INCREASING THE EFFICIENCY OF MULTIPLE-USE INVENTORY PROCEDURES

by

Roger Dean Stuck II

Dissertation submitted to the Graduate Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY
in
Forestry and Forest Products

APPROVED:

H.E. Burkhart, Chairman

W.A. Flick

W.A. Leuschner

H.S. Mosby

M.R. Reynolds

April, 1976
Blacksburg, Virginia

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to Dr. H.E. Burkhart for his continual criticism and suggestions.

I would also like to thank the U.S. Forest Service for their financial aid and cooperation. A special thanks to Brenda Ayers, Bill Bottomly, Bob Colona, and Bob Fish of the Jefferson National Forest who helped in compiling information for this study.

Finally, I would like to thank my parents and fellow graduate students for their encouragement.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS ii

TABLE OF CONTENTS iii

LIST OF TABLES vi

LIST OF FIGURES vii

Introduction 1

Objective 3

UNIT PLANNING ON THE NATIONAL FOREST IN THE EAST 4

Organization 4

Inventory 5

Engineering 6

Fire and Law Enforcement 6

Lands 7

Landscape Architecture 7

Recreation 8

Soils 9

Timber 11

Wildlife 12

Use of Computers in Unit Planning 13

STUDY AREA 16

Organization and Description 16

Unit Planning Process 17

ANALYSIS OF INVENTORY DATA CURRENTLY BEING COLLECTED . . 20

Procedures 20

Page

Simple Correlation Coefficients 21

Multiple Linear Regression 22

Factor Analysis 23

Data 27

Results and Discussion 29

A FRAMEWOFK FOR OPTIMAL ALLOCATION OF INVENTORY RESOURCES FOR MULTIPLE-USE PLANNING 41

Cost-Loss Functions for Inventories Where One Attribute is Observed 46

Case I: Single Inventory, Single Decision . . . 48

Case II: Single Inventory, Multiple Decisions . 54

Case III: Multiple Inventories, Single Decision 56

Case IV: Multiple Inventories, Multiple Decisions 59

Cost-Loss Functions for Inventories Where Multiple Attributes are Observed 63

Decisions Made more than One Time 71

APPLYING THE COST-LOSS MINIMIZATION PROCEDURES: A CASE STUDY 73

Decisions for a Unit Plan 73

Timber 74

Wildlife 75

Recreation 76

Decisions Requiring Inventory Data 77

Determination of Sampling Intensity 78

Page

Obtaining Variance Estimates 80

Cost of Inventories 83

Determination of λ 84

Development of k Vectors 91

Determination of T's 93

Minimization 96

Discussion 106

RECOMMENDATIONS FOR MODIFICATIONS IN DATA HANDLING AND PLANNING PROCEDURES ON THE JEFFERSON NATIONAL FOREST 108

Inventory 108

Data Handling 109

Planning and Procedures 120

LITERATURE CITED 122

VITA 125

LIST OF TABLES

Table Page

1 Correlation coefficient matrix of the Inventoried data on the High Knob Planning Unit 30

2 Summary of multiple linear regression analysis on inventoried data for the High Knob Planning Unit 34

3 Rotated factor matrix (Oblimin Bi-quartimin method) with communalities for the High Knob Planning Unit 35

4 Rotated factor matrix (Orthogonal Varimax method) with communalities for the High Knob Planning Unit 37

5 The variance-covariance matrix of the five attributes from the timber inventory 82

6 A summary of inventory costs (dollars) for four inventories on the High Knob Planning Unit 85

7 k vectors for Unit Planning for the case of multiple attributes measured at each sampling sampling point 94

8 k vectors for Unit Planning for the case of a single attribute measured at each sampling point 95

9 Summary of results from the four methods of determining sample size assuming covariances equal zero 104

LIST OF FIGURES

Figure		Page
1	The estimated loss in dollars from making a timber harvesting decision with various amounts of sampling error (years) in the age of the stand. Also shown is a curve for the loss per squared error (years^2) . . .	53
2	Graphic representation of the sampling cost, expected loss, and total cost functions for the timber harvesting decision example . . .	55
3	The relationship of six decisions made for Unit Planning and the inventories that aids in making the decisions. The decision number corresponds to numbers on page 77 . . .	79
4	Plots of squared error, z^2, against monetary loss. If the assumption $L(n)=\lambda z^2$ is correct then the plots will be linear. Also graphed is the regression through the points with its value of λ. Plots (a) through (f) are associated with decisions 1 through 6, respectively	92
5	Portion of High Knob Planning Unit's Land Status map. In Map A, 0 code (.) is used for private land and 9 code (■) is used for Forest Service Land. In Map B, 0 code (.) is used for private land and 1 code (,) is used for Forest Service Land . . .	117

INTRODUCTION

Growth in the world population has increased demand on forest land for timber products, as well as forest related products such as recreation and wildlife. These demands necessitate the use of multiple-use management.

In the Multiple-Use-Sustained Yield Act, the U.S. Congress (1960) directed the Forest Service to manage the National Forest for "...all the various renewable surface resources...so that they are utilized in the combination that will best meet the needs of the American people; ...and harmonious and coordinated management of the various resources, each with the other, without impairment of the productivity of the land, with consideration being given to the relative values of the various resources, and not necessarily the combinations of uses that will give the greatest dollar return or greatest unit output." To follow this directive the U.S. Forest Service has introducted the concept of Unit Planning for the National Forest in the East. The Unit Plan is "a total management plan for all the social, economic, natural resources, and other environmental situations found within a ..." homogeneous land area such as a watersted or isolated owership called a Unit (U.S.D.A., Forest Service, 1970b).

Navon (1971), Halterman (1972), and Porter (1974) have developed models to help make multiple-use planning decisions, but little attention has been directed toward multiple-use inventory. Most managers still conduct inventories as if they were dealing with only one product. Planners commonly combine information from several inventories to formulate management plans. Under the current Forest Service procedure the Unit Planning staff directs questions that are necessary to construct a Unit Plan to the appropriate section staff within the Supervisor's Office. The section staff conducts the necessary inventories to answer the questions assigned to them. In many cases general "rules of thumb" are used to determine sampling intensity.

It is hypothesized that there are two major areas for improvement in multiple-use inventory. First, each section staff conducts their own inventory; thus inventories from different sections may overlap, either by collecting data on the same variable or by measuring variables that are highly correlated. There could be considerable savings of both time and money by eliminating sampling of the same variables and by predicting highly correlated variables from one another using statistical procedures.

The second area for possible improvement is more efficient allocation of sampling units. When using only "rules of thumb" to determine the number of sampling units, many inventories may be too intensive -- which increases the cost -- or not intensive enough, in which case the data are not sufficiently precise to make an adequate decision. By finding improved methods of determining the sampling intensity, costs should be reduced.

Objective

The objective of this study was to increase the efficiency of inventory and data handling procedures for use in unit plans. To accomplish this objective the study was divided into three main sections. First the minimum number of variables needed for the information now being collected was analyzed. Second a framework for optimal allocation of inventory resources for multiple-use planning was developed. Finally recommendations were made for modifications in the data handling and planning procedure of the U.S. Forest Service to increase efficiency in multiple-use planning. The methodology developed was tested with a case study on a planning unit of the Jefferson National Forest in Virginia.

UNIT PLANNING ON THE NATIONAL FOREST IN THE EAST

Organization

Before one can evaluate the Unit Planning procedure of the U.S. Forest Service it is necessary to look at present organization and inventory responsibilities for unit planning. To handle the complex planning that is necessary to manage the forest land under the Sustained Yield Act of 1960, the National Forests in the East introduced the Unit Planning System. The Unit Planning concept is stated in _The System for Managing the National Forests in the East_ (U.S.D.A., Forest Service, 1970b), which is commonly called "The System". "The System" divides the National Forest in the East into "'regions' each of which has its own peculiar make-up of people, topography, climate, problems, and opportunities for development" (U.S.D.A., Forest Service, 1970b). Guidelines are given for forming and executing the Unit Planning System. "The System" also sets guidelines for writing the "Guide" for each "region." "The 'Guide' provides the broad objectives, policy, and direction to be followed by all Forest Service administrative units within a 'region'" (U.S.D.A., Forest Service, 1970b).

Each "region" is subdivided into "units" defined as an isolated ownership or a watershed. The plan for each unit on a forest is the responsibility of an interdisciplinary Unit Planning Team in the Forest Supervisor's Office. The Team works with each functional staff gathering the necessary information for the unit plan. Conducting public meetings before and after the completion of the plan is also the responsibility of the Team. When all the necessary information has been collected, the Team coordinates the writing of the unit plan. The unit plan is an "...all-inclusive on-the-ground plan for carrying out the objectives, policy, and directives in the 'Guide'" (U.S.D.A., Forest Service, 1970b). The District Ranger, with support from the Forest Supervisor and the Regional Forester, is responsible for the execution of the unit plan. The sum of all unit plans for a "region" plus the "Guide" make up the "Region Master Plan."

Inventory

Each functional staff in the Forest Supervisor's Office is responsible for gathering and recording certain types of information. The following is a brief description of the information gathered by each functional staff and the inventory process used.

Engineering

The engineering staff conducts no field inventory. They act as a support unit for the other functional areas. Their job includes boundary line surveys, surveys, plans, estimates, and construction supervision of roads, trails, and other improvements such as buildings, sewage disposal, and water systems. The engineering staff keeps records on existing and proposed roads, trails, and bridges with such information as length, width, surface, right-of-way ownership, and maintenance responsibilities. They also keep records on all improvements, including dams, with information on size, condition, and maintenance requirements.

Fire and Law Enforcement

This functional staff is responsible for guiding enforcement of the laws and regulations on the forest and protecting the property of the forest. To fulfill this job the staff keeps information on travel speeds on various roads, travel time between various points, location of fire fighting equipment, past history of fires, slope of the terrain, and monetary loss that would occur from a fire.

Lands

The lands staff also conducts no field inventory. Their job consists of land ownership adjustment (including purchase and exchange), settling land claims, acquiring road and trail right-of-ways, administering special use land permits, and mineral management. A land status atlas is kept showing Forest Service ownership and an atlas is also kept showing the status of right-of-ways. The lands staff keeps some geological records but these records are a composite of existing data gathered by the Regional Geologist and no new field work is involved.

Landscape Architecture

The Southern Region has recently instituted a distance zone and sensitivity level concept. These distance zones are located with respect to distances from prominent viewing points on the planning unit. The method of determining sensitivity areas is described in <u>Forest Service Handbook</u> (U.S.D.A., Forest Service, 1973b). The landscape architect and district prescriber also use the booklet <u>National Forest Landscape Management, Volume I</u> (U.S.D.A., Forest Service, 1973d) as a reference in such cases as harvest area layouts and road layouts.

Recreation

The duties of the recreation staff are varied because recreation means different things to different people. The Recreation Information Management system, or RIM as it is commonly called, is a computer oriented system that keeps records about people, places, and things over periods of time. This system contains existing and past information and can be used to help make projections into the future.

As a starting point for developing new recreation areas the National Forest Recreation Survey (NFRS) was conducted on the Jefferson National Forest in 1961. In this survey information was compiled for all potential recreation sites (areas with slope less than fifteen percent). Using a point system, a quality rating was given to each site. Also a list of instructions and estimates on cost for developing the area were included. This survey was conducted to aid the planners in choosing recreation sites as they were needed.

Composite Plans have been written for most of the Forest. In this plan information about traffic flow, economic situation of the area, and projected demands were combined with proposed development areas and an analysis of each area was included. Also included was a schedule of development and its cost per year.

The Forest is in the process of listing all historical, cultural, and archaeological sites on its lands. This list is being compiled by consulting with personnel in the field, public contact, and literature searches.

The Forest has been instructed to inventory and develop a plan for off-road vehicles. In the future some means of inventorying sites for off-road vehicle use will have to be devised taking into account soil erosion, noise pollution, damage to wildlife, and conflicts with other activities.

Soils

The soils specialist interacts with the different functional staffs within the Forest Supervisor's Office to determine if their management suggestions can be performed in an area without impairment of soil productivity. To perform this job the soils specialist conducts a survey to determine Ecological Management Units (E.M.U.) on a planning unit using as a guide the Soils Resource Guide (U.S.D.A., Forest Service, 1972).

The E.M.U. consists of five characteristics: Land Forms, Soil Source, Soil Texture, Water Regime, and Soil Modifiers. The first, Land Forms, is defined on a topographic map without going to the field. The remaining four characteristics are defined by walking over an area and occasionally taking a soils boring up to thirty inches in depth. When the survey is complete, a map of E.M.U.'s, generally twenty acres or more in size, is drafted.

The soils specialist determines E.M.U. boundaries by observing changes in characteristics such as timber overstory, understory, ground vegatation, and surface rocks. A notebook, in addition to the map, is kept with special information about an area such as the location of sink holes, rock outcrops, and material observed in streams that might be useful in making management decisions.

Once the specialist has divided the planning unit into E.M.U.'s he writes a critique of each E.M.U. The critique contains a brief description of the soils and notes to each functional staff as to problems they will have to consider in making their management decisions.

The survey that is described above is a "phase I" inventory. A "phase II" inventory is conducted when a more intensive soil description is needed such as when constructing a recreational area. In the "phase II" survey the area is divided into a hundred foot grid with a soils boring being taken at each intersection. This survey is more costly than a "phase I" and is only performed on small areas.

The soils data are computerized in a system called Soils-File-8 at Georgia Tech, Atlanta, Georgia. This system ties the E.M.U. to the timber stands within the various timber compartments.

Timber

The timber staff handles the management and sale of timber products from Forest Service land. To accomplish this job it is necessary to conduct an inventory of the timber resources on the Forest. The staff tries to prescribe timber management needs on a tenth of the Forest every year.

The prescriber begins the inventory by using aerial photographs to divide a planning unit into compartments using easily distinguishable boundarys such as roads, streams, and ridges. Each compartment is then divided into stands. Cruise lines and sampling points are laid out according to the <u>Compartment Prescription Handbook</u> (U.S.D.A., Forest Service, 1970a) and their locations are recorded on both the aerial photograph and the Stand Tally Sheet (U.S.D.A., Forest Service Form R8-2400-48).

The cruise is conducted by district personnel. At each sampling point, information is recorded on the stand tally sheets using the <u>Compartment Prescription Field Book</u> (U.S.D.A.,Forest Service, 1971a) as a guide. After the cruise is completed, information from each stand in a compartment is averaged and recorded on the Stand Prescription Summary (U.S.D.A., Forest Service Form R8-2400-36) along with other general information about the compartment. The prescription is written in the field by certified prescribers who follow guidelines in the Compartment Prescription Handbook (U.S.D.A., Forest Service, 1970a). The District Ranger reviews and submits the prescriptions to the Forest Supervisor. The Supervisor reviews and approves the prescription with the aid of his resource staff. Records on stands and compartments are stored in a computer system called Continuous Inventory of Stand Condition (CISC) in Fort Collins, Colorado.

When the timber on an area is to be sold, an inventory of the area is conducted to estimate the volume of merchantable wood. The type of sampling method varies depending on the acreage of the sale and the value of the timber.

Wildlife

The wildlife staff conducts a wildlife habitat survey on different survey units within a planning unit. A survey unit consists of several timber compartments within a planning unit; these survey units are generally about ten thousand acres. The survey is conducted according to the Fish and Wildlife Survey Handbook (U.S.D.A., Forest Service, 1973a). U.S.D.A., Forest Service Form R-8 2600-3 is completed on each survey unit. Such information as present wildlife population levels and trends, timber overstory, wildlife food, key wildlife areas, ponds, lakes, reservoirs, streams, facilities available to sportsmen, access, limiting factors for various wildlife species, and soils are obtained on each survey unit.

Information on soils and facilties available to sportsmen are obtained from the soils specialist and recreation staff, respectively. Since the survey unit encompasses several timber compartments the necessary information on timber overstory is obtained from the timber staff's inventory.

The wildlife staff consults with local groups and persons who have specific knowledge on special wildlife situations. The local groups also help determine the location of any endangered plant or animal species.

The wildlife staff uses as a guide the *Wildlife Habitat Management Handbook* (U.S.D.A., Forest Service, 1971b) for managing wildlife and determining the compatability of wildlife with management decisions on the planning unit. The *Wildlife Habitat Management Handbook* also specifies standards and requirements for habitat development used in the wildlife habitat survey.

Use of Computers in Unit Planning

During the past two years Region 8 has gradually been introducing the use of computers in the unit planning procedure. The U.S. Forest Service has been using computers to store various types of information for selected functional staffs. A modified version of the Harvard Grid Program (Sinton and Steinitz, 1971), commonly called "Grid", is being used to produce computerized maps.

The "Grid" procedure consist of dividing the unit into squares called "grid cells." Each grid cell is assigned a code between 0 and 9 depending on the value of the factor being mapped. These coded values are input to the Grid program and a "grid map" is produced.

The use of the Grid system has several advantages over using hand-drawn maps. First grid maps can be stored on the computer and updated periodically with little difficulty and cost. The maps can also be reduced in size to fit into a report without losing detail in the shading of the various grid cells and map coordinates. The most important advantage of the Grid system is the ability to produce overlays of several grid maps (U.S.D.A., Forest Service, 1975) in a few minutes and at a low cost. The overlay maps can be used to test for conflicts between different management alternatives. Using the grid overlay system can considerably reduce the time spent on testing management alternatives.

Personnel of the Forest Service are becoming accustomed to computer usage. The more accustomed they become, the more computers will aid them in the unit planning process.

STUDY AREA

Organization and Description

The High Knob Unit of the Jefferson National Forest was used as the study area for this project. The 71,188 acre unit, located on the southern portion of the Clinch Ranger District in Lee, Scott, and Wise counties of Virginia, is part of the National Forests of the Appalachians. The National Forests of the Appalachians is one of seven "regions" established by the *System for Managing the National Forests in the East* (U.S.D.A., Forest Service, 1970b). The National Forests in the East consist of all National Forest land east of the "100th Meridian." The unit planning procedure on the High Knob Unit is guided by "The System" and the *Guide for Managing the National Forest in the Appalachians* (U.S.D.A., Forest Service, 1973c). Management activities on the unit are carried out by the District Ranger of the Clinch Ranger District of the Jefferson National Forest, which is part of Region 8 of the U.S. Forest Service.

The climate of the area is moderate most of the year with temperature extremes ranging from -10 to 0 degrees in the winter to 90 to 100 degrees in the summer. Precipitation averages 50.37 inches per year.

The main soil types are loams, sandy loams, and sandy clay loams with surface soils ranging between four and eight inches deep in most locations. The land is generally steep with many slopes over 50 percent.

The High Knob Unit is in the heart of the coal mining area of Virginia. There are numerous deposits of coal beneath the Unit with some mining presently in progress. Deposits of natural gas have also been found on the Unit.

Cove and upland hardwoods are the predominant timber species. Deer, grouse, squirrel, and turkey are the major game species on the Unit. Both warm and cold water fishing are available in the mountain streams and lakes.

There are numerous outdoor activites available on the High Knob Unit including fishing, hunting, camping, picnicing, and hiking. Many of these activites are provided through the developed recreational areas such as High Knob Lake, Hanging Rock Picnic Area, and Bark Camp Lake and Recreation Area.

Unit Planning Process

Prior to writing the High Knob Unit Plan a public meeting was arranged and held on August 14, 1973, at Big Stone Gap, Virginia which is near the Unit. At this meeting individuals and representatives of groups were given an opportunity to present facts and opinions that they felt should be considered in forming a management plan for the Unit.

The Unit Planning Team, using the "Guide" for the Appalachians (U.S.D.A., Forest Service, 1973c) and a memo from the Region 8 Unit Planning Team (U.S.D.A., Forest Service, 1974a), gave directives to the various functional staffs regarding the information needed for the Unit Plan. In February, 1975, the team decided that they had sufficient information for the Unit Plan. A week-long meeting was held between the Unit Planning Team of the Jefferson, representatives of each functional staff, the Forest Supervisor, and the District Ranger of the Clinch.

During the first three days of the session, specific problems on the High Knob Unit were considered and a management directive for each was formed. Comments made at the public meeting were also reviewed, one by one, and a decision on each comment was made. Each functional staff presented an alternative management plan for the High Knob Unit during the morning of the fourth day. The remainder of the week was spent with each functional staff writing the section of the Unit Plan for their function.

Following this week-long meeting, the Unit Planning Team assembled a draft of the High Knob Unit Plan. After the draft was completed, each functional staff reviewed the plan to insure that no conflicts had developed between different management objectives. Once the Unit Planning Team decides that all conflicts in management are resolved, another public meeting is held to review the plan. If there are no conflicts brought out by the public, or after any conflicts are resolved, the Unit Plan is drafted along with a Draft Environmental Impact Statement (DEIS). This draft Unit Plan and the DEIS are sent to the Region 8 Office for distribution to the Council on Environmental Quality (CEQ), to the regional unit plan review group, and to all participants in the planning process. Any comments are reviewed and incorporated into the final plan and Environmental Impact Statement.

ANALYSIS OF INVENTORY DATA CURRENTLY BEING COLLECTED

The current set of variables being inventoried for Unit Planning was examined for association between variables and combinations of variables. Even if one assumes that all variables now being observed are used in the Unit Planning process, if there is a high degree of association between two variables then one can possibly be eliminated since both are providing essentially the same information. The objective was to determine the minimum number of variables that can be observed and still retain essentially the same amount of information as is now provided by observing the full set of variables.

Procedures

Three statistical methods -- simple linear correlation, multiple linear regression, and factor analysis -- were used to evaluate relationships between variables currently being inventoried. Simple linear correlation coefficients were used to describe the linear association between two variables. The linear association between a variable and a combination of the other variables was examined with multiple linear regression. Factor analysis was used to find common explanation for the intercorrelations of the set of observed variables. Using these three statistical methods all possible linear associations can be examined. If a high degree of association between variables is shown, then perhaps simplifications can be made in the inventory process and still provide the information needed for Unit Planning.

Simple Correlation Coefficients

The correlation coefficient is a measure of linear association between two random variables, X and Y. The simple correlation coefficient for a sample of size n and observations (x1,y1), (x2,y2),...,(xn,yn) is defined by Draper and Smith (1966) as:

$$r\ xy = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{[\sum_{i=1}^{n} (x_i - \bar{x})^2]^{1/2}\ [\sum_{i=1}^{n} (y_i - \bar{y})^2]^{1/2}}$$

r xy lies between -1 and 1. If r xy = 1, x and y are said to be perfectly positively correlated, that is, large values of x are perfectly related to large values of y. If r xy = -1, the variables are said to be perfectly negatively correlated, that is, large values of x are perfectly related to small values of y. If r xy = 0, x and y are uncorrelated or in other words they are not linearly associated.

If more than two variables are of interest, the correlation coefficients can be formulated in a matrix. The correlation matrix is a square symmetric matrix of the form:

$$\begin{vmatrix} 1 & r_{12} & r_{13} & \dots & r_{1p} \\ r_{21} & 1 & r_{23} & \dots & r_{2p} \\ r_{31} & r_{32} & 1 & \dots & r_{3p} \\ & \dots & & & \\ r_{p1} & r_{p2} & r_{p3} & \dots & 1 \end{vmatrix}$$

where

$r_{ij} = r_{ji}$ $(i,j = 1,2,\dots,p \quad i \neq j)$ is the correlation coefficient between variable i and variable j.

r_{ii} $(i = 1,2,\dots,p)$ is 1 because of the perfect correlation of variable i with itself.

Multiple Linear Regression

Stepwise multiple linear regression procedures were used to compute multiple correlation coefficients between selected variables that are being inventoried and sets of other measured variables. The multiple correlation coefficients were calculated because the simple correlation between a variable and each of two other variables may, in both cases, be near zero, whereas the multiple correlation may be near -1 or 1.

A sample prediction equation from multiple linear regression can be written as (Draper and Smith, 1966):

$Y = b0 + b1\ X1 + b2\ X2 + \ldots + bk\ Xk$

where:

Y = dependent variable

b0, b1, ..., bk = sample estimates of the true regression coefficients, B0, B1, ..., Bk

X1, X2, ..., Xk = independent variables.

The multiple correlation coefficient (R) and the percentage of variation "explained" by the regression equation (R-square) were examined after each step. The multiple correlation coefficient is the correlation between an dependent variable and the independent variables in the regression equation. The value of the multiple correlation will lie between -1 and 1. If the multiple correlation coefficient is -1 or 1 there is perfect association between the dependent variables and the combination of independent variables. For further information on regression analysis refer to Draper and Smith (1966).

Factor Analysis

Factor analysis is a mathematical process that will identify a set of variables in terms of a smaller set of variables called factors if there is linear association between variables in the original set. The factors are hypothetical variables but may be strongly related to the variables in the original set. This analysis was used instead of principal component analysis because the results from principal component analysis are such that all components are needed to predict the original variables, whereas factor analysis does not require the use of all factors for prediction. Also, principal component analysis examines all the variance, whereas, factor analysis examines only the variance after the specific variance for each variable is removed.

Factor analysis is a very complex mathematical process. To perform factor analysis correctly it takes both a person experienced in factor analysis and a person in the field of science to which factor analysis is being applied. The author, with the assistance of Robert Schulman (1), used factor analysis in this study.

The BMD Program 8M (Dixon, 1974) was used to perform factor analysis for this study. The factor anlysis procedure begins with a matrix of simple correlation coefficients. Factors are extracted from this matrix.

---

(1) Robert Schulman is Assistant Professor of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

There are an infinite number of ways to factor a correlation matrix. The number of factors extracted using the BMD program is based upon the eigenvalues of the correlation matrix. For this study, the number of factors extracted was equal to the number of positive eigenvalues. This extraction process produces a matrix that represents the extent the variables and hypothetical factors are related. The matrix is called the matrix of unrotated factor loadings.

Communality is a measure of the overlap between a variable and k factors. If the communality is 1 the variables can be predicted perfectly with a weighted form of k factors. The variable has nothing in common with the factors or other variables if the communality is 0. The communality is the sum of squares of the loadings from the matrix of unrotated factor loadings over the k factors. It is only this portion of the total variance that factor analysis is concerned with.

The BMD program calculates the percentage of variation explained by the k factors. In most cases, this value should be relatively close to one for factor analysis to be of much use.

The unrotated factor loading matrix is not the best choice for interpretation According to Comrey (1973) the reason is that most methods of extracting factors are designed to extract approximately as much variance as possible with each successive factor. By rotating the matrix a mathematically equivalent matrix to the unrotated matrix is formed. There are many methods of rotation, and there is little agreement between users as to which method is best. Different rotations produce different solutions. In this study two methods of rotation were used, the orthogonal varimax and the oblimin bi-quartimin. The only difficulty in interpretation of the oblimin solution is that for any given factor a variable contributes to it in two ways -- directly through its loadings on that factor and indirectly through its loadings on correlated factors. In short, two matrices are needed -- the matrix of factor loadings and the matrix of correlation between variables and factors. Using orthogonal rotation these two matrices are the same. Comrey (1973) developed a set of criteria for interperting an orthogonally rotated matrix. Once the user decides which variables are associated with a factor, the user must determine what the factor represents. If the rotated matrix has high correlations between the variables and the factors and the communality has a high value then the factors can be estimated by measuring which seem to be highly correlated with them instead of the original variables to provide essentially the same information.

Factor analysis is an area where there is much flexibility as to procedures and interpertations. There are an infinite number of solutions that can be derived from the same data set. For further information on factor analysis refer to Harmen (1967), Comrey (1973), and Lawley and Maxwell (1971).

Data

A list of inventoried variables for Unit Planning observed by the functional staffs was compiled. The direct overlap of inventories, that is when the same variable is observed by more than one functional staff, will be discussed in the chapter on recommendations. Variables observed in inventories for Unit Planning were: Soil Source, Soil Texture, Water Regime, Soil Modifiers, Slope, Site Index, Stand Condition, Forest Type, and Wildlife Featured Species.

Soil Source, Soil Texture, Water Regime, and Soil Modifiers are four of the five soil characterists of the Forest Service E.M.U. Soil Source identifies the origin of the soil material. Soil Texture gives the size of the soil particles and whether the soil is a loam, sand, clay, or a combination of the three. The amount of moisture generally found in the soil is described by Water Regime. Soil Modifiers identifies critical characterics of the soil such as extreme steepness, stoninesss, and erosiveness.

Slope and Site Index need no explanation. Stand Condition identifies the age of timber stands by describing them in terms of immature, mature, or damaged poletimber and sawtimber. Areas of non-stocked land and regenerated land are also identified. Forest Type is determined by species composition of the forest stands. Wildlife Featured Species identifies the wildlife species that is being managed on an area.

These nine variables were mapped using the Harvard Grid System (Sinton and Steinitz, 1971). Each cell of a grid map was assumed to be a sampling unit. Cells with the same coordinates on each map were assumed to be paired sampling units. Under these assumptions there were 3293 sampling units for the High Knob Unit.

Three additional variables were mapped and used in the analysis of inventory data currently being collected. These variables were Area, Land Forms, and Geological Features. When examining the data using the simple correlation procedure, the data were stratified on each of these three variables. When multiple linear regression and factor analysis were used to examine the data, these variables were assumed to be part of the set of observed variables. Areas represented six divisions of the High Knob Planning Unit -- each division being a different mountain or ridge of mountains. These divisions were felt necessary because of the size of the Planning Unit and the change in the environment that seemed to occur from one end of the Planning Unit to the other. Land Forms, the fifth variable in the Forest Service E.M.U., identifies benches, coves, flood plains, strip mines, or slopes (elevation and aspect). Geological Features describe the geological formations in the cell.

Results and Discussion

A simple linear correlation coefficient matrix was developed for the twelve variables being observed by functional staffs by using the 3293 sampling units on the High Knob Unit. The largest correlation coefficient was -0.7834 between Area and Wildlife Featured Species. This high value can be explained. Wildlife Featured Species are designated on large blocks of land that happened to correspond closely to the High knob Unit division into Areas. For another unit with another division into Areas, this correlation coefficient might be lower. Therefore, no definite conclusions can be made. The remaining correlations were all low, the largest being -0.3975 between Soil Source and Soil Texture (Table 1). Assuming that all data observed are important to Unit Planning, this analysis indicated that all variables currently being observed should

Table 1. Correlation Coefficients Matrix of the Inventoried Data on the High Knob Planning Unit.

	Land Forms	Soil Source	Soil Texture	Water Regime	Soil Modif.	Slope
Land Forms	1.0000	0.1111	0.0310	0.2966	-0.0877	0.0698
Soil Source		1.0000	-0.3975	0.2155	0.0451	0.2123
Soil Texture			1.0000	-0.0609	-0.0597	-0.1663
Water Regime				1.0000	0.2665	0.0264
Soil Modif.					1.0000	0.1035
Slope						1.0000
Geology						
Site Index						
Stand Cond.						
Forest Type						
W/L sp.						
Area						

Table 1. Correlation Coefficients Matrix of the Inventoried Data on the High Knob Planning Unit (continued).

	Geology	Site Index	Stand Cond.	Forest Type	W/L Species	Area
Land Forms	0.1614	-0.0368	0.0327	-0.0192	-0.0397	-0.0322
Soil Source	0.1253	0.1025	-0.0867	0.0347	-0.1854	0.2945
Soil Texture	-0.0820	-0.0140	0.0022	0.0812	0.2067	-0.2579
Water Regime	0.1189	0.0287	-0.0149	-0.0189	0.0747	-0.1468
Soil Modif.	-0.1247	0.0209	0.0013	-0.0447	-0.0059	-0.0469
Slope	-0.0302	0.1912	-0.1333	0.0801	-0.0089	0.0260
Geology	1.0000	-0.0453	0.1130	0.0536	-0.3272	0.1475
Site Index		1.0000	-0.1192	0.0714	0.1010	-0.0698
Stand Cond.			1.0000	-0.0623	-0.1105	0.0150
Forest Type				1.0000	0.1351	-0.1449
W/L sp.					1.0000	-0.7834
Area						1.0000

be retained. If two variables are closely associated (high correlation coefficient), then one can be eliminated because observing both adds little information.

Multiple linear regression technique were used to examine the relationship between individual variables and combination of other variables. Except Area, Land Forms, and Geology, each of the twelve variables now observed by functional staffs was, in turn, taken as the dependent variable. Area, Land Forms, and Geology were not dependent variables because they were not inventoried and thus their association with a combination of the other variables is of no importance in this study. Also, each variable was expressed only to the first power; no higher powers of variables or transformations of variables were considered.

When Soil Source was the dependent variable, the first independent variable to enter this regression equation was Soil Texture. The R-square value was only 0.1580. The next variable to enter the equation was Area and the multiple correlation coefficient was 0.4444 (R-square of 0.1975). The R-square value after all eleven independent variables entered the regression equation was 0.3026. The conclusion was that Soil Source could not be adequately described as a linear combination of the other variables.

Using each of the remaining variables as the dependent variable produced similar results and conclusions. The results are summarized in Table 2.

Factor analysis was used to help identify a set of factors that can supply essentially the same information as the nine variables needed for Unit Planning. Land Forms and Geology were included in the analysis but Area was not used as a variable in the initial factor analyses. Nine factors were extracted from the correlation matrix. This is little savings in number of variables when the initial set contained eleven variables. This savings seems even less impressive when only 43.36 percent of the total variance was explained by the nine factors. The value of the communalities for this data ranged from 0.1757 to 0.6672. These values are relatively low when the communality for a perfect prediction is one.

The unrotated factor loadings matrix was rotated by two methods, oblimin bi-quartimin and orthogonal varimax. There are no criteria for evaluating the results of the oblimin rotation but trends in the results could be examined. The rotated factor matrix by oblimin method is presented in Table 3. Comparing Table 3 with Table 4, which is the rotated factor matrix by the orthogonal method, the matrices show similar trends. Using Comrey's (1973) criteria for evaluating the orthogonally rotated matrix and discussing the results with Dr. Schulman, the conclusion was that no gains could be made by using factor analysis on this

Table 2. Summary of Multiple Linear Regression Analysis on Inventoried Data for the High Knob Planning Unit.

Dependent Variable	First Step		Second Step		R^2 Eleven Indep. Var.
	Independent Variable	R^2	Independent Variable	R^2	
Soil Source	Soil Texture	0.1580	Area	0.1975	0.3026
Soil Texture	Soil Source	0.1580	Area	0.1797	0.2041
Water Regime	Land Forms	0.0880	Soil Modif.	0.1742	0.2540
Soil Modif.	Water Regime	0.0710	Land Forms	0.1015	0.1474
Slope	Soil Source	0.0451	Site Index	0.0741	0.1185
Site Index	Slope	0.0365	W/L sp.	0.0471	0.0653
Stand Cond.	Slope	0.0178	W/L sp.	0.0302	0.0602
Forest Type	Area	0.0210	Slope	0.0280	0.0583
W/L sp.	Area	0.6137	Geology	0.6595	0.6768

Table 3. Rotated Factor Matrix (Oblimin Bi-quartimin Method) with Communalities for the High Knob Planning Unit.

Variable	Factor				
	1	2	3	4	5
Land Forms	-0.0227	-0.0416	-0.0394	-0.0455	-0.1218
Soil Source	0.6445	-0.1042	-0.2434	0.0317	0.0922
Soil Texture	-0.6578	0.0578	-0.1112	-0.0031	0.1301
Water Regime	0.1204	0.0129	0.1150	0.4889	0.2031
Soil Modif.	0.0002	0.0397	-0.0192	0.6785	-0.0913
Slope	0.1470	-0.0046	0.1075	0.0827	-0.5775
Geology	0.0460	-0.6141	0.1776	-0.0750	0.0116
Site Index	0.0134	0.0332	-0.0850	-0.0029	-0.0932
Stand Cond.	-0.0261	-0.0927	-0.0970	0.0265	0.0528
Forest Type	-0.0273	0.0038	-0.0285	-0.0248	-0.0451
W/L sp.	-0.1547	0.6222	0.1133	-0.0080	0.0403

Table 3. Rotated Factor Matrix (Oblimin Bi-quartimin Method) with Communalities for the High Knob Planning Unit (continued).

Variable	Factor				Communality
	6	7	8	9	
Land Forms	0.7096	-0.0606	-0.0440	0.0397	0.4610
Soil Scurce	0.1178	0.0836	0.0853	-0.0976	0.5523
Soil Texture	0.0990	0.0498	0.1056	-0.0456	0.4801
Water Regime	0.4665	0.0410	-0.0361	-0.0629	0.6672
Soil Modif.	-0.0674	-0.0146	-0.0261	0.0483	0.4791
Slope	0.1235	0.2083	0.0873	-0.1029	0.4025
Geology	0.1014	0.0302	0.1410	0.1058	0.5039
Site Index	-0.0522	0.4659	0.0188	-0.0638	0.2322
Stand Cond.	0.0273	-0.0756	-0.0326	0.3858	0.1757
Forest Type	-0.0432	0.0052	0.5203	-0.0184	0.2658
W/L sp.	0.0466	0.0907	0.1755	-0.0382	0.5501

Table 4. Rotated Factor Matrix (Orthogonal Varimax Method) with Communalities for the High Knob Planning Unit.

Variable	Factor				
	1	2	3	4	5
Land Forms	-0.0027	-0.0776	0.6528	-0.1008	-0.0230
Soil Source	0.6605	-0.1326	0.1723	0.0531	0.0835
Soil Texture	-0.6457	0.0870	0.0726	-0.0333	0.1264
Water Regime	0.1536	-0.0135	0.5832	0.4498	0.0165
Soil Modif.	0.0362	0.0592	-0.0384	0.6800	-0.0498
Slope	0.1832	0.0050	0.0638	0.0867	0.0854
Geology	0.0754	-0.6219	0.1776	-0.1028	0.1403
Site Index	0.0397	0.0512	-0.0201	0.0130	0.0660
Stand Cond.	-0.0396	-0.1050	0.0147	0.0094	-0.0590
Forest Type	-0.0296	0.0106	-0.0142	-0.0339	0.5044
W/L sp.	-0.1960	0.6172	0.0618	-0.0152	0.2218

Table 4. Rotated Factor Matrix (Orthogonal Varimax Method) with Communalities for the High Knob Planning Unit (continued).

Variable	Factor				Communality
	6	7	8	9	
Land Forms	0.1059	-0.0523	0.0497	0.0416	0.4610
Soil Source	0.0317	0.1341	0.1379	-0.1448	0.5523
Soil Texture	-0.1393	0.0275	0.1074	-0.0389	0.4801
Water Regime	-0.2020	0.0943	-0.2155	-0.0676	0.6672
Soil Modif.	0.0841	0.0069	0.0204	0.0180	0.4791
Slope	0.5280	0.2228	0.0127	-0.1464	0.4025
Geology	-0.0654	0.0080	-0.1583	0.1423	0.5039
Site Index	0.1161	0.4404	-0.0063	-0.1250	0.2322
Stand Cond.	-0.0760	-0.1207	0.0007	0.3728	0.1757
Forest Type	0.0460	0.0608	-0.0048	-0.0569	0.2658
W/L sp.	-0.0711	0.1378	-0.2166	-0.0807	0.5507

set of variables. Similar results were produced when the data were stratified on Area, Land Forms, and Geological Features and the analysis repeated.

The conclusion was that the present set of variables cannot be reduced and still provide the same information. The results could be due to the type of data used in the analysis. Use of the discrete grid data as sampling units instead of using continuous variables could have an effect on the results. The procedures used are for continuous data. Among the data set analyzed, in most cases the variables were categorized on a uniform interval. An exception is slope where one category was 51 - 60 percent and another was 61 plus percent. The lack of uniform intervals makes the assumption of continuous variables even less plausible. Such variables as Forest Type, that describe the forest overstory species composition, cannot be considered continuous with any real justification because the number assigned to each category is arbitrary and renumbering could produce a different outcome of the analysis.

The grid cells were used as sampling units in the analysis because actual field observations were not available. Before concluding that no reduction in the number of variables can be made, the variables should be examined using actual field data. The results may be different from this study's results because of the problems in the data discussed above.

In this study only linear associations were examined, there is also the possibility that nonlinear association may exist. Techniques for examining categorized data might also be used, because the U.S. Forest Service puts most of their information into discrete categories and the techniques used in this study were for continuous data.

A FRAMEWORK FOR OPTIMAL ALLOCATION OF INVENTORY RESOURCES FOR MULTIPLE-USE PLANNING

In the past, the basic approach to multi-resource inventories has generally been to collect large amounts of information from all possible sources. This process has several shortcomings. First, the information gathered may or may not be appropriate to make the decisions necessary for planning. Second, the cost of data collection may be higher than necessary because not all the information is used.

With the rising cost of inventories and the more complex decisions that must be made, inventories with well defined objectives are necessary. According to Dyer(1974) there are two basic approaches to develop an inventory system for resource planning. The first starts with the decision-maker and develops an inventory system that collects the information necessary for making decisions. This approach eliminates the cost of gathering data that are of no interest.

The second approach begins with the inventory design and progresses towards the decision-maker. This approach does not necessarily provide the manager with the best information for decision making. Also this type of inventory design tends to produce some data that are of no interest.

The first of these two approaches to inventory design is the best approach from the decision makers' point of view. Using this approach, Hamilton (1975) listed the following four steps that should be followed to design a resource inventory: (1) state the objective; (2) define the population to be sampled and data to be collected; (3) state the sampling precision required; and (4) select a sampling design to collect the data. These steps basically depend on the answer to the questions: (1) What decisions are to be made with the aid of the inventoried data?; (2) What data are needed to make the decisions?; and (3) What is the cost involved with making decisions from the data?

Questions (1) and (2) should be reasonably straight forward to the decision maker. The answer to question three is of great importance in determining the precision requirements of the inventory. In the past, resource managers have used rules-of-thumb to determine sampling intensity. In many cases a given percentage of the population was observed, which often led to inventories that were too intensive or not intensive enough depending on the variation in the population and the use of the data. Past experience was also used to set precision levels, but experience is of little use unless the variance of the population being sampled and the use of the inventory data are the same as those for which the experience was gained.

With the tremendous increase in sampling costs and the tightening of most budgets, determination of the sample size is very important. In this study a procedure to determine optimal sample size for multiple-use inventories as a function of the value of the information, rather than a function of an arbitrary set level of precision, was formulated and evaluated. This procedure was first introduced by Blythe (1945) for a single inventory and single decision. Blythe investigated a problem of determining sample size for sampling logs on timber sales of the U.S. Forest Service. The sample conducted by the Forest Service determined the price paid for the timber sale. The Forest Service wanted to spend as little as possible on the sample (i.e., conduct a small sample) but the purchaser wanted the probability of an overestimate to be very small to reduce the probability of overpayment for the timber. Blythe's suggestion was to sample to the point where the cost of sampling the next log was equal to the reduction in the risk of overpayment in monetary terms. The major problem with this approach is expressing the risk of overpayment in terms of dollars. The approach is essentially one of minimizing a cost of sampling plus a monetary loss function.

Yates(1960) examined the idea of minimizing a cost-loss or cost-risk function to determine the optimal sampling level. A basic problem with this method is determining the expected loss function. The decision-maker may know little of the loss expected from making decisions with data containing sampling error. Expressing this loss in monetary terms is a problem in itself. Cochran (1963) and Baron(1973) also examined the theory of this idea.

Grundy, Healy, and Rees (1956) worked with a related problem that arises in management. When a new process is introduced, management usually conducts experiments to determine whether to invest capital in the new process or to continue using the old process. They suggested determining the optimal sampling level by minimizing the total risk function. That is, minimize the cost of experimentation and the expected loss due to making a wrong decision.

Smith (1965) suggested an approach to determine the optimal size for quality control. His problem related to sampling parts to determine whether to accept or reject a lot. Smith presented a model that accounts for costs of inspection and the monetary consequence of making a wrong decision from a sample.

Hamilton (1970) evaluated four methods for determining sample size: (1) maximum utility; (2) marginal utility; (3) inventory control; and (4) minimum cost plus loss. He concluded that the first three methods were all variations of the cost-loss function. Hamilton used two problems in his evaluation, one a timber sales appraisal problem and the other an allowable cut determination problem. Combinations of three types of loss functions, squared error loss, absolute value loss, and strictly linear loss, were evaluated but no one combination was found to be best for general use. Hamilton concluded that expressing the loss function was the most difficult task because the decision-maker knows little of monetary loss incurred from making decisions.

The approach used in this study to design a sampling scheme for multiple-use planning will begin with the decision maker and progress towards a user oriented system. This approach was chosen so that cost would be minimized by collecting only the information needed by the decision makers. In this study, instead of using some rule-of-thumb to set the precision, the minimization of a cost-loss function associated with the multiple-use decisions was used. This approach determines the sample size based upon the value of the information rather than some arbitrary level of precision that may or may not be appropriate for the decisions being made from the data. According to Hamilton (1975) three basic needs must be met to use the cost-loss approach. First, a cost of sampling function must be developed. Second, the distribution of possible outcomes of the inventory must be evaluated. Third, a loss function that describes the monetary losses that are expected to occur at different levels of sampling must be determined.

To complete the development of an inventory system an efficient sampling scheme must be chosen. The actual sample design to be employed will be highly dependent on local conditions, so no attempt will be made to develop this particular phase of multiple-use inventory.

The main intent of this section of the study was to develop and evaluate the cost-loss approach to determine sampling size for multiple-use inventories. During the remainder of this section the general methodology for developing cost and expected loss functions for multiple-use decisions is discussed. A case study on the High Knob Unit of the Jefferson National Forest was used to evaluate the methodology developed and will be discussed in the next chapter.

Cost-Loss Functions for Inventories Where One Attribute is Observed

Timber inventory will be used as an example of the application of a cost-loss minimization procedure for sample intensity determination. Cochran (1963) presented a development of the cost-loss minimization procedure in which he assumed the variance of the estimate was inversely proportional to the sample size, n, and the cost of sampling was a linear function of n. These same assumptions will be applied in this example. Also in this section it will be assumed that only one attribute is observed at each sampling location of each inventory. For this study, the monetary loss incurred from making a decision with data containing an overestimate or underestimate of the same magnitude will be assumed equal; that is, symmetric loss functions will be assumed.

For example, a land owner may harvest his timber stand at the age that maximizes mean annual increment (MAI), which we will also assume is the age that maximizes net revenue. When deciding whether or not to harvest a given timber stand, an estimate of the stand age is needed. Therefore, an inventory is conducted to estimate stand age. Making a harvesting decision with the estimated age may result in a monetary loss to the land owner. Suppose the stand is actually younger than the estimated age, a monetary loss would be incurred from loss in growth potential because maximum MAI was not obtained. If the stand is actually older than the estimated age, the monetary loss would be incurred from loss in potential growth because of harvesting after the point of maximum MAI.

Case I: Single Inventory, Single Decision

The simplest application of the cost-loss procedure occurs when one decision is made with information from one inventory. Let:

$l(z)$ = the monetary loss that will be incurred in a decision through an error of amount z in the estimate needed to make the decision.

$f(z,n)$ = the frequency distribution of z, which for a specific sampling method will depend on the sample size, n.

Hence the expected loss is:

$$L(n) = \int l(z)f(z,n)dz$$

If l(z) is defined as the simple loss function

$$l(z) = \lambda z^2$$

where λ = a constant

then

$$\begin{aligned} L(n) &= \int \lambda z^2 f(z,n)dz \\ &= \lambda E(z^2) \end{aligned}$$

Let the error, z, be defined as the difference between the estimated value and the true value, that is:

$$z = \hat{\bar{Y}} - \bar{Y}$$

where

$\hat{\bar{Y}}$ = estimated value of the mean of Y

$\hat{\bar{Y}}$ = the true mean of Y

then

$$\begin{aligned} L(n) &= \lambda E(z^2) \\ &= \lambda E((\hat{\bar{Y}} - \bar{Y})^2) \\ &= \lambda V(\hat{\bar{Y}}) \end{aligned}$$

Where $V(\hat{\bar{Y}})$ is the variance of the estimate of $\bar{Y}$ and the $E(\hat{\bar{Y}})$ equals $\hat{Y}$. If simple random sampling with replacement is assumed:

$$L(n) = \frac{\lambda S^2}{n}$$

where

S^2 = population variance of the sampling units

The larger the sample size the lower the expected loss will be but the higher the cost of sampling. A simple cost function is:

$$C(n) = c_0 + c_1 n$$

where

c0 = cost of overhead

c1 = cost of sampling one unit

n = number of sampling units

The cost-loss function is defined as:

$$C(n) + L(n) = c_0 + c_1 n + \frac{\lambda S^2}{n}$$

Using differentiation, the cost-loss function is minimized when n is:

$$n = \sqrt{[\lambda S^2/c_1]}$$

Simple random sampling with replacement was assumed to simplify the notation. In most applications sampling is performed without replacement; however, the finite population correction is often near one and, in those cases, it is ignored.

In this study, to determine a value for λ for a given decision, the decision-maker was consulted. He was questioned as to what the monetary loss would be from making a decision with data containing varying amounts of error.

In the squared error loss function

$$l(z) = \lambda z^2$$

$l(z)$ is in dollars, and z is in the unit of measure of the inventory. If the squared error loss function is appropriate, a linear relationship exists between monetary loss (in dollars) and z^2 (in units^2), and λ is the slope coefficient of this straight line. Linear regression can be used to estimate λ. The dependent variable is $l(z)$, which is estimated by the decision-maker, and the independent variable is z^2. The regression equation is forced through the origin since it is assumed that when there is zero error in the data there is zero loss incurred in using the data for decision making.

The example used earlier regarding the determination of harvesting a timber stand based upon a sample estimate of age will be continued. Assume, the decision-maker estimated the following monetary losses for this decision at the specified sampling errors:

z	$l(z)$
0	0
1	10
2	40
3	90
4	160
5	250

We can express $l(z)$ as simply being proportional to z^2, i.e. $l(z)=\lambda z^2$ and estimate λ graphically as in Figure 1 or as the linear regression slope coefficient. Applying regression through the origin, λ was determined to be 10. Further, assume S^2 is 100 years^2, c0 is 100 dollars, and c1 is 10 dollars and that simple random sampling without replacement will be employed. With a linear sampling cost function, the cost-loss function to be minimized is

$$c_0 + c_1 n + \frac{\lambda S^2}{n}$$

The value of n that minimizes this equation is

$$n = \sqrt{}[\lambda S^2/c_1]$$

Substituting assumed values for this example results in

$$n = \sqrt{}[10(100)/10]$$
$$= 10$$

If 10 sampling units are observed, the total cost of the inventory and of the monetary cost expected to occur from making a decision with the information is 300 dollars.

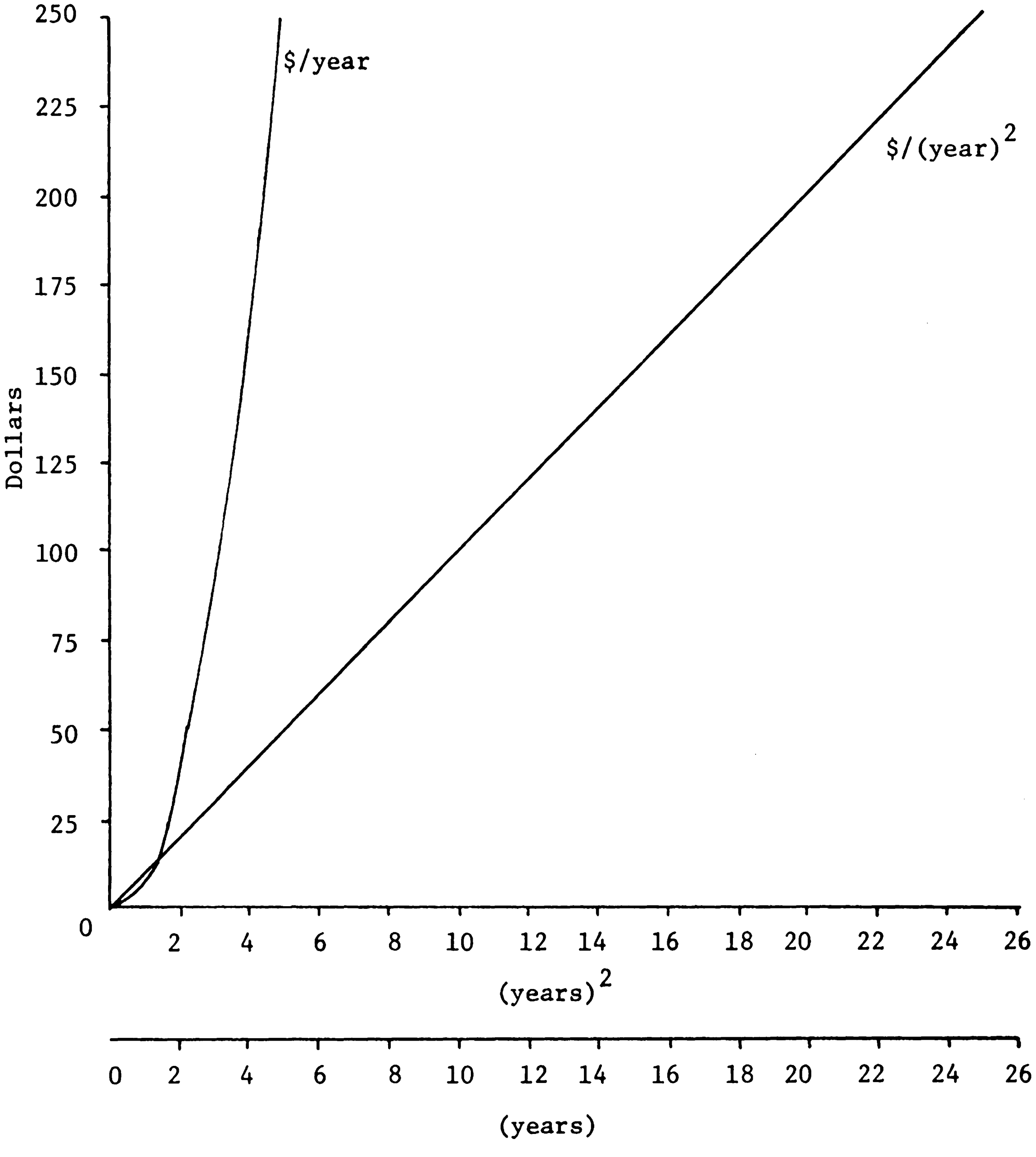

Figure 1. The estimated loss in dollars from making a timber harvesting decision with various amounts of sampling error (years) in the age of the stand. Also shown is a curve for the loss per squared error (years^2).

This is the minimum cost plus expected loss that can be achieved under the stated assumptions. A graphic representation of the cost, loss, and cost plus loss functions for this example is shown in Figure 2.

Case II: Single Inventory, Multiple Decisions

In most multi-resource situations the information from an inventory is used to make several decisions. The theory developed above will be extended to the case of one inventory to obtain information for several decisions. Suppose m decisons are to be made from one sample, then there would be m possible loss functions and the cost plus expected loss function to minimize would be:

$$C(n) + L_1(n) + L_2(n) + . . . + L_m(n)$$

If a linear cost function, simple random sampling with replacement, and squared error loss are assumed, the cost-loss function to be minimized may be expressed as

$$c_0 + c_1 n + \frac{\lambda_1 S^2}{n} + \frac{\lambda_2 S^2}{n} + . . . + \frac{\lambda_m S^2}{n}$$

Differentiating with respect to n and setting equal to zero, the value of n that minimize the cost-loss function can be

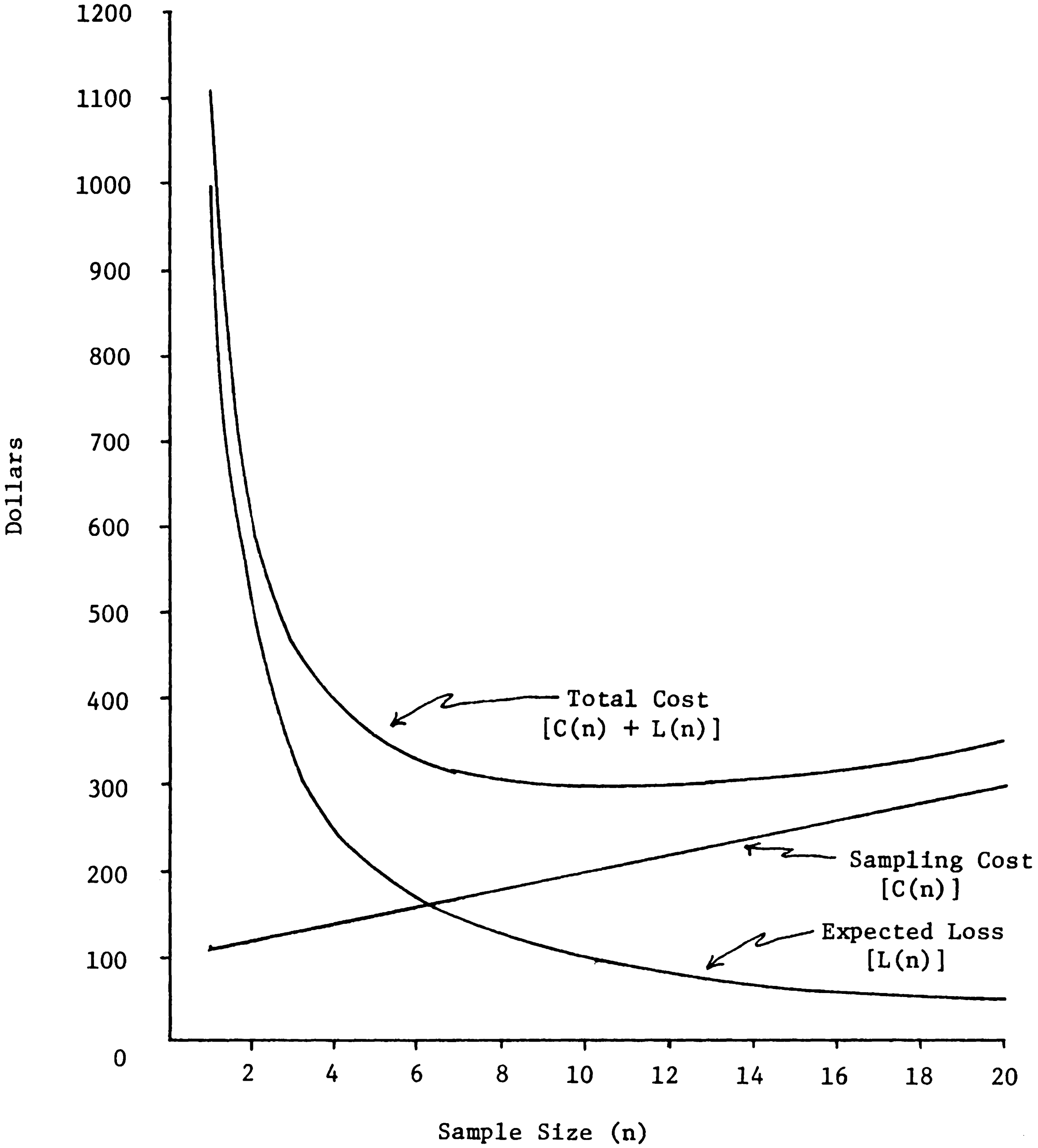

Figure 2. Graphic representation of the sampling cost, expected loss, and total cost function for the timber harvesting decision example.

obtained as follows:

$$c_1 - \lambda_1 S^2 n^{-2} - \lambda_2 S^2 n^{-2} - \ldots - \lambda_m S^2 n^{-2} = 0$$

$$S^2 n^{-2} \quad (\lambda_1 + \lambda_2 + \ldots + \lambda_m) = c_1$$

$$n^{-2} = c_1/[S^2(\lambda_1 + \lambda_2 + \ldots + \lambda_m)]$$

$$n = \sqrt{}[S^2(\lambda_1 + \lambda_2 + \ldots + \lambda_m)/c_1]$$

$$n = \sqrt{}[S^2\lambda'/c_1]$$

where:

$$\lambda' = \sum_{i=1}^{m} \lambda_i$$

Case III: Multiple Inventories, Single Decision

Another extension of the simple one inventory-one decision case occurs where information from several inventories is necessary to make one decision. Suppose there are q independent inventories and one decision to be made from data from all q inventories. The cost-loss function to be minimized may be written

$$C(n_1, n_2, \ldots, n_q) + L(n_1, n_2, \ldots, n_q)$$

where:

$$\begin{aligned} L(n_1, n_2, \ldots, n_q) &= \int \lambda z^2 f(z, n_1, n_2, \ldots, n_q)dz \\ &= \lambda E(z^2) \\ &= \lambda V(\hat{\bar{Y}}) \end{aligned}$$

where:

$$E(z^2) = V(\hat{\bar{Y}})$$

$\hat{\bar{Y}}$ - is estimated from data from all inventories.

The symbol $\hat{Y}$ represents an estimate of the decision variable which is an integration and synthesis of information from the q inventories. For convenience and because no general theory for combining information from several inventories for multiple-use decisions exists, suppose $\hat{Y}$ is a linear function of the independent $\hat{\bar{y}}$'s, that is:

$$\hat{\bar{Y}} = k_1\hat{\bar{y}}_1 + k_2\,\hat{\bar{y}}_2 + \ldots + k_q\hat{\bar{y}}_q$$

where:

ki (i=1,2,...,q) is a weight given the ith inventory item in making the decision. (Note that the ki weights have units such that ki times $\hat{\bar{y}}i$ is in units of $\hat{Y}$.)

Then:

$$V(\hat{\bar{Y}}) = k_1{}^2 V(\hat{\bar{y}}_1) + k_2{}^2 V(\hat{\bar{y}}_2) + \ldots + k_q{}^2 V(\hat{\bar{y}}_q)$$

For simple random sampling with replacement, the variance of $\hat{\bar{Y}}$ may be expressed

$$V(\hat{\bar{Y}}) = (k_1{}^2 S_1{}^2/n_1) + (k_2{}^2 S_2{}^2/n_2) + \ldots + (k_q{}^2 S_q{}^2/n_q)$$

Further assuming a linear cost function for all inventories, the cost-loss function can be written as:

$$c_{01} + c_{11}n_1 + c_{02} + c_{12}n_2 + \ldots + c_{0q} + c_{1q}n_q + \lambda[(k_1{}^2 S_1{}^2/n_1) + (k_2{}^2 S_2{}^2/n_2) + \ldots + (k_q{}^2 S_q{}^2/n_q)]$$

Taking the q partial derivatives with respect to the ni's (i=1,2,...,q) and setting them equal to zero, the values of the ni's that minimize the cost-loss function can be obtained as follows:

$$\frac{d}{dn_1} = c_{11} - \lambda\, k_1{}^2 S_1{}^2 n_1{}^{-2} = 0 \quad \rightarrow \quad n_1 = \sqrt{[\lambda k_1{}^2 S_1{}^2/c_{11}]}$$

$$\frac{d}{dn_2} = c_{12} - \lambda\, k_2{}^2 S_2{}^2 n_2{}^{-2} = 0 \quad \rightarrow \quad n_2 = \sqrt{[\lambda k_2{}^2 S_2{}^2/c_{12}]}$$

$$\vdots$$

$$\frac{d}{dn_q} = c_{1q} - \lambda k_q{}^2 S_q{}^2 n_q{}^{-2} = 0 \quad \rightarrow \quad n_q = \sqrt{[\lambda k_q{}^2 S_q{}^2/c_{1q}]}$$

Case IV: Multiple inventories, Multiple Decisions

The basic framework can be extended to the case of several independent inventories to make several decisions. Suppose there are q independent inventories to be conducted and m decisions to be made and that the data gathered in each of the q inventories will be used as an aid in making each of the m decisions. The q linear cost functions may be written as:

$$\begin{aligned} &c_{01} + c_{11}n_1 \\ &c_{02} + c_{12}n_2 \\ &\vdots \\ &c_{0q} + c_{1q}n_q \end{aligned}$$

The m expected loss functions can be obtained as

$$\begin{aligned} L_1(n_1, n_2, \ldots, n_q) &= \int l_1(z_1)f_1(z_1, n_1, n_2, \ldots, n_q)dz_1 \\ L_2(n_1, n_2, \ldots, n_q) &= \int l_2(z_2)f_2(z_2, n_1, n_2, \ldots, n_q)dz_2 \\ &\vdots \\ L_m(n_1, n_2, \ldots, n_q) &= \int l_m(z_m)f_m(z_m, n_1, n_2, \ldots, n_q)dz_m \end{aligned}$$

Let

$$z_1 = \hat{\bar{Y}}_1 - \bar{Y}_1$$

and suppose $\hat{\bar{Y}}$ is a linear function of the independent $\hat{\bar{y}}$'s, therefore:

$$\hat{\bar{Y}}_1 = k_{11}\hat{\bar{y}}_1 + k_{12}\hat{\bar{y}}_2 + \ . \ . \ . + k_{1q}\hat{\bar{y}}_q$$

where:

k ji (i=1,2,...,q; j=1,2,...,m) is the weight of the i th inventory item in making the j th decision.

Substituting for the first expected loss function, we obtain

$$\begin{aligned} L_1(n_1, n_2, \ . \ . \ ., n_q) &= \int \lambda_1 z_1^2 f_1(z_1, n_1, n_2, \ . \ . \ ., n_q) dz_1 \\ &= \lambda_1 E(z_1^2) \\ &= \lambda_1 V(\hat{\bar{Y}}_1) \\ &= \lambda_1 [k_{11}^2 V(\hat{\bar{y}}_1) + k_{12}^2 V(\hat{\bar{y}}_2) + \ . \ . \ . \\ &\qquad + k_{1q}^2 V(\hat{\bar{y}}_q)] \end{aligned}$$

If simple random sampling with replacement is assumed, then:

$$\begin{aligned} L_1(n_1, n_2, \ . \ . \ ., n_q) &= \lambda_1 [(k_{11}^2 s_1^2/n_1) + (k_{12}^2 s_2^2/n_2) \\ &\qquad + \ . \ . \ . + (k_{1q}^2 s_q^2/n_q)] \end{aligned}$$

For linear cost functions, simple random sampling with replacement, and squared error loss functions, the cost-loss function is:

$$C_{01} + C_{11}n_1 + C_{02} + C_{12}n_2 + \ldots + C_{0q} + C_{1q}n_q + \lambda_1[(k_{11}{}^2 S_1{}^2/n_1)$$
$$+ (k_{12}{}^2 S_2{}^2/n_2) \ldots + (k_{1q}{}^2 S_q/n_q)] + \ldots + \lambda_m[(k_{m1}{}^2 S_1{}^2/n_1)$$
$$+ (k_{m2}{}^2 S_2{}^2/n_2) + \ldots + (k_{mq}{}^2 S_q{}^2/n_q)]$$

The above expression may be simplified by making use of matrix notation as follows:

$$C_{01} + C_{11}n_1 + C_{02} + C_{12}n_2 + \ldots + C_{0q} + C_{1q}n_q + \lambda_1[\underline{k}_1{}' Q\underline{k}_1]$$
$$+ \lambda_2[\underline{k}_2{}' Q\underline{k}_2] + \ldots + \lambda_m[\underline{k}_m{}' Q\underline{k}_m]$$

where:

$\underline{k}j$ = a vector of weights used in making the j th decision. (j=1,2,...,m)

Q = a matrix where:

$\frac{S_i{}^2}{n_i}$ = diagonal elements. (i=1,2,...,q)

0 = otherwise

Taking the q partial derivatives with respect to the ni's (i=1,2,...,q) and setting them equal to zero, the following set of equations is obtained:

$$\frac{d}{dn_1} = c_{11} - \lambda_1 k_{11}^2 S_1^2 n_1^{-2} - \lambda_2 k_{21}^2 S_1^2 n_1^{-2} - \ldots - \lambda_m k_{m1}^2 S_1^2 n_1^{-2} = 0$$

$$\frac{d}{dn_2} = c_{12} - \lambda_1 k_{12}^2 S_2^2 n_2^{-2} - \lambda_2 k_{22}^2 S_2^2 n_2^{-2} - \ldots - \lambda_m k_{m2}^2 S_2^2 n_2^{-2} = 0$$

$$\vdots$$

$$\frac{d}{dn_q} = c_{1q} - \lambda_1 k_{1q}^2 S_q^2 n_q^{-2} - \lambda_2 k_{2q}^2 S_q^2 n_q^{-2} - \ldots - \lambda_m k_{mq}^2 S_q^2 n_q^{-2} = 0$$

Solving for the values of the ni's that minimize the cost-loss function one obtains:

$$n_1 = \sqrt{}(S_1^2/c_{11})(\lambda_1 k_{11}^2 + \lambda_2 k_{21}^2 + \ldots + \lambda_m k_{m1}^2)$$

$$n_2 = \sqrt{}(S_2^2/c_{12})(\lambda_1 k_{12}^2 + \lambda_2 k_{22}^2 + \ldots + \lambda_m k_{m2}^2)$$

$$\vdots$$

$$n_q = \sqrt{}(S_q^2/c_{1q})(\lambda_1 k_{1q}^2 + \lambda_2 k_{2q}^2 + \ldots + \lambda_m k_{mq}^2)$$

In summary, the cost-loss functions were developed assuming independent inventories, simple random sampling with replacement, and linear cost functions. It was assumed that only one attribute was observed in each inventory. A symmetric squared error loss function was assumed.

Cost-Loss Functions for Inventories Where Multiple Attributes are Observed

In the development of the cost-loss framework in the previous section it was assumed that only one attribute was observed in each inventory. For most inventories on forested land, more than one attribute is observed at each sampling location. The theory developed in the previous section will be extended to account for multiple attributes being observed in each of q inventories to make m decisions.

When observing more than one attribute at each sampling location of an inventory there are four basic ways to determine an optimal sample size. First, one may assume a single variance equal to the largest variance of the individual attributes for that inventory. The optimal sample size can then be obtained from procedures discussed in the previous section. If the largest variance is used for computing sample sizes, all attributes, except the attribute with largest variance, will be over sampled thus increasing the sampling cost. The expected loss will be reduced as a result of these larger sampling intensities but not by the magnitude of the increase in sampling cost.

A second method of determing an optimal sample size is to use the variance of the most important attribute in the inventory as the variance for all attributes. Methods assuming one attribute per inventory can then be used to determine the optimal sample sizes. The major shortcoming of this method is that some attributes may be over sampled while others may be under sampled.

The optimal sample size that minimizes a cost-loss function can also be determined if one assumes that each attribute will be observed at each sampling location and that the variance of each attribute will be considered in the calculation of the sampling intensity. This approach will determine a sample size for an inventory that reflects both the variance of each attribute and the weight of importance of the attributes, without necessitating calculation of a sample size for each attribute.

A fourth method of determining sample size is to determine the sample size for each attribute instead of a sample size for each inventory. This approach will be formulated in this section. Suppose there are q inventories, ai (i=1,2,...,q) attributes in the i th inventory, and m decisions, then the q linear cost functions are:

$$
\begin{array}{l}
c_{01} + c_{111}n_{11} + c_{112}n_{12} + \cdots + c_{11a_1}n_{1a_1} \\
c_{02} + c_{121}n_{21} + c_{122}n_{22} + \cdots + c_{12a_2}n_{2a_2} \\
\cdot \\
\cdot \\
\cdot \\
c_{0q} + c_{1q1}n_{q1} + c_{1q2}n_{q2} + \cdots + c_{1qa_q}n_{qa_q}
\end{array}
$$

Writing the above equations in more compact summation notation gives:

$$c_{0i} + \sum_{h=1}^{ai} c_{1ih} n_{ih} \qquad (i = 1, 2, \ldots, q)$$

Suppose that information from all attributes in the q inventories is used in each of the m decisions, then the m expected loss functions can be expressed as:

$$L_1(n_{11}, n_{12}, \ldots, n_{qa_q}) = \int l_1(z_1) f_1(z_1, n_{11}, n_{12}, \ldots, n_{qa_q}) dz_1$$
$$L_2(n_{11}, n_{12}, \ldots, n_{qa_q}) = \int l_2(z_2) f_2(z_2, n_{11}, n_{12}, \ldots, n_{qa_q}) dz_2$$
$$\vdots$$
$$L_m(n_{11}, n_{12}, \ldots, n_{qa_q}) = \int l_m(z_m) f_m(z_m, n_{11}, n_{12}, \ldots, n_{qa_q}) dz_m$$

Let:

$$z_1 = \hat{\bar{Y}}_1 - \bar{Y}_1$$

Suppose $\hat{\bar{Y}}$ is a linear function of the $\hat{\bar{y}}$'s of the attributes, then:

$$\hat{\bar{Y}}_1 = \underline{k}_{11}\hat{\underline{\bar{y}}}_1 + \underline{k}_{12}\hat{\underline{\bar{y}}}_2 + \ldots + \underline{k}_{1q}\hat{\underline{\bar{y}}}_q$$

where:

$\underline{k}$ ji's = a vector, ai in length, of weights of the i th inventory in making the j th decision. (i=1,2,...,q; j=1,2,...,m)

ai = the number of attributes observed in the i th inventory.

Thus the squared error loss function for the first decision is:

$$\begin{aligned} L_1(n_{11}, n_{12}, \ldots, n_{qa_q}) &= \int \lambda_1 z_1^2 f(z_1, n_{11}, n_{12}, \ldots, n_{qa_q}) dz_1 \\ &= \lambda_1 E(z_1^2) \\ &= \lambda_1 V(\hat{Y}_1) \\ &= \lambda_1 [\underline{k}_{11}' V(\hat{\underline{y}}_1)\underline{k}_{11} + \underline{k}_{12}' V(\hat{\underline{y}}_2)\underline{k}_{12} + \ldots \\ &\qquad + \underline{k}_{1q}' V(\hat{\underline{y}}_q)\underline{k}_{1q}] \end{aligned}$$

where $V(\hat{Y}i)$, $(i=1,2,\ldots,q)$ is a variance-covariance matrix for the i th inventory.

If simple random sampling with replacement is assumed $V(\hat{Y}i)$ is a matrix, Qi, where the diagonal elements equal:

$$S_{ih}^2/n_{ih} \qquad (h = 1,2, \ldots, a_i)$$

where Sih^2 is the variance of the hth attribute in the ith inventory. The off-diagonal elements equal:

$$S_{ih(il)}\left(\frac{2}{n_{ih} + n_{il}}\right) \qquad (h, l = 1, 2, \ldots, a_i;\ h \neq l)$$

and Sih(il) is the covariance between the hth and lth attributes of the ith inventory. Then:

$$L_1(n_{11},n_{12}, \ldots, n_{qa_q}) = \lambda_1[\underline{k}_{11}'Q_1\underline{k}_{11} + \underline{k}_{12}'Q_2\underline{k}_{12} + \ldots$$
$$+ \underline{k}_{1q}'Q_q\underline{k}_{1q}]$$

The cost-loss function (assuming linear cost functions, simple random sampling with replacement, and squared error loss functions) is

$$\sum_{i=1}^{q}(c_{0i} + \sum_{h=1}^{ai} c_{1ih}n_{ih}) + \sum_{j=1}^{m} (\lambda_j \sum_{i=1}^{q} \underline{k}_{ji}'Q_i\underline{k}_{ji})$$

Sets of equations and unknowns will be formed representing each inventory by taking the first partial derivative with respect to each n ih and setting the partial equal to zero.

Each set will contain equal numbers of equations and unknowns but no solution can be obtained analytically for the n ih's. The set of equations for the ith inventory is

$$
\begin{aligned}
\frac{d}{dn_{i1}} = {} & c_{1i1} + \lambda_1[-k_{1i1}{}^2 S_{i1}{}^2 n_{i1}{}^{-2} - 2k_{i1i}k_{1i2}S_{i1(i2)}(n_{i1}+n_{i2})^{-2} \\
& - \ldots - 2k_{1i1}k_{1ia_i}S_{i1(ia_i)}(n_{i1}+n_{ia_i})^{-2}] + \ldots \\
& + \lambda_m[-k_{mi1}{}^2 S_{i1}{}^2 n_{i1}{}^{-2} - 2k_{mi1}k_{mi2}(n_{i1}+n_{i2})^{-2} - \ldots \\
& -2k_{mi1}k_{mia_i}S_{i1(ia_i)}(n_{i1}+n_{ia_i})^{-2}] = 0
\end{aligned}
$$

.
.
.

$$
\begin{aligned}
\frac{d}{dn_{ia_i}} = {} & c_{1ia_i} + \lambda_1[-k_{1ia_i}{}^2 S_{ia_i}{}^2 n_{ia_i}{}^{-2} - 2k_{1i1}k_{1ia_i}S_{i1(ia_i)}(n_{i1}+n_{ia_i})^{-2} \\
& - \ldots - 2k_{1ia_i-1}k_{1ia_i}S_{ia_i-1(ia_i)}(n_{ia_i-1}+n_{ia_i})^{-2}] + \ldots \\
& + \lambda_m[-k_{mia_i}{}^2 S_{ia_i}{}^2 n_{ia_i}{}^{-2} - 2k_{mi1}k_{mia_i}S_{i1(ia_i)}(n_{i1}+n_{ia_i})^{-2} \\
& - \ldots -2k_{mia_i-1}k_{mia_i}S_{ia_i-1(ia_i)}(n_{ia_i-1}+n_{ia_i})^{-2}] = 0
\end{aligned}
$$

It may be possible to obtain an approximate solution by using iterative techniques such as the Newton-Raphson procedure for solving a system of nonlinear equations (Ralston, 1965), but these would require a large amount of computational effort.

Another approximation can be obtained if the covariances between attributes are assumed to be zero. The resulting equations are the same as those developed for q inventories with one attribute to make m decisions, except that there is a larger number of equations. These equations are

$$n_{11} = \sqrt{(S_{11}^2/c_{111})(\lambda_1 k_{111}^2 + \lambda_2 k_{211}^2 + \ldots + \lambda_m k_{m11}^2)}$$

$$n_{12} = \sqrt{(S_{12}^2/c_{112})(\lambda_1 k_{112}^2 + \lambda_2 k_{212}^2 + \ldots + \lambda_m k_{m12}^2)}$$

$$\vdots$$

$$n_{qa_q} = \sqrt{(S_{qa_q}^2/c_{1qa_q})(\lambda_1 k_{1qa_q}^2 + \lambda_2 k_{2qa_q} + \ldots + \lambda_m k_{mqa_q}^2)}$$

In practice all attributes are commonly observed at all sampling locations. This procedure is logical since the overhead cost and cost of initially establishing a sampling location are generally very large relative to the cost of observing a given attribute once at the location. If all attributes are to be observed at each sampling location, then all n ih's are equal (to ni). If the affects of the variance of each attribute on the sample intensity for an inventory is wanted without calculating a sample size for each attribute, the cost-loss function to minimize is:

$$\sum_{i=1}^{q} (c_{0i} + c_{1i} n_i) + \sum_{j=1}^{m} (\lambda_i \sum_{i=1}^{q} \underline{k}_{ji}' Q_i \underline{k}_{ji})$$

where Qi is a variance-covariance matrix and the diagonal elements are:

$$S_{ih}^2/n_i \qquad (h = 1,2, \ldots, a_i)$$

and off-diagonal elements equal 0.

Taking the q partial derivatives with respect to the ni's (i=1,2,...,q) and setting them to zero, the values of the ni's that minimize the cost-loss functions are found to be:

$$n_1 = \sqrt{\{[\lambda_1(k_{111}^2 S_{11}^2 + k_{112}^2 S_{12}^2 + \ldots + k_{11a_1}^2 S_{1a_1}^2)}$$

$$+ \ldots + \lambda_m(k_{mq1}^2 S_{q1}^2 + k_{mq2}^2 S_{q2}^2 + \ldots$$

$$+ k_{mqa_q}^2 S_{qa_q}^2)]/c_{11}\}$$

$$\vdots$$

$$n_q = \sqrt{\{[\lambda_1(k_{1q1}^2 S_{q1}^2 + k_{1q2}^2 S_{q2}^2 + \ldots + k_{1qa_q}^2 S_{qa_q}^2)}$$

$$+ \ldots + \lambda_m(k_{mq1}^2 S_{q1}^2 + k_{mq2}^2 S_{q2}^2 + \ldots$$

$$+ k_{mqa_q}^2 S_{qa_q}^2)]/c_{1q}\}$$

In summary, independent inventories conducted by simple random sampling were assumed when developing the cost-loss functions. The cost function was assumed to be linear and the variance of the estimate was assumed to be inversely proportional to the sample size. The loss functions were assumed to be squared error functions with the monetary loss for overestimation and underestimation of the same magnitude being assumed equal. Also the error was assumed to be a linear combination of the errors of the individual attributes. When the attributes are not assumed to be independent, it is not possible to obtain an analytical solution to the resultant set of equations.

The four cost-loss functions developed in the previous section are special cases of the cost-loss function developed for inventories where multiple attributes are observed. The cost-loss function for the multiple attribute case can be shown to be mathematically equivalent to each special case under the assumptions used to develop each of the four cases.

Decisions Made more than One Time

When management plans are made for large land areas it may be necessary to make any given decision many times. If these decisions are independent, then the expected loss is simply the sum of the expected losses for each time the decision is made. For example, if the first decision is made T1 times and the expected loss is the same for each independent decision the the expected total loss is

$$\sum_{t=1}^{T_1} L_1(n_{11}, n_{12}, \ldots, n_{qa_q})$$

In the special case of multiple inventories and multiple decisions this total expected loss is

$$T_1L_1(n_{11},n_{12}, \ldots, n_{qa_q}) = T_1\lambda_1[\underline{k}_{11}'Q_1\underline{k}_{11} + \underline{k}_{12}'Q_2\underline{k}_{12} + \ldots + \underline{k}_{1q}'Q_q\underline{k}_{1q}]$$

Assuming linear cost functions, simple random sampling with replacement, and squared error loss functions, the cost-loss function is:

$$\sum_{i=1}^{q}(c_{0i} + \sum_{h=1}^{a_i} c_{1ih}n_{ih}) + \sum_{j=1}^{m}(T_j\lambda_j\sum_{i=1}^{q}\underline{k}_{ji}'Q_i\underline{k}_{ji})$$

APPLYING THE COST-LOSS MINIMIZATION PROCEDURE: A CASE STUDY

The user-oriented approach to developing a sampling scheme and the minimization of cost-loss functions to determine sampling intensity were evaluated using a case study. The case study was conducted on the High Knob Planning Unit of the Jefferson National Forest. The three major steps in this evaluation were: 1) to list the decisions necessary for Unit Planning; 2) to determine the input information needed to make these decisions; and 3) to determine the optimal sampling intensity that will minimize the total cost of obtaining data plus expected cost involved in using the data to make decisions.

Decisions for a Unit Plan

The resource management decisions in a Unit Plan can be grouped into three production categories -- timber, wildlife, and recreation. All remaining functional areas in the Supervisor's Office act as support units for these three production areas. Inventories from both production and support function groups are, however, necessary to obtain the required information to make Unit Planning decisions.

Timber

Timber management decisions center around developing a list of forest stands that might be harvested. The list of stands is the nucleus for a timber management program. From this list, the stands for harvest are selected, and a more intensive inventory of these stand is made to determine volume information.

The decisions for timber management are listed below in chronological order. As the timber manager progresses through this list of decisions, he eliminates some forest stands from possible harvest, places restrictions on harvesting others, or decides to harvest others without restrictions. The timber management decisions are:

1. Will the U.S. Forest Service harvest timber on the Unit?
2. Are there restrictions that preclude timber harvesting on part or all of the Unit? (Examples of a totally restricted area for timber harvest are wilderness areas or legislated watersheds where no harvest is allowed under any circumstances.)
3. What timber stands within the Unit have timber products that could be scheduled for harvest?
4. Are there partial restrictions on the stand that would remove it from general timber producing land? (A stand with partial restrictions would be located in such areas as developed recreational areas, special

management areas, and around the Appalachian Trail. In these areas, a cut would be made only in case of disease, storm damage, insects, or maintenance.)

5. Are there limited restrictions that dictate the type of timber harvest? (Limited restrictions on type and method of harvest are placed on some stands because of slope, water quality, wildlife, productivity, and undeveloped recreation.)
6. Can access be obtained to the forest stand at a reasonable cost?
7. What type of logging system and cutting system are suited to the stand?
8. Is the cost of logging the stand under the type of system selected less than the expected return from the timber?

The first three decisions are made for the Planning Unit. The remaining decisions are made for all forest stands remaining after the first three decisions are made.

Wildlife

The wildlife program is based on the concept of providing wildlife for recreational purposes. The decisions that must be made to accomplish this goal for unit planning are listed below.

1. What wildlife species are to be featured on the unit?

2. Which streams are to be stocked with trout?
3. What habitat improvements must be made to feature the wildlife and fish species desired?

Recreation

Recreation planning decisions consist of formulating a combination of developed and dispersed activities that most closely meet the demands of the public. The list of decisions that are necessary to accomplish this task follows.

1. What improvements to existing facilities are need?
2. Does the unit need additional recreational activities; if yes, what kind?
3. Can the Forest Service justify these additional activities?
4. Are new developed facilities required; if yes, what kind of facilities and where should they be located?
5. Are there areas with unique combinations of activities or characteristics that are recognized for dispersed recreational uses; if yes, what improvements (if any) are required to make these areas available to the public?
6. Are there areas of cultural interest, such as historical and archealogical sites, that need protection or identification?

7. Are there areas where public use should be restricted?

Decisions Requiring Inventory Data

Many of the decisions listed above are not based on field inventories. Some of the decisions are policy decisions while others are made from existing information sources. Listed below are examples of decisions that rely on inventory data and the inventory information that is required to more soundly base the decision.

1. What timber stands within the Unit have timber products that could be scheduled for harvest?
    a. stand age
    b. forest type
    c. stand condition class
    d. site index
    e. operability
2. Are there visual restrictions that dictate the type of timber harvest?
    a. distance zones, sensitivity levels
3. Are there soil restrictions that dictate the type of timber harvest?
    a. soil modifiers
4. What wildlife species are to be featured on the Unit?
    a. stand age

    b. forest type

5. Which streams are to be stocked with trout?

    a. trout reproduction

6. Are there areas where public use should be restricted?

    a. soils modifiers

The data needed to make the above decisions are obtained from inventories by four functional staffs. As part of the timber inventory, stand age, stand condition class, forest type, site index, and operability are determined. The wildlife inventory identifies trout reproduction. The soil modifiers are inventoried by the soils specialist. The Landscape Architect inventories the distance zones and sensitivity levels. Figure 3 illustrates the relationship of the inventories and the decisions.

Determination of Sampling Intensity

As stated earlier the sampling intensity was determined as a function of the value of the information observed rather than a function of some rule-of-thumb. The minimization of cost-loss functions was used in this determination. In the previous sections it was concluded that four inventories of which one had five attributes must be conducted to collect information to make six decisions for Unit Planning. In order to use the cost-loss approach to

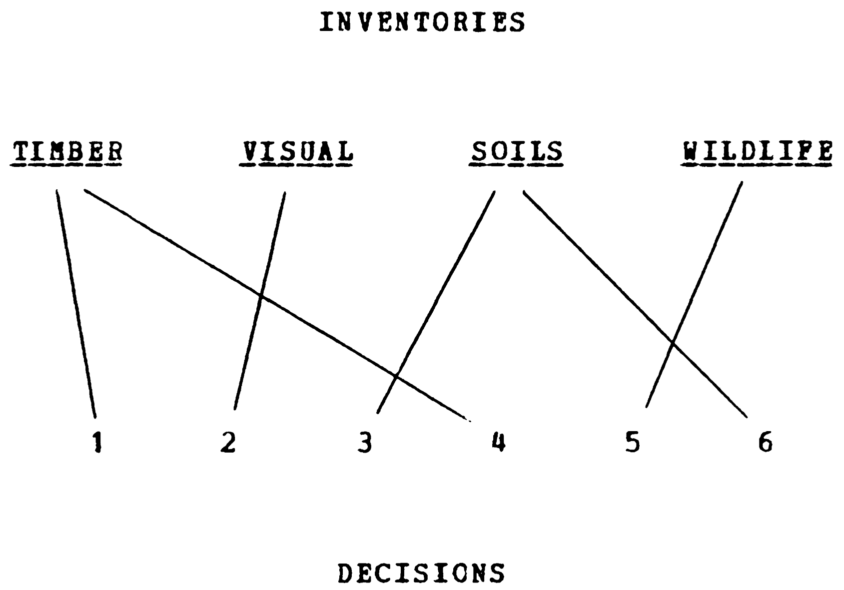

Figure 3. The relationship of six decisions made for Unit Planning and the inventory that aids in making the decision. The decision number corresponds to numbers on page 77

determine sample size for this example five types of information, (1) variance-covariance estimations; (2) cost figures; (3) λ determinations; (4) k vector values; and (5) T values, must be developed. The staff of the Jefferson National Forest, as the decision-makers, were consulted in obtaining this information.

Obtaining Variance Estimates

For an inventory with only one attribute, only a single estimate of the variance is needed but for an inventory with more than one attribute the variance of the attributes and the covariances between attributes must be estimated. The grid maps of the High Knob Unit were used to estimate variances and covariances. Each of the 3293 cells was considered to be a sample unit and the cells from different maps with the same coordinates were considered to be paired sampling points. All necessary maps were available except Age and Operability. With the help of the Timber Staff at the Jefferson National Forest headquarters, estimates of the maps were made using information from the Forest Type map, Stand Condition Class map, and Site Index map. In application, one should acquire variances from previous inventories, if possible. In this case study, it was not possible to acquire variance figures from field observations, because the Forest Service discards field data after it is summarized.

Five maps were used to estimate variances and covariances for the timber attributes. Site index was coded into 10 categories, the mid-point of the code class was used in the variance-covariance computations. This same procedure was followed for age. For the remaining maps -- Forest Type, Stand Condition Class, and Operability -- the grid code value for that attribute was used for variance computations. The variance-covariance estimates for timber are shown in Table 5.

The variance of Soil Modifiers and Distance Zones-Sensitivity Levels were similarly obtained from grid maps. Their variances were respectively, 3.5 and 6.9. No covariance existed between these variables because independence was assumed between inventories and, in this case, there was only one attribute observed per inventory.

The wildlife inventory included a determination of whether or not streams were stocked with trout reproduction. The observed variable was assumed to be binomial where success was presence of trout reproduction and failure was absence of trout reproduction. From the grid map of streams, 224 cells contained streams; of these, 99 were classified as native trout streams. Native trout streams are streams that contain trout reproduction. Therefore the probability of success was 99/224 for the High Knob Unit. Using the

Table 5. The variance-covariance matrix of the five attributes from the timber inventory.

	Stand Age	Forest Type	Stand Cond.	Site Index	Operabi.
Stand Age	293.6003	-0.8981	-28.7789	49.6406	24.6263
Forest Type		1.1357	-0.1488	1.0747	-0.1846
Stand Condition			5.0232	-3.9073	-1.5470
Site Index				277.6184	9.8564
Operability					4.8555

binomial variance formula the variance of the wildlife inventory was estimated as 55.2.

Cost of Inventories

The timber staff of the Jefferson National Forest estimated their timber inventory cost $ 0.35 per acre. They inventory approximately ten percent of the 674,000-acre Forest each year. The High Knob Unit contains 71,188 acres which is approximately ten percent of the total Forest acreage.

It was assumed that a sampling unit was equal to an acre because most information was available on an acre basis. Of the $0.35 spent to sample an acre, $0.20 was assumed to be overhead cost. The total overhead cost was $0.20 per acre times 71,188 acres in the High Knob Unit or $14237.60. The remaining $0.15 spent to sample an acre was assumed to be the cost of measuring the five attributes on each sample plot. The cost of measuring each attribute was assumed equal, therefore the cost per sample observation of measuring an attribute was $0.03.

The staff of the Jefferson also estimated the cost of inventorying the visual and soils characteristics of a unit at $0.35 per acre. For each inventory, twenty cents of the value was considered to be overhead. Ten percent of the Forest, which is approximately equal the size of the High

Knob Unit, 71,188 acres, is inventoried every year. Therefore, the overhead for the High Knob Unit is $14,237.60. Fifteen cents was the cost of observing an acre or sample unit.

Two men conducted the wildlife inventory. They observed two two-hundred-foot segments per mile of stream. Six segments can be observed per day. The yearly salary of the two men was assumed to be $28,000 and there are 260 working days per year, thus the salary per day is $107.69. General overhead was assumed to be 50 percent of the salary plus $20 travel cost per day. The overhead cost was $73.83 per day. It was estimated that two days would be spent inventorying the High Knob Unit, for a total overhead of $147.69.

Six stream segments are observed in one day at a cost of $107.69, for a sampling unit cost of $17.96. This value was assumed to be the cost of observing one "plot". The costs of the four inventories are summarized in Table 6.

Determination of λ

A λ was determined for each decision. The decision-maker was consulted as to the monetary loss that would be incurred from making decisions with information containing a sampling error of z. When more than one attribute was used to make a decision, a unit of measure for z had to be determined that related all the attributes. For

Table 6. A summary of inventory costs (dollars) for the four inventories on the High Knob Planning Unit.

	Timber	Visual	Soil	Wildlife
Overhead	14,237.60	14,237.60	14,237.60	147.68
Cost/Sample Unit	0.15	0.15	0.15	17.96
Cost Observe Attribute One	0.03			
Cost Observe Attribute Two	0.03			
Cost Observe Attribute Three	0.03			
Cost Observe Attribute Four	0.03			
Cost Observe Attribute Five	0.03			

example, the decision of which wildlife species is to be featured in an area uses Stand Age and Forest Type. The unit of measure of z was determined to be habitat.

When a decision is referred to by number, the number corresponds to the decisions listed in the previous section entitled "Decisions Requiring Inventory Data". Decision 1 refers to determining timber stands for possible harvest. The five attributes observed in the timber inventory were used to make this decision. z was defined in terms of thousand board feet of timber and the losses associated with various values of z were estimated as:

z	l(z)
0	$ 0.00
1	141.65
2	141.65
3	500.00
4	500.00
5	500.00

The timber staff estimated the monetary loss to be zero if the error was one to two thousand board feet for site indexes greater than 70 and $500 for site indexes less than or equal to 70 for the same error. A weighted average, based upon site index, determined the loss as $141.65. For errors of three to five thousand board feet the loss was $500 regardless of site index. A regression line was fitted using z^2 as the independent variable and loss as the dependent variable. The line was forced through the origin since the data with no error would enable a decision to be made with no loss. The slope of this regression line, 26, is λ for decision 1, which timber stands to harvest.

For the decision related to determining whether the timber stand had some visual restrictions that would dictate a particular harvesting system, the z value was defined in terms of visual classes. The losses associated with various values of z were estimated to be:

z	l(z)
0	$ 0
1	9,000
2	12,000
3	18,000

The landscape architect estimated that if he misclassified a stand by one classification unit the cost to remedy the results of the harvest would be one and one half times the cost of logging. Similarly he estimated the loss from errors of two and three classification units to be, respectively, two and three times the logging cost. The logging cost was estimated to be $6000 for an average size harvest area of 26 acres in the mountains. This figure was obtained from data available at Virginia Polytechnic Institute and State University. Using regression techniques $\lambda 2$ was found to equal 2,235.

What timber stands have soil characteristics that dictate restrictions on timber harvest is decision 3. z was defined in terms of soil classes. The losses associated with various values of z are:

z	l(z)
0	$ 0
1	6,000
2	12,000
3	12,000

For a z value of one the monetary loss to repair the damage from harvest was the cost of logging. For z values of two and three, the estimated loss was twice the logging cost. The cost of logging was assumed to be $6000 for an average size harvest area of 26 acres in the mountains. λ3 was determined to equal 1,653.

What wildlife species to feature was decision 4. As explained earlier, z was defined in terms of habitat classes. If z equals one, the loss for deer would be zero because of their wide range of movement but the squirrel population would decrease drastically. The Jefferson strives for one squirrel per acre that produces 7 hunter days per acre per year (U.S.D.A., Forest Service, 1971b), or 91 hunter days for an average stand of 26 acres per year. It was assumed that a z value of one would reduce the squirrel population by one half. According to a Forest Service study (U.S.D.A., Forest Service, 1974b) 30.6 percent of the hunters on National Forest lands in the Southeast placed abundance of game as the most important factor in their hunting, 33.4 percent placed it second, and 15.8 percent placed it third. It was assumed that if the squirrel population was reduced by 50 percent then all the hunters that placed abundance of game as most important would not hunt, and that forty and twenty percent, respectively, of the remaining two groups would be lost to hunting. Thus, 47.12 percent of the hunters would cease to hunt on the National Forest if the squirrel population were reduced by a half. The Forest Service study also showed the average small game hunter places a value of $43.64 on his day of hunting. Therefore, 91 hunter days per stand times the percentage of hunters lost times the value of a hunter day times the weight for High Knob Unit equals the loss for z equals 1 ($901.38). This loss would occur for ten years thus increasing the total loss to $9,013.80.

For a z value of two no loss would occur from deer but for squirrels the population was projected to drop to zero. Therefore, the loss would be the weighted value of the number of hunter days times the value of a hunter day times ten years or $19,129.50. A summary of the losses for various z values are

z	l(z)
0	$ 0
1	9,014
2	19,130

$\lambda 4$ was calculated to be 5030.

Whether or not to stock a stream with trout is the fifth decision. Reproduction of trout per mile of stream were the units for z. The loss for various z values is

z	l(z)
0	$ 0
1	50
2	75

For a z value of 1, a stream is stocked when it does not need to be. In this case, additional time is spent sampling and testing to determine why the stream does not have reproduction only to find it stocked. The loss per stream was estimated by Forest Service personnel to be $50. If z is two, the loss was estimated to be $75. The value of λ_5 was 21.

Are there areas where public use should be restricted because of soil characteristics is decision 6. It was assumed that this restriction would generally occur from people moving from one place to another over trails. z is defined in soil classes. If z equals one, only general maintenance occurs, which costs $40 per mile. The trail would have to be reworked for $650 per mile, if z was two. For a z of three, the trail would have to be relocated at a cost of $2000 per mile. A summary of the losses for various z values are

z	$l(z)$
0	$ 0
1	40
2	650
3	2,000

Using regression techniques $\lambda 6$ was found to equal 220.

In Figure 4, z is plotted against dollars loss for each decision. Also the regression line is shown with its value of λ.

Development of k vectors

The decision-makers of the Jefferson National Forest were questioned as to the weight they placed on information needed to make a decision. Considering the case for multiple attributes observed per sampling point, a set of four vectors for each of the six decisions must be developed. Each vector of a set will represent an inventory. For each attribute not needed to make a particular decision, a zero is entered in the appropriate vector element. For the attributes contributing to making a decision the decision-maker placed a number representing the weight of that attribute with respect to the other attributes contributing to making the decision.

For example, to make the decision what stands to schedule for possible harvest the five attributes from the timber inventory are needed. The vectors for the remaining three inventories are vectors of zeros. The decision-maker

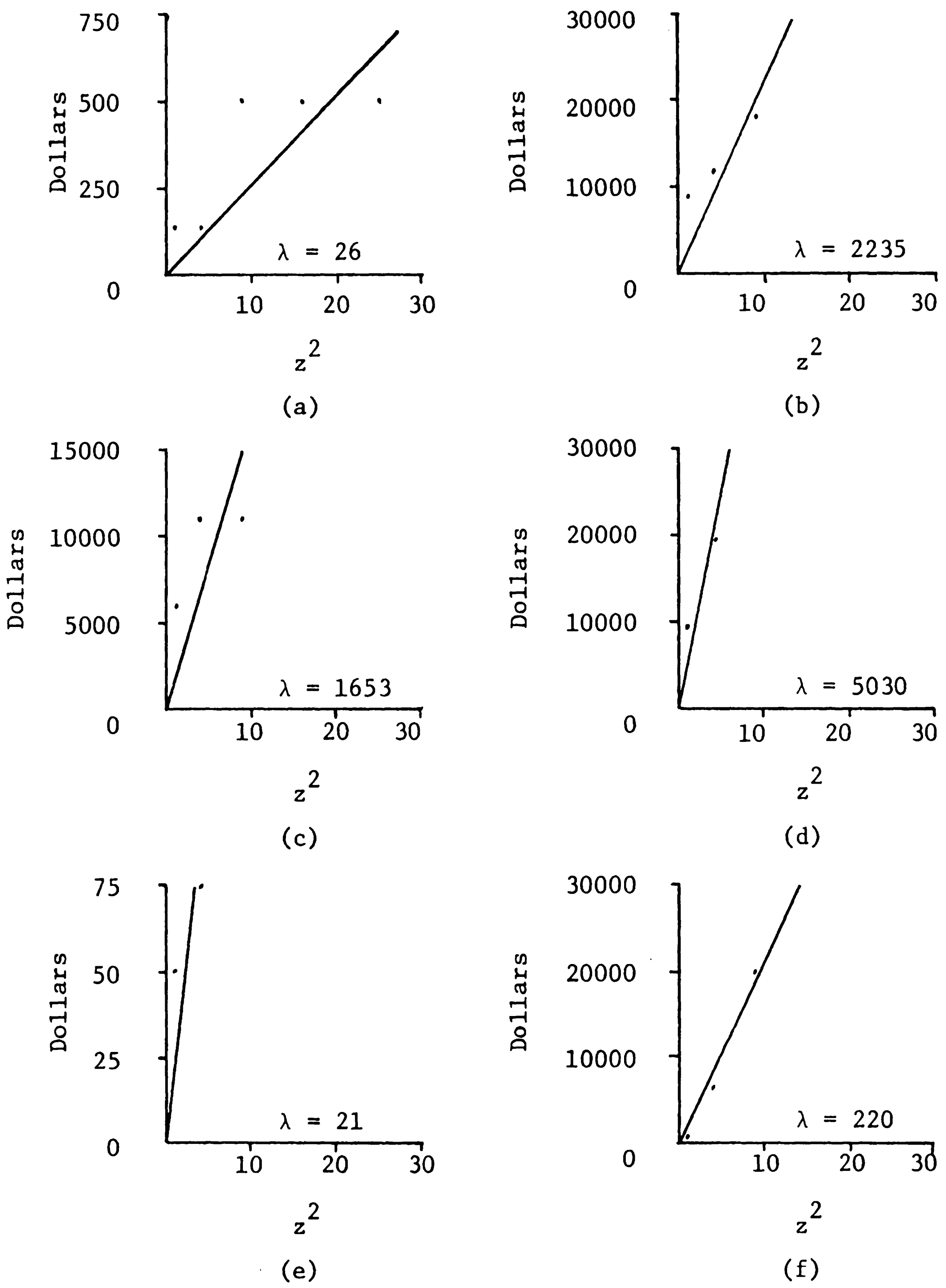

Figure 4. Plots of squared error, z^2, against monetary loss. If the assumption $L(n)=\lambda z^2$ is correct then the plots will be linear. Also graphed is the regression through the origin with its value of λ. Plots (a) through (f) are associated with decisions 1 through 6, respectively.

considered Stand Condition Class and Site Index twice as important and Operability three times as important as Stand Age and Forest Type. Therefore, Stand Age and Forest Type would have a value of one, Stand Condition Class and Site Index would have a value of two, and Operablilty would have a value of three. The k vectors for the High Knob Unit are listed in Table 7.

If only one attribute per inventory was assumed then only one vector would be developed for each decision. Each element in the vector would represent the weight of importance of that inventoried item in making the decision. The k vectors for the latter case are given in Table 8 for the High Knob Unit.

Determination of T's

A T value for a decision is the number of independent times the decision is made on the Unit. For the timber decision regarding selecting stands for potential harvest, the expected loss was developed for a stand of timber. The average size cut on the High Knob Unit has been 26 acres and was assumed to be the size of an average stand of timber. Assuming there were no areas on the High Knob Unit that preclude timber harvesting, then the decision is made for each stand on the 71,188-acre Unit. T1 is calculated as 2,738. For decision 4, what wildlife species to feature, T4

Table 7. k vectors for Unit Planning for the case of multiple attributes measured at each sampling point.

	Timber	Visual	Soils	Wildlife
Decision 1	1 1 2 2 3	0	0	0
Decision 2	0 0 0 0 0	1	0	0
Decision 3	0 0 0 0 0	0	1	0
Decision 4	1 1 0 0 0	0	0	0
Decision 5	0 0 0 0 0	0	0	1
Decision 6	0 0 0 0 0	0	1	0

Table 8. k vectors for Unit Planning for the case of a single attribute measured at each sampling location.

	Timber	Visual	Soils	Wildlife
Decision 1	1	0	0	0
Decision 2	0	1	0	0
Decision 3	0	1	1	0
Decision 4	1	0	0	0
Decision 5	0	0	0	1
Decision 6	0	0	1	0

was calculated in the same manner for the 71,188 acres, thus T4 equals 2,738.

The T values for the decisons regarding the limited restrictions, visual and soils, will be the same because each decision is made for timber stands scheduled for harvest. Ten percent of the High Knob Unit is to be harvested every ten years which is the length of time the Unit Plan covers. The assumption was made that the decisions would be made on each average size timber stand on a tenth of the Unit's acreage. Thus T2 and T3 equal 274.

Based on data from grid maps, it was estimated that decision 5, which streams to stock with trout, would be made four times for the High Knob Unit.

For decision 6, what areas to limit public use, T6 was estimated to be 34. This figure is based upon an estimate of 34 miles of trails on the High Knob Unit which was obtained from grid maps.

Minimization

There are several basic approaches to minimizing the cost-loss functions with the information developed in the previous sections. These approaches depend on the assumptions made concerning the variances-covariances structure of the inventories and whether or not all attributes of an inventory will be observed at each sampling location.

The first approach is to assume that the variance for an inventory is equal to the largest variance of the attributes observed in that inventory and that all attributes will be observed at all sampling locations. This is a conservative approach to the minimization problem, because, all other factors being equal, the larger the variance the larger the sample size. Also, the remaining attributes within that inventory may be over sampled, thus giving more precision than necessary.

Of the four inventories conducted to obtain information for Unit Planning, multiple attributes were recorded at each sampling location only for the timber inventory. The largest variance of the attributes for this inventory was 293.6 for Stand Age. Using this as the variance, the equations developed earlier for q inventories with one attribute to make m decisions were used to determine the sample sizes that minimize the cost-loss functions. The sample size to minimize the total cost plus loss are

$$n_i = \sqrt{(S_i^2/c_{1i})(T_1\lambda_1k_{11}^2 + T_2\lambda_2k_{21}^2 + \cdots + T_m\lambda_mk_{m1}^2)}$$
$$(i = 1,2, \ldots ,q)$$

where, for this example, q equals four and m equals six. The sample size for timber can be determined by substituting the appropriate information as follows:

$$\begin{aligned} n_1 &= \sqrt{[293.6/0.15][2738(26)(1) + 274(2235)(0)} \\ &\quad \overline{+ 274(1653)(0) + 2738(5030)(1)} \\ &\quad \overline{+ 4(21)(0) + 34(220)(0)]} \\ &\approx 164609 \end{aligned}$$

This sample size is more than twice the population size for the High Knob Unit. Sampling with replacement is being assumed, but in reality sampling without replacement would probably be used. Therefore, when the sample size was greater than the population size, a one hundred percent sample was assumed with a variance of zero. The sample size for timber is then 71,188. The sample sizes for the Visual, Soils, and Wildlife inventories were calculated to be 5293, 3281, and 16, respectively.

The cost of the overhead for the four inventories was $42,860.49. Sampling costs were determined by summing the cost of observing one sample unit times the number observed for the inventory. This value was found to be $12,251.66.

The expected loss for the j th decision was determined using the following equation:

$$\begin{aligned} L_j(n_1,n_2, \ldots, n_q) &= T_j\lambda_j[(k_{j1}{}^2S_1{}^2/n_1) + (k_{j2}{}^2S_2{}^2/n_2) \\ &\quad + \ldots + (k_{jq}{}^2S_q{}^2/n_q)] \end{aligned}$$

For the decision regarding whether or not to stock a stream with trout, the loss was determined by substitution as follows:

$$\begin{aligned} L_5(n_1,n_2,n_3,n_4) &= 4(21)[(0)(0.00/71188) + (0)(6.8/5293) \\ &\quad + (0)(3.5/3281) + (1)(55.2/16)] \\ &= 8.00 \end{aligned}$$

The loss for decisions 1,2,3,4, and 6 are $0.00, $793.82, $469.89, $0.00, and $8.00, respectively, for a total expected loss of $1561.73. The total cost-loss at this sampling intensity was $56,673.88.

A minimization was accomplished by assuming all variances of an inventory to be equal to the variance of the most important attribute and assuming all attributes will be observed at each sampling location. In the timber inventory, operability was considered the most important attribute, thus the variance was set equal to 4.9. Since only the timber sample size was affected, the only changes in the results from the previous minimization were the sample size for timber, the sampling cost, and the expected loss from decisions one and four. Due to the decreased variance estimate, the sampling size for timber was reduced to 21,169 sampling units. The expected losses for decisions one and four were recalculated as $16.33 and $3,158.89, respectively because of the low variance assumed for the timber information. The sampling cost was reduced to $4,748.81 and the total cost-loss for inventory at this sampling level was reduced to $52,346.25.

In another method of determining sampling intensity, the assumption was made that each attribute of an inventory would be observed at each sampling location. The variance of each attribute was used in the calculation of sampling intensity. The timber inventory was the only inventory considered in the case study that had multiple attributes observed and thus is the only sample size affected by these assumptions. The sampling intensity for timber that minimizes the cost-loss function was calculated to be 166,607. This is greater than the population size of the High Knob Unit thus a one hundred percent inventory will be assumed (i.e., a sample size of 71,188). The total cost-loss for an inventory at this sampling level is $56,673.88.

In a fourth method of determining sampling intensity, sample size is obtained for each attribute. Covariances between attributes are assumed to be zero. The solution can be obtained using the same equations developed for q inventories with a single attribute to make m decisions. The variance of each attribute is used in the sample size estimation for that attribute. As before the only calculations affected were those involving the timber inventory. The sample size for the timber attributes were calculated as 368,076; 22,892; 6,905; 51,333; and 10,183, respectively. The sample size for the first attribute was greater than the population size; as before, the sample size was set equal to the population size, 71,18,8 and the variance was set equal to zero. The cost of sampling was $6,448.49 and the total cost-loss was $53,610.13.

A fifth approach to determine sampling intensity is to determine the sample size for every attribute instead of for each inventory and to use estimated covariance values between attributes of the same inventory. When the first partial derivatives were taken and set equal to zero the only sample sizes affected were those for timber. These five partial derivatives produced the following five equations with five unknowns:

$$\begin{aligned}\frac{d}{dn_{11}} = {} & 0.03 + 26(2738)[-293.6n_{11}^{-2} -2(-0.9)(n_{11}+n_{12})^{-2} \\ & -4(-28.8)(n_{11}+n_{13})^{-2} -4(49.6)(n_{11}+n_{14})^{-2} -6(24.6)(n_{11}+n_{15})^{-2}] \\ & +5030(2738)[-293.6n_{11}^{-2} -2(-0.9)(n_{11}+n_{12})^{-2}] = 0\end{aligned}$$

$$\begin{aligned}\frac{d}{dn_{12}} = {} & 0.03 + 26(2738)[-1.1n_{12}^{-2} -2(-0.9)(n_{11}+n_{12})^{-2} \\ & -4(-0.1)(n_{12}+n_{13})^{-2} -4(1.1)(n_{12}+n_{14})^{-2} -6(-0.2)(n_{12}+n_{15})^{-2}] \\ & + 5030(2738)[-1.1n_{12}^{-2} -2(-0.9)(n_{11}+n_{12})^{-2}] = 0\end{aligned}$$

$$\begin{aligned}\frac{d}{dn_{13}} = {} & 0.03 + 26(2738)[-5.4n_{13}^{-2} -4(-28.8)(n_{11}+n_{13})^{-2} -4(-0.1)(n_{12}+n_{13})^{-2} \\ & -8(-3.9)(n_{13}+n_{14})^{-2} -12(-1.5)(n_{13}+n_{15})^{-2}] = 0\end{aligned}$$

$$\begin{aligned}\frac{d}{dn_{14}} = {} & 0.03 + 26(2738)[-277.6n_{14}^{-2} -4(49.6)(n_{11}+n_{14})^{-2} -4(1.1)(n_{12}+n_{14})^{-2} \\ & -8(3.9)(n_{13}+n_{14})^{-2} -12(9.9)(n_{14}+n_{15})^{-2}] = 0\end{aligned}$$

$$\begin{aligned}\frac{d}{dn_{15}} = {} & 0.03 + 26(2738)[-4.9n_{15}^{-2} -6(24.6)(n_{11}+n_{15})^{-2} -6(-0.2)(n_{12}+n_{15})^{-2} \\ & -12(-1.5)(n_{13}+n_{15})^{-2} -12(9.9)(n_{14}+n_{15})^{-2}] = 0\end{aligned}$$

These equations could not be solved analytically but a numerical solution was attempted using iterative techniques of the Newton-Raphson procedure. Various starting values were used (including the sample sizes determined for the attributes when the covariances were assumed to be zero) but a solution to these equations could not be found. A solution might exist, but the cost and time spent finding a solution may be more than the cost-loss savings that would be gained over using other methods of determining sampling intensity.

A summary of the findings for four methods of determining sampling intensity that assumed the covariance to be zero is found in Table 9. The difference in the sample size calculated for the timber inventory caused the difference in total cost-loss between the four methods. The sample size for the method assuming the largest variance was the conservative estimate thus the higher cost-loss. The same cost-loss was calculated for method three which assumes each attribute will be measured at each sampling location and accounts for the variances of each attribute while not calculating a sample size for each attribute. The cost-loss for the two methods are equivalent since both calculated the sample size for timber as greater than the population size and a one hundred percent inventory was assumed. The method assuming the variance of the inventory equal to the variance of the most important attribute had the lowest cost-loss because the variance of the most important attribute in this case was very low. When sample sizes were calculated for each attribute the cost-loss was between the cost-loss of the other methods. This phenomenon occurred because the sample size for each attribute was determined and the sample size that minimized the cost-loss function did not over or under sample some attributes as the previous methods did.

The changes in the cost-loss values between the various methods of determining sample sizes were relatively small.

Table 9. Summary of results from the four methods of determining sample size assuming covariances equal zero.

	All Attributes Sampled At All Locations		All Attributes Sampled At All Locations	Not All Attributes Sampled At Each Location
	Largest Variance Used	Most Important Variable's Variance Used	Individual Attributes Variance Used	
<u>Sample Size</u>				
Timber	71,188	21,169	71,188	
Stand Age				71,188
Forest Type				22,892
Stand Cond.				6,905
Site Index				51,333
Operability				10,183
Visual	5,293	5,293	5,293	5,293
Soils	3,281	3,281	3,281	3,281
Wildlife	16	16	16	16

Table 9. Summary of results from the four methods of determining sample size assuming covariances equal zero (continued).

	All Attributes Sampled At All Locations		All Attributes Sampled At All Locations	Not All Attributes Sampled At Each Location
	Largest Variance Used	Most Important Variable's Variance Used	Individual Attributes Variances Used	
Loss (dollars)				
Decision 1	0.00	16.33	0.00	2,056.17
Decision 2	793.82	793.82	793.82	793.82
Decision 3	469.89	469.89	469.89	469.89
Decision 4	0.00	3,158.89	0.00	683.25
Decision 5	290.02	290.02	290.02	290.02
Decision 6	8.00	8.00	8.00	8.00
Total Expected Loss	1,561.73	4,736.95	1,561.73	4,301.15
Sampling Cost	12,251.66	4,748.81	12,251.66	6,448.49
Overhead Cost	42,860.49	42,860.49	42,860.49	42,860.49
Cost-Loss	56,673.88	52,346.25	56,673.88	53,610.13

These small changes indicate that any of the methods of determining sampling intensity from a minimization of a cost-loss function may be useful for establishing a guide to sampling intensities.

Discussion

Determining sampling intensity based upon the value of the information rather than some rule-of-thumb seems to be a better approach. There are some difficulties that arise. One of the most prominent was the difficulty of estimating a monetary loss that would occur from making a decision based on information containing various amounts of sampling error. It is beneficial for decision-makers to contemplate these expected losses regardless of what methods are ultimately used for determining sampling intensities.

In this study, a squared error loss function was assumed. Loss functions may be of some other functional form, depending on the decision involved. Hamilton (1975) has studied other loss functions for simple cost-loss functions. The framework developed in this study should be expanded to incorporate Hamilton's loss functions.

The squared error loss function was assumed to be symmetric. This assumption does not often hold true. An overestimation may produce no loss in some cases while an underestimation may produce a considerable loss. Developing nonsymmetric loss functions should be considered for future study.

In conclusion, the cost-loss minimization procedure appears to be adaptable to multi-resource inventory situations. Following through the process of specifying what decisions need to be made, what data are needed to make these decisions, and what are the expected losses from making decisions with sample data should benefit resource managers by helping eliminate superfluous data gathering.

RECOMMENDATIONS FOR MODIFICATIONS IN DATA HANDLING AND PLANNING PROCEDURES ON THE JEFFERSON NATIONAL FOREST

The High Knob Unit Plan has been under development for over two years. During this time, the author followed closely the various planning phases. Comments and recommendations from an outsiders point of view were made continuously throughout the planning process. Many of these comments and recommendations are included in this chapter. The Unit Planning staff of the Jefferson National Forest implemented many of these recommendations at the time they were made.

Inventory

In this study, procedures were developed to evaluate the inventory system of the Jefferson National Forest with a case study on the High Knob Unit. The currently inventoried variables were examined. It was assumed that all variables were needed for Unit Planning because if they were not needed then the variable should not be inventoried. The analysis showed no high degree of association between variables and/or combinations of variables. Therefore, the conclusion was that all variables should continue to be observed to provide information necessary for Unit Planning.

The results of the analysis could have been a function of the data taken from grid maps and should be repeated using actual field data.

A cost plus loss minimization procedure was developed to determine sampling intensities for Unit Planning purposes. Cost of sampling is an important part of this analysis. The Forest Service averages the cost of inventory on a per acre basis, which makes separation of overhead costs difficult. Information on a sample unit basis would facilitate analyses of sampling costs.

The sample sizes used by the different inventories were unavailable on the High Knob Unit. If this type of information was available the sampling intensity calculated in this study could be compared with the actual sampling intensity used. Also, the total cost-loss for each sampling intensity could be compared. The cost-loss approach to determine sample intensity can be used either to determine sampling intensity for an inventory or as a comparison against existing procedures.

Data Handling

With the increasing amount of information needed for managing and planning, many organizations have instituted the use of computers. The Forest Service is no exception.

The soils, recreation, and timber staffs have stored data on computers for several years. The Unit Planning Team began using a variation of the Harvard Grid System (Sinton and Steinitz, 1971, and U.S.D.A., Forest Service, 1975) in 1973 for storing and mapping information. The High Knob Planning Unit was the first unit on the Jefferson National Forest for which a computerized data base was used in the Unit Planning process. The use of computers has been greatly increased on the Jefferson National Forest with the addition of a computer terminal in January of 1975. The terminal enables the staff to connect with the Forest Service's computer installation in Fort Collins, Colorado.

Using computers to map and store information and to test for conflicts entails some changes in the data handling process. The Continuous Inventory of Stand Conditions or CISC system at Fort Collins stores information about timber stands. The major problem with this system is that the data from different timber compartments can only be summarized by Ranger Districts or Planning Units. When information is wanted on a subdivision of a Ranger District or Planning Unit the staff reverts to the compartment summary sheets and does the summaries by hand. The Wildlife functional staff uses timber information on their Wildlife Survey Units. The Survey Units follow timber compartment boundaries and may encompass several compartments. The staff may spend several days summarizing timber data from the compartment summary sheets to determine stand composition (species and age) for a Survey Unit. If the CISC system was programmed to allow summaries on specified compartments, the time spent summarizing could be greatly reduced. This modification in CISC should also save some time for the timber staff. This change cannot be made by the Jefferson National Forest personnel but is a change the Regional Office in Atlanta must institute.

The timber staff classifies browse use into two categories: none to moderate use and heavy use. This information is collected at the time the District personnel are evaluating the stands for a stand perscription. The Wildlife staff also collects browse-use information on their one- to two-day visit of a Wildlife Survey Unit. Their classification of browse use is light, moderate, and heavy.

The two staffs should agree on one classification system for browse use. The timber staff should collect the information because of their more thorough coverage of an area. The information could be summarized using the CISC computer system. Also, the timber staff might collect information on the quality of browse for the Wildlife staff.

This project helped establish the use of the Harvard Grid System on the Jefferson National Forest. Since the Fall of 1973 many problems developed in the original gridding process used by the Jefferson National Forest, some problems were solved and others are still under consideration. Hopefully, in this section an improved gridding process for a Planning Unit will be outlined and suggestions for improvements over the process used for the High Knob Planning Unit will be made.

The first step is to determine a grid cell size. According to unpublished studies of the U.S. Forest Service the acreage size of cells makes little difference in area estimation. The choice of cell size should be based on a convenient sized computer output. A cell size that creates a computerized map where two or three computer sheets must be placed together to make a complete map is not desirable. A map of less than 120 columns (cells horizonally) has many advantages.

The grid should be drawn on the mylars used to produce all base maps, thereby permanently locating the grid cells on all maps. This was not done on the High Knob Unit. Instead a transparent grid overlay was placed over a base map each time a cell needed to be located. This overlay stretched after some use, thus mislocation of many of the cells resulted. Also, with the grid coordinates reproduced on all base maps, the location of a grid cell would be made easier by not having to use the transparent overlay each time a cell must be located.

The first map to be coded is the Land Status Map. This map shows which cells are in the Planning Unit. Within the Planning Unit the private land and the Forest Service land are designated by different codes. The Land Status Map is the most important map that will be gridded. All the other maps will be based upon this map. It cannot be emphasized enough that this map must be checked and rechecked to determine if the land status of a cell is coded correctly.

This cell code verification was not done initially on the High Knob Planning Unit and consequently several problems were created. After several other maps were gridded it was found that they did not match. The boundaries were not the same, one map had a cell coded as in the Planning Unit and another map had the same cell coded as outside the Unit. This made the overlay maps useless. Also the staff was continuously changing the maps that had previously been gridded to try and obtain grid maps that matched.

If the Land Status Map was determined to be correct, it could be compared to all other maps. On the High Knob Planning Unit a computer program was developed to compare the Land Status Map with each map gridded. If a cell was coded as Forest Service land on the Land Status Map the program checked for a code in that cell on the grid map. If a cell was outside the unit on the Land Status Map the program checked for a blank in that cell on the grid map.

A general computer program to check boundaries on a Planning Unit should be developed and placed on the Fort Collins Computer System. The information to correct the grid-coding mistakes could then be easily obtained and the corrections made from the Supervisor's Office.

After the Land Status Map is finalized the next step is to determine a list of maps to grid. On the High Knob Planning Unit twenty-one maps were gridded. They were Land Status, Land Forms, Soil Source, Soil Texture, Water Regime, Soil Modifier, Slope, Transportation, Streams, Key Wildlife Habitat Areas, Wildlife Featured Species, Forest Type, Stand Condition, Site Index, Ten Year Stand Selection, Land Class, Distance Zones and Visual Sensitivity Levels, Fire Size, Fire Damage Class, Geology, and Mineral Rights. A judgement of whether or not this information is the most important to grid depends on the special problems of a planning unit. The High Knob Planning Unit is in an area of intense strip mining. This adds special management problems such as ownership of mineral rights. Gridding this information was helpful in determining management alternatives on the High Knob Unit, but on other units mineral rights information may be unimportant in management planning. A decision regarding which information to grid should be made separately for each planning unit. An additional map to consider is one for endangered species, plants and animals.

When making grid maps, the variable class interval must be carefully chosen so that the mapped information will have maximum utility. An example of the importance of interval selection was found on the High Knob Planning Unit. When Slope was mapped the slope percentages were divided into seven classes, 0 to 5 percent, 6 to 10 percent, 11 to 15 percent, 16 to 25 percent, 26 to 50 percent, 51 to 60 percent, and 61 plus percent. This coding system could not give all the information needed for planning. A critical point in slope information for logging systems is 35 percent which could not be determined with the present coding system. A much more useful coding system is 0 to 8 percent, 9 to 15 percent, 16 to 35 percent, 36 to 50 percent, and 50 plus percent. This coding system contains all the critical breaking points in slope for the functional staffs, 8 percent for engineering, 15 percent for recreation, and 35 percent and 50 percent for timber. Determining the proper coding system is essential to obtain the most from a grid map.

Each cell in a unit is represented by a code between zero and nine for a particular map. The computerized map produced converts these codes into different shadings. The larger the code the darker the shading. When less than ten codes are used, the maximum difference in numeric value of the codes should be used to produce the greatest contrast on the resultant map. For example, on the Land Status Map there are only two codes, Forest Service land and private land. The maps obtained if "0" is the code for one and "9" is the code for the other will have the most contrast in shading on the maps. If the codes are "0" and "1" there will be difficulty in distinguishing between the two classes. The two coding systems discussed above are shown in Figure 5.

When coding more than one grid map from a base map a problem can occur that is best explained by an example. When coding the E.M.U.'s, and the boundary between two E.M.U.'s crosses a cell, the decision as to which E.M.U. the cell will represent must be the same for all five maps that form the E.M.U. If this decision is not consistently made, cells may have fictitious E.M.U. numbers. The easiest way to solve this problem is to simultaneously code all maps from the same base map.

There are two basic procedures to code data from a base map to computer cards. In the first procedure, the codes are written on grid coding forms and then keypunched onto computer cards . There are chances for translation mistakes in coding onto the forms and then again in keypunching. The second method uses mark sensitive forms which are read into a machine that creates a card deck. Using this method there

```
■■■■■■■■■■■■...    ,,,,,,,,,,,,...
■■■■■■■■■■■■■..    ,,,,,,,,,,,,,..
■■■■■■■■■■■■...    ,,,,,,,,,,,,...
■■■■■■■■■■■....    ,,,,,,,,,,,....
■■■■■■■■■■..■■.    ,,,,,,,,,,..,,.
..■.....■■.■■..    ..,.....,,.,,..
...............    ...............
..■■■..■■......    ..,,,..,,......
■■■■■■■■.......    ,,,,,,,,.......
■■■■.■■........    ,,,,.,,........
■■■............    ,,,............
■■..■■■......■■    ,,..,,,......,,
..■■■.■........    ..,,,.,........
■■■■■■.........    ,,,,,,.........
■■...■■..■■■■..    ,,...,,..,,,,..

      (A)                (B)
```

Figure 5. Portion of High Knob Planning Unit's Land Status Map. In Map A, 0 code (.) is used for Private Land and 9 code (■) is used for Forest Service Land. In Map B, 0 code (.) is used for Private Land and 1 code (,) is used for Forest Service Land.

is only one place for translation mistakes, coding onto the mark sensitive forms.

The cost for the keypunching system was approximately 18 cents per map line and averaged 60 mistakes per map deck on the High Knob Planning Unit. On maps of comparable size, the cost for the mark sensor method was 3 cents per map line and averaged 10 mistakes per map deck. The Forest Service presently uses the keypunch method but should investigate the latter method for future use.

The Transportation Map for the High Knob Planning Unit contained information on roads and trails. When a cell contained both a road and a trail, the cell was coded as a road. This procedure created problems in trying to follow trails from one point to another and in determining a trail head. For other units, two maps, one for roads and one for trails, would be more useful.

The Land Class Code Maps have several codes for "others." At the Unit Planning session for the High Knob Unit three staff members spent over an hour trying to decipher what these areas were. The classification should be made clearer and probably a new code should be established especially for each area presently specified as "others."

A grid map should be constructed for special management areas. Such areas are proposed wilderness areas and city watersheds. When testing for conflicts in management planning by using the computer overlays system, conflicts in these areas can then be detected.

The Land Status, Transportation, and Streams Maps were the only information gridded for both Forest Service and private land within the Planning Unit. With little additional effort slope and geology could also be gridded for the entire Unit. This information would be helpful in possible land acquisitions and management planning near Forest Service boundaries. A third code could be added to the Land Status Map for possible future acquisitions. As many maps as possible should contain information on these cells. Planning for future acquisitions would be greatly aided.

A transparent overlay for the grid maps is very useful when determining locations for cells. The overlay should show towns, roads, and streams.

The computer overlay system added to the Harvard Grid System is very useful in determining location of cells with certain characteristics. It is also helpful in determining conflicts between management decisions. The staff is gaining skill at the use of this system but practice in its use is necessary to obtain the full potential of the system. The staff must realize the information obtained from the computer is no better than the input information to the computer.

The grid mapping system can be updated using the computer terminal at the Forest Supervisor's office. The majority of the maps such as the soils maps will stay unchanged over time. The maps that must be updated periodically could be done in approximately two hours. When an activity such as timber harvesting changes a cell code, the change should be recorded when the activity takes place. The updating process should be done annually.

Planning and Procedures

Many benefits would occur if it were possible to shorten the total planning time frame. The High Knob Plan began prior to the August 14, 1973 public meeting and is not completed as of this writing. The arrangement of the data into the grid maps for the first time was very time consuming but it was not a major factor in time delays. Some delays were because the Unit Planning Team has other assignments besides planning. Another area that can cause delays, not necessarily on the High Knob Unit, is the lack of necessary information before beginning the formulation of a Unit Plan. The possibility of the postponement of the completion of the plan because of this type of delay can be reduced with careful planning. The information needed could have been gathered simultaneously with other information without any delay.

Planning for a given unit over a long time period causes time loss due to the needs for reviewing information and problems because of memory loss. Also difficulty is found in keeping information in perspective.

The week-long meeting of the Jefferson National Forest staff to compile the High Knob Unit Plan in February, 1975 was very useful in directing attention to the problems of that Unit. I believe the meeting should be reduced to three days with a follow up meeting two weeks later for one or two days. This time would give the staff an opportunity to obtain much of the additional information that was found lacking after the first three days of meetings.

LITERATURE CITED

Baron, D.P. 1973. Point estimation and risk preference. Joun. of Am. Stat. Assoc. vol. 68. pp. 944-950.

Blythe, R.H. 1945. The economics of sample size applied to the scaling of saw logs. Biom. Bull., 1,67-70.

Cochran, W.G. 1963. Sampling Techniques. John Wiley and Sons, Inc., New York. 413pp.

Comrey, A.L. 1973. A First Course in Factor Analysis. Academic Press, New York. 316pp.

Dixon, W.J., Ed. 1974. BMD Biomedical Computer Programs. University of California Press, Los Angeles, California. 773pp.

Draper, N.R. and H. Smith. 1966. Applied Regression Analysis. John Wiley and Sons, Inc., New York. 407pp.

Dyer, A.A. 1974. Multi-resource data systems for land use planning. Inventory Design Analysis, Edited by W.E. Frayer, George B. Hartman, and R. Bower. SAF Workshop on Inventory Fort Collins, Colorado. pp. 162-169.

Grundy, P.M., M.J.R. Healy, and D.H. Rees. 1956. Economic choice of the amount of experimentation. Jour. Roy. Stat. Soc., B, 18, 32-55.

Halterman, B.G. 1972. Using linear programming to integrate timber management and wildlife management planning. M.S. Thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia. 87p.

Hamilton, D.A., Jr. 1970. Precision requirements for some information in timber management decisions. Ph.D. Dissertation, Iowa State University, Ames, Iowa. 114pp.

__________. 1975. Precision requirements for forest management inventories. Paper presented Biometric Society Meeting, Corvallis, Oregon. 25pp.

Harmen, H.H. 1967. Modern Factor analysis. Second edition. The University of Chicago Press, Chicago. 474pp.

Lawley, D.N. and A.E. Maxwell. 1971. Factor analysis as a statistical method. American Elsevier Publishing Company, Inc., New York. 153pp.

Navon, D.I. 1971. Timber RAM...a long range planning method for commercial timber lands under multiple use management. U.S.D.A., Forest Service, Pacific South West Forest and Range Exp. Station, Berkeley, Calif. Research Paper PSW-70, 22p.

Porter, J.R. 1974. Using linear programming to improve unit planning on the Appalachian National Forests. M.S. Thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia. 112p.

Ralston, A. 1965. A First Course in Numerical Analysis. McGraw-Hill Book Company, New York. 578p.

Sinton, D. and C. Steinitz. 1971. GRID user's reference manual, Version 3. Lab. for Computer Graphics and Spatial Analysis, Harvard Univ. Cambridge, Massachusetts. 35p.

Smith, B.E. 1965. The economics of sampling inspection. Industrial Quality Control, Vol. 21, pp. 453-458.

U.S. Congress. 1960. United States at Large, Public Law 86-517. United States Government Printing Office, Washington, D.C.

U.S.D.A., Forest Service. 1970a. Compartment Prescription Handbook. Forest Service Handbook, Southern Region, FSH 2409.21d, Atlanta, Georgia.

__________. 1970b. System for managing the National Forests in the East. Forest Service, Eastern and Southern Regions, Atlanta, Georgia. 24p.

__________. 1971a. Compartment Prescription Field Book. Southern Region. Atlanta, Georgia.

__________. 1971b. Wildlife Habitat Management Handbook. Forest Service Handbook, Southern Region, Atlanta, Georgia. 140p.

__________. 1972. Soils Resource Guide. Forest Service, Southern Region, Atlanta, Georgia. 48p.

__________. 1973a. Fish and Wildlife Survey Handbook. Forest Service Handbook, Southern Region, FSH 2609.21R, Atlanta, Georgia.

__________. 1973b. Landscape Management Handbook. Forest Service Handbook, Southern Region, FSH 2309.22R-8, Atlanta, Georgia.

__________. 1973c. Guide for managing the National Forests in the Appalachians. Forest Service, Eastern and Southern Regions, Atlanta, Georgia. 34p.

__________. 1973d. National Forest landscape Management, Volume I. Agriculture Handbook Number 34, U.S. Government Printing Office, Washington, D.C. 76p.

__________. 1974a. Memo from Region 8 Unit Planning Team, Atlanta, Georgia.

__________. 1974b. Executive Summary, Economic Survey of Wildlife Recreation, National Forest Service. Southern Region, Atlanta, Georgia. 102p.

__________. 1975. GRID II. Forest Service, Southern Region, Atlanta, Georgia. 29p.

Yates, F. 1960. Sampling methods for censuses and surveys. Charles Griffin and Co., London, Third Ed. pp. 292-293.

The vita has been removed from the scanned document

INCREASING THE EFFICIENCY OF MULTIPLE-USE INVENTORY PROCEDURES

by

Roger Dean Stuck II

(ABSTRACT)

Two approaches were used to evaluate multiple-use inventories. The first approach was to analyze inventory data currently being collected. Assuming that all variables being inventoried are necessary in making decisions for multiple-use management, simple correlation, multiple linear regression, and factor analysis techniques were employed. Data from the Jefferson National Forest in Virginia were analyzed but no strong association between variables currently being observed was detected.

The second approach was to define the decisions that are to be made and the variables necessary to make these decisions. Once the variables have been defined, sampling intensity must be determined. The decision-maker was not required to define the sampling intensity based on experience or a rule-of thumb, but rather a cost-loss function for multiple inventories and multiple decisions was minimized. The cost of collecting data will increase as sample size increases, while the expected monetary loss that will occur to the decision-maker from making decisions with

inventory data will decrease. Minimizing the cost-loss function determines the sampling intensity that will provide the lowest total (cost plus loss) monetary cost to the decision-maker. Methods were developed to determine necessary information for the cost-loss function, and several methods of minimization were evaluated.