimiting values of the variable must be entered. For DBH and other continuous variables, the delimiters are the minimum and maximum levels. For discrete variables, codes for specific characteristics are used in the subpopulation definitions. Though in most cases only one variable is used as a delimiter, the data structure is capable of recording several delimiting variables and their chosen values.

The type of sampling rule used to select the trees in each subpopulation defines the probability with which a tree is sampled. For fixed-area samples, the probability of a tree being sampled depends only on its presence or absence. All trees in the defined area that meet the subpopulation definitions are measured, so the sampling probability is proportional to frequency. With geometric point and line sampling, tree size characteristics are used to define the probability of tree selection.

Corresponding to each sampling scheme is an expansion constant used in converting the counts tallied to per-acre stand attributes. For fixed-area plots, the expansion constant is the reciprocal of plot area while for variable-radius points the expansion constant is BAF.

Several examples will help clarify how this expansion constant is used. To calculate trees per acre for a fixed-area plot, the expansion constant is multiplied by the number of trees counted on the plot. But for variable-radius points, the expansion constant is used to calculate a conversion factor for each tree DBH, that is, $BAF/(BA \text{ for the tree})$. Then this conversion factor is multiplied by the number of trees counted with that DBH and summed for all DBH classes to get trees per acre.

SUBSAMPLING TREE CHARACTERISTICS

The third element describes the rules specifying which measurements are to be taken on a particular sampled tree. The method we use is to enter values for each tree, however estimated, for each attribute, along with an indicator of their reliability. This indicator, besides noting whether trees were directly measured or estimated, can also describe the general precision associated with each directly measured attribute. The precision will have an impact on how much trust one has in the stand attributes calculated using those measurements. For example, whether heights were measured using a clinometer with ± 1 foot precision or using ocular estimation with ± 10 foot precision would have an impact on the inferences made from stand attributes based on heights.

LINKING SUBPOPULATIONS TO SAMPLING RULES

To specify compilation procedures to be used at a particular date, it is necessary to link the subpopulation being sampled, the sample rule used, and the date when sampling began to each plot/subplot identifier. Each plot/subplot description includes the total design description of which it is a part. Then, a linking variable in the plot/subplot description references the specific subpopulation that is sampled, the sampling rule, and the date the design was initiated for this plot/subplot. Whenever design changes are made, such as how a subpopulation is sampled, a new set of these linking variables must be added to the description so that analysis procedures can be changed accordingly.

VARIABLES USED IN DESIGN DESCRIPTION

The variables used to describe spatial relations are:

- Plot number
- Subplot number

The variables used to define subpopulations are:

- Subpopulation number
- Variables for delimiting a subpopulation
 - Minimum value or first code for each delimiting variable
 - Maximum value or second code for each delimiting variable

The variables used to describe sampling rules are:

- Sampling rule number
- Variable for defining sampling probability
- Expansion constant
- Number of samples

The linking variables are:

- Subpopulation sampled
- Sampling rule used
- Date design initiated

ILLUSTRATION OF THE DATA STRUCTURE

To demonstrate the capability of the data structure, we will describe an actual sampling design with the design variables. The design we will use for the demonstration is the one used in the Managed Stand Survey in Region 6 of the National Forest Systems (U.S. Department of Agriculture, 1987). A narrative description of this design is: A cluster of five concentric plots is located within a similarly treated area. Each of the concentric plots consist of two plots each: a 1/100-acre plot for counts of live trees <1.0 inches DBH, and a 1/20-acre plot for measurement of live and dead trees ≥ 1.0 inches DBH. Dead trees are measured as an indicator of wildlife habitat. There is an additional area around the second 1/20-acre plot to record additional live trees with a size of ≥ 9.0 inches DBH. Fewer characteristics are measured on the trees in this area than on those similarly sized trees in the 1/20-acre plot. The area of this plot is 3/20 acre, the "donut-shaped" area between a 1/5-acre plot and a 1/20-acre circular plot. Figure 1 is a schematic drawing of this design for one stand.

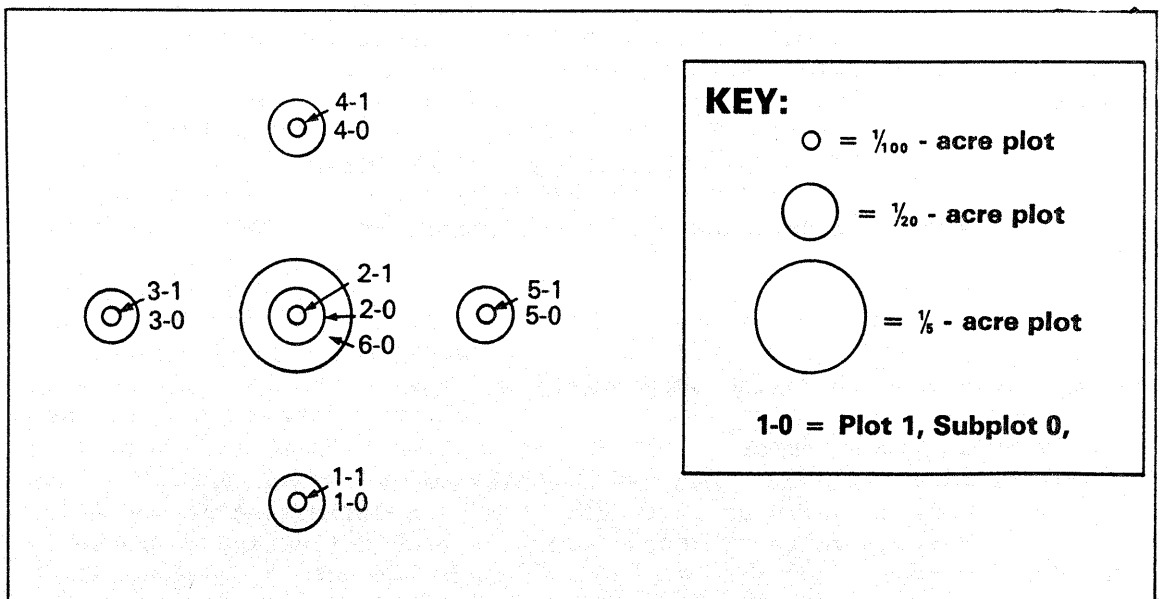


Figure 1. Schematic drawing of example design.

There are four separate subpopulations represented in this design. They are defined in Table 1.

TABLE 1. Definition of subpopulations for example design.

Subpopulation Number	1	2	3	4
1st variable for				
subpopulation definition	DBH	DBH	DBH	DBH
Minimum or 1st code	9.0"	9.0"	1.0"	0.0"
Maximum or 2nd code	99.9"	99.9"	8.9"	0.9"
2nd variable for				
subpopulation definition		TC	TC	
Minimum or 1st code		2	0	
Maximum or 2nd code			2	

TC = Tree class, where 0 represents live trees and
2 represents dead trees.
Tree class codes used as subpopulation defining
variables only when codes other than live are used.

Three sampling rules are represented in the design and are described in Table 2.

TABLE 2. Definition of sampling rules for example design.

Sampling rule number	1	2	3
Variable defining probability	FQ	FQ	FQ
Expansion constant	20	6.667	100
Number of samples	5	1	5

FQ = Frequency

And finally, the numbering scheme for the plots and subplots within the design is shown. A linking variable, included with the numbering, links the numbering scheme to the sampling design description (i.e., 2-1-1987 means subpopulation 2 is sampled using sampling rule 1 beginning in 1987).

TABLE 3. Numbering scheme and associated linking variables for example design.

Plot number	Subplot number	Linking variables
1	0	1-1-1987, 2-1-1987, 3-1-1987
1	1	4-3-1987
2	0	1-1-1987, 2-1-1987, 3-1-1987
2	1	4-3-1987
3	0	1-1-1987, 2-1-1987, 3-1-1987
3	1	4-3-1987
4	0	1-1-1987, 2-1-1987, 3-1-1987
4	1	4-3-1987
5	0	1-1-1987, 2-1-1987, 3-1-1987
5	1	4-3-1987
6	0	1-2-1987

In compiling stand attributes using this design, several things must be considered. First, because the stand is subsampled it is possible to calculate a variance estimate for each attribute of interest, but the covariance between the plots in each nest must be included in the overall variance estimate. And second, instead of calculating the contribution of each plot to the stand attribute estimate, it may be better to calculate the contribution of each subpopulation because trees in the nested plots will move from one subpopulation to another.

LITERATURE CITED

- Curtis, R. O. 1983. Procedures for establishing and maintaining permanent plots for silvicultural and yield research. United States Department of Agriculture, Forest Service, General Technical Report PNW-155, 56 p.
- United States Department of Agriculture, Forest Service. 1933. Sample plots in silvicultural research. Circular 333. 88 p.
- United States Department of Agriculture, Forest Service. 1987. Field instructions for managed stand surveys, Region 6. Forest Inventory Section, Division of Timber Management, Region 6, Portland, Oregon. 178 p.
- Western Forestry and Conservation Association. 1977. A set of recommended parameters and measurement instructions for field records of permanent growth plots. In: Reimer, D. R., chairman. Standards of measure and data sharing: report of the Committee of Standards on Measure and Data Sharing (COSMADS). Portland, Oregon. 7 p.