United States
Department of
Agriculture

Forest Service

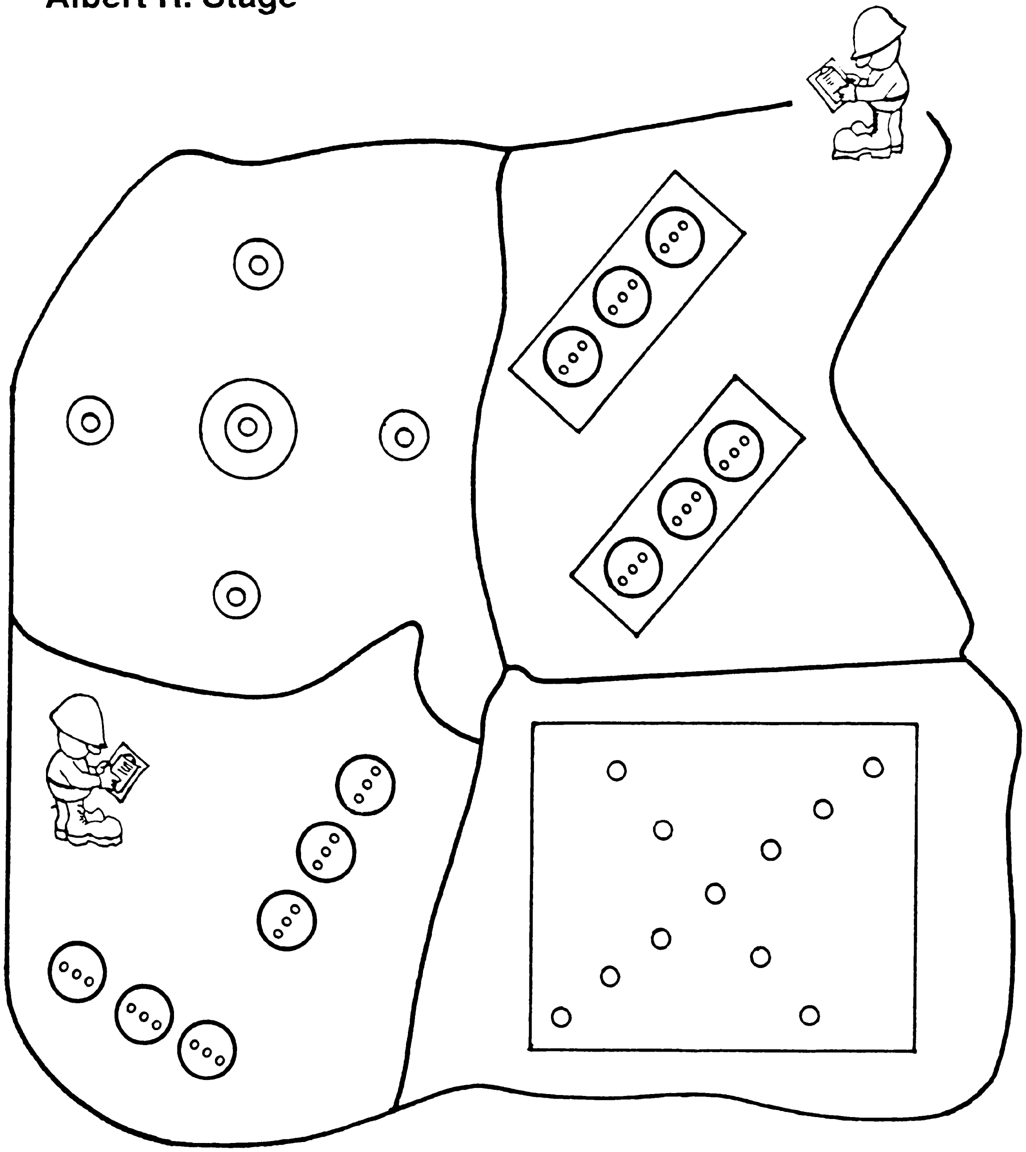
**Intermountain
Research Station**

General Technical
Report INT-247

June 1988

A Data Structure for Describing Sampling Designs To Aid in Compilation of Stand Attributes

**John C. Byrne
Albert R. Stage**



THE AUTHORS

JOHN C. BYRNE is a forester with the Intermountain Research Station stationed in Moscow, ID.

ALBERT R. STAGE is a project leader and principal mensurationist with the Intermountain Research Station stationed in Moscow, ID.

RESEARCH SUMMARY

Maintaining permanent plot data with different sampling designs over long periods within an organization, and sharing such information between organizations, requires that common standards be used. Standards for measuring trees and describing stand conditions have been addressed, but procedures for describing underlying sampling designs within a stand have been ignored. We propose a data structure composed of just those variables and their relationships needed to compile stand attributes and their variances. Elements of this data structure include: (1) the spatial relationship of plots within the stand; (2) the definition of the subpopulations of trees sampled, the rules used to sample each subpopulation, and the linking variables that tie the subpopulation definitions and rules together; and (3) the accuracy with which the measurements on a sampled tree are taken. The data structure is able to incorporate changes in design that commonly occur through the life of a set of permanent plots. Using seven actual or proposed designs, from simple to complex, the data structure is compared to one of the better systems now available for managing permanent plot data. Whereas that system is inadequate in describing some of the more complex designs and any changes in design, the proposed data structure can fully describe all of the designs. We include applications of the data structure for designing of sampling schemes and for cataloging data.

CONTENTS

Intermountain Research Station
324 25th Street
Ogden, UT 84401

A Data Structure for Describing Sampling Designs To Aid in Compilation of Stand Attributes

John C. Byrne
Albert R. Stage

INTRODUCTION

Sharing of data on tree growth between organizations offers many opportunities for improving knowledge of forest yield capabilities. However, merging data from diverse sources requires that common definitions and standards be used. While standards for measuring trees and for describing stand conditions have received much attention (Curtis 1983; USDA Forest Service 1933), standardized procedures for describing variations in the underlying sampling designs within a stand have been ignored. As a consequence, opportunities for strengthening data bases or reducing costs have been forgone. Designs other than fixed-area plots have often been rejected as too inconvenient to process. Furthermore, the data needed for calculation of sampling variances have not been provided. Therefore, the weight to give to a particular stand in a merged analysis has not been available.

Differences in sampling design may be due to changing objectives and personnel within and between organizations, to changing forest conditions, or to innovation in sampling techniques. Designs are often made efficient for describing the stand at the time of installation. But at the time of remeasurement both organizational objectives and stand conditions will be different. So, a design that is appropriate at one stage of development may be inefficient for subsequent remeasurement. For example, large, fixed-area plots initially installed in stands close to rotation age have proven to be cumbersome when, through insects, disease, or regeneration treatments, the older stand is gradually replaced by regeneration. Some of these variations in sampling procedures in remeasured plot data may render the data worthless, while others complicate compilation of the data. Many of these problems could be avoided by appropriate modifications of the initial design.

We propose a data structure for describing designs that facilitates use of data from a wide variety of designs. This data structure consists of just those variables and their relationships needed to compile stand attributes and their variances. Our criterion is that any design that provides a valid basis for inference should be describable within this data structure. Conversely, mensurationists may be able to avoid some of the more troublesome pitfalls by considering how a proposed measurement procedure or a change in design would be represented in this data structure. This system is intended to apply to both routine inventory and silvicultural research data.

The data structure consists of variables that describe spatial relations, definition and sampling of subpopulations, and procedures for subsampling of tree attributes. This data structure does not define estimation techniques that might be used for statistical inference.

Compilation procedures based on the proposed data structure should be able to process data from a wider variety of designs than can be processed with present protocols. (An outline for such a procedure is available from the authors upon request. Write to Forestry Sciences Laboratory, Intermountain Research Station, Forest Service, U.S. Department of Agriculture, 1221 South Main St., Moscow, ID 83843.)

Finally, to illustrate our proposal, we apply it to seven existing or proposed designs.

EXISTING DATA STRUCTURES

The Weyerhaeuser Company and the Pacific Northwest Forest and Range Experiment Station undertook a notable effort to organize a common data base for Douglas-fir stand yield tables. The Western Forestry and Conservation Association sought to extend this initial effort and organized a committee (Committee on Standards of Measure and Data Sharing, COSMADS) to carry out the work. This committee set standards for permanent sample plots and cataloged the existing permanent plots in the Pacific Northwest (Arney and Curtis 1977; Western Forestry and Conservation Association 1977). At this time the COSMADS plot catalog is maintained by the British Columbia Forest Service.

In the early 1980's, Curtis and Clendenen (1981) began to coordinate a system of computer programs, initially written for the Douglas-fir stand yield work, for use in maintaining and accessing permanent-plot data bases. The Plot Data Management System (PDMS) is the outgrowth of that effort. In PDMS, one set of records describes plot attributes, including plot design. An associated set of records describes individual tree attributes, indexed by a plot identification variable.

Both PDMS and COSMADS were originally designed to deal with separate fixed-area plots only. Of the large number of permanent plots that we maintain at the Intermountain Research Station, only a small percentage can be classified as separate fixed-area plots. Many of the plots are variable-radius plots or have subsamples that are used to sample different parts of a population, or the plot design has changed over time. A recent update of PDMS describes variable-radius plots and two subsampling schemes (cluster layouts and concentric plots). And recently, COSMADS has added the capability of describing variable-radius plots.

ELEMENTS OF A 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 variables 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. Stand variables used for summarization include volume per acre, basal area per acre, increment in volume, change in top height, and so forth. Even if the data are to be used to represent growth relations for individual trees, stand variables are still required to define density or other variables influencing growth. When the stand is inventoried with multistage sampling within the stand, variances can be calculated to characterize internal variability of the stand and to estimate precision of the stand variables.

Designs are characterized by three elements. One, the spatial layout, determines how measurements taken at one place are related or are to be used with measurements taken nearby. The second element describes the emphasis given to sampling different members of the population. For example, the subpopulations of large trees or scarce species might be sampled on larger plots, while small trees or common species might be sampled on one or several smaller plots. The required information is the sampling probability associated with each tree in the list. The third element describes how tree characteristics are measured for each sampled tree. For example, were tree heights measured with a clinometer or were they estimated from a diameter at breast height-height regression?

The advent of geometric point and line sampling and of sampling with arbitrary probability has made conventional ways of describing sampling designs inadequate if not totally inappropriate. Whereas conventional specification of inventory design focused on description of "the plot" or "plots" as pieces of ground, the procedures we propose focus 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 area being sampled to defining the sampling unit as a group of trees, drawn from the population of all trees in the area being sampled. In this data structure, the role of the "plot," if used at all, is to define the clustering of the sample trees. In this perspective, we follow the concepts used by O'Regan and Palley (1965).

For fixed-area plots, either conceptual framework is adequate. However, with variable-probability sampling designs such as point sampling, plot-oriented definitions lead to such concepts as "ongrowth," "nongrowth," and their logical corollary, "offgrowth." These concepts have been used as though they are characteristics of the trees when in fact they are artifacts of the sampling design. They are unnecessary in the system we propose. Their role in compilation has been supplanted by the variables in the proposed data structure.

Spatial Relations

The concept of a "stand" as a uniquely treated, biologically uniform area is a key reference point in the hierarchy of spatial relations that extends from an individual tree upward to the complex mosaic of site and vegetation that comprise a forest. Levels of the spatial hierarchy within the "stand" then are used to describe the internal spatial variability of the stand. Internal variability, whether described by a stem map or by variation between plots, may be useful for purposes of explaining stand growth, for localizing effects on individual trees, and for calculating variances of stand variables. In research studies, the experimental unit would correspond to our "stand" and represents a single degree of freedom. Then, spatial relations between stands describe the patterns of replication and blocking that are usually thought of as describing an experimental design. Even though this paper will be concerned only with levels of spatial hierarchy within the "stand," we recognize that an installation numbering system is required to describe the relations between stands in an experimental design or to identify stands to an administrative unit in an inventory.

Sampling Subpopulations

Contribution to stand totals from various subpopulations is an important stand attribute. Subpopulations of particular interest are often defined by diameter, species, or age class.

Varying Emphasis—When some subpopulations are more interesting than others, sampling procedures may be optimized by varying the sampling probabilities among the subpopulations. The distribution of sampling probabilities is controlled by two factors of the design: the number of sampling points within the stand and the rules that determine which trees are measured at each point. If the stand and plot are synonymous (except for the necessary buffer strips) and all trees within the plot are tallied, then all sampling probabilities are equal to unity. However, if the stand is too large for complete enumeration, it may be sampled at a number of points at which trees are included with constant probability as with fixed-area plots, or with probability proportional to basal area as with variable-radii plots, or proportional to arbitrary attributes of the trees as in 3-P sampling (Grosenbaugh 1958).

When inclusion of a tree in the sample is defined by the geometrical relations between tree location and its dimensions—as it is with geometrical point and line samples using d.b.h. or height in comparison to an angle gauge—then sampling probabilities are defined implicitly for all trees by the geometry of the angle gauge (Grosenbaugh 1958). However, for sampling with arbitrary probabilities, the sampling probability is required for every tree in the population (Grosenbaugh 1963).

Finally, the sampling probabilities for a tree in the tally at a sample point may be a function of both tree and plot geometry and of categorical attributes of the tree such as species, whether live or dead, or sound or cull.

Implications for Estimating Components of Growth—Growth analysis often requires compilation of change in the **population** between two examinations, composed of ingrowth, accretion, and mortality (Gilbert 1954). Procedures for estimating change between remeasurements of permanent plots have been a long-standing area of confusion. Most of the confusion surrounding growth calculations for variable-sized plots arises from: (1) the possibility of estimating the accretion component based on either the starting or ending sampling probabilities; and (2) the attempt to constrain sample-based estimates to have the same relations between growth components as would be obtained from complete enumeration of a finite population (Martin 1982). When changes in design are considered, the confusion is even worse (Chapman 1985). Each component depends on the period to which it applies and on the population (or subpopulation) to which it applies.

Of the three components, ingrowth is unique in that estimates can be based only on the end-of-period sampling probabilities. Accretion and mortality can be estimated using either start-of-period sampling probabilities or end-of-period probabilities. Indeed, the equality of sampling probabilities from period to period is one of the attractions of fixed-area plots for which sampling probabilities are constant.

Subsampling Tree Characteristics

This third element of design describes the rules that specify the measurements to be taken on a particular sampled tree. We had two alternatives. In one, the data structure might specify how an attribute is to be estimated from the subsample. For example, one could specify a regression estimator with its coefficients, or specify a probability-based estimator with its attendant subsampling probabilities.

The other alternative, which we chose, is simply to enter values for each tree, however estimated, for each attribute along with an indicator of the attribute's reliability. This indicator, besides noting whether trees were directly measured or estimated, can also describe the general accuracy associated with each directly measured attribute. The accuracy will influence the trust one has in the stand attributes calculated using those measurements. For example, whether heights were measured using a clinometer with ±1-foot accuracy or using ocular estimation with ±10-foot accuracy would have an impact on the inferences made from stand attributes based on heights. Also, rules could be established to relate measurements taken with less accuracy to ones with more accuracy from the same stand. If crown ratios were measured directly on some trees with a clinometer and ocularly estimated on the rest of the trees, a rule could be derived to adjust the ocular estimates based on the more accurately measured crown ratios.

PROPOSED DATA STRUCTURE FOR DESCRIBING SAMPLING DESIGNS

In this section, we will define the pertinent variables that are necessary in detailing each of the three design elements.

Spatial Relations

The first design element, the spatial relationship of the plots in a stand, is described by the numbering scheme of the plots. The important consideration in recording spatial relations is describing the proximity of plots used to sample the various subpopulations. When one plot is nested within another (a subplot), covariance between stand attributes for the subpopulations must be included in the overall variance estimate of the stand when computing variance of the estimates for the total population. For variance calculations, and for finding numbered trees during remeasurement, it is essential to know which samples are within related samples used to describe another subpopulation.

The plot numbering scheme also serves a vital role in data management. The fundamental unit in the numbering scheme within the stand is the "plot," which is defined as the collection of trees included within a single sampling unit. Each plot is given a number that must be a part of the unique identifier for each individual tree record such that the tree records always can be associated with records describing the plot. When sampling units are nested, subordinate levels of numbering are needed: subplots, sub-subplots, and so forth. Our work with a variety of complex designs suggests that one level subordinate to plot (subplot) is sufficient for most designs. With more complex plot designs, additional levels of subordination could use columns of the subplot field. All sampling units within a nest have the same plot number but different subplot numbers. Subplot number 0 is assigned to the sample that is not internal to any other plot. Once plot and subplot numbers are created, they must never change. Several sampling designs that are described in a later section of this paper include the numbering scheme.

Using the numbering scheme to describe spatial relations only indicates plots within other plots, which is sufficient for most purposes. If it is necessary to know more about the proximity of plots, more information would be needed. Indicating coordinates (x, y) for each plot within a stand would provide a means for determining the distance to other plots such that spatial autocorrelations could be represented.

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 the same stand, trees will often grow from one subpopulation into another. Proper compilations of per-acre values require 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 only occur once in each subpopulation.

The type of sampling design may have implications on the tree numbering. When sampling is done without replacement, no one tree will occur in more than one sample. But when sampling is done with replacement, as with a cluster of variable-radius plots in a stand, a tree may be included in more than one plot. Points too closely clustered, or a basal area factor (BAF) that is too small, will increase the likelihood of this undesirable situation happening. However, when a tree does occur in two samples as described above, the tree should be treated as two trees with different numbers.

The variables used to describe spatial relations are:

Plot number = An integer number P_i that identifies a cluster of sample trees; $i = 1, \dots n$ where n is the total number of plots in the stand.

Subplot number = An integer number S_j that identifies a cluster of sample trees that is nested within plot P_i; $j = 0, 1, \dots m_i$ where m_i is the total number of subplots within plot P_i. Subplot number 0 indicates a set of trees that is within plot P_i and not within any subplot.

Sampling Subpopulations

The totality of trees to be included in stand statistics may be divided into subpopulations. The division may be motivated by the intrinsic interest in the subpopulation or by considerations of sampling efficiency. In the data structure we are specifying, each subpopulation must be mutually exclusive and, in the aggregate, encompass the whole population. Furthermore, each subpopulation is to be sampled with one or more simple sampling rules. By simple, we mean that the sampling procedure can be described by the number of samples and the rule that establishes the sampling probability for each individual tree (that is, plot area for fixed-area plots, BAF for variable-radii plots, and so forth). Subpopulations sampled with more than one sampling rule also require that each tree be identified with the rule that brought it into the sample. Because the defined subpopulations are associated with particular sampling rules, the definition must be based on attributes as they exist at the time of sampling.

In defining each subpopulation one must record the tree characteristics 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 (d.b.h.) limits. For vertical point sampling schemes, height is the delimiting variable. In some studies the presence or absence of a certain tree characteristic—for example, presence of a particular disease organism or whether the tree is alive or dead—may define the subpopulations. Once the delimiting variable is recorded, the delimiting values of the variable must be entered. For d.b.h. or height limits, the delimiters are the minimum and maximum level. For discrete attributes, codes for the specific characteristics are used in the subpopulation definitions. Though in most cases only one variable is used as a delimiter, the system should be capable of recording several delimiting variables and their appropriate values. An example of using more than one variable to define a subpopulation would be when a "plot" is used to measure both live and dead trees greater than 9.5 inches d.b.h., instead of the assumed case of only live trees greater than 9.5 inches. Usually, tree death is recorded and then measurement is stopped on the tree, but in this example, continuous measurement of dead trees is necessary for quantifying possible wildlife habitat. Both d.b.h. and a tree code (for live and dead) would be necessary to define the subpopulation.

The variables used to define subpopulations are:

Subpopulation number = An integer number K that identifies a subpopulation of trees in a stand; $K = 1, \dots p$, where p is the total number of subpopulations that make up the population of trees in the stand.

Variable for delimiting a subpopulation = A tree characteristic used in delimiting the trees that belong in subpopulation K. Designated as V, V=1,... t, where t is the total number of variables used in delimiting subpopulation K. The tree characteristic can be either continuous (d.b.h., height, and so forth) or discrete (tree class code, tree condition code, species code, and so forth) variables. Possible codes for commonly used variables are:

D = d.b.h. limits
H = height limits
T = tree class codes (alive, dead, and so on)
C = tree condition codes (insect and disease damage and so on)
S = species codes.

Minimum value or first code for variable V = A delimiting value for variable V, used in defining subpopulation K. For continuous variables, the minimum value (that is, a minimum d.b.h. limit) and for discrete variables an inclusive value (a specific code, such as tree class codes where 0 = live trees and 2 = dead trees).

Maximum value or second code for variable V = A delimiting value for variable V, used in defining subpopulation K. For continuous variables, the maximum value and for discrete variables, an additional inclusive value (if necessary).

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 (whether plots or strips), 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 constant. With geometric point and line sampling, tree size characteristics are used to define the probability of tree selection. With horizontal point sampling (that is, variable-radius plots), basal area is used as the tree characteristic, while with horizontal line sampling, tree d.b.h. is used. With vertical point and line sampling, height is used as the tree size characteristic for determining whether a certain tree should be included in a sample. See Grosenbaugh (1958) and Oderwald (1981) for a more detailed explanation of the various point and line sampling schemes.

Corresponding to each sampling scheme is an expansion constant used in converting the counts tallied to per-acre stand attributes. For a fixed-area plot, the expansion constant is the reciprocal of plot area. Multiplying the number of trees counted on a plot by this expansion constant results in trees per acre. For horizontal point samples, the expansion constant is BAF. To calculate trees per acre for horizontal point samples, BAF is used to calculate a conversion factor for each tree d.b.h., that is, BAF/(BA for the tree). Then this conversion factor is multiplied by the number of trees counted with that d.b.h. and summed for all d.b.h. classes to get trees per acre. Likewise, for the other point and line sampling schemes, there are appropriate expansion constants (Grosenbaugh 1958).

When the sampling probabilities cannot be calculated from plot and tree factors, as in 3-P or other arbitrary sampling probability schemes, then an explicit variable defining the sampling probability must be provided for each tree.

The variables used to describe sampling rules are:

Sampling rule number = An integer number R that identifies the set of variables (that is, variable for defining sampling probability, expansion factor, and number of samples) that form a sampling rule; $R = 1, \ldots q$ where q is the total number of sampling rules used through the life of the sampling design.

Variable for defining sampling probability = The tree characteristic used in determining the probability of a tree being sampled. Possible codes for commonly used variables are:

CON = constant, for fixed-area plots or strip plots
BA = basal area, for a horizontal point sample
DBH = diameter at breast height, for horizontal line sample
HTS = square of height, for a vertical point sample
HT = height, for a vertical line sample
ARB = subjective definition of sampling probabilities (3-P).

Expansion constant = The inverse of the sampling probability used in converting counts tallied to a per-unit area basis. The expansion constants for the variables listed above are:

For CON, expansion constant = 1/plot area or 1/strip area
For BA, expansion constant = BAF
For DBH, expansion constant = horizontal line factor (HLF)/line length
For HTS, expansion constant = vertical point factor (VPF)
For HT, expansion constant = vertical line factor (VLF)/ line length
For ARB, sum of variable proportional to probabilities.

Number of samples = Total number of samples used with sampling rule R.

Knowing the stand area is often important in designs where the stand area is represented only by a sample. The relation between sampled area and stand area allows the finite population correction factor to be included in the sampling error. Although traditional experimental studies have a 100 percent sample of the population in each stand area except for a buffer strip, in operational studies, a 100 percent sample of such a population is impractical.

Subsampling Tree Characteristics

Associated with each tree characteristic measurement in the tree records, there should be a field that designates whether or not that characteristic was directly observed, and if directly observed the accuracy of that measurement. This could easily be done with a simple set of codes for each tree characteristic. Possible codes for each of three commonly measured tree characteristics are listed below.

D.b.h.

0 = not directly observed
2 = ±0.1 inch (D-tape or caliper)
5 = ±1 inch (Biltmore stick)
7 = ±2 inches (ocular estimate to 2-inch d.b.h. class)

Height

0 = not directly observed
2 = ±0.1 foot (dendrometer)
4 = ±0.5 foot (telescoping height pole)
6 = ±1 foot (clinometer, Abney level, Haga altimeter)
9 = ±10 feet (ocular estimation)

Crown ratio

0 = not directly observed
2 = ±1 percent (measured heights with clinometer, and so forth)
6 = ±5 percent (ocular estimate)

Codes for other measurement accuracies could easily be added for any of the tree characteristics.

LINKING SUBPOPULATIONS TO SAMPLING RULES

To completely specify compilation procedures to be used for a particular examination 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. The date indicates when changes in plot design, such as how a subpopulation is sampled, have occurred so that analysis procedures can be changed accordingly. Each plot and subplot description includes the total design description of which it is a part. Then, a field in the plot and subplot description references the specific subpopulation that is sampled, the sampling rule, and the date the design was initiated for this plot and subplot. Whenever design changes are made, a new set of these linking variables must be added to the description in order to facilitate compilation procedures. These linking variables are:

Subpopulation sampled = Subpopulation number K sampled at plot P_i and subplot S_j.

Sampling rule = The sampling rule R used to sample subpopulation K at plot P_i and subplot S_j.

Date design initiated = The calendar year in which subpopulation K was initially sampled with sampling rule R at plot P_i and subplot S_j.

DESIGN CHANGES AND THEIR IMPLICATIONS

Design changes can be the result of changes in subpopulation definitions or of changes in the sampling rule applicable to a subpopulation. When subpopulation definitions or sampling rules or both change, new plot and subplot identifiers may be needed. These identifiers must be added to each tree record.

Because trees can "move" from one subpopulation to another in successive periods, it is possible for a tree to "leave the sample." However, to calculate accretion, it is necessary to record the tree attributes at subsequent examinations even though the tree would not be included by the rules in effect at that later period. Trees "leaving the sample" because of design changes that create an additional level of subplots are assigned a zero subplot number at that level to indicate that they are no longer included by the current sampling rules. Designs that permit trees to leave the sample should be diligently avoided. However, the procedures we have described would permit valid compilations, although the field work and the compilation program are rendered more cumbersome.

The following example is used to describe how to handle design changes. The design (design 3 in the examples) has one subpopulation that existed from 1933 through 1974. All trees ≥1.0 inches d.b.h. were sampled on a square 0.25-acre plot. During this time all trees were in plot 26, subplot 0. Due to a change in the composition of the stand from one of larger trees to one with a thick stand of smaller trees, the sampling scheme for smaller trees was changed at the 1984 remeasurement. Two new subpopulations were created. Subpopulation 2, consisting of trees ≥3 inches d.b.h., is sampled with the entire 0.25-acre plot. Subpopulation 3, consisting of trees 1.0 to 3 inches d.b.h., is sampled with 11 plots of $^1/_{300}$ acre each within the original 0.25-acre plot. Now the numbering scheme changes to plot 26, subplot 0 (≥3 inches d.b.h.) and plot 26, subplots 1 to 11 (1.0- to 3-inch d.b.h.). From 1984 on, plot 26, subplot 0 is used to sample subpopulation 2 (this is indicated with the linking variables previously described). All trees in subplot 0 before 1984 remain in subplot 0 after 1984 except for those 1.0- to 3-inch trees that are within one of the 11 plots of $^1/_{300}$ acre each. Those trees have their subplot numbers changed to indicate in which subplot they now reside. Those trees that are 1.0- to 3-inch d.b.h., were tallied previous to 1984, and were not in one of the $^1/_{300}$-acre plots in 1984 are retained in subplot 0 for calculation of accretion. But if stand attributes were compiled in 1984, these trees would not be used in the compilations. Only when they reach a size ≥3 inches will they be included in stand calculations again.

ILLUSTRATION OF THE PROPOSED DATA STRUCTURE

Seven sampling designs will be used to demonstrate the capabilities of the proposed data structure. Designs 2, 3, 4, and 7 are actual or proposed designs in use at the Intermountain Research Station. Design 1 is used in a research project at the University of Montana (Sweet 1987) and was provided through cooperative work in compiling a permanent plot data base for the Inland Northwest Growth and Yield Cooperative (INGY). Design 5 is the Managed Stand Survey in the Pacific Northwest Region of the National Forest Systems (USDA Forest Service 1986b). Design 6 is used by the Northern Region of the Forest Service for its growth-monitoring plots (USDA Forest Service 1986a)

These seven designs illustrate a progression from simple to complex. Design 1 is a simple fixed-area plot. Design 2 illustrates a change in the population sampled on a fixed-area plot, while design 3 shows a change in the sampling rule for part of the population. Design 4 documents the sampling of a stand (that is, clustering), with one level of subsampling within each plot in the stand, while design 5 shows additional complexity in a stand subsampling design. Design 6 illustrates multiple subsamples within each plot of a cluster design, and design 7 illustrates larger plots added to design 6 after its establishment.

The description of each sampling design consists of several parts. First is a narrative description, along with a sketch of the design (sketches are not included for the simpler designs). The symbols used in the sketches are shown in figure 1. Next, the design is described using the proposed data structure. Finally, the numbering scheme for the plots and subplots within the design is shown. A linking variable is included with the numbering. This linking variable is not part of the unique identifier for each sample (as plot and subplot are), but instead links the numbering scheme to the sampling design description (example: 2-1-1984 means subpopulation 2 is sampled using sampling rule 1 beginning in 1984).

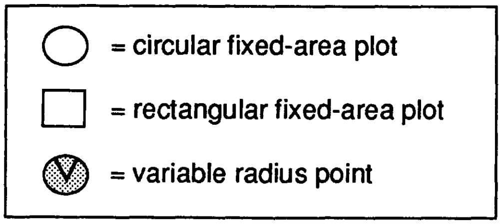


Figure 1—Symbols used in design sketches.

DESIGN 1

A study with an experimental design of two blocks with four treatments within each block and two replications of each treatment per block. Each treated plot of the same size, a square 0.2-acre plot. All trees ≥0.5 inch d.b.h. are measured within a plot. (Swan Valley Ponderosa Pine Study.) In the descriptions, only the description for one plot is given. Because each plot consists of a distinct treatment, each plot is treated individually.

Description using proposed data structure:

Subpopulation number	1
1st var. for subpop. definition	D
Minimum or 1st code	0.5
Maximum or 2nd code	99.9
2nd var. for subpop. definition	
Minimum or 1st code	
Maximum or 2nd code	

\- -

Sampling rule number	1
Variable defining probability	CON
Expansion constant	5
Number of samples	1

Numbering scheme:	**Plot #**	**Subplot #**	**Linking variable**
	1	0	1-1-1984

DESIGN 2

A 2.0-acre square plot where the diameter limit of trees measured changes after one measurement. In 1922, all trees ≥3.6 inches d.b.h. were measured on the plot, but in measurements from 1927 to 1952 all trees ≥0.6 inch d.b.h. were measured on the plot. (Plot 1 of the Intermountain Research Station Permanent Sample Plot Collection.)

Description using proposed data structure:

Subpopulation number	1	2
1st var. for subpop. definition	D	D
Minimum or 1st code	3.6	0.6
Maximum or 2nd code	99.9	99.9
2nd var. for subpop. definition		
Minimum or 1st code		
Maximum or 2nd code		

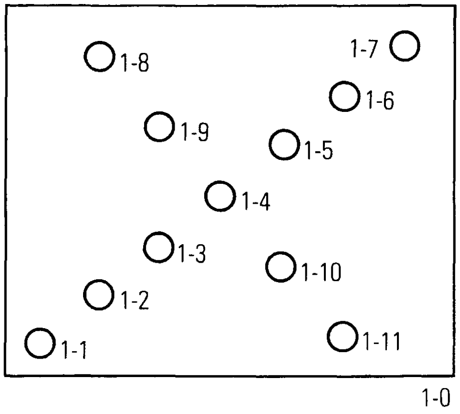
\- -

Sampling rule number	1
Variable defining probability	CON
Expansion constant	2
Number of samples	1

Numbering scheme:	**Plot #**	**Subplot #**	**Linking variable**
	1	0	1-1-1922, 2-1-1927

DESIGN 3

A 0.25-acre square plot in which the sampling of the smaller tree population has changed after the plot has been established for many years. From 1933 to 1974, all trees ≥1.0 inches d.b.h. were measured on the plot. Due to a change in the composition of the stand from one of larger trees to one with a thick stand of smaller trees, the sampling scheme for smaller trees was changed at the 1984 measurement. All trees ≥3.0 inches d.b.h. were measured on the plot, but trees 1.0 to 2.9 inches d.b.h. were subsampled using 11 plots of $^{1}/_{300}$ acre each, distributed systematically within the larger 0.25-acre plot. All trees 1.0 to 2.9 inches d.b.h. were measured and tagged on these $^{1}/_{300}$-acre plots. See drawing below. (Plots 26, 28, 148, 157, 160 in Intermountain Research Station Permanent Sample Plot Collection.)



Description using proposed data structure:

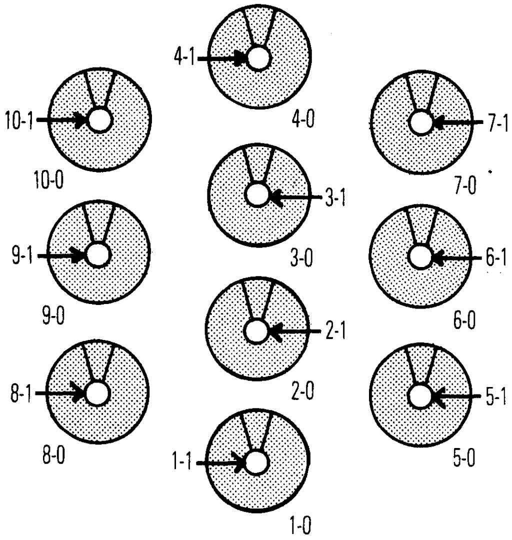
Subpopulation number	1	2	3
1st var. for subpop. definition	D	D	D
Minimum or 1st code	1.0	3.0	1.0
Maximum or 2nd code	99.9	99.9	2.9
2nd var. for subpop. definition			
Minimum or 1st code			
Maximum or 2nd code			

Sampling rule number	1	2
Variable defining probability	CON	CON
Expansion constant	4	300
Number of samples	1	11

Numbering scheme:	**Plot #**	**Subplot #**	**Linking variable**
	1	0	1-1-1933, 2-1-1984 ($^{1}/_{4}$-acre plot)
	1	1	3-2-1984($^{1}/_{300}$-acre plot)
	1	2	3-2-1984
	1	3	3-2-1984
	1	4	3-2-1984
	1	5	3-2-1984
	1	6	3-2-1984
	1	7	3-2-1984
	1	8	3-2-1984
	1	9	3-2-1984
	1	10	3-2-1984
	1	11	3-2-1984

DESIGN 4

A cluster of 10 BAF 20 variable radius plots, each with a $^{1}/_{300}$-acre subplot for measurement of regeneration, scattered through a similarly treated area. Variable radius plots used to measure trees ≥5.0 inches d.b.h. and $^{1}/_{300}$-acre plot used to measure trees from seedlings to 4.9 inches d.b.h. There are 35 stands in this study. Drawing of one stand is shown below. (USDA Forest Service Region 1 Birgenheier Study and older Forest Survey Standard Design.)



Description using proposed data structure:

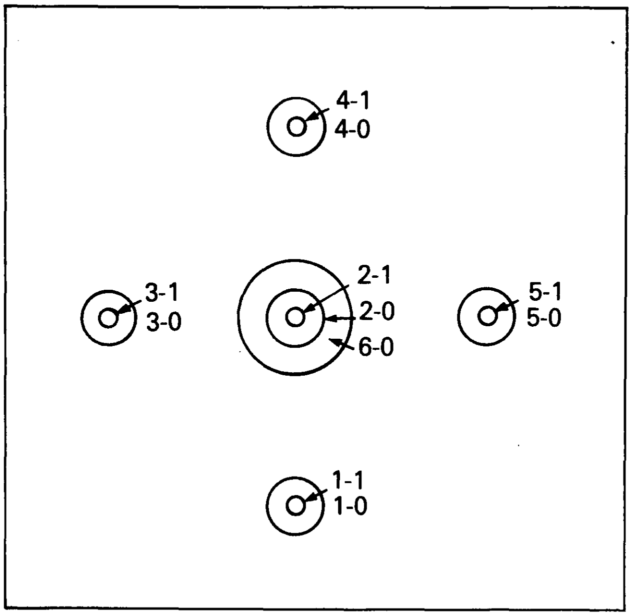
Subpopulation number	1	2
1st var. for subpop. definition	D	D
Minimum or 1st code	5.0	0.0
Maximum or 2nd code	99.9	4.9
2nd var. for subpop. definition		
Minimum or 1st code		
Maximum or 2nd code		

Sampling rule number	1	2
Variable defining probability	BA	CON
Expansion constant	20	300
Number of samples	10	10

Numbering scheme:	**Plot#**	**Subplot#**	**Linking Variable**
	1	0	1-1-1970 (var. radius plot)
	1	1	2-2-1970 ($^{1}/_{300}$-acre plot)
	2	0	1-1-1970
	2	1	2-2-1970
	.	.	.
	.	.	.
	.	.	.
	10	0	1-1-1970
	10	1	2-2-1970

DESIGN 5

A cluster of five concentric plots is located within a similarly treated area. Each of the concentric plots consists of two plots: a $^{1}/_{100}$-acre plot for counts of live seedlings and trees <1.0 inch d.b.h., and a $^{1}/_{20}$-acre plot for measurement of live trees ≥1.0 inch d.b.h. 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 the size of ≥9.0 inches d.b.h. 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 and a $^{1}/_{20}$-acre circular plot. The drawing shows one stand. Many such stands in the study. (USDA Forest Service Region 6 Managed Stand Survey.)



Description using proposed data structure:

Subpopulation number	1	2	3	4
1st var. for subpop. definition	D	D	D	D
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 var. for subpop. definition		T	T	
Minimum or 1st code		2	0	
Maximum or 2nd code			2	

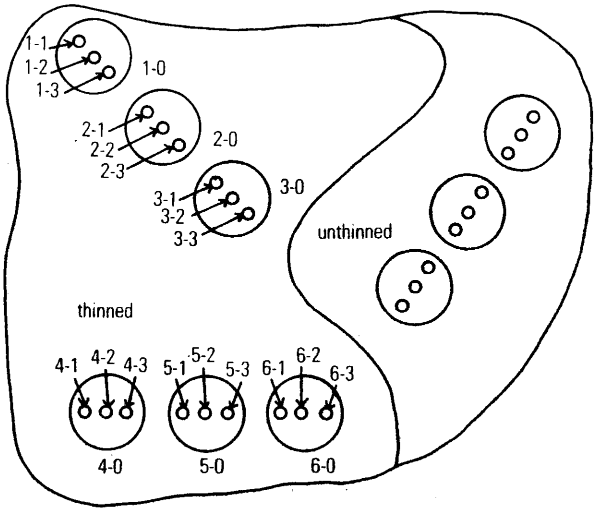
\- -

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

Numbering scheme:	**Plot #**	**Subplot#**	**Linking variable**
	1	0	1-1-1987,2-1-1987,3-1-1987 ($^{1}/_{20}$-acre)
	1	1	4-3-1987($^{1}/_{100}$-acre)
	2	0	1-1-1987,2-1-1987,3-1-1987
	2	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 ($^{3}/_{20}$-acre)

DESIGN 6

An installation with two treatments, one thinned, the other a control. Each area has the following basic design: a strip of three $^{1}/_{20}$-acre plots, located randomly in the area. Within each $^{1}/_{20}$-acre plot there is a subsample of three $^{1}/_{300}$-acre plots for measurement of trees from seedlings to 2.9 inches d.b.h. The $^{1}/_{20}$-acre plot is used to measure trees ≥3.0 inches d.b.h. The thinned area has two such strips (six $^{1}/_{20}$-acre plots) and the control has one strip (three $^{1}/_{20}$-acre plots). See drawing below. Only the thinned stand for one installation is described below. Many installations of this sort in the study. (USDA Forest Service Region 1 Monitoring Plots.)



Description using proposed data structure:

Subpopulation number	1	2
1st var. for subpop. definition	D	D
Minimum or 1st code	3.0	0.0
Maximum or 2nd code	99.9	2.9
2nd var. for subpop. definition		
Minimum or 1st code		
Maximum or 2nd code		

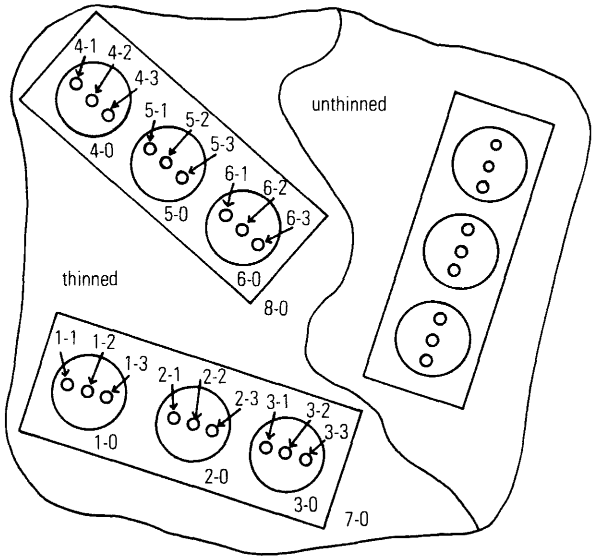
\- -

Sampling rule number	1	2
Variable defining probability	CON	CON
Expansion constant	20	300
Number of samples	6	18

Numbering scheme:	**Plot #**	**Subplot #**	**Linking variable**
	1	0	1-1-1984 ($^{1}/_{20}$-acre plot)
	1	1	2-2-1984 ($^{1}/_{300}$-acre plot)
	1	2	2-2-1984 ($^{1}/_{300}$-acre plot)
	1	3	2-2-1984 ($^{1}/_{300}$-acre plot)
	2	0	1-1-1984
	2	1	2-2-1984
	2	2	2-2-1984
	2	3	2-2-1984
	.	.	.
	.	.	.
	.	.	.
	6	0	1-1-1984
	6	1	2-2-1984
	6	2	2-2-1984
	6	3	2-2-1984

DESIGN 7

A larger plot is laid out around each set of three plots described in design 6. This additional plot is established at the time of the second measurement. This plot, with area of $^{17}/_{20}$ acre (1.0 acre minus $^{3}/_{20}$ acre), samples only trees that have died since the last measurement. Once a tree is recorded as dead, it is painted so that it will not be sampled again. See drawing below. Only the thinned stand is described below. (Proposed mortality overlay on USDA Forest Service Region 1 Monitoring Plots.)



Description using proposed data structure:

Subpopulation number	1	2	3
1st var. for subpop. definition	D	D	T
Minimum or 1st code	3.0	0.0	2
Maximum or 2nd code	99.9	2.9	
2nd var. for subpop. definition			
Minimum or 1st code			
Maximum or 2nd code			

- -

Sampling rule number	1	2	3
Variable defining probability	CON	CON	CON
Expansion constant	20	300	1.176
Number of samples	6	18	2

Numbering scheme:

Plot#	Subplot #	Linking variable
1	0	1-1-1984 (1/20-acre plot)
1	1	2-2-1984 (1/300-acre plot)
1	2	2-2-1984 (1/300-acre plot)
1	3	2-2-1984 (1/300-acre plot)
.	.	.
.	.	.
.	.	.
6	0	1-1-1984
6	1	2-2-1984
6	2	2-2-1984
6	3	2-2-1984
7	0	3-3-1988 (17/20-acre plot)
8	0	3-3-1988 (17/20-acre plot)

The proposed data structure is general enough to effectively describe all of the example designs. It should be flexible enough to also work with most other sampling designs. The Plot Data Management System (PDMS) was able to adequately describe only designs 1 and 4. Application of PDMS to the other designs was limited by the inability to describe changes in plot design, subsamples other than concentric designs, population definitions other than diameter limits, and designs that have differing numbers of plots of different sizes. When designs cannot be described adequately within a data structure, one would have to rely on comment fields to explain the missing facts about the design. This is inefficient and precludes preparation of analysis programs that can use the design information to compile stand statistics.

APPLICATIONS

Applications of this system could be in the field or in cataloging of data.

Sampling Design and Field Procedures

Although this data structure can be used to describe complex designs, this power should be used sparingly. Simplicity of design and field procedures should be paramount. These suggestions should be followed:

1. Numbers of subpopulations associated with different sampling rules should be kept small, again, to keep field procedures simple and easy to remember. The situation in design 5 illustrates this point. Trees ≥ 9.0 inches d.b.h. are sampled with five 1/20-acre plots and also by one donut-shaped 3/20-acre plot that is concentric with one of the 1/20-acre plots. This extra plot unduly complicates the design if its sole purpose is to obtain additional growth sample trees. If additional sample trees are needed, it might be simpler to increase the size of the present 1/20-acre plots, thus eliminating the need for the extra odd-shaped plot.

2. For cluster-sampling, spacing between plots should be considered in relation to plot size to avoid having the same tree in more than one plot. A cluster of 10 BAF 10 points, with plot centers 50 feet apart, in a stand with average d.b.h. of 10 inches, would result in many trees occurring in more than one plot.

3. Sampling rules and subpopulations should be so related that ingrowth is not sampled with geometric variable probability rules. Ingrowth should enter the sample through a constant probability sample (a fixed-area plot). Otherwise, the size of the ingrowth trees would have to be determined as of the previous measurement to establish whether the tree is actually ingrowth or an additional tree in the new sample. Remeasurement intervals should be related to growth rates such that ingrowth will only be possible into the subpopulation containing the smallest trees. For example, if a BAF 20 point is used as a permanent plot, a 1/300-acre plot, concentric with the point center, should be used to

measure ingrowth 1 to 3.5 inches d.b.h., and the BAF 20 point be used to sample trees ≥3.5 inches d.b.h. Remeasurement intervals of 10 years or less in most regions would be adequate because a tree is not likely to grow more than 2.5 inches d.b.h. (or, 3.5 minus 1.0 inches) in that period.

4. The possibility of a tree "leaving the sample" previously mentioned would imply that a previously tagged tree should not be remeasured at some examinations, unless it might be ingrowth into a subpopulation of particular interest. Hence, the relations between the design variables for each subpopulation should be considered with regard to their effect on the field examination procedures. If the d.b.h. range for the $^1/_{300}$-acre plot mentioned in item 3 above were 1 to 3 inches instead of 1 to 3.5 inches, there exists a possibility that a tree being tagged as ingrowth at one measurement would not be "in" in either $^1/_{300}$-acre plot or BAF 20 point at a subsequent remeasurement.

5. The guidelines for numbering plots, especially nested plots, and trees that were previously addressed should be diligently adhered to. Without this careful documentation, compilation of stand attributes for more complex designs may become unduly complicated.

Cataloging of Data

For purposes of preparing a general catalog of data, complete description of sampling designs may not be necessary. There are currently many endeavors to catalog available permanent plot information on a specific subject or in a specific region for the purpose of data sharing and future planning of new permanent plot installations. Such catalogs do not require all known information about pertinent plots. They only need a general description for purposes of planning and summarization. Recording of information on specific plot numbers and linking variables can be deferred until stand attributes must be calculated for analysis procedures.

When collecting general permanent plot information from a wide variety of sources, it is often best to provide a survey form that specifically describes the needed information. Survey forms should be designed in as simple and uncluttered a fashion as possible so as not to bog the responder down in unnecessary detail or terminology but still obtain the pertinent information. A format for providing general sampling design information consistent with our data structure is provided below:

Use the following table to provide information on the sampling design for a stand. If the design has changed over time, fill out one table for each design.

Description of a Sampling Design

Year Sampling Began ________

	SAMPLING RULES[1]			
	1	2	. . .	*n*
PLOT TYPE[2]				
EXPANSION CONSTANT[3]				
NUMBER OF SUCH PLOTS				
KIND OF TREES SAMPLED WITH THIS PLOT TYPE[4]				

[1] Use as many columns as needed to describe your design. n = the number of sampling rules in the design.

[2] Plot type:

CON = fixed area plot or strip
BA = horizontal point sample
DBH = horizontal line sample
HTS = vertical point sample
HT = vertical line sample
ARB = subjective definition of sampling probabilities (3-P)

[3] Expansion constant:

If type = CON, then 1/plot area or 1/strip area
If type = BA, then BAF
and so forth

[4] Examples: Trees 1 to 3 inches d.b.h.
dead trees
and so forth.

REFERENCES

Arney, J. D.; Curtis, R. O. 1977. Code legend for standardized permanent plot records. In: Reimer, D. R., chairman. Standards of measure and data sharing: report of the Committee of Standards of Measure and Data Sharing (COSMADS). Portland, OR: Western Forestry and Conservation Association. 18 p.

Chapman, R. C. 1985. Square pegs and round holes: proceed only with caution. In: Van Hooser, D. D.; Van Pelt, N., eds. Proceedings, growth and yield and other mensurational tricks: a regional technical conference; 1984 November 6-7; Logan, UT. Gen. Tech. Rep. INT-133. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station: 15-18.

Curtis, R. O. 1983. Procedures for establishing and maintaining permanent plots for silvicultural and yield research. Gen. Tech. Rep. PNW-155. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Forest and Range Experiment Station. 56 p.

Curtis, R. O.; Clendenen, G. W. 1981. Plot data management system (PDMS). Study Plan. Olympia, WA: U.S. Department of Agriculture, Forest Service, Pacific Northwest Forest and Range Experiment Station, Forestry Sciences Laboratory. 17 p.

Gilbert, A. M. 1954. What is this thing called growth? Stn. Pap. 71. Broomall, PA: U.S. Department Agriculture, Forest Service, Northeastern Forest Experiment Station. 5 p.

Grosenbaugh, L. R. 1958. Point-sampling and line-sampling: probability theory, geometric implications, synthesis. Occas. Pap. 160. New Orleans, LA: U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station. 27 p.

Grosenbaugh, L. R. 1963. Some suggestions for better sample-tree-measurement. In: Proceedings, Society of American Foresters; 1963 October 20-23; Boston, MA. Washington, DC: Society of American Foresters: 36-42.

Martin, G. L. 1982. A method for estimating ingrowth on permanent horizontal sample points. Forest Science. 28(1): 110-114.

Oderwald, R. G. 1981. Point and plot sampling —the relationship. Journal of Forestry. 79(6): 377-378.

O'Regan, W. G.; Palley, M. N. 1965. A computer technique for the study of forest sampling methods. Forest Science. 11(1): 99-114.

Sweet, M. 1987. [Personal communication.] University of Montana, Missoula, MT.

U.S. Department of Agriculture, Forest Service. 1933. Sample plots in silvicultural research. Circ. 333. Washington, DC. 88 p.

U.S. Department of Agriculture, Forest Service. 1986a. Timber management data handbook, Region 1. Chapter 600, Special Surveys. Missoula, MT: U.S. Department of Agriculture, Forest Service, Region 1, Division of Timber Management. 24 p.

U.S. Department of Agriculture, Forest Service. 1986b. Field instructions for managed stand surveys. Portland, OR: U.S. Department of Agriculture, Forest Service, Region 6, Forest Inventory Section, Division of Timber Management. 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 of Measure and Data Sharing (COSMADS). Portland, OR: Western Forestry and Conservation Association: Appendix B, p. 1-7.

Byrne, John C.; Stage, Albert R. 1988. A data structure for describing sampling designs to aid in compilation of stand attributes. Gen. Tech. Rep. INT-247. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 20 p.

Maintaining permanent plot data with different sampling designs over long periods within an organization, and 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 composed of just those variables and their relationships needed to compile stand attributes and their variances. The data structure is able to incorporate changes in design that commonly occur through the life of a set of permanent plots. Using the data structure, seven actual or proposed designs, from simple to complex, are adequately described. Applications of the data structure for designing sampling schemes and for cataloging data are included.

KEYWORDS: permanent plots, data management systems, subsampling

INTERMOUNTAIN RESEARCH STATION

The Intermountain Research Station provides scientific knowledge and technology to improve management, protection, and use of the forests and rangelands of the Intermountain West. Research is designed to meet the needs of National Forest managers, Federal and State agencies, industry, academic institutions, public and private organizations, and individuals. Results of research are made available through publications, symposia, workshops, training sessions, and personal contacts.

The Intermountain Research Station territory includes Montana, Idaho, Utah, Nevada, and western Wyoming. Eighty-five percent of the lands in the Station area, about 231 million acres, are classified as forest or rangeland. They include grasslands, deserts, shrublands, alpine areas, and forests. They provide fiber for forest industries, minerals and fossil fuels for energy and industrial development, water for domestic and industrial consumption, forage for livestock and wildlife, and recreation opportunities for millions of visitors.

Several Station units conduct research in additional western States, or have missions that are national or international in scope.

Station laboratories are located in:

Boise, Idaho

Bozeman, Montana (in cooperation with Montana State University)

Logan, Utah (in cooperation with Utah State University)

Missoula, Montana (in cooperation with the University of Montana)

Moscow, Idaho (in cooperation with the University of Idaho)

Ogden, Utah

Provo, Utah (in cooperation with Brigham Young University)

Reno, Nevada (in cooperation with the University of Nevada)