CR 151243

SPACE SCIENCES LABORATORY

(NASA-CR-151243) FORESTRY APPLICATIONS PROJECT/TIMBER RESOURCE. SAM HOUSTON NATIONAL FOREST INVENTORY AND DEVELOPMENT OF A SURVEY PLANNING MODEL Final Report, 15 Nov. 1974 - 31 Jul. 1976 (California Univ.) G3/43
N77-21517
HC A07/MF A01
Unclas
22922

FORESTRY APPLICATIONS PROJECT/TIMBER RESOURCE

SAM HOUSTON NATIONAL FOREST INVENTORY

AND

DEVELOPMENT OF A SURVEY PLANNING MODEL

Final Report
14 July 1976
Space Sciences Laboratory
Series 17, Issue 55

UNIVERSITY OF CALIFORNIA, BERKELEY

University of California
Space Sciences Laboratory
Berkeley, California 94720

Final Report
for
NASA Contract 9-~~14552~~ 14452

Forestry Applications Project/Timber Resource
Sam Houston National Forest Inventory
and
Development of a Survey Planning Model

Principal Investigator

Dr. Robert N. Colwell

Project Scientist

Stephen J. Titus

Period Covered

11/15/74-7/31/76

14 July 1976

Space Sciences Laboratory Series 17, Issue 55

PREFACE

Objectives. The task I objective was to develop and implement, in a new geographic area, timber inventory procedures using data sources including Landsat, conventional aerial photography, and direct measurements. The task II objective was to develop and begin testing a planning model for evaluating sampling and remote sensing measurement factors as they interact in sampling-estimation systems.

Scope of work. The Sam Houston National Forest (145,332 acres) in Texas was surveyed to estimate parameters, as nearly as practicable, required by the U.S. Forest Service Forest Survey. The planning model used a multivariate approach to sample survey design, a population representation based on Landsat and supporting data, and concentrated on sampling systems which could utilize remote sensing data sources.

Conclusions. Because of vegetation and terrain homogeneity, Landsat data sources were not useful in identifying vegetation class in the detail required by the Forest Service. However, for less detailed classifications Landsat may still be useful. The survey system utilized Forest Service vegetation stratification data and large scale sample photography in a system which estimated many of the parameters required by Forest Survey. Growing stock volume was estimated as 384 million cubic feet with a relative error estimate of 7.8 per cent. Shortcomings of the inventory system as well as estimates provided in addition to those required by Forest Survey are discussed.

The planning model developed here represents a formalization of the interrelationships which must be considered in planning sample surveys for geographic multivariate applications. Verification tests and an application of the model to the Quincy Ranger District in California have been completed. In management applications where a cell-by-cell population model, such as the Landsat-based model developed in this study, is available, the planning model permits evaluation of the effects of a number of factors as they interact in a particular sampling context. These include 1) size and shape of sample units, 2) selection probability, 3) precision and probability levels, 4) sampling strategy, and 5) cost variables. Finally,application of the model could easily be extended to other disciplines using geographic populations such as agriculture, range, forestry, and regional planning.

Summary of recommendations. With additional training and specifications of variables required for measurement, the kind of sampling system used in this study should be able to provide estimates of all parameters required by Forest Survey. A joint application by both Forest Service and research personnel would provide the necessary details to bring this approach into operational use. Additional work on the planning model is needed to 1) increase computational efficiency and the size of populations and parameter sets which can be evaluated by the model, 2) incorporate additional capabilities, especially more sampling and measurement alternatives, and 3) further test and evaluate the model with changing assumptions and input parameters.

TABLE OF CONTENTS

Page

List of Figures

List of Tables

1.0 INTRODUCTION

The Remote Sensing Research Program (RSRP) at the University of California, Berkeley, has been involved in remote sensing-aided inventory systems research since 1973, first through the ERTS investigation program and most recently the Forestry Applications Project. This research has been directed towards solving the problem of meeting informational needs of the resource managers utilizing remote sensing data sources including satellite data, conventional aerial photography, and direct measurement on the ground in such combinations as needed to best achieve these goals. It is recognized that sampling plays an important role in generating relevant information for managing large geographic populations. The central problem, therefore, is to define the kind and amount of sampling and the place of remote sensing data sources in that sampling system to do the best possible job of meeting the manager's informational needs.

Perhaps the most significant recent event which shows the increasing importance of implementing comprehensive procedures for describing and monitoring wildland systems, is the passage of the Forest and Rangeland Renewable Resources Planning Act (RPA) of 1974 and related legislation such as the Multiple Use-Sustained Yield Act of 1960 and the National Environmental Policy Act. This legislation is directed primarily towards management activities of the U.S. Forest Service, but the implications for other federal and state agencies as well as industrial and nonindustrial private lands are present. In particular, the RPA specifies procedures to be followed in developing comprehensive policies and programs with respect to wildland management and requires frequent resource assessments to evaluate progress towards these ends. The necessity to consider all the varied uses and products associated with these lands has shown the need for extensive broadly-based information on a variety of wildland characteristics including forests, range, fish and wildlife, recreation and wilderness, land and water, and human and community development. It is not at all clear what type of information production procedures will evolve to meet these needs, but the strong implication is that sampling systems will play a significant role and that multiple characteristics will be of interest (Hyde, 1976, p. 283). With this premise, a systematic procedure for evaluating sampling alternatives, where the objective is to estimate several parameters, is of obvious importance.

1.1 Contract Background

Work under the current contract included two separate but related tasks. Task I involved timber inventory of the Sam Houston National Forest in Texas, and Task II was the development of a multivariate survey design planning model. Chapter Two reports on the survey of the Sam Houston NF. Chapter Three describes the development of the planning model including verification tests and application to the Quincy Ranger District of the Plumas NF in California. Chapter One provides background on the informational needs of wildland resource managers and Chapter Four summarizes and makes recommendations for future research.

Task I also involved developing procedures for a "level II" inventory of small selected parcels within the Sam Houston NF. This work as well as a part of Task II concerned with developing a plan to test the planning model in the same area was not possible since Landsat MSS data were not capable of providing the kind of population data required by the planning model. Instead, planning for a "level II" inventory is considered in a general context in Chapter Three as a possible application of the planning model in areas where such data is available.

1.2 Wildland Management and Information Systems

A wildland resource system may be partitioned into two interrelated subsystems: the forest ecosystem and the management system, as shown in Figure 1.

The ecosystem is the physical entity to be managed. It is defined by geographic limits and implies inclusion of the atmosphere above and the interior of the earth below. An example would be the ecosystem specified by the boundaries of a typical Forest Service Ranger District. Two aspects of the ecosystem are of major concern to the forest manager: site -- components including atmosphere, climate, soil, geology, and topography; and community -- the living component of the system, plants and animals. In addition, the ecosystem itself must be considered as part of a larger system, which has a certain influence and creates the environment within which the ecosystem exists. Macroclimate, fauna, and other natural processes from this larger system affect the ecosystem, and conversely the ecosystem may affect the larger system.

The management system consists of man's activities and efforts to manipulate and control the ecosystem. Historically, he has manipulated the ecosystem to acquire the things he needed or wanted with little regard for the long term consequences of his actions. However, as the scale of demands on the ecosystem has increased so has understanding of ecosystem complexities and interactions. The effects of past activities have shown that often the short-term achievement of needs has adversely affected long term availability of other necessary resources. Consequently both our evaluations of the ecosystem as it exists, and our view of what the system should look like must be considered as we evaluate treatments which might be applied to the ecosystem. "Treatment" is an all-inclusive term meaning "an activity carried out by man to alter the system". It could be a complex maintenance and harvesting plan extending over 5 to 10 years, or it could be the opposite extreme, no activity. The desirability, or necessity, of applying treatments depends on (1) the apparent state of the system and (2) the desired state of the system. The latter is a transformation of policy, philosophy, and political objectives

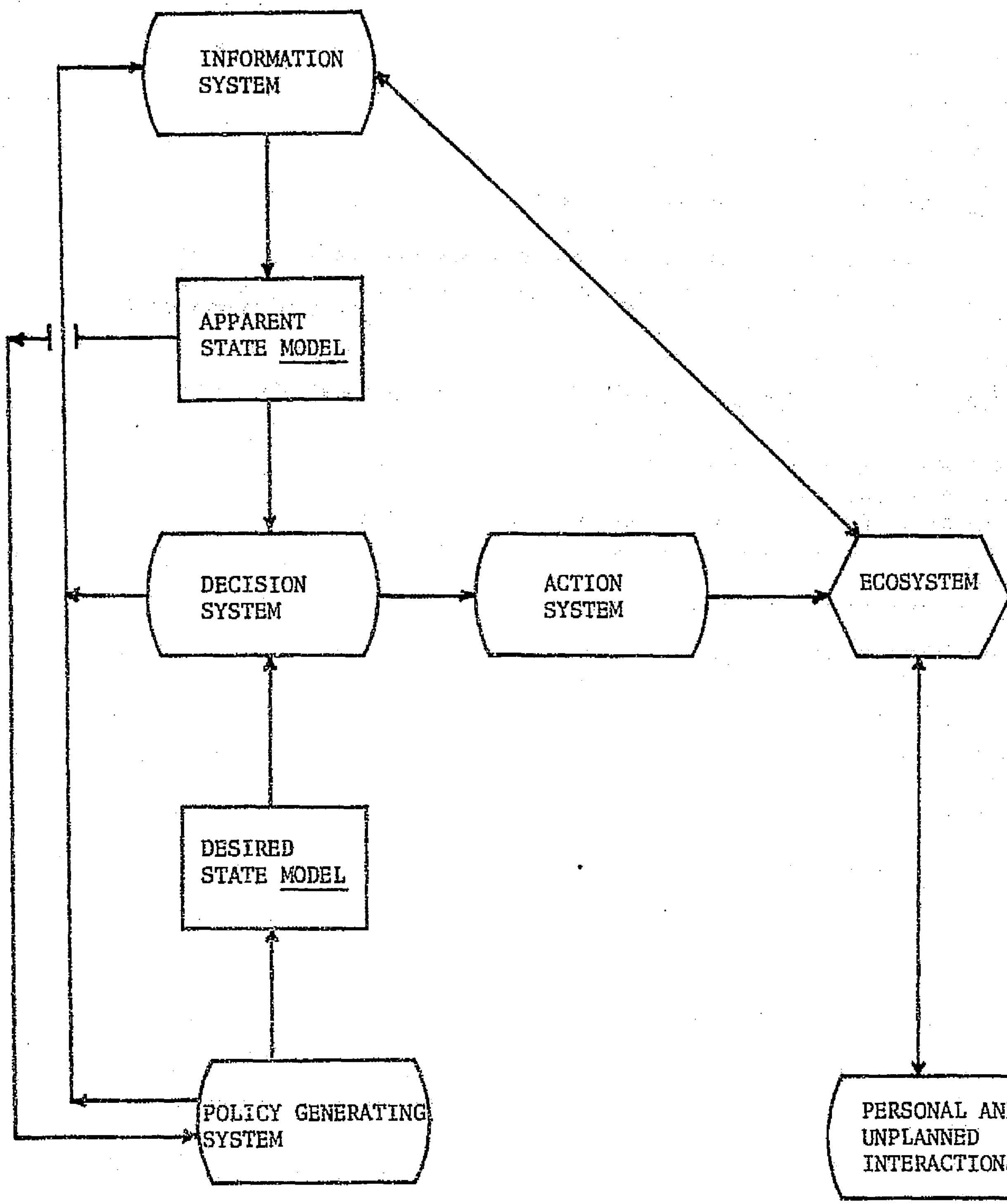


Figure 1. MANAGEMENT SYSTEM and the object of management, the ecosystem.

into some specific desired or optimal state of the system. A comparision of the apparent with the desired state allows evaluation of the need for treatments to remove or reduce existing discrepancies. Additional decision-making techniques may then be used to determine the operational specifications of treatments. The apparent state is based on an interpretation of the actual state; however, this interpretation is imperfect and subject to errors of various types, including possible bias. The term apparent state recognizes this potentially important distinction between what we think is there and what is actually there. Further, while the ecosystem includes all the varied components of the physical and biological complex, the desired and apparent states typically include only a limited number of particular aspects or components of the overall system.

Both the ecosystem and the management system may be viewed as composed of a number of subsystems. In Figure 1, the management system is shown as interacting with the ecosystem primarily through the information system and direct applications of treatments. It is the information system which provides policy-makers and decision-makers with estimates of current conditions so that informed management decisions may be made and appropriate treatments applied to the ecosystem.

1.2.1 The Information System: A Major Management Subsystem

The information system is linked with the ecosystem through data gathering, as well as the decision and action system that utilizes the information provided. The kind of information required by the decision system determines the nature of the information system and the kinds of activities which are necessary to produce the desired information products. If well-established decision models exist, the specific information requirements will be known. On the other hand, in the early stages of management, when little is known about the ecosystem and treatment alternatives are not known or specified, then the requirements may be vague. In this case the description phase may amount to obtaining information believed to be important in understanding the system prior to determining treatment possibilities.

Major activities and components of an information system, Figure 2 , relate to data acquisition, handling and storage, information production, and maintenance. The common factor linking all the components is the data base itself. In addition to all of these, an internal control or management function is also essential for larger systems.

Data Acquisition is the first step in a series of activities which ultimately provides information products. As with the other activities it is important that acquisition be focused on obtaining data which is appropriate for the decisions that must be made by the manager. Many types of acquisition are possible ranging from almost continuous monitoring to

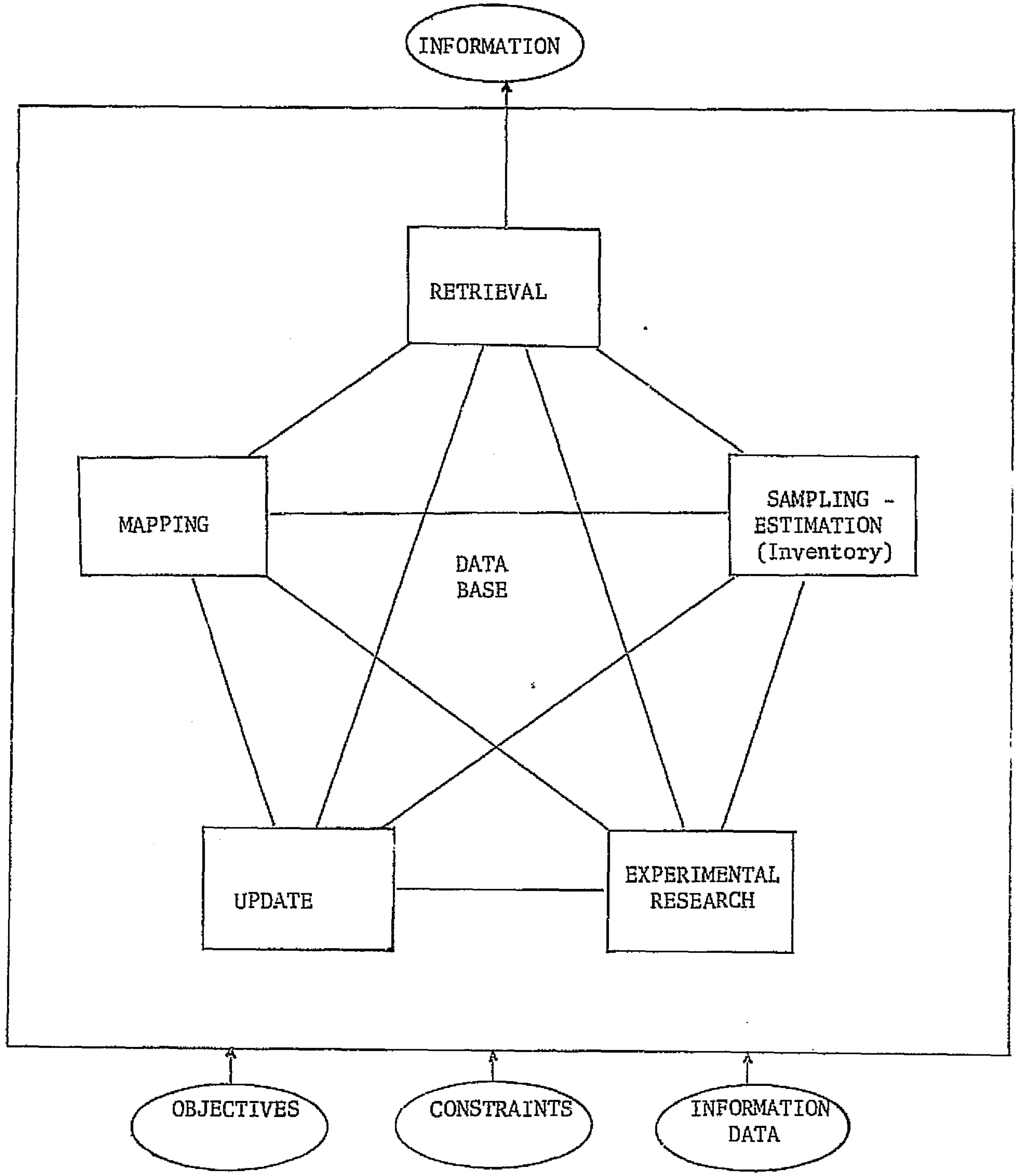


Figure 2. An Information System

periodic sample observations or complete enumeration. However, the quality, amount, and relevance of the data base must be managed within economic bounds; hence the importance of developing procedures for the use of sampling techniques.

Data handling and storage has received considerable attention in literature on geographic information systems (Tomlinson, 1972). Many options exist for data manipulation and storage; the important point however, is that the particular data structure selected must be flexible and that a good geographic reference system be incorporated as an integral part of the structure. This is particularly significant where special purpose summaries require combining data from several sources. Such a requirement would arise, for example, in answering the question, where are lands with 1) slope less than 35%, 2) soil type A, and 3) less than 1320 feet from the nearest road?

Maintenance and update of data files are particularly important so that when new information requirements arise maximum advantage is taken of currently available data. Sample survey design can be much more efficient if useable data files are maintained. All too often, however, the basic data is lost or confused as time passes, requiring costly independent sampling. A real commitment to information system maintenance is required to provide data continuity for long term management use.

Information production includes analysis, reduction, and display of output products in a convenient format for use by decision-makers. Usually both graphic and tabular output is required and some indication of accuracy or precision of information is desirable. A stratified sample could provide all these if the strata are mapped as an overlay to a management base map, and the sample estimates are tabulated as a complement to the strata map. Prognosis for future developments could be handled either as an information system function or as part of decision system activities.

A data base is the common thread which ties the whole system together, and must relate on the one side to management needs and on the other to basic ecosystem characteristics. Specific types of information required vary for different management areas, but all basically relate to particular ecosystem components, and to the time, place, and manner in which treatment activities will be applied.

Based on its definition, three primary informational components, or dimensions, of an ecosystem are evident: 1) time, 2) location, and 3) characteristics, or profiles of information. A time dimension is necessary since it influences all of man's activities. Geographic location within the system is useful since activities are planned for specific parts of the system. Finally, characteristics, or profiles of information, are required for use in management planning and decision-making. An information profile is simply some characteristic of interest, such as vegetation type, elevation, slope class, soil type, or any other variable which is location specific. These three dimensions define a model of the system, portrayed graphically in Figure 3. Time and profiles may be considered on an ordinal scale, e.g. time t_1, t_2 etc. and profile 1,2, etc. with appropriate explicit definitions for each.

Definition of ecosystem profiles may be accomplished using a hierarchical structure to whatever level of resolution is desired. Given specific management activities (which may also be defined hierarchically) it is possible to relate profiles and activities through an "association" matrix (Leopold, et al, 1971) which may identify influences or reference the nature and magnitude of the influence. Table 1 gives an example of an association matrix for a wildland ecosystem. The character in each cell of the table determines the relative nature of relationships between the profile and activities. It would also be possible to use the cell entry as a "pointer" to a more specific quantitative statement of the relationships.

1.2.2 Data Acquisition and Information Production

Information production is the process of translating basic data elements into a useable summary format, and methods are largely dependent on the manner in which basic data is acquired. Data acquisition may be accomplished in many ways, and opportunities for error are associated with each. As will be seen in the following paragraphs, sampling is one technique that permits control of errors while at the same time providing a measure of precision associated with information products. The population size and ease of direct measurement as well as other considerations, such as time and money available for making measurements, influence the decision on what acquisition methods might be used. Three possible sources of error affect information production: specification, measurement, and sampling error.

Specification errors occur in cases where some incorrect decision in the planning phase is made which means that although sampling and measurement errors may be small the required information is still not provided. This situation can arise if there is insufficient communication between people who use information and those involved with its production. Hurried planning efforts often prevent full analysis of information needs. Changes in requirements during the information production phase, and occasionally misunderstanding can lead to improper specification of what is needed. In addition, it is possible that incomplete specifications of the sampling population could lead to a survey of the wrong subjects. Use of an inappropriate statistical model would also be a specification error since even though estimates might be provided, they would be inherently wrong.

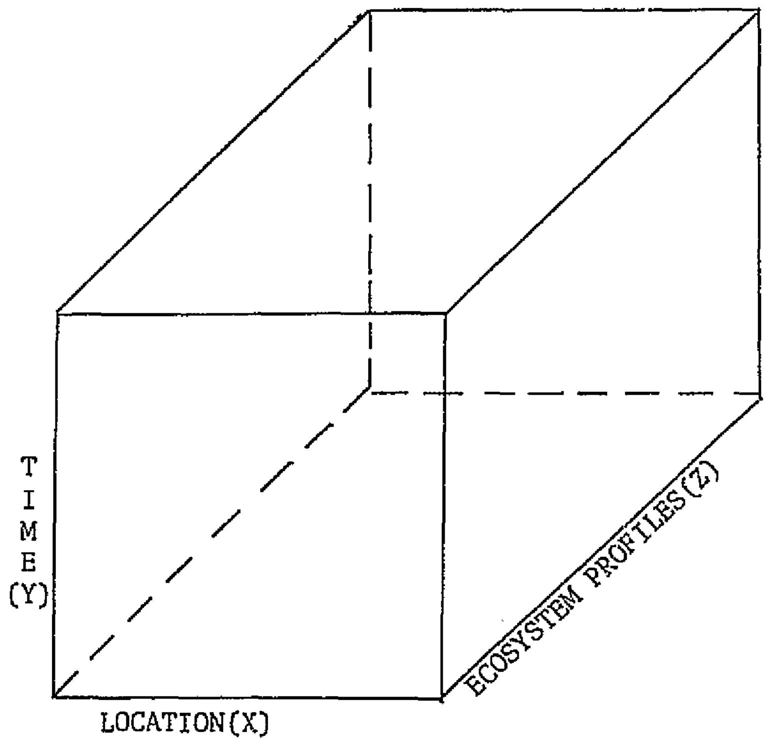


Figure 3. A Three Dimensional Data Base and Model for an ECOSYSTEM

Location: Geographic Location within the System, e.g. UTM Identifier (or code); likely varying with Point, line and area data

Time: Historic records
Present State Data
Future Prognosis

Profiles Specific data profiles of interest to one or more management areas as determined by Management profiles association matrix (Including precision rating)

Table 1. Management - Information Profiles Association Matrix

			Management Areas					
			Watershed	(Harvest) Timber	Cattle (Yield) Range	Wildlife	Recreation	Fire Control
ECOSYSTEM PROFILES	Inputs:	Climate	P	S	P	P		P
		Water			P	P		P
		Animals				P		
		Veg				P		
		Sediment	P					
	Site:							
	Cultural:	Roads	P	P			P	P
		Trails					P	P
		Population		P			P	P
		Ownership	P	P	P		P	P
	Site:							
	Natural:	Landform	P	P	P			
		Slope	P	P	P			
		Aspect	P	P	P			
		Elev	P	S	P	P		
		Soil	P	P	P			
		Geology	P		P			
		Water	P	P				
		Aesthtics		P			P	
		Uniqueness		P			P	
	Vegetation:							
		Timber	P	P		P	P	P
		Range				P	P	P
		Brush	P			P	P	P
	Animal:							
		Big Game				P	P	
		Small Game				P	P	
		Fowl				P	P	
		Insects						
	Outputs:							
		Water	P	P				
		Animal						
		Veg						
		Sediment		P				

P - Primary Relationship
S - Secondary Relationship

Measurement error occurs to some degree because of instrumentation limits or human errors in recording data. The first can be measured, but errors in using the instrument are very difficult to evaluate. Errors can be made in taking readings, recording data, transferring data, and summarization. Also some items might be omitted completely. A final unpredictable factor is the physical and emotional state of the individuals making measurements.

Sampling error is introduced in those cases where only a part of the population is examined. Inferences about the population based on a single sample may or may not provide a good representation of the true characteristics. However, sampling error can frequently be measured since it is possible, at least theoretically, to take repeated samples from a population and evaluate variability in inferences.

All error components can be controlled to some degree. However, increased accuracy and precision often results in higher costs, hence there is an economic interface with the level of error desired in a measurement system. Specification errors can be eliminated by careful consideration of the information needs and appropriate translation into data requirements and statistical models. Instrument error can be kept in line with the magnitude of other error components by selection of appropriate measurement equipment. For example, several possibilities for tree volume measurement are available including diameter tape and volume tables, relascope, dendrometer, and cut-and-measure. Human errors can be controlled to some degree by selection of who is to make measurements, the training given them, and the environment in which they work. Double sampling and field checking of measurements can help to estimate and reduce these errors, but they are the most difficult to evaluate.

Sampling error is usually measured by the variance of an estimator and is a function of sample size. Intuitively, the larger the sample the more precise the estimates are. At the limit, the entire population is measured and sampling error is zero. Measurement error is present even with sample data, but since these data sets are usually much smaller than the population more care can be taken to control measurement error. Judicious allocation of the number and kind of sample observations helps insure that the required information is obtained for the least cost.

Considering these error sources and opportunities for their control, it is obvious that in many cases sampling is an excellent tool for information production. It is usually less expensive than complete enumeration and permits one to reduce the measurement error substantially. By considering tradeoffs between costs and precision of the information derived, sampling error can be controlled to the degree required (or to the limit of budget resources). Population inferences are then "better" than would be obtained with complete enumeration since measurement errors are substantially reduced and sampling error is estimated.

Sampling represents an important first step in a sequence of activities which ultimately may have a significant impact on the ecosystem. The effort that goes into designing a sampling system should therefore reflect the relative significance of decisions which are to be made. With the importance of sampling and its place in management clearly established, the central problem of this study is now considered: how to design the best possible sampling system, within the limits of objectives and constraints imposed by management and the ecosystem.

2.0 SAM HOUSTON NATIONAL FOREST INVENTORY

The Sam Houston National Forest (SHNF) is located in the southern pine region and was selected by NASA as a second area for potential application of remote sensing data in a forest inventory. Initially, the orientation was to be toward meeting the needs of the national forest managers for timber management planning data. Midway through the project but before the survey design had been finalized the emphasis was shifted from local requirements to national data needs as outlined by the Forest Survey Handbook (FSH 4809.11). This change in emphasis was made in response to the requirements for frequent evaluations of the forest resource as required by the Resources Planning Act, of 1974. At the same time there are two related aspects which were addressed in this study as well. First was the need for broad based resource surveys rather than timber or range surveys conducted separately, and second was the possibility that local manager's needs may still be met, partially, if not completely, in the same process that acquired data required by the Forest Survey.

2.1 Estimation Objectives

Forest Survey estimation requirements are currently under revision, but the most recent available version of the Forest Survey Handbook was used as a guide for implementation in this project (FSH 4809.11 dated February 1972, Amendment No. 6). Estimates are required for the entire forest and include 1) areas, 2) number of trees, 3) volume, 4) growth, and 5) mortality with each estimate broken down into various subclasses based on species, size class, tree class, and several others. (For the most definitive statements of the requirements reference to the handbook is required.) A number of output tables are given with specifications for each. These same identifiers are used in the output from this project. Evaluations are included for each case identifying what was provided in addition to requirements, as well as shortcomings and possibilities for their correction, in an operational system. The major additional item provided by this survey is an indication of the precision achieved for each estimate.

2.2 Development of the Sampling System

Given the estimation objectives outlined above a sampling system was constructed which would, as nearly as practicable, meet those objectives by taking maximum advantage of auxiliary information available. The potential auxiliary data that were considered included satellite data, aerial photography data, Forest Service type delineations, and precise yield data for trees on the SHNF.

2.2.1 Population and Parameter Specification

The SHNF including only Federal lands constitutes the population of interest. Table 2 summarizes the set of tables which contain estimates of parameters as required by the Forest Survey. In addition to the 28 tables (broken down broadly into area statistics, forest estimates, product outputs, and projections), are generalized type maps. For a variety of reasons not all the tables are produced by this study and for tables produced the specifications given by FSH 4809.11 are not always obtained. Where discrepancies occur they are identified and discussed.

2.2.2 Auxiliary Variables Available

The aim of this study was to design a survey to take maximum advantage of available auxiliary data. Several data sources were available including remote sensing data of the type utilized in the Plumas 1974 inventory (Titus et al, 1975, Colwell, 1974). Landsat MSS data provided useful information in the first stage of the Plumas inventory. For the SHNF, however, after considerable effort was expended in attempts to relate spectral data to vegetation classes (timber type/condition classes) it was concluded that the uniformity between vegetation classes as sensed by the satellite prevented meaningful discriminant analysis of the data. This conclusion was supported by a subjective comparison of terrain, vegetation, and climatic conditions on the SHNF as contrasted with the Plumas. On the SHNF terrain is very flat, climate is humid and wet, and vegetation is very homogeneous. Even differences between hardwood and softwoods were difficult to detect on large scale aerial photography except at certain times of the year. Forest Service vegetation data were available, including maps and acreages by a detailed vegetation classification. Large scale photography of the type used in the Plumas study enabled measurement on a selected sample area. Individual tree photo measurements were considered impractical because of vegetation density and difficulties of seeing the ground for parallax measurements. Species separations other than hardwood-softwood were impossible; even on the ground species differences for pines can be difficult to detect. Instead, stem counts and percent crown closure of conifers, hardwoods, and snags were taken as well as an average difference of parallax which was relatively well correlated with stand height in a small pilot study. Finally, precise on-the-ground measurements of volume are costly to obtain so typically local volume tables are used. A previous study for the SHNF (Baker, 1975) provided dendrometer measurements for a sample of trees. The volume relationships shown in Baker's study were so good that additional dendrometer measurements as part of the SHNF inventory were considered unnecessary. These local volume relationships use DBH as the independent variable and predict cubic foot volume of the tree.

2.2.3 Sampling Technique and Estimators

The sampling system for the Plumas 1974 inventory was basically a stratified multistage system in which Landsat data were used to make

Table 2. Forest Survey Tabular Output Summary

Forest Survey Table Number	Description
1	Area by land classes
2	Area of commercial forest land by ownership
3	Area of commercial forest land by stand size and ownership
4	Area of commercial forest land by stand volume and ownership
5	Area of commercial forest land by stocking classes and tree class
6	Area of commercial forest land by stand condition and ownership
7	Area of commercial forest land by site and ownership
8	Area of commercial forest land by forest type and ownership
9	Area of commercial forest land by forest type
10	Number of growing stock trees on commercial forest land by species and diameter class
11	Net volume on commercial forest land by softwood and hardwood
12	Softwood net volume of growing stock and sawtimber by ownership
13	Growing stock volume by species and diameter class
14	Sawtimber volume by species and diameter class
15	Sawtimber volume by species and quality class
16	Growing stock net growth and removals of growing stock by species
17	Growing stock net growth and removals of growing stock by ownership
18	Sawtimber net growth and removals by species
19	Sawtimber net growth and removals by ownership
20	Mortality of growing stock and sawtimber by species
21	Softwood mortality of growing stock and sawtimber by ownership
22	Mortality of growing stock and sawtimber by causes
23-27	Product output tables
28	Projections

the stratification as well as define the primary sampling units, (Colwell, 1974). For the SHNF the use of Landsat data was not feasible so the Forest Service stratification was utilized instead. For this reason as well, sampling units consisted of variable size clusters located at random within the forest areas and with allocation among strata proportional to area. The clusters were defined by location of 0.4 acres circular plots on large scale aerial photos acquired September 20-21, 1975.

The number of photo plots in the cluster was determined by the proportion of the line which fell in the strata of interest up to a maximum of 10. A subsample was taken from the set of photo plots for direct measurement of variables by a field crew. Figure 4 shows schematically the various activities; a technical summary of the system including the estimates is provided in Appendix A.

2.2.4 Measurement Procedures

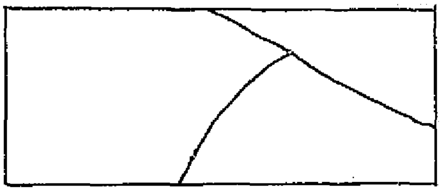
Measurements were made to summarize the stratification scheme, to obtain photo variables, and to obtain direct measurement of variables of interest on the ground. Stratification data were summarized directly from Forest Service records. Photo measurements were made of average parallax difference (a proxy for height), number of trees by various categories and percent crown closure by the same categories. Individual tree measurements were not considered necessary or practical owing to the extreme uniformity of vegetation conditions. Ground measurements included tree species class, diameter, ten year radial growth (softwoods only) bark thickness (softwoods only), and tree quality class. Site class, stand age and average height were recorded as well as necessary administrative data. Detailed measurement procedures are summarized.

2.3 Analysis Procedures

Estimates and tabular output products were generated using a series of FORTRAN programs to summarize the basic data and generate estimates. Two computers were available for the actual computer processing of the inventory data: (1) the large computer complex (CDC 7600) at the Lawrence Berkeley Laboratory, which was used for processing the Plumas inventory, or (2)a Data General dedicated computer using a real time disc operating system (RDOS) recently acquired by RSRP for machine processing of MSS data. The latter system was chosen for the analysis since this project offered an opportunity to evaluate the utility of such a system for a medium scale analysis effort involving a large quantity of numerical data. The dedicated computer system offered the advantage of almost real time turnaround of processing as well as the potential for cost savings over the larger computer facility. Computational limits of 3-4 significant digits were encountered unless double precision computing was used. For this research study single precision computations were used.

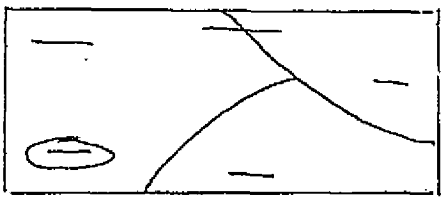
Schematic

Activity



1) Stratification

A partitioning into units of relatively homogenous vegetation. Based on USFS data.



2) A random cluster sample within each stratum where cluster sample units consist of up to 10 plots along a flight line.

Measurements of Photo Variables made for each of the 0.4 acre photo plots

3) A random sample of 0.4 acre plots from the set of all photo sub plots for direct measurements of characteristics on the ground

Figure 4. SHNF Inventory Overview

2.4 Results

Using the format and notation of the Forest Survey Handbook, the inventory results are presented in Appendix B. Eight tables are included and are numbered, according to the handbook tables references, 10, 11, 12, 13, 14, 16, 18, and 20. They relate to number of trees (10), volume (11,12,13,14), growth and removals (16,18), and mortality (20). Each table also includes relative error, (the standard error of the estimate divided by the estimate), for each table entry. For comparisons of these estimates as they relate to the USFS specifications, see section 2.5.1. Tables 23-28 relate to production outputs and projections and are not derived from the same sources as the standing timber statistics. Tables 1-9, 15, 17, 19, 21, and 22 are not included for various reasons as outlined in section 2.5.1.

Table 3 shows a summary of estimates for the five basic parameters of interest. Precision of the estimates compares favorably with the goals of Forest Survey (five percent relative error per billion cubic feet of volume, growth, and removals) which are given for very large areas. Goals for smaller areas, such as a single national forest would be expected to be larger. Both growth and volume relative errors are less than 10 percent and error for number of trees is about 5 percent. The estimates for both removals and mortality are much higher than the other estimates because both are sporadically distributed over the area and the sampling scheme is not oriented towards obtaining this kind of information.

Additional estimates would be available on a stratum basis since these estimates are obtained in order to calculate overall estimates. They have not been included here since the primary interest is on Forest Survey requirements.

2.5 System Evaluation

The success of the Sam Houston inventory is indicated in the previous section. In addition, however, it is necessary to compare this inventory with the Forest Survey requirements as well as evaluate the cost and effectiveness of the survey. Finally various components of the system are evaluated to determine what directions future research efforts might take.

2.5.1 Comparison with Forest Survey

FSM 4809 provides very detailed specifications for various parameters to be estimated. As nearly as practical these specifications were the objective for the SHNF inventory. A number of items were not provided by this study in as complete a detail as required by FSM 4809. Volumes were summarized as gross rather than net because

Table 3. Sam Houston National Forest 1976, Summary of Estimates and Comparison with Simple Random Sampling

Parameters of Interest	Stratified Ratio			Simple Random			
	Estimate Mean/Acre	Estimate Total	Relative Error(%)	Estimate Mean/Acre	Estimate	Relative Error(%)	Increase in Relative Error(%)
Area (from USFS records)	-	145,332	-	-	-	-	-
Number of trees	160	23,300,000	4.93	165	23,000,000	8.74	47
Growing stock volume (5" + DBH, Ft^3)	2642	384,000,000	7.76	2810	408,000,000	9.71	24
Sawtimber volume (11" + DBH, Ft^3)	1975	287,000,000	10.8	-	-	-	-
Volume growth (10 year, $Ft.^3$)	76	11,100,000	6.84	81	11,800,000	8.52	25
Volume removed (10 year, $Ft.^3$)	56	8,070,000	55.1	55	7,990,000	67.8	14
Mortality volume (10 year, $Ft.^3$)	41	6,020,000	74.4	33	4,800,000	92.5	25

assessing defect was difficult for the field crew which was from California and had only a limited amount of time for training. Through the use of local foresters this difficulty could be removed and on-the-ground measurements of net volume obtained. Most tables requiring species breakdown use several hardwood classes. Species identification was difficult because of the time of year when field work was conducted and because of unfamiliarity of the field crew with local species types. For these reasons only two species groups were considered, hardwoods and conifers. Log grade as well was not evaluated because of the time required to train a field crew. Growth, removals, and mortality were all evaluated at least partially. The growth estimates do not include two components of net growth as defined by forest survey: 1) ingrowth (small trees which pass the 5.0 minimum diameter limit) and 2) increment in trees which were cut, died or became culls. Also growth, removals, and mortality estimates are 10 year rather than annual values.

A number of Forest Survey tables were not included in the output from this study. FS tables 1-3, 8 ,9 require area data readily available from the USFS vegetation records. FS tables 4,5, and 6 require areas by volume class, stocking class, and stand condition class. These items were not incorporated into photo interpretation results but could be generated given definitions of the various classes. FS table 7 requires area by site class and these data are best obtained from another source since once obtained they are not likely to change. FS tables 17, 19, 21 and 22 require estimates already available in other tables; these tables represent summaries by ownerships and this study only included USFS lands. FS table 15 requires volume by quality class and the difficulties of obtaining quality data have been mentioned. Since permanent plots were not available and it was considered undesirable to use an increment borer on hardwoods, growth was not evaluated for hardwoods.

Permanent plots which are remeasured over time are generally considered desirable for estimating change. The possibility exists that large scale photographs of specific sample locations could be used for specific location of the sample areas for photo acquisition at a second point in time. If the details of precise location could be worked out, measurement of photo characteristics might correlate well to change in volume etc. It is even possible that some electronic device might be utilized for precise location of certain permanent locations. However, even without such elaborate measures a short interval between measurements, for example five years, would enable relatively easy location of sample areas using previously obtained imagery.

2.5.2 Cost-Effectiveness

Cost effectiveness comparisons between this survey system and USFS procedures in the Southern Region are difficult due to our unfamiliarity with local practices and even Forest Survey procedures in that area.

Independently of this, however, statements relating to approximate cost in dollars and time are possible. All figures given in this section ignore manpower and other costs associated with other tasks of this contract. In addition, the effort expended in attempting to extract useful information from the Landsat MSS data is excluded. Report preparation costs and problems associated with changes in the objectives of the study are also excluded. Considering manpower, flying time, photographic supplies, computer processing costs, and administrative and overhead costs, the survey cost was approximately $25,000 if software development is included and roughly $20,000 excluding software development. Manpower time requirements were 8.5 man months for survey activities, and 1.5 man months for software development. The ratio of ground plot cost to photo plot cost was roughly 5:1 excluding per diem and travel to and from the SHNF and including much greater labor costs than the Plumas inventory expenses. If all activities could have proceeded with the best possible timing the estimates could have been provided in about a 6 month time frame.

2.5.3 Stratification

As already pointed out the initial objective was to utilize Landsat MSS data in the stratification of the study area. Instead, existing USFS vegetation classification results were used. That classification showed ten different combinations of vegetation and condition. It is interesting to note that fifty-five percent of the area occurred in one stratum, namely loblolly pine immature saw timber. The second largest stratum included only eleven percent of the area and the smallest stratum was only 0.07 percent of the area. Hardwoods occupied only five percent of the area. For sampling purposes some stratum had to be grouped together. With such a skewed distribution in the stratification, consideration should be given to possible alternatives for a better stratification. On the other hand this does tend to confirm earlier observations on the general uniformity of the area.

2.5.4 Sampling Techniques and Design Considerations

Results of the survey in terms of relative precision estimates are satisfactory for all except removals and mortality estimates. For these parameters some other allocation, stratification or selection probabilities might prove more useful. It is also possible that unless low precision limits for these parameters are acceptable it may prove more reasonable to conduct separate surveys or consider alternative sources of information for removals. The use of ratio estimators was based on theoretical results and it would be desirable to consider post survey analysis of the relationships implied to determine if other estimators are more appropriate. See also section 2.5.7.

2.5.5 Large Scale Photo Measurements

Aside from the uniformity of the vegetation complex, two other factors contributed to problems in interpreting the large scale, small format/photos. First, in the mixed hardwood type, the season of year in which the supporting photography is obtained is critical and is dependent upon what characteristics are to be estimated. For this study, photo acquisition occurred in September when many of the hardwoods still had their leaves and were easily confused with the softwoods. It should be noted, however, that if the photos had been acquired later in the year after the hardwoods had dropped their leaves they could not have been separated from the snags in the forest. Second, because a small format (35mm) camera was used to acquire the large scale photos, stereo triplets were required to ensure that complete stereo coverage of the sample plot, located in the center of the middle photo, was obtained. For a number of the plots two stereo set ups were required, viz., photos 1 and 2 and photos 2 and 3, to complete the interpretation of the plot this was not only time consuming but required the added effort of maintaining a double recording procedure. This problem might be solved if the photo plots were relocated in the overlap area.

As shown in Table 3 the relative increase in the estimated error when only simple random sampling was used was dependent on which variable was estimated. It is possible that some other photo variables or combinations of variables might lead to even greater efficiencies for the ratio approach. Previous cutting (harvesting or thinning), was not evaluated on the photos, but both types of activities could be identified and subjectively at least, it may be possible to estimate relative amounts of timber removed.

2.5.6 Ground Measurements

The only significant problem encountered with ground measurements was that of physically locating the sample areas on the ground. Experienced photo interpreters were able to overcome this problem reasonably well. Variables which were required by Forest Survey but not measured were excluded largely due to limited time for training the field crews.

2.5.7 Photo-Ground Relationships

Empirical relationships between five photo variables and five ground variables for the 29 paired plots are summarized in Table 4. The correlation coefficients show only moderate relationships between photo and ground variables except for number of trees which is much stronger. However the usefulness of photo variables in the sampling system depends on both correlation and the relative variability of the photo variables. Coefficients of variation are included in the table to provide the additional information needed to determine if ratio estimation would be an improvement over simple random sampling. According to Cochran (1963) ratio estimation will provide greater precision if

$$\rho > \rho^{\circ} \text{ where } \rho^{\circ} = \frac{1}{2}\frac{CV(x)}{CV(y)} \text{ and CV is the coefficient of variation.}$$

For the data in Table 4, (1-% con.) and growth yield $\rho^{\circ}=0.7835$ and for (DP) and growth $\rho^{\circ}=0.3543$. Thus since in the first case $\rho=0.4641$ these would be no gain and in the second case $\rho= 0.4014$ so there would be a gain in using ratio estimation. Based on this comparison DP was used as the photo variable in a ratio estimate of growth. Similar calculations are possible for the other photo variables.

Table 4. Correlations Between Photo and Ground Variables for 29 Plots in the SHNF.

		PHOTO					GROUND					Coefficient of Variation %
		# Trees	DP	1 - % Con.	# Snags	1 - % CC	# Trees	Volume	Growth	Mortality	Removals	
G	# Trees	.7510					1.000					39.1
R	Volume	-.0703	.5004				.3476	1.000				43.4
O	Growth	.2244	.4014	.4641			.5220	.6201	1.000			38.1
U	Mortality	.0223	-.0221	-.1673	.4266		-.0405	.3343	-.2598	1.000		280.9
N	Removals	.1094	-.2687	-.1462	.1823	.2320	.0748	.3152	-.2478	-.0892	1.000	413.5
D												
P	# Trees	1.000										50.6
H	DP	.3592	1.000									27.0
O	(1-%Con.)	.8718	-.1554	1.000								59.7
T	(# Snags)	.4134	-.2303	.1890	1.000							34.5
O	(1-%cc)	-.5597	-.1882	-.7845	-.1771	1.000						136.4

DP = Average difference in parallax

1-% Con = 1 minus per cent crown closure of conifers
Snags = the number of dead trees
1-% cc = 1 minus per cent crown closure expressed in hundredths, (all species)

3.0 SURVEY PLANNING MODEL

Survey planning is by nature not restricted to one particular kind of application or population type. The work described in this chapter was undertaken in an attempt to better define the procedures required to design sampling systems utilizing remotely sensed data; hence much of what was developed is presented in those terms. At the same time the sampling background and conceptual development are independent of these considerations. The population model used, however, is strongly remote sensing oriented in that it implies considerable information about the population in a digital format which can be analyzed by computers. This kind of information has only recently become available through the ERTS and Landsat programs. Early results have shown that these digital spectral data can be analyzed and provide useful land use data with resolution elements of roughly one acre. This chapter contains 4 major sections which 1) present the sampling problem framework and planning model conceptual development, 2) describe the components and functions of the planning model, 3) present the verification tests and application results, and 4) discuss potential applications to small wildland populations.

3.1 Sampling and the Survey Design Problem

Logical questions when considering a sampling task are: How do we design an optimal sampling system and what does "optimal" mean? In answering these questions, three topics are discussed in this section. First, basic approaches to univariate sample design are included for historical background. Second, sampling for estimation of several parameters is discussed as an increasingly important aspect of sampling wildland populations. Third, the conceptual framework for a multivariate survey planning model is presented. Orientation of the model is towards geographically referenced populations. The population model used in the planning context is derived from multispectral data obtained from the Landsat satellite and combined with supporting data acquired by direct measurement of specific variables.

3.1.1 Univariate Survey Design

Survey design is the process of putting together appropriate sampling and measurement techniques to estimate a parameter of interest for a particular population. According to Raj (1968, p. 31) the fundamental principle of sample design is related to the fact that:

> "With every sampling and estimation procedure is associated the cost of the survey and the precision (measured, say, in terms of the mean square error) of the estimates made. Only those procedures are considered from which an objective estimate of the precision attained can be made from the sample itself. And the procedures should be practical in the sense

> that it is possible to carry them through according to desired specifications. Out of all these procedures of sample selection and estimation (called sample design), the one to be preferred is that which gives the highest precision for a given cost of the survey or the minimum cost for a specified level of precision. This is the guiding principle of sample design." (Raj, 1968, p. 31)

Thus the objective is a sampling system which will estimate the parameter of interest with the greatest precision and the least cost subject to whatever constraints may be imposed by the management system. This formulation, however, assumes perfect knowledge of exactly what is needed and how precisely it is needed, or equivalently how much it is worth, in the context of decisions which are to be made using the information. This assumption, however, is rarely, if ever, met and considerable additional research is needed on this important question. In addition, it must be pointed out that a minimize cost objective ignores many benefits which might be associated with one system compared to another over and above the cost effectiveness similarities.

The objective can be translated into specific terms for sampling so that a "cost effective" system would provide the desired sample estimates for the least possible cost, subject to time and other constraints. For design purposes, the obvious performance measure is the cost of the system which meets all the specified constraints. The decision rule, depending on whether cost or precision is considered fixed, is to select the system which gives minimum cost subject to a specified precision level, or select the system which gives minimum variance subject to fixed cost. In a broader decision-making sense however, it must be noted that considerations in addition to system cost will necessarily be incorporated into the final decision on a particular alternative. These factors would include technical constraints such as the time of year when measurements can or must be made, availability of necessary measurement equipment, training and/or experience levels required for various techniques, and capital costs associated with setup for particular sampling or measurement techniques; social factors relating to motivation and willingness to alter existing procedures; and political factors relating to organizational structure, function, and responsibility. Additionally, the decision may be based on an array of solutions with varying assumptions about the nature of the population and measurement relationships.

Without any knowledge of population conditions survey design plays a minor role and sampling is largely restricted to random sampling. As observed by Mahalanobis (1952, p. 40):

> "It is only when some previous information (which may be only approximate in nature) is available about the field that the problem of the sample-design becomes important. The object then is to use the available information in the most effective way to prepare an improved sample-design (in the sense that it would be an improvement over the best design that would otherwise be possible) so that it can be reasonably expected to reduce the cost of the survey as much as possible without sacrificing the accuracy, or alternatively, reducing the margin of error (or uncertainty) to the greatest possible extent for the same expected cost."

Remotely sensed data, as from aircraft and spacecraft, provide auxiliary information which is inexpensive relative to direct measurement costs and is often related to the variables of interest. This kind of data offers the possibility of improved survey design as well as the opportunity to minimize the need for costly direct measurements, particularly when dealing with large geographic populations.

Identification of alternative sampling strategies depends on the nature of the population, the parameter of interest, cost or precision constraints, and other institutional factors. The system will be greatly influenced by the definition of 1) the sampling frame, 2) the sampling units, and 3) supplementary information and its availability. Within the same problem context, variations are possible in the way the sampling frame and units are formed; this influences the amount and type of auxiliary information which determines the appropriateness of different sampling techniques. Previous sampling experience and an understanding of the nature of the population and problem context aid the designer in limiting the set of feasible systems to a small number of likely candidates. Since selecting some subset of alternatives offers the possibility of excluding the best one, considerable care must be used in selecting the set of alternatives for detailed evaluations.

Given a set of alternative sampling systems, it is still necessary to determine the optimum allocation of effort for each.

The decision rule stated earlier requires minimizing cost (or variance) subject to a constraint on precision (or cost). For any given sampling alternative it is possible to evaluate both the cost and precision in terms of mathematical relationships, which are functions of sample size. Then, using the techniques of LaGrangian multipliers the combination of sample sizes which will yield the minimum cost, while still achieving the desired precision levels, may be determined. After repeating this procedure for each of the alternative designs, the one with lowest cost may be identified.

With the population specified and estimation objectives defined, the design problem for any system is solved by fixing:

1) the sampling frame and stratification
2) the sample units and measurements required
3) the selection probability scheme and procedure
4) the estimator functions
5) the variance estimator functions
6) the sample size and allocation.

Success of the optimization procedure is dependent on how closely population conditions are reflected by approximations used in allocating resources. Since estimates of population conditions are subject to considerable uncertainty, it may be desirable to adjust them to reflect optimistic or pessimistic attitudes towards their value. Several methods can be used to modify these estimates to account for uncertainty (Duerr et al., 1975).

Considerable research has been done on theoretical comparisons between sampling techniques and methods (Cochran, 1963; Raj, 1968; Sampford, 1962; Hansen, Hurwitz, and Madow, 1953). Most discussions of survey design are, however, devoted to listing and describing the various techniques (Raj, 1974). Ek (1968) and Schreuder, et al., (1963, 1971) have made comparisons among a number of sampling techniques using forest populations, but design and allocation of effort were not included in their work.

Aldred (1971) is one of very few who has attempted to put a number of techniques together with a particular population representation in a planning model. His work was limited to univariate methods and did not consider any multistage sampling techniques, but it provides an excellent example of a systematic approach to the problem of univariate survey design. The performance criteria, P, was defined as the variance of the estimator $\hat{\mu}, V(\hat{\mu}k)$, where k represented the identifier of a particular alternative sampling procedure. The objective for each design was to minimize P subject to a fixed budget constraint; a linear budget function was assumed. For each design the least-cost allocation of samples was used in a simulation of sampling from a population model to evaluate both sampling methods and their interactions with various aspects of the measurement process.

3.1.2 Design for Estimation of Several Parameters

Even though most surveys are designed for only one parameter of interest, data are usually collected on other variables as well, since from the manager's viewpoint there is increasing need for information on a number of characteristics. It is logical to acquire this data as much as possible within the context of one sample survey since a large segment of wildland survey cost goes to preparations and travel time. The incremental cost associated with additional measurements on selected sample units is quite small by comparison. On this basis it is logical to ask: how can we design a survey to estimate several parameters, considering cost and/or precision desired for all the parameters estimated? To solve this problem requires use of the basic decision-making procedures used in the one variable design problem but with new objectives and performance criteria.

When discussing statistical procedures dealing with several variables, certain definitions of terminology are desirable to avoid confusion. "Multivariate", taken in its general sense, simply means several, or more than two, variables; however, it also refers to a body of statistical analysis procedures dealing with the multiple variables as a set (Kendall and Stuart, 1968, p. 239). The key feature of multivariate data is that a vector of observations is associated with each sample element in the data set. Multivariate analysis makes inferences based on the set of variables as a whole rather than separately. On the other hand there are cases where several variables are considered but the data sets are definitely not multivariate in nature, and these may logically be referred to as "multiparameter" data sets. For example, if data on gross yield for a set of N trees are available, but net yield is only recorded for a small subsample of the N trees, then the data is not multivariate since there is no vector of observations. The data sets would more reasonably be referred to as multiparameter. The terminology problem is further complicated by the analysis procedures which may be used to generate estimates and make inferences based on the data sets. Multivariate data sets could be analysed using either the usual univariate estimation methods, with each variable being treated independent of the others, or they could be analysed as a set using standard multivariate procedures (Morrison, 1967). The survey planning model developed here has been termed a multivariate model since the sample data set is in fact a vector of observations for each sample unit. It must be qualified, however, since analysis procedures for evaluation of alternative designs are based on univariate estimation methods for each of the variables. Multivariate analysis could be incorporated if simultaneous inferences (Miller, 1966) were needed for management decisions. However, at this point the decision problems have not reached a level of complexity where simultaneous inferences are useful. In this study, the term multivariate will be used in its general sense and will not refer to the very specific multivariate statistical analysis context.

3.1.2.1 Objectives and the Decision Criteria

Since sampling for several variables is a natural extension of the univariate case, optimization objectives and the decision criteria reflect most of the limitations outlined in the previous section on univartiate methods. These included 1) the possibility that the set of alternatives considered in the design process might not contain the true least-cost alternative, 2) the assumption that the precision desired reflects the utility of information derived, 3) the use of a decision criteria which does not evaluate benefits, as well as cost, of one system as compared to the other, and 4) the need to include exogenous variables in deciding which alternative is "best".

One objective, which evolves naturally from univariate procedures would be to minimize cost subject to variance constraints for some or all of the parameters to be estimated. Others have been suggested including 1) minimize cost plus loss (Cochran, 1963, p. 120) where loss is a weighted function of variances for each parameter; 2) a variation on the approach of Wensel (1974) which is to minimize a weighted function of relative variances subject to a constraint on cost; and 3) minimize cost subject to constraints on the generalized variance, the determinant of the covariance matrix (Chakravarti, 1954; Ghosh, 1958, p. 162; Arvanitis and Afonja, 1971). Only the first approach places precision constraints individually on one or all the characteristics to be estimated. Each of the others implies use of rather novel performance criteria or constraints. If, for example, appropriate weights could be specified, the composite variance function of Wensel's approach could be used. However, definition of loss functions, weighting factors and the meaning of the generalized variance require additional evaluation and testing to determine how they might best be used, if at all, in the context of current decision-making and management activities.

Minimizing cost subject to constraintson variances, or providing estimates of each parameter within desired precision limits, is useful in the current management context since parameters are generally not all of equal importance. Additionally, managers can usually specify precision levels which reflect the relative importance of the different parameters. This objective and the implied performance criteria, system cost, will be assumed in the later development of the survey planning model. The decision rule will be to select the alternative design with the lowest cost.

3.1.2.2 Allocation of Effort

After specifying the precision desired for each variable and the minimum-cost objective, several approaches can be taken to determine the best allocation of effort. Univariate methods could be used to make a sample allocation determination for each parameter and then by some process derive a sampling procedure which would provide for all required estimates. Examples could include 1) separate independent samples for each parameter, 2) several independent samples to survey groups of related characteristics, 3) a single sample but with measurements taken only on the minimum required number of sample units for each parameter so that certain units would not be measured for all variables, and 4) a sample selecting the maximum number of units required for any variable and still measuring all characteristics on all units. These univariate procedures offer some control on precision, but do not necessarily minimize the cost. Further, opportunities for simultaneous inferences are limited and strict multivariate analysis procedures are not possible except for the fourth alternative which generates a multivariate data set.

Assuming that all characteristics are measured on each sample unit, a multivariate allocation based on a mathematical minimization procedure, referred to as programming (Hillier and Lieberman, 1967), and discussed extensively in later sections, may be employed. Using that approach, the objective is to minimize the cost of the survey, expressed as a mathematical function of sample sizes, subject to constraints (usually variance) for each parameter of interest. Programming is the mathematical procedure of finding that allocation of sample sizes which yields the lowest cost. Huddleston et al. (1970) describe application of this technique to stratified sampling and compare it to two other allocation procedures. Their example showed that with an allocation based on the fourth univariate method listed above, over-sampling occurred and estimates were more precise than required. Using another allocation, based on the average of sample sizes for each variable by strata, sampling failed to meet all precision constraints. In Huddleston's stratified example, allocation by the programming technique yielded the lowest cost while still meeting all constraints.

Comparing the programming allocation approach with the third univariate method is difficult due to problems of identifying differential sampling ratesand incremental measurement costs associated with each variable for the univariate allocation method. For a stratified example comparison is possible by considering the general form of cost functions for both methods of allocations. Consider the two functions for total variable cost (TVC) for the programming (PRO) and separate (SEP) allocation alternatives:

$$\mathrm{TVC(PRO)} = \sum_{h=1}^{L} (n_h c_{ha} + n_h \sum_{i=1}^{M} c_{hi})$$

$$\mathrm{TVC(SEP)} = \sum_{h=1}^{L} (n_{hmax}(c_{ha} + c_{hb}) + \sum_{i=1}^{M} c_{hi} n_{hi})$$

where L = the number of strata

M = the number of parameters

n_h = the sample size by programming for the h^{th} strata

n_{hi} = the sample size by optimum allocation for the h^{th} strata and variable i

n_{hmax} = max $(n_{hi}, i=1,\ldots,M)$

c_{ha} = access cost for units in strata h

c_{hb} = cost of implementing differential sampling rates in each strata

c_{hi} = incremental measurement cost for variable i in strata h.

The second term in the cost function will be greater for the programming approach since all variables are measured for each sample unit. But, the first term will be greater for the separate approach since 1) the maximum number of sample units for each variable must be visited in each strata and 2) there is the additional cost due to differential sampling rates. The actual values of the cost coefficient determines which alternative is less costly; however, if access costs are high compared to measurement costs and the c_{hb} term increases as the number of variables, it is likely that the programming approach will be the most cost-effective way to allocate effort.

3.1.3 Formulation of the Multivariate Survey Planning Model

Given the important role of sampling for resource management and the increasing emphasis on multivariate sampling applications, it is evident that a formal planning mechanism is needed to document and improve the selection of techniques and procedures to be used in meeting estimation objectives. One approach to providing such a mechanism for multivariate survey design is the planning model developed in this study.

A basic function of the model is to systematically evaluate alternative sampling schemes using available population and cost data. Results of this evaluation include estimates of cost and implementation specifications for each of the alternatives. One significant benefit of the planning model approach is that assumptions, inputs, and evaluation techniques are all formalized and subject to review, modification, or improvement.

Two assumptions are crucial to the model development. First, advance data is needed for each of the variables of interest. In addition, the data must be geographically referenced in detail so that it can be combined and summarized as required by sampling alternatives. This kind of population data, even if it is somewhat crude, allows flexibility in evaluating, among other things, size and shape of sampling units and methods of stratifying the population. Second, the goal is to minimize cost subject to precision constraints for each of the parameters to be estimated. Generally a survey budget is fixed in advance by administrative considerations, but desired precision and expected cost must be consistent with overall budgeting limitations. A minimize-cost objective is useful even if a budget is fixed since it will show the expected survey cost. If this cost is greatly different from the fixed budget, either above or below, the budget or precision limits need to be modified, otherwise money would be spent acquiring extra data not really needed, or the full budget would be spent and the result would be far from meeting precision objectives.

Figure 5 shows the basic components of the planning model. Advance population data and problem specifications are the primary inputs required, and a central control mechanism is needed to coordinate the other model components. The complexity of the population model depends on the nature

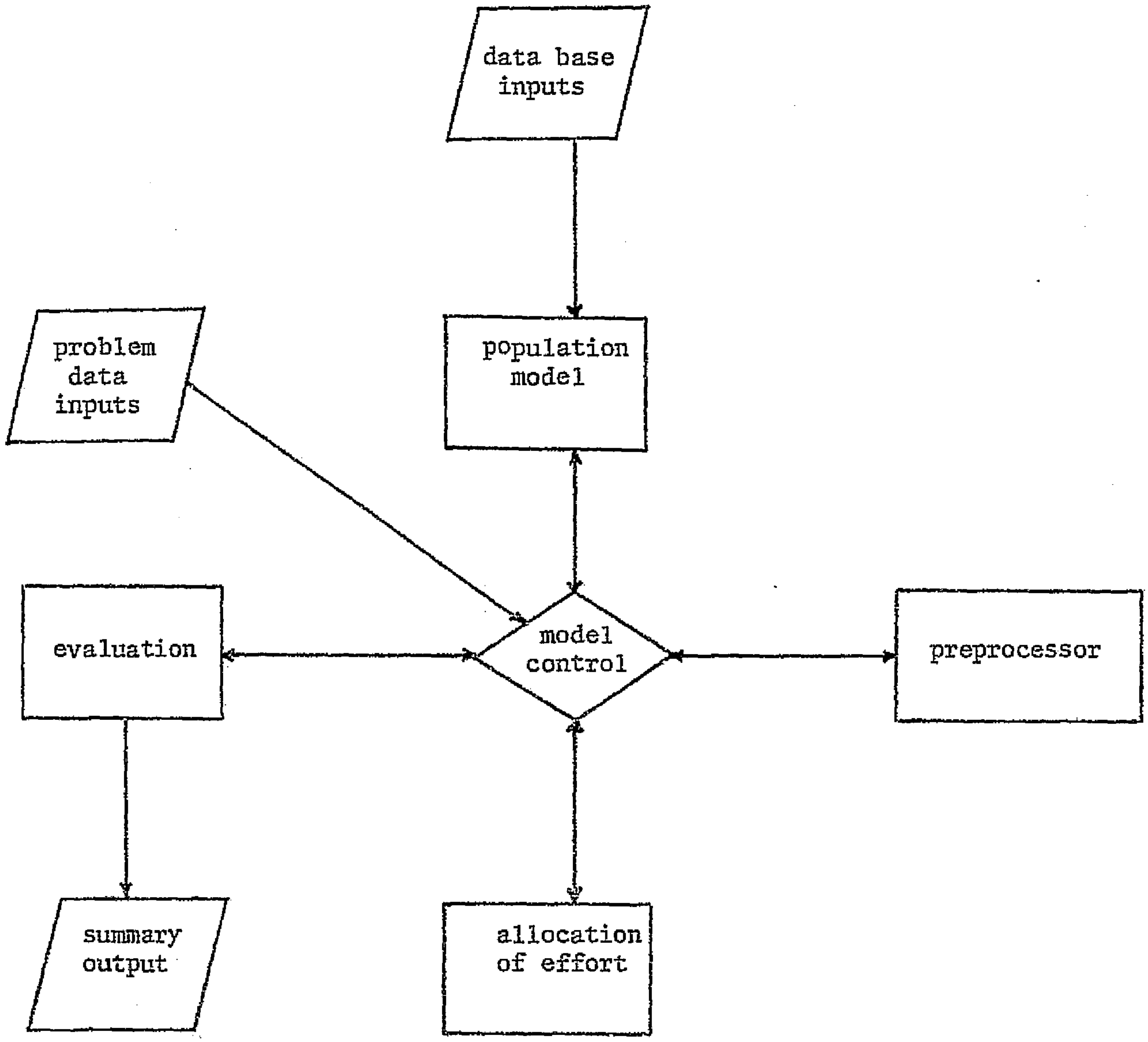


Figure 5. Basic planning model components.

of available population data. In an extreme case the data could be complete so that only an input device would be required; on the other hand it is more likely that several data sources would be combined to form a population model, possibly with simulation of missing or incomplete data elements. A preprocessing activity summarizes population data in a specific form required for a particular sampling alternative. Then, given the summary data on the population and other inputs, the allocation of effort component identifies, for a particular alternative, the sample sizes which minimize cost while meeting precision constraints. Aside from model control which coordinates and governs the scope of the model, allocation of effort is the most complex component, with preceeding components preparing data for its use and subsequent components using its results. Finally, in preparation for making summary outputs an evaluation of the set of alternatives is made to rank them by cost.

Because of the large amount of data required and the complexities of the solution algorithm, implementation of the model requires access to a large scale computer.

3.1.3.1 The Population Representation

A population representation, such as described in Chapter One provides a very important starting point for the planning model. It must preserve the spatial distribution characteristics, a key feature of geographic populations, which may have a significant impact on the sampling methods used. In addition, if the data base has been well designed and maintained, available and pertinent data for the population will be readily accessible. With this data as a starting point, simulation techniques may be used to fill any data gaps.

Most geographic data base structures use either the polygon or grid approach. The first uses a series of polygons, with coordinates of vertices in some convenient system, to code information on locations of the data elements. The major advantage is that less data storage space is required, but this is offset by the requirement for more complicated analysis methods. If the area is partitioned into mutually exclusive cells, data elements may be assigned to each unit in the population. Size, number and shape of the cells determines how much detail is reflected in the data base. A 40 acre cell size would give a very different representation than a 4 acre size. A major disadvantage of the grid system is that more storage is required since each element in the population must be identified and labeled; however, this is offset by simpler analysis methods.

Simulation is often useful since numerical data relating to the parameters to be estimated are based on relatively small sets of sample data for various classes of cells within the area. The cell assignment may be known for each element in the population, but estimates of specific parameters are usually based on limited sample data. A cell-by-cell class assignment, or stratification, is an excellent tool for evaluating spatial distribution as it influences sampling methods. In this study such a digital data base is available as the result of a discriminant analysis of multispectral data, acquired by satellite. Each cell, roughly one acre in size, is assigned to one class in a land use and vegetation classification scheme. Complementing this partitioning into land uses and vegetation classes are sample data providing estimates of specific parameters of interest for each of the classes.

With appropriate assumptions about the probability distribution of the variables over the population it is possible to simulate individual cell values with average characteristics maintained and spatial distribution influences preserved. This technique seems very useful where population sampling units are likely to be clusters of cells rather than individual cells since the variability due to simulation is reduced by clustering the cells. The most obvious distribution for simulation is the multivariate normal which is often not too realistic, but is easily simulated. Other distributions derived from the multivariate normal including multivariate chi-squared and t could be simulated to account for some types of departures from normal. With a cell-by-cell population representation, summaries are easily made to obtain variance components required for the sampling alternatives to be evaluated.

3.1.3.2 Allocation of Effort

Answering the questions, How to sample?, and what cost?, are the goals of the allocation of effort component. It requires as input summary data and produces as output, for a given alternative sampling system, the sample sizes and allocation as well as implementation cost. To achieve this goal requires, more specifically, 1) mathematical representations of variance relationships for a specific sampling scheme and cost of sampling activities, 2) estimates for cost coefficients, 3) estimates of population variability, and 4) a method for finding the least cost allocation of sampling resources. Mathematical models for cost and variance relationships can be obtained for particular sampling schemes. Cost of various sampling activities can be obtained from previous experience. Nonlinear programming, a mathematical procedure for determining an optimal allocation of resources subject to certain constraints, provides a means for allocating sampling resources. The general nonlinear problem has not been completely solved; however, most sampling problems fall into a special class of problems for which solutions are possible. Success or failure of the planning process depends largely on how closely the inputs to the planning model correspond to the actual relationships in the population to be sampled.

In summary, the planning model utilizes a population representation which is derived from the data base of the information system and it incorporates as much of the basic data as is possible including simulation of certain detailed values on a unit-by-unit basis. This representation then forms the basis for generating specific inputs for an allocation model, a convex nonlinear programming problem formulated for a specific sampling system. The resulting allocation of effort minimizes the associated cost function. With these basic elements, the scope of the model is limited only by the alternative systems which can be formulated for allocation and by the details of cost or other comparisons which are possible among alternatives.

3.2 The Multivariate Survey Planning Model for Geographically Referenced Populations

From a technical viewpoint the planning model utilizes a series of input data sets and is composed of several computer algorithms which create the detailed population representation and enable evaluation of various sampling alternatives. This section outlines the components of the planning model including data inputs, the population representation, and the resources allocation technique as well as the process of developing an optimum sampling system. Figure 6 summarizes the sequence of activities which leads to development of optimum sampling system. These activities relate to 1) generating a population representation, 2) identifying alternative sampling systems, 3) allocating resources for each, and 4) selecting the best alternative. A brief summary of the computer software package based on the planning model is included as Appendix C.

3.2.1 Planning Model Data Requirements

A number of planning model inputs are required to define the problem context and sampling alternatives to be evaluated. The success of any design effort depends on the amount and quality of advance information about the population. At the outset the exact limits of the population must be defined so that the sampling frame may be identified or constructed. For geographic populations this means delineating boundaries, especially any areas to be excluded from consideration because of physical or other considerations. Examples would include such things as water bodies, ownership boundaries, urban areas, or special areas reserved from management for political or ecological reasons.

The specific parameters to be estimated and relationships among them may have a significant impact on the allocation of sampling effort. If the variables are highly correlated the sample allocations may not be substantially different from a univariate allocation. On the other hand, lack of any strong relationship will mean that one of the parameters will require more sampling than the others, and the cost of acquiring this larger sample will be the lower limit on implementation cost. In any case, however, the parameters to be estimated must be defined explicitly so that measurement procedures and population estimates used in the planning phase are consistent.

Figure 6. Generalized sequence of planning model activities.

specify population and parameter set

↓

generate population elements

↓

identify feasible alternative sampling systems

↓

summarize population representation for input to allocation process

↓

allocate resources for each alternative

↓

compare system costs and select the best system from the set of alternatives

↓

summarize specifications for the best system

Along with this definition, the desired precision associated with each parameter must be determined since these form the major constraints which the system must be designed to meet. Of course, the actual implementation of a sampling system based upon this design may or may not achieve the desired precision. This could be caused by errors in assumptions about the population characteristics, or by the possibility that an "unrepresentative" sample is obtained. Specification of desired precision must take this possibility into account since the level of precision specified affects the likelihood of a "wild" sample.

With the sampling population limits defined, and parameters specified, the next requirement is an estimate of the population variability for each of the parameters. Since this input data is dependent on the nature of the population representation it is discussed in the next section. Finally, sampling alternatives and the associated activity costs must be provided so that overall system costs can be evaluated. Absolute cost values are not necessary as long as relative relationships are maintained between alternatives.

3.2.2 The Population Representation

Several possibilities exist for population representations which would be useful in survey design. Aldred (1971) for example, used a large sample of data obtained from medium scale aerial photography in his univariate planning model. With an interest in stratification, clustering, and a set of parameters, however, the approaches that can be used are more limited. This study utilizes inputs from several sources and a random variable simulator to obtain a spatially referenced population representation which can then be stratified and clustered in many ways.

The input sources for estimates of population conditions include interpretations of digital multispectral data obtained by satellite as well as complementary sampling data for specific parts of the population.

The simulation of individual population elements assumes a multivariate normal distribution which is very convenient for simulation, but may not be indicative of actual distributions which are often nonnormal and skewed. Simulation of multivariate nonnormal distributions might be desirable in some cases but is considered a refinement which could be incorporated later if data on distributions of the variables becomes readily available.

The Landsat data base provides a cell-by-cell model of the population where each cell is roughly one acre in size. Estimates of the actual parameters by strata are provided from other sources which could include a special pilot survey or previous surveys in the same or a similar population.

Perhaps the single most important planning model assumption, and the one least subject to verification is that the population representation is a reasonable reflection of the true population relationships. Sampling based on a design developed using this planning model would provide one possibility for evaluating sampling system population model, but even this is subject to sampling variability which may still not give a true representation of actual population conditions. Thus, even though the sampling system found to be optimum relative to the planning model is used, this system may not be optimum with respect to the actual population.

3.2.2.1 Population Simulation

The population simulation is based on 1) a spatially referenced gridded data base, 2) a vegetation or land use class assignment for each data cell, and 3) parameter estimates for each of these vegetation and land use classes. A set of parameter values is simulated for each cell in the population.

The Landsat (NASA, 1971) multispectral data structure provides a convenient representation of large geographic areas (100x100 nautical miles) for manipulation using computer systems. The data base may be viewed as a Cartesian coordinate system with each pair of coordinates representing the location of a single data cell. These cells are commonly called "pixels", or "picture elements". The Landsat system acquires four items of spectral reflectance data for each pixel and provides the starting point for generating the population representation. This data base is very convenient for computer manipulation and may be related to any number of other "profiles", or levels of information such as elevation, aspect, coordinate systems, ownership, or other institutional and organizational data which are referenced to the Landsat coordinate system.

In order to relate the population limits to the Landsat data base, the MAPIT data handling system (RSRP, 1976) has been utilized. Among other things, MAPIT permits the intersection of two or more profiles of information so that a particular study area can be defined and extracted from the larger full scene base. This is accomplished by translating a representation of the study area in one coordinate system to that of the Landsat data base. A least squares procedure is commonly used to fit a mathematical function using a set of control points with coordinates observed in both systems and well distributed over the study area. These same control points can also be used to determine the empirical average cell (pixel) size for a particular study area since there are inherent variations in scale for each different Landsat scene.

With the study area physical limits defined, and the basic spectral data available from Landsat, the problem is to identify the land use or vegetation class for each pixel from some appropriately defined classification scheme. Dealing with wildland populations this scheme would include such classes as water, agricultural, urban, and various classes of vegetation related as nearly as possible to major vegetation types and conditions. Several techniques for accomplishing this classification are available and they all involve the statistical techniques of discrimination and classification (Kendall and Stuart, 1968, p. 314). Discriminant analysis would assign to each pixel a class from a predefined set of classes. Classification would group the elements into clumps or classes which would be as distinct from each other as possible. These classes however, would be obtained independently of any predefined scheme. Additional work would then be required to relate classification results, sometimes referred to as "computer classes," with the predetermined class structure.

Given the classification (meaning either discrimination or classification) results, and the estimates of the population average parameter set (by strata) individual parameter sets are simulated for each pixel in the population. Simulation is based on the assumption that the parameter set for each strata has a multivariate normal distribution with parameters, mean vector and covariance matrix, determined by class assignment. Loosely speaking, the simulation is of a stratified multivariate normal distribution with the classes representing strata and each pixel preassigned to a class. This simulation, following a technique outlined by Naylor et al, (1968), preserves the relationships among the variables which are represented by the covariance matrix. Additionally, certain classes may not be of interest for sampling, e.g. water, urban, agricultural areas, etc., and these may be effectively deleted by giving a zero vector for the parameter set. Estimates of the mean vector and covariance matrix for each of the land use or vegetation classes are usually based on some kind of sampling procedure either previously on the same area or another similar area. Since these advance estimates are based on sample data which may not even be from the same population area, it may be desirable to modify them to account for either optimistic or pessimistic attitudes about their quality or validity. No provisions exist in the planning model for changing those variability estimates, rather they must be modified prior to input.

3.2.2.2 Preparation for Sampling and Allocation of Effort

From the population simulation, an abbreviated population representation is derived which includes several elements of the sampling system and also helps to minimizes the variability due to simulation while at the same time allowing incorporation of spatial relationships. This abbreviated representation provides the population variance components necessary for allocation of sampling resources. Rectangular or square clusters are formed from the basic internal data base structure and these become primary sampling units (PSU's) and the sequence of PSU's becomes the sampling frame.

A stratification of PSU's is introduced at this point based on counts of pixels in the various classes of interest, excluding water and others not of interest. The strata for any particular PSU then becomes the class which has a plurality of pixels. This straightforward interpretation of PSU conditions was utilized in the model since it closely resembles the point-by-point classification results. If the overall classification scheme meets management needs, then it is inferred that the stratification of PSU's by this method would also meet management needs.

Selection probabilities are derived for computing the population variance components. Either "equal" or "proportional to size" probabilities for selection of PSU's may be used. The "size" variable is the number of pixels in "good" classes, classes of interest. Equal selection probabilities are assumed for subsamples within PSU's. Population variance components are then derived using these probabilities for both between and within PSU's.

3.2.3 Allocation of Effort

For each alternative sampling system the optimum allocation of effort is that which minimizes cost subject to a number of constraints. The usual LaGrangian minimization techniques of univariate survey design become more and more awkward to deal with as the number of constraints increases, and for this reason other methods need to be considered.

3.2.3.1 The Allocation Problem Formulation

To overcome computational difficulties associated with multiple constraints, mathematical programming techniques have been developed. The general programming allocation problem is formulated as follows (Hillier and Lieberman, 1967):

$$\text{minimize } f(x_1,x_2,\ldots,x_n)$$

subject to

$$\begin{array}{c} g_1(x_1,x_2,\ldots,x_n)\geq 0 \\ \vdots \\ g_m(x_1,x_2,\ldots,x_n)\geq 0 \end{array}$$

where the $(x_i \mid i=1\ldots n)$ are the decision variables, allocation values, and f, the objective function, and $(g_j \mid j = 1\ldots m)$ the constraints, are given functions of the decision variables. The well known linear programming formulation is a special case where the functions are all linear. For the general formulation the Kuhn-Tucker conditions (Hillier and Lieberman, 1967 p. 575; Fiacco and McCormick, 1968, p. 20) establish necessary but not sufficient conditions for location of a local minimum. If convexity assumptions are met (Hillier and Lieberman, 1968), that is f is convex and the g_j are concave functions with all functions continuous and differentiable, then meeting the Kuhn-Tucker conditions constitutes sufficient conditions for a local minimum, and a local minimum for the convex problem is a global minimum (Fiacco and McCormick, 1968, p. 90).

Kokan and Khan (1967) have shown that stratified random sampling variances and a typical linear cost function form a convex set. In addition, the techniques of two stage and double sampling have been included as other examples of the application of convex programming to sampling problems (Kokan, 1963). Intuitively it follows that sampling problems will generally meet these assumptions of convexity; however, any violation of these assumptions is important and checks for convexity have been incorporated in the planning model.

Wildland sampling applications of nonlinear programming have been largely limited to replacement strategies for continuous forest inventory (Hazard, 1974) and the work of Arvanitis (1971) for stratified sampling using a generalized multivariate variance function.

3.2.3.2 Example of An Allocation Problem

Stratified samplied is a logical example for demonstrating the nonlinear programming problem formulation for a sampling problem. Several different approaches have been used in the literature (Ghosh, 1958; Kokan, 1963; Kokan and Khan, 1967; Chatterjee, 1968; Huddleston, Claypool, and Hocking, 1970; Arvanitis, 1971). This formulation is included here as a basic example and because the numerical data given by Huddleston et al. is used as a test problem for one component of the planning model. The objective function to be minimized is

$$f = \sum_{h=1}^{L} (c_{h1}n_h + c_{h2}\sqrt{n_h})$$

where c_{h1} = measurement cost for a unit in stratum h

c_{h2} = travel cost between units in stratum h

L = the number of strata

The primary constraints are a set of M variance expressions, one for each parameter to be estimated:

$$V(\hat{Y}_j) = \sum_{h=1}^{L} N_h^2 S_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) \leq V_j, \quad j = 1,2,\ldots,M.$$

In terms of the general formulation this becomes

$$V_j - V(\hat{Y}_j) \geq 0, \; j - 1,2,\ldots,M.$$

Minimum and maximum sample size constraints would also be imposed to make the solution set bounded.

3.2.3.3 The Allocation Algorithm

Many computer algorithms have been developed for the nonlinear optimization problem (Hillier and Lieberman, 1967) but none are completely satisfactory (Fiacco and McCormick, 1968). The allocation algorithm used for the planning model, SUMT, is based on the theoretical work of Fiacco and McCormick (1968) and was developed by Research Analysis Corporation (Mylander, et al, 1971). It is a computer software package which utilizes the functions as well as their first and second partial derivatives to identify the $(X_i^o | i = 1 \ldots n)$ which minimizes a composite function (P) made up of the objective function and the constraints. Hillier and Lieberman (1967) provide a relatively intuitive description of the way the algorithm works. Basically the constraints form a penalty function which is forced to go to zero and then the minimum of P is the minimum of F subject to all constraints.

To implement SUMT the user must specify a number of parameters which control the operation of the program, but more important three FORTRAN subroutines must be provided to evaluate the various functions and their first and second partial derivatives. Since specifying these functions is often a source of error in implementation, numerical analysis methods are available at the user's option to approximate these partials either as a check on the user supplied subroutines or to be used in place of them.

3.2.4 Development of an Optimum Sampling System

Development of an optimum system is comprised of three interrelated activities: 1) identification of alternative feasible systems; 2) optimization within alternatives; and 3) selection of an optimum sampling system. The planning model provides a convenient tool for examining and optimizing a number of alternative systems. Qualifications on the term "optimization", outlined in chapter two, apply to the planning model since a limited number of alternative sampling strategies are available.

3.2.4.1 Identification of Alternative Feasible Systems

Alternative sampling systems are identified as particular combinations of cluster size and shape, stratification method, selection probabilities, and sampling technique. The first three factors influence the summary population representation and the last factor relates to the particular optimization formulation. Rectangular or square clusters up to a maximum of 90 x 90 pixels are permitted. As described in the earlier section on preparation for sampling, only one stratification scheme is available in the current version of the model; selection probabilities can be either equal or proportional to size. The model allows three sampling strategies; stratified, stratified two stage, and stratified two stage with double sampling.

Stratified sampling assumes direct measurement of all units within a cluster and is therefore impractical except for very small cluster sizes. Stratified two stage sampling eliminates the need to measure the entire sampling unit by introducing a subsample to estimate sampling unit characteristics. Stratified two stage with double sampling assumes that in addition to subsampling within PSU's, the set of secondary sampling units (SSU's) over all PSU's is measured to evaluate an auxiliary variable (x) which is related to the ultimate variable of interest (y). If the relationship is reasonably good and these x variables are relatively inexpensive to evaluate, at least when compared to the cost of evaluating y, then a large number of SSU's may be sampled and a smaller subset of all the SSU's sampled for direct measurement of y, following the classical double sampling procedures (Cochran, 1963).

Utilizing a regression model based on the paired (x,y) observations, an estimated characteristic y is obtained for the larger set of unpaired observations. Remote sensing data can often provide the auxiliary data in a very cost effective manner. This is true especially if the relationships between x and y are reasonably good and the cost of photo measurement is small compared to direct ground measurement. A linear relationship is assumed in the planning model. This is the method used in the Plumas National Forest inventory of 1974 (Titus et al., 1975; Colwell, 1974) where the first stage units were rectangular clusters of 225 pixels (45 x 5 pixels), second stage units were 0.4 acre circular plots located within PSU's, and the third stage units were a subset of the set of all second stage units. The strength of the relationship between the two variables is reflected in the correlation between the two, an input to the model. Currently, the same correlation is assumed for all parameters and strata but this could be expanded to include differences between parameters and/or strata.

Associated with each of the sampling methods is a cost function which determines the cost of implementation given the various sample sizes. All cost functions include terms for cost of direct measurement and also travel cost between units using a square-root-of-sample-size term based on the rationale presented by Hansen, Hurwitz, and Madow (1953, p. 272). These cost functions include only variable costs associated with each alternative since fixed costs do not affect the allocation procedure. Appendix D shows the explicit cost functions associated with each alternative technique.

3.2.4.2 Optimization of Feasible Alternatives

Optimization of an alternative system requires appropriate summary population data for the particular sampling framework chosen and a nonlinear problem formulation for the basic sampling strategy. Of the three sampling strategies included in the planning model, stratified sampling has been outlined earlier. In addition, expressions for the first and second partials of the objective and each constraint function are required. Similar formulations are required for stratified two stage and stratified two stage with double sampling. Second partials for the latter alternative were very cumbersome to evaluate algebraically so numerical differencing methods were utilized instead. A summary of the programming formulations, including first and second partials for all three strategies is given in Appendix D. Given the appropriate programming formulation and population summary data, the SUMT algorithm is executed to obtain the optimum allocation of effort and its associated cost.

3.2.4.3 Selection of an Optimum Sampling System

The planning model can evaluate one or more alternative sampling systems and will rank the alternatives by cost values. Sampling allocations are summarized and constraint values and other solution parameters are provided which may be useful in making the final decision among alternatives. As has been pointed out earlier cost comparisons alone will not provide the final word in selecting a best alternative. Fixed cost requirements for different systems may be significant and would have to be considered. Since each population representation and the other model inputs may be quite different in different applications it will generally be necessary to consider several different precision levels and cost coefficients until sensitivities of the model to various inputs are determined. Even then it may be necessary to document the relationship between cost and precision constraints and evaluate the factors which are limiting the least-cost solution so that managers will become aware of the importance of careful consideration of precision limits.

Finally, but often of most importance are the factors other than strictly design considerations which are likely to affect the choice of a particular system. Practical limitations such as the amount of work which can reasonably be accomplished in an 8-hour work day may affect cluster sizes or sample size and allocation. Related to this would be overall time constraints for survey completion. Other factors have been mentioned earlier including expertise and training requirements, and organizational structure.

3.2.4.4 Specifications for the Optimum Sampling System

The results of the planning phase need to be summarized in the form of a plan for implementation of the optimum sampling system. This very important aspect of survey work is often neglected and can seriously affect the overall success of the survey. A complete implementation plan is a necessary aid for focussing effort and ensuring that objectives are met with a minimum of compromise. It should include a statement of objectives, desired products, as well as summaries of all population data utilized in the planning phase and should address the following topics in considerable detail:

1) estimation objectives
2) summary of planning phase
3) sampling system description
4) implementation procedures and specifications
5) analysis procedures.

3.3 Model Verification and Application

For model construction, testing, and application, an area in northern California was utilized because as a result of prior research the kind of data required by the planning model was available. This section is divided into three major parts. The first describes the study area and construction of the population representation. Discussion of verification tests and application results comprise the other two parts.

3.3.1 The Study Area and Data Inputs

Previous research work in the Plumas National Forest lead to the selection of the Quincy Ranger District as the study area. As a result of a 1974 survey for the entire National Forest, conducted utilizing Landsat multispectral data, large-scale photography, and ground measurements (Titus et al, 1975), a data base of the type required for the planning model was available. A stratified three stage sampling system was used with the Landsat data aiding in making the stratification and generating selection probabilities for first stage sample units. The second and third stages utilized large scale photo and ground plots respectively for successively more detailed measurements. Federal lands in the Quincy Ranger District area, approximately 196,000 acres comprised the study area, as shown in Figure 7.

This area is a good representative of the northern Sierra Nevada mountains of California with an elevational range of 2000-7000; including both the western and eastern slopes of the mountain range.

The gridded data base structure generated by the Landsat (then called ERTS) multispectral scanner system (NASA, 1971) was used in the population model. All other data sources were transformed to the Landsat base. This base is made up of a set of picture elements (pixels) which contain a vector of specific observations made by the scanner at a particular point in time. The on-the-ground area covered by each pixel varies according to parameters associated with the satellite orbit and its relationship to the area scanned. For the Plumas NF it has been determined empirically that the area represented by each pixel is a little more than one acre. Individual cells are referenced by an (x,y) coordinate where each point is one pixel. To identify the study area explicitly and to delete all internal blocks of private ownership, an ownership base map was digitized into a local (x,y) coordinate system. This local system was then transformed to the Landsat system using mathematical functions based on least squares regression analysis of a series of control points located in both coordinate systems. The results of this intersection of ownership and multispectral data are shown in

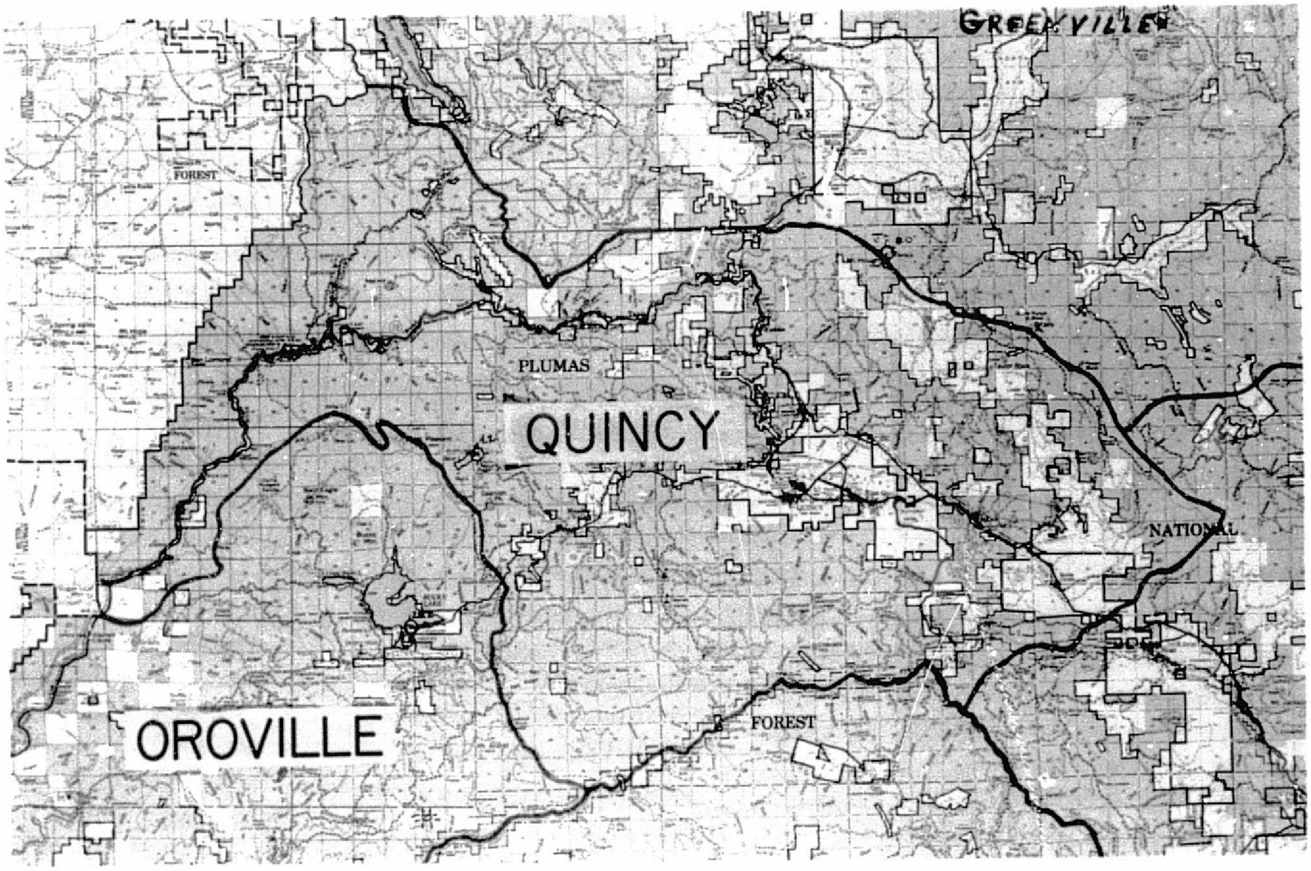


Figure 7. The Plumas National Forest in California

Figure 8, a color composite of three spectral bands for the Quincy Ranger District, including roughly 172,000 pixels and representing 196,000 acres. The blacked out areas represent non-federal lands.

With the source data and study area combined, a maximum likelihood discriminant analysis (RSRP, 1976) was performed to assign each pixel in the population to a vegetation or land use class according to a structure determined by the training areas. As required by discriminant analysis, training areas or examples of particular classes of vegetation and land use are required on which to base the rules for assignment of individuals to specific classes. For the study area two separate analyses were made, because of the Landsat data structure, each with 50 to 60 examples of various vegetation and land use categories. Further aggregation of vegetation classes into 17 broad vegetation classes was made by interpreting each of the training areas in light of the desired stratification scheme which included three classes of vegetation (mixed conifer, east side pine, and true fir) and within each vegetation class, five classes of vegetation condition, (regeneration, immature, mature, overmature, and poorly stocked) were recognized. Two other classes were specified, hardwoods and a residual of all other classes. Results of this class aggregation are shown in Figure 9.

Associated with each of the classes are a number of variables important for management. Depending on the particular ecosystem component to be managed the set could be limited to a few major variables. For example, the 1974 inventory in the study area concentrated on variables important for timber management, and therefore, many variables useful in managing such components as recreation, watershed, wildlife, or range were not included. Still, a basic approach to a number of resource values depends on the vegetation components and their condition and distribution. For this reason it is considered appropriate to use as parameters, in the verification and application of the planning model, three characteristics which describe the tree component of vegetation. The three are number of trees, basal area, and basal area growth. The last two are highly correlated with the volume and volume growth. At the same time they are basic physical measures of the vegetation component which could be related to other management areas. Number of trees combined with these gives an indication of the relative size of the trees. Selection of these parameters is not to imply in any way that the planning model is limited to timber related parameters. If data is available any parameter set could be used. The sample data from the 1974 inventory provided estimates of the mean vector and covariance matrix for each vegetation class (Appendix F).

In addition to the cell-by-cell vegetation class data and the parameter estimates by class, other data inputs are also required. However, since this application is primarily experimental, no desired precision levels were specified initially; instead they were considered as variables to be included in the analysis. Cost data for the 1974 inventory (Appendix E) were used to generate coefficients for the cost

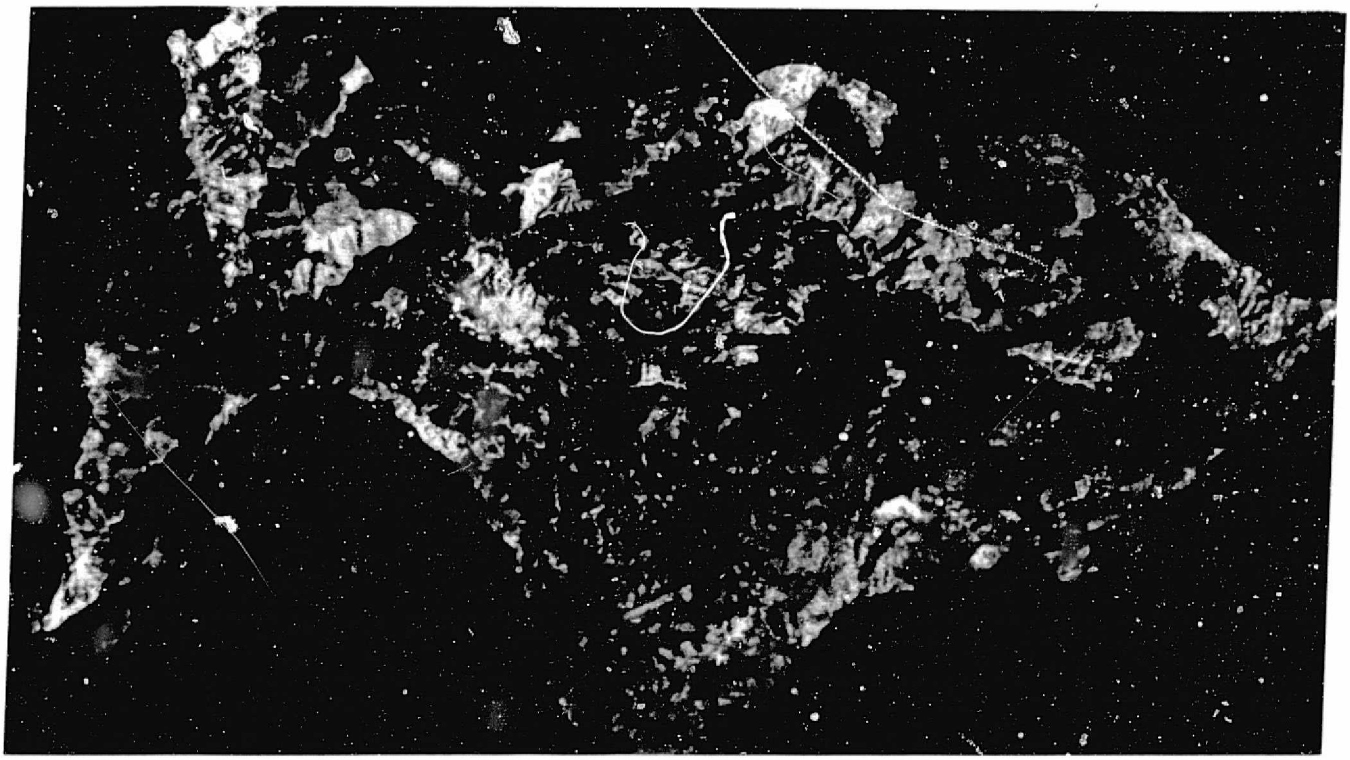
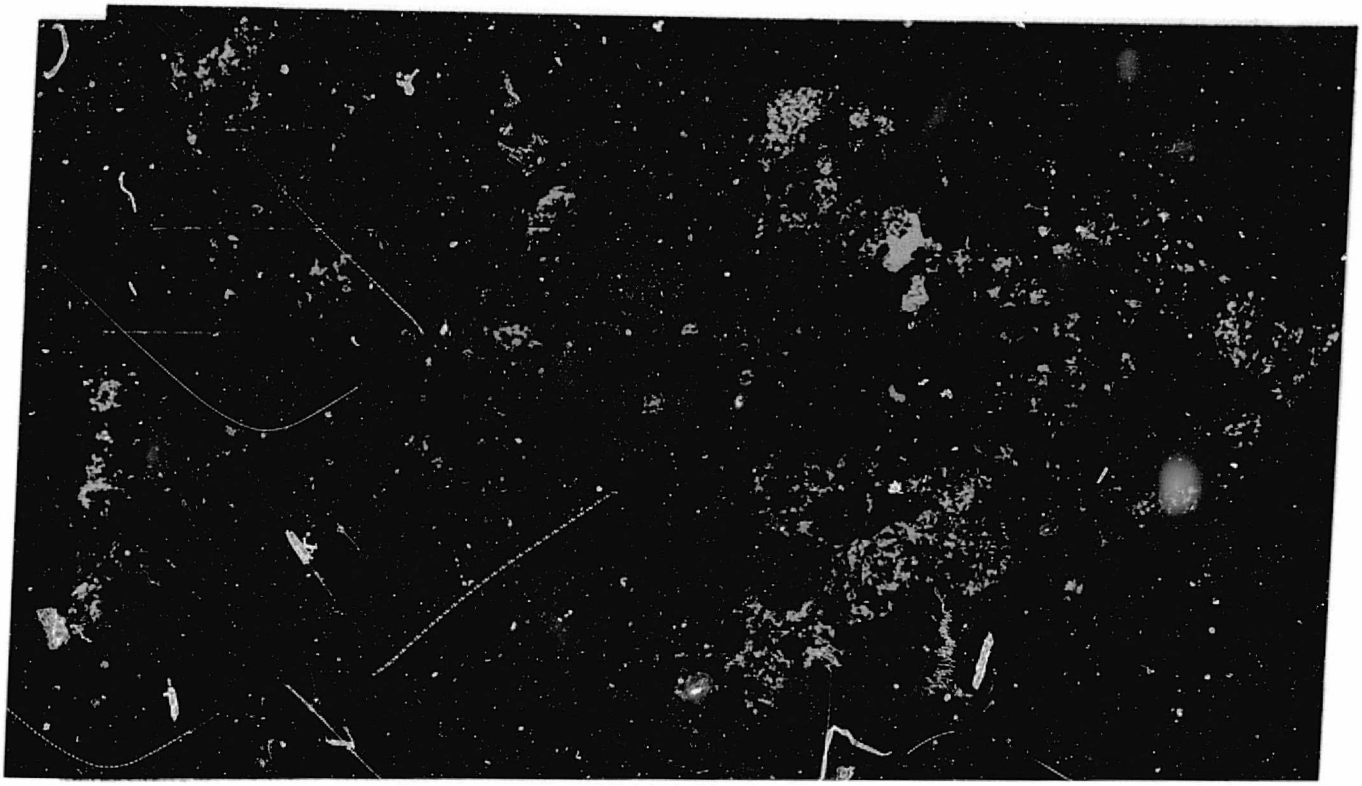
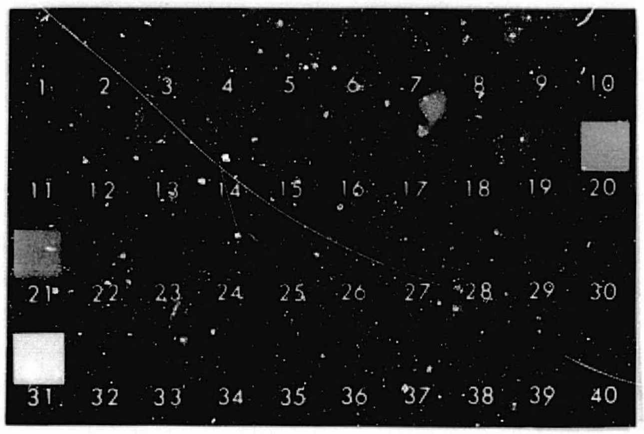


Figure 8. Landsat MSS data color composite of three bands for the Quincy Ranger District.





LEGEND:

CODE	COLOR	INTERPRETATION
1	Black	Fill
4	Orange	Hardwoods
7	Light Green	Mixed conifer - immature
10	Dark Blue	Water
11	Grey	Mixed conifer - regeneration
20	Yellow	Non-forest
27	Light Blue	True fir - mature
30	Lighter Blue	Mixed conifer - mature
36	Dark pink	True fir - poorly stocked
40	Pink	Mixed conifer - poorly stocked

Figure 9. Discriminant analysis results for the Quincy Ranger District.

functions used with the different sampling strategies. As with the precision levels, cost was treated as a variable, and a number of values were used. Derivation of these coefficients as well as precision factors are discussed in the section on application results.

3.3.2 Verification Tests

With a study area defined and population data inputs available, it is desirable to verify the proper functioning of the various planning model components. In particular, for the population model component, verification of the simulation procedure and its influence on allocation of effort is necessary. For the allocation of effort component it is important to verify the functioning of the experimental SUMT algorithm and evaluate assumptions made with respect to the form of estimators used in alternative sampling strategies.

In generating the population model for the Quincy Ranger District, over 172,000 individual parameter sets were simulated. To verify the accuracy of the simulation, averages of simulated parameter values in each vegetation class were compared with the input parameter values (summarized in Appendix F). Table 5 summarizes the average values of the simulated parameters by class number; class number from Table 5 is equivalent to the TCC code listed in Appendix F. Differences between the average simulated value and the input parameter values are on the order of less than one per cent.

To see how much the particular sequence of random numbers used in the population simulation affected the overall results, two separate sequences were used to generate two population models. These models were then used in a series of allocations for various alternative sampling systems. The results were evaluated in terms of differences in overall system cost. For all alternatives the differences in objective, or cost, function values was less than one per cent. Differences in the related allocation of effort were also very small. (Table H-1 and H-4, Appendix H.)

The experimental SUMT algorithm, a major component of the planning model, was tested using a numerical agricultural example cited in Huddleston et al. (1970) for a stratified random sampling application. In the example there are 15 strata and 7 characteristics to be estimated. The objective is to minimize cost subject to variance constraints for each of the parameters. This same problem has been solved using a different algorithm, FCDPAK, developed by M. J. Best at the University of Waterloo, Canada. FCDPAK has options for using several approaches to solving the allocation problem. The feasible conjugate direction technique (Hillier and Lieberman, 1967) was used for the agricultural test problem. The results of this test are summarized in Table 6, giving the published, FCDPAK, and SUMT allocations. This shows that FCDPAK and SUMT give

Number of Points in Each Result Class, and Means for Result Class Parameters

Class No.	Name	Total Pts.	Means		
0	Null	0			
1	One	0	0.000	0.000	0.000
2	Two	1044	81.610	55.745	4.958
3	Three	10896	158.865	117.376	12.723
4		0	0.000	0.000	0.000
5	Five	4257	36.472	27.874	2.515
6	Six	9865	84.573	81.591	5.837
7	Seven	51001	108.563	75.895	7.022
8	Eight	19852	132.544	65.485	7.932
9		0	0.000	0.000	0.000
10	Ten	12104	40.962	22.591	2.425
11		0	0.000	0.000	0.000
12		0	0.000	0.000	0.000
13		0	0.000	0.000	0.000
14		0	0.000	0.000	0.000
15		0	0.000	0.000	0.000
16	Sixteen	30919	28.786	17.864	1.798
17	Seventeen	32512	0.000	0.000	0.000

Number of points in illegal old classes = 2
Number of points in illegal new classes = 0

Total number of points processed from NOVAMAP = 420000

Total number of points in result classes = 172450

Table 5 . Population Simulation results summary

Table 6. Allocation Algorithm Test Problem Results

		n_h	
h	PUBLISHED	FCDPAK	SUMT
1	232	232	232
2	103	96	96
3	87	75	75
4	117	104	104
5	150	157	157
6	207	200	200
7	106	104	104
8	92	83	88
9	87	87	87
10	369	443	443
11	98	110	110
12	33	32	32
13	83	77	77
14	27	27	27
15	49	49	49
TOTAL	1840	1880	1880
COST	39300	40900	40912

essentially identical allocation but these are both different from the published optimum. Further analysis showed that the variance constraints for the third and sixth parameters are violated using the published optimum sample sizes. The variance values are 523,692 and 26.447 respectively compared with upper limits of 507.49 and 25.98 respectively. These results were obtained by using the published optimum as initial points for both FCDPAK and SUMT with the result that the initial point was infeasible.

Associated with any sampling system is the physical procedure for selecting the sample, making measurements, etc. Estimators are usually based on assumptions about the particular way in which sample units are selected. If the actual technique varies from the assumed technique there may be changes in the performance of estimators derived from the sample. Sampling from finite populations is one case where there is opportunity for deviations from the usual assumptions, since formulations of estimators are frequently based on sampling with replacement or from large populations. The formulations for variable probability sampling, one alternative selection procedure used in the planning model, are particularly important since in actual practice sample unit selections are often made without replacement of units. This means actual selection probabilities change for each selection. To investigate this effect on small and medium size finite populations a sampling simulator was constructed. The results of applying the simulator to two different populations, one with 12 units and the other with 115 units are summarized in Appendix G. For both populations the average value of $\hat{Y}$, the estimator of the population total based on repeated sampling of 15 units for the large population and 4 units for the small population, was not significantly different from the true population total.

The average value of variance estimators ($\hat{V}(\hat{Y})$) however, showed more valiability. In particular for the small 12-unit population the average value of $\hat{V}(\hat{Y})$ was about 50 per cent larger than the actual variance of the estimates (Appendix G). In addition, estimators based on sampling without replacement (including the general PPS estimator but with selection without replacement) consistently outperformed (lower variance) estimators based on with replacement sampling.

All this confirms what was expected based upon theoretical considerations. It also documents the usefulness of the usual PPS estimator (which assumes with replacement sampling) when sample selection is actually without replacement, with the caution that the variance estimates will be biased.

3.3.3 Application Results

With the population model specified and the planning model available it is possible to evaluate the sampling effort required to meet desired precision levels for each of a number of alternative sampling systems. It is obvious that time and money constraints on the planning phase, as well as the limitations of the planning model, will limit the total number of alternatives which might be evaluated. Nonetheless, by careful selection of few alternatives to be examined, it should be possible to gain valuable insights about the best, at least in a suboptimal sense, way to allocate sampling resources in those situations where we must act on the basis of incomplete and uncertain information. In addition, the ability to vary a number of factors which specify the nature and performance of a particular sampling system enables the analyst to interact with the manager to test alternative sampling system formulations.

3.3.3.1 Factors Evaluated

Analysis of sampling systems for the Quincy Ranger District is based on alternatives obtained as combinations of the following factors: 1) level of confidence and precision, 2) cost, 3) sampling strategy, 4) selection probability, 5) cluster size, and 6) correlation between measurement variables. Table 7 provides a detailed outline of factors evaluated in this case study. However, not all combinations of all factors were evaluated. Two additional factors, the population model and relationships among the parameter set, obviously would affect conclusions about which design might be best. Taking these as given limits the generality of any conclusions which might be made.

Cost and precision might be factors which would be specified in advance, however, uncertainty as to what the real sampling costs are or what precision level is desired, often makes it necessary to evaluate the impact on design varying both cost and precision levels.

The cost alternatives listed in Table 7 require some explanation of their derivation. All cost coefficients are based on the actual costs of the Plumas 1974 inventory, and they differ primarily in the amount of travel and per diem costs included. For stratified two-stage (STRAT2S) sampling the total costs per unit were based on the number of PSU's evaluated and for stratified two-stage with double (STR2SD) sampling the costs were based on the number of ground plots visited. Cost levels one and two represent the average cost per unit for each term independent of the form of the cost function. In level two, travel and per diem costs are deleted. In level three travel cost and per diem are combined with measurement costs. Cost level 4 includes the affect of the form of the function used in deriving cost coefficients. For example, in STRAT2S the coefficient is derived as

$$TC = C_3 \sqrt{n^*}$$

$$3149 = C_3 \sqrt{30}$$

$$575 = C_3$$

Table 7. Factors and descriptions of factor levels utilized in application of the Planning model to the Quincy Ranger District.

Factor	Level	Description
1. Selection Probability	1	Equal probability
	2	Probability proportional to size
2. Sampling Strategy	1	STRAT2S - stratified two stage sampling
	2	STR2SD - stratified two stage with double sampling
3 Cost (STRAT2S)	1	$C_1 = 20$, $C_2 = 23$, $C_3 = 200$
	2	$C_1 = 20$, $C_2 = 23$, $C_3 = 125$
	3	$C_1 = 20$, $C_2 = 147$, $C_3 = 0$
	4	$C_1 = 20$, $C_2 = 23$, $C_3 = 575$
Cost(STR2SD)	1	$C_1 = 50$, $C_2 = 7$, $C_3 = 15$, $C_4 = 11$, $C_5 = 175$
	2	$C_1 = 50$, $C_2 = 7$, $C_3 = 15$, $C_4 = 11$, $C_5 = 100$
	3	$C_1 = 50$, $C_2 = 7$, $C_3 = 15$, $C_4 = 111$, $C_5 = 0$
	4	$C_1 = 50$, $C_2 = 7$, $C_3 = 15$, $C_4 = 11$, $C_5 = 778$
4. Correlation (STR2SD)	1	$\ell = 0.8$
	2	$\ell = 0.9$
5 Random Number Sequence	1	Random number seed 1
	2	Random number seed 2
6. Probability Level for Confidence Statements	1	$t = 1.00$, $1-\alpha = .68$
	2	$t = 1.96$, $1-\alpha = .95$
7. Allowable Error	1	$\pm$ 20% for each parameter
	2	$\pm$ 10% for each parameter
8. PSU Size	1	60 x 6 pixels
	2	40 x 4 pixels
	3	10 x 10 pixels

Where

TC = 3149 = total cost of n* PSU's in Plumas 74, excluding Travel and per diem cost

n* = 30 = number of PSU's sampled in Plumas 74.

The overall set of numerical results for the Quincy Ranger District are summarized and tabulated in appendix H, but will be discussed and evaluated in the following paragraphs.

3.3.3.2 Analysis, Phase I

The initial analysis effort considered cluster size to be fixed (60 x 6 pixels) and generated allocations based on combinations of selection probabilities (equal and PPS), cost (two versions), correlations between photo and ground variables (0.8 and 0.9), sampling method (STRAT2S, STR2SD), level of confidence and precision (combinations of t = 1.00 and 1.96, and allowable standard error = 0.2 and 0.1 for each of the three parameters).

The results of evaluating these 32 alternatives helped to further substantiate the overall functioning of the planning model since a number of trends expected based on sampling theory were confirmed. Increasing the precision requirements resulted in substantially increased cost values regardless of the levels of other factors. However, the magnitude of the increase did depend on the other factors. For example, increasing precision from the lowest level (t = 1.00, AE = 0.2) to the highest (t = 1.96, AE = 0.1) required slightly more than three times the cost for PPS selection but more than five times the cost with equal selection probabilities. As expected increased correlation between variables (0.8 to 0.9) used for regression relationships lead to a decrease in sampling costs though by only about 5 per cent.

The difference between the two cost alternatives was a lower travel cost term in one case and the results as expected reduced overall cost. However, changing the cost alternative did not change the relative ranking between alternatives. STR2SD consistently required lower costs for all levels of the other factors.

3.3.3.3 Analysis Phase II

Based on these results the correlation coefficient was set at 0.8 and was not varied in subsequent analysis. The middle two precision levels and one cost combination were dropped. Two additional cost alternatives and two additional clusters sizes (40 x 4, and 10 x 10 pixels) were added to the analysis.

The second phase of the analysis yielded additional insights into the relationships among the alternatives. Sampling with selection probabilities proportional to size consistently required a lower cost than equal probabilities and the relative differences between the two increased with larger PSU size. On the other hand, given equal selection probabilities, smaller PSU sizes yielded lower cost values. In changing from low to high precision limits the marginal cost increase is generally less with the systems utilizing more auxiliary information, that is with PPS selection and double sampling. Cost coefficients can change the rankings among alternatives, especially for low precision limits. Combining travel costs with measurement costs for both sampling strategies leads to the conclusion that STR2SD is better in all cases examined here. However, treating travel costs as a separate term in the cost function leads to the conclusion that STRAT2S is better for low precision requirements.

The third cost alternative attributes a large cost to travel and seems to give a more realistic cost value as compared with actual costs required for the 1974 survey. This alternative confirms other interpretations since STRAT2S shows substantially less cost than STR2SD for low precision and PPS sampling and slightly less cost for high precision. With equal selection probabilities STR2SD in slightly better for high precision limits.

3.3.3.4 Sensitivity Analysis

Four quantitative factors, correlation, probability level, allowable error, and PSU size, can be evaluated in terms of percentage changes in the objective functions as the factor level is changed. Tables 8 through 11 show results of such a comparision over the different levels of the remaining factors.

Percentage reduction in cost when correlation was increased from 0.8 to 0.9, Table 8, was variable depending on other factors, particularly selection probability and allowable error. For the PPS selection method the reduction was about 6 per cent for a 0.1 allowable error and 3 per cent for 0.2 allowable error. For equal selection probabilities the reduction was about 3 per cent regardless of allowable error.

PSU size, Table 9, appears to be affected primarily by the selection probability method. With PPS selection, there is generally a decrease in cost as PSU size increases. On the other hand with equal selection probability there is an increase in cost as PSU size increases. The percentage changes are quite variable with no obvious patterns with respect to the other factors.

STRATEGY	PROB	COST	PLEVEL	AE	F.9/F.8
STR2SD	PPS	1	1	.2	-.03
				.1	-.06
			1.96	.2	-.06
				.1	-.07
		2	1	.2	-.04
				.1	-.06
			1.96	.2	-.06
				.1	-.06
	EQL	1	1	.2	-.03
				.1	-.04
			1.96	.2	-.04
				.1	-.03
		2	1	.2	-.04
				.1	-.04
			1.96	.2	-.04
				.1	-.03

Table 8. Decimal percentage change in cost function, F, when correlation changes from 0.8 to 0.9 for different levels of other factors.

STRATEGY	PROB	COST	PRECISION	F60/F40	F40/F10
STR2SD	PPS	2	1	.00	.01
			4	-.03	-.04
		3	1	.02	.02
			4	-.02	-.03
		4	1	.01	.00
			4	-.01	-.02
	EQL	2	1	.07	.03
			4	.13	.04
		3	1	.05	.02
			4	.11	.03
		4	1	.01	.01
			4	.08	.02
STRAT2S	PPS	2	1	-.05	-.04
			4	-.03	-.02
		3	1	.00	.01
			4	.00	.01
		4	1	-.07	-.07
			4	-.08	-.04
	EQL	2	1	-.09	.02
			4	.03	.02
		3	1	.09	.05
			4	.01	.04
		4	1	.08	.00
			4	.03	.00

Table 9. Decimal percentage change in cost function, F, when PSU size changes from 10 x 10 to 40 x 4 and from 40 x 4 to 60 x 6 for different levels of other factors.

STRATEGY	PROB	COST	CORREL.	P.LEVEL	F1/F2
STR2SD	PPS	1	.8	1.00	.55
				1.96	1.13
			.9	1.00	.50
				1.96	1.12
		2	.8	1.00	.75
				1.96	1.38
			.9	1.00	.70
				1.96	1.38
	EQL	1	.8	1.00	.96
				1.96	1.73
			.9	1.00	.94
				1.96	1.76
		2	.8	1.00	1.23
				1.96	1.99
			.9	1.00	1.22
				1.96	2.02
STRAT2S	PPS	1		1.00	2.10
				1.96	2.36
		2		1.00	2.29
				1.96	2.51
	EQL	1		1.00	1.92
				1.96	1.98
		2		1.00	2.11
				1.96	2.13

Table 10. Decimal percentage change in cost function, F, when allowable error changes from .2 to .1 for factors.

STRATEGY	PROB	COST	CORREL.	AE	F1.[illegible]/F1.00
STR2SD	PPS	1	.8	.2	.52
				.1	1.10
			.9	.2	.48
				.1	1.08
		2	.8	.2	.71
				.1	1.33
			.9	.2	.67
				.1	1.33
	EQL	1	.8	.2	.91
				.1	1.66
			.9	.2	.89
				.1	1.69
		2	.8	.2	1.17
				.1	1.91
			.9	.2	1.16
				.1	1.94
STRAT2S	PPS	1	--	.2	2.00
				.1	2.25
		2	--	.2	2.17
				.1	2.39
	EQL	1	--	.2	1.83
				.1	1.89
		2	--	.2	2.00
				.1	2.03

Table 11. Decimal percentage change in cost function, F, when probability level, for confidence statements, changes from 68% to 95% for different levels of other factors.

Since the ratio of level two to level one of both probability level and allowable error is roughly two, the percentage change in F is about the same for the two factors, Tables 10 and 11. The response appears to be independent of correlation, but it is affected by the level of other factors including cost, selection probability, and strategy. In particular the STRAT2S strategy requires a 200 per cent increase in cost. Less than 200 per cent is observed with STR2SD, and even less using PPS selection probabilities.

3.3.3.5 Optimal Model Output

The only conclusion which holds over all levels of factors examined here is that PPS sampling requires a lower cost than equal probability sampling, and that a large PSU is better with PPS sampling while a small PSU size is better with equal probability sampling. These conclusions must be qualified by the limits of this particular population model and the range of PSU sizes evaluated. However, the implication seems to be that regardless of relationships among the variables themselves, the larger PSU sizes strengthen the relationship between the size variable used for the selection probability and the magnitude of parameter totals by PSU. Conclusions about which alternative system is best must be qualified by the particular factors and levels which are utilized.

Taking the set of inputs and model parameters which seem to be as realistic as possible within the limits of this analysis one may select an optimum system. With this system it is possible to complete the planning process by preparing a detailed implementation plan. Table 12 presents the summary model output and appendix J is an abbreviated implementation plan based on the assumptions and specifications of the planning model. In particular the assumed specifications for system performance are 1) probability level for confidence statements at 95 per cent, 2) allowable error of $\pm$10 per cent for each parameter, 3) correlation at 0.8 for the double sampling alternative and 4) cost factor level 4 as shown in Table 7. With these assumptions in addition to those of the population representation, the optimum system is identified from the possible alternatives (Appendix H, tables H-1 to H-6). System cost is $23,190 (assuming cost coefficients are in dollar units), and its basic elements are 1) PSU size of 60 x 6 pixels, 2) selection probabilities proportional to size, and 3) stratified two stage with double sampling. Because of the small sizes of strata 2, 5 and 6 they would be combined or included in the larger strata for implementation. Additional details on the sampling system as it would be implemented are given in Appendix J.

SUBROUTINE SUMSIZE -- STRATIFIED TWO STAGE WITH DOUBLE SAMPLING

PPS selection

IX,IY,T,RHO,AERR(3)

50	6	1.9600	.8000	.1000	.1000	.1000
$(C_{hi},i=1,5)$ 50.000	7.000	15.000	11.000	778.000		

ORIGINAL PAGE IS OF POOR QUALITY

I,DIST,VAR,AVAR,PSE,CIHWPC

1	.123597E+08	.362577E+12	.602144E+06	.487183E+01	.954879E+01
2	.825718E+07	.177473E+12	.421276E+06	.510193E+01	.999979E+01
3	.812636E+06	.171639E+10	.414293E+05	.509814E+01	.999235E+01

h	N_h	n_{h^*}	$n_{h'}$	n_h	$n^*_h n'_h$	$n_{h'},n_h$
1	23	4.797	5.377	.794	25.793	3.809
2	2	.352	8.529	8.529	3.002	3.002
3	20	2.732	4.847	1.307	170.969	3.571
4	347	35.706	10.913	.292	389.660	10.426
5	10	.363	8.280	8.280	3.006	3.006
6	7	.310	9.676	9.674	3.000	2.999
7	92	10.949	7.526	.451	82.402	4.938
8	229	22.898	9.513	.324	217.829	7.419
		78.106	737.920	39.174	737.920	39.174

POINT= 123 DCTT= .9262419E-02 RHO= .3727067E+00 MAGNITUDE= .3918439E+05 PHASE= 2

F= .2319000E+05 P= .2317101E+05 G= .2315898E+05 RSIGMA= -.2960078E+02 H= 0.

THE CURRENT VALUE OF X IS

.4796657E+01	.3521079E+00	.2731758E+01	.3570556E+02	.3627778E+00	.3104086E+00
.1094893E+02	.2289797E+02	.5376668E+01	.8529326E+01	.4847370E+01	.1091309E+02
.8279782E+01	.9675767E+01	.7525936E+01	.9512572E+01	.7936034E+00	.8528914E+01
.1306599E+01	.2920042E+00	.8279700E+01	.9674404E+01	.4508485E+00	.3242896E+00

THE CONSTRAINT VALUES

NOT INCLUDING THE NON-NEGATIVITIES

.1649987E+00	.2340038E+01	.3160646E+00	.2512500E-01	.2656508E+01	.3078703E+01
.8046359E-01	.3930518E-01	.1832110E+00	.1144648E+00	.2035197E+00	.8885528E-01
.1179984E+00	.1035732E+00	.1300961E+00	.1023463E+00	.1074087E+01	.5665867E-05
.5589896E+00	.7332975E+01	.1186340E-05	.1456678E-04	.2085166E+01	.2978549E+01
.3507497E+11	.7479985E+07	.2627914E+07	.9979152E+02	.9715996E+02	.9963394E+02
.9997199E+02	.9724349E+02	.9677844E+02	.9990867E+02	.9995633E+02	.9981401E+02
.9988275E+02	.9979370E+02	.9990837E+02	.9987922E+02	.9999665E+02	.9986713E+02
.9999488E+02	.9873992E+02	.9998275E+02	.9923471E+02	.9657539E+02	.9987927E+02
[illegible]	.9778196E+02	.9691633E+02	.2119098E+00	.1031710E-02	.1595765E+00
.7122626E+00	.1228936E-02	.1004994E-02	.3922582E+00	.5959896E+00	

Additional discussion of the detailed model solution output is useful since with it one can interpret the effects of the various constraints on the solution as well as other factors including the stratification scheme. A representative summary of both input and model output are included in Appendix I. This example has 8 different strata and associated with each is a sample size for the number of PSU's (n*), the number of SSU's (n'), and the number of direct measurement units (n). The first page, labeled SUBROUTINE READIN, provides inputs of population conditions and other problem specifications required by SUMT. The next portion of output, labeled SUMT VERSION 4, shows the initial parameters and starting values for SUMT and then skips to POINT 194 which represents the final solution point. The estimate (the current value of x) with the cost function (F) and its lower bound (G) as well as second and first order estimates of the solution are given. Then the LaGrange multipliers, the partials of the objective function (F) with respect to both the X_i and the constraints are provided. Mylander et al (1971) provide detailed descriptions of the SUMT outputs. Finally SUBROUTINE SUMRIZE shows the sample size allocation by strata and summarizes other model inputs.

A measure of the relative tolerance for the solution estimate may be obtained by evaluating (F-G)/F. Also the LaGrange multipliers for the constraint functions may be utilized to evaluate the relative sensitivity of the solution to each of the different constraints.

In this example constraints 53 and 57 which require that $(n^*)\ (n) \geq 3$ for strata two and six, and constraints 18 and 22, which require that $n' \geq n$ have the largest values. The final sample sizes confirm that these constraints are limiting the solution size for strata 2 (class 5) since $(n^*)\ (n) = 3.01$ and $(n') = (n) = 2.762$. A similar situation holds for strata 6 (class 10).

The implication from all this is that the stratification procedure could be improved since there are several very large strata and several which are very small. A better scheme would be one in which the sizes of strata are more equal. Possible remedies include different stratification methods in the planning model or a review of the population classification results on which the population model is based.

The possibility of several stratification alternatives always raises the question: should stratification be made to reduce variance or to provide for estimates of particular subpopulations? The answer is usually that they are both important. A compromise procedure would be to use stratification to reduce variance as much as possible, and then use "domain", or subpopulation, estimation procedures (Durbin, 1956; Cochran, 1963; Singh and Tessier, 1976) to obtain estimates for those subpopulations which are important for management purposes. While error terms associated with domain estimators are generally larger since there is no control on sample size, similar problems occur with stratified sampling since sample sizes in each strata are generally small. In addition the number of estimates of subpopulations often includes many more categories than could be reasonably incorporated in a single stratification scheme. For instance, in inventories for timber management, estimates are frequently desired by site, by administrative units, by stand conditions, by species groups, by size classes, by soil type and, perhaps others as well.

3.4 Planning for a Level II Inventory

According to the statement of work, level II areas are 10-200 acre parcels which are of particular interest, presumably for management treatments. The precision desires are specified as plus or minus ten per cent at the 95 per cent level of probability for timber volume information. Populations of this small size are significantly different than level I areas (50,000 acres and larger) since not only are we dealing with a finite population but a small finite population. In addition these precision limits are quite high and the immediate implication is that a large amount of sampling will be required. Assuming a level I inventory has occurred and that Landsat classification results are available, the planning model would be useful as an initial indicator of the intensity of sampling required. The problem with the planning model is that simulations for populations this small would have to be replicated to overcome the variabilities associated with such a small population size. In addition different kinds of techniques,especially ratio and regression sampling,would have to be incorporated with very small cluster sizes and the option of no clustering would also need to be evaluated. The types of auxiliary data available and their relationships to volume data would need to be examined. If no ground sampling effort was allowed there would be potential bias in the procedures which could not be evaluated. Using ground sampling procedures and large scale photography for auxiliary data with a ratio or regression technique would still require fairly large sample sizes for this high precision level. Stratification of these small populations might be possible using medium or small scale imagery, but Landsat classification would likely be prohibited because of problems in locating precisely small sampling units. While this possible application of the planning model would be a promising avenue for research activities it may be more profitable to concentrate on gaining experience and expanding the planning model to somewhat larger populations.

4.0 CONCLUSION

4.1 Sam Houston National Forest Inventory

Results of this inventory application have shown that while Landsat data were not useful because of extreme homogeneity of topographic and vegetative conditions, remote sensing data from large scale photographs did provide timely and cost-effective estimates of the kinds of parameters currently required by the Forest Survey. The estimated relative errors for volume and growth were near the desired Forest Service levels. Mortality and removals estimates had larger error, but the Forest Survey does not state a desired precision level for these estimates. By considering some of the opportunities for improvement which are outlined in Section 2.5 and with appropriate training for field crews and photo interpreters, it should be possible to expand and improve the system beyond the demonstration phase to allow estimation of all the characteristics of interest to the Forest Service.

4.2 Sample Survey Planning Model

The planning model developed here represents a formalization of the interrelationships which must be considered in planning sample surveys for multivariate applications. As with any model, assumptions are required in order to simplify the "real world" complexities so that a workable model can be formulated. But even with assumptions, some of which require additional research to evaluate their importance, the planning model provides an important mechanism for evaluating the various factors, data and components of a sample survey system. Without this formal planning mechanism, survey design is largely relegated to experience, hunches and guesses. This is especially true when large multi-purpose surveys are contemplated.

A significant requirement for the successful implementation of a planning model of this kind is the availability, in a useable form, of considerable prior information about the population to be surveyed. A "data base management system" of the type referred to in the introduction, is probably the only way to insure that this kind of information is accumulated and maintained over time. If this level of information is available and its quality is improved as time passes (and more data are acquired), the planning effort becomes more effective and the efficiency of the sampling effort should increase. The Landsat population model used in the Quincy Ranger District study meets this requirement and represents a unique application of a remote sensing data source to planning for sample surveys.

4.2.1 Quincy Ranger District Application

In terms of the case study application to the Quincy Ranger District specific conclusions are possible, keeping in mind that extrapolating beyond the limits of this example is not without risk. The only trend which was consistent over all levels of other factors was that PPS sampling with larger

PSU sizes produced consistently lower cost than equal probability sampling. The cost function and precision limits used were the most important factors in determining which sampling alternative was least costly.

More effort needs to be devoted to defining the true sampling costs, both the level of costs and their functional relationships to sampling activities. Following these factors in importance were sampling strategy and selection probability, but they also interact with cost and precision.

The optimum solution was relatively insensitive to the random number sequence used for simulation, PSU size, and correlation, at least in the range of values considered in this application.

Effects on the optimum design of changes in the population representation and relationships among the parameters were not examined. Either or both could have potentially significant impacts on actual results and also affect the sensitivities of the various other factors influencing the ranking of alternatives.

In order to select an optimum system for the Quincy Ranger District certain model parameters must be specified in advance. Since this application was basically experimental the following parameters were assumed to be the most realistic:

1) probability level for confidence statements at 95 per cent

2) allowable error at 10 per cent for each parameter

3) correlation at 0.8 for the double sampling alternative

4) cost factors as shown in level 4 of table 7

With these parameters and the population representation, the least-cost system, among the alternatives considered, was stratified two stage with double sampling with PSU size at 60 x 6 pixels and selection probabilities proportional to size.

4.2.2 Implications for Management and Survey Planning

Survey design requires interaction between the resource managers and survey planners to arrive at a cost effective sampling system which meets the managers objectives. The planning model provides a formal mechanism for linking the two sides with each of the key links between the two well defined. With the decision model specified on one side and the planning model on the other, the chances of successfully designing a sample survey to meet the decision-maker's needs are substantially greater, especially as the decision process and information requirements become more complex.

In management applications where a cell-by-cell population model, such as the Landsat model developed in this study, is available, the planning model permits evaluation of the effects of a number of sampling factors as they interact in a particular sampling context. Variations in both size and shape of clusters used as PSU's can be examined. Both equal and PPS sample selection probabilities may be used. Any level of precision, both allowable errors and probability level, may be specified, and any of three sampling strategies may be employed. In addition, the interactions among all these factors may be evaluated. Finally, cost coefficients and their functional relationships in determining system cost are quite flexible since coefficients for each term are specified as model inputs. While the planning model will not solve all design problems, its modular nature and use of a detailed population model provides the opportunity for incorporating additional capabilities, assumptions, and factors which are, or may become, important in the planning process.

Finally, it should be obvious that this approach to planning for survey design is not limited to wildland applications or more generally even to geographic populations. The general approach to planning and design extends to all areas of sampling work. Differences in population structure and complexity would lead to different approaches to various components of the planning model. However, geographic populations offer many applications in different subject areas including agriculture, range management, forestry, regional planning, and even urban planning. The key to application in any of these subject areas is the availability of a sufficiently detailed data base.

4.2.3 Model Limitations

Several model limitations exist, but most could be reduced or eliminated with further development work. The current planning model requires a population model based on a land use class assignment for each population element and supplementary parameter data for each class. This requirement is based on a need to be able to generate population variability components under different sampling designs. In some cases it may be possible to obtain estimates of variance components from prior surveys with subjective modifications to make them applicable to particular populations. Then, specific sampling systems could be evaluated to determine the best allocation of effort. However, evaluating and comparing several alternatives would be difficult unless variances could be expressed as functions of the factors which would differ among the alternative schemes.

The planning model only allows three basic sampling strategies, one of which is impractical for most large geographic populations. In future studies with this and other population models, additional alternative sampling techniques should be investigated and incorporated.

The cost of evaluating an alternative with the planning model is subject to rapidly increasing cost as the complexity of the sampling strategy increases. Finally the comparison among alternatives is based strictly on the variable costs associated with this population model. Incorporation of other factors in the decision process is left to the decision-makers.

4.2.4 Recommendations

Opportunities for improving the utility of the planning model can be viewed as 1) refinements, 2) additional capabilities, and 3) further application and testing. Refinements would include 1) investigating the possibility that more efficient or improved nonlinear allocation algorithms are available; 2) individualizing the formulations of inputs for each different parameter (for example, allowing variable correlation for each parameter and stratum) and 3) considering possibilities for increasing the number of variables in the parameter set which can be incorporated into the model.

Additional capabilities might be desirable such as 1) the ability to stratify by different methods, 2) additional sampling strategies, and 3) the ability to work with different population representations (such as aircraft multispectral data).

Further application and testing is desirable to 1) make the software package more user oriented, 2) apply the model to other types of problems, and 3) evaluate the sensitivity of the model to varying population model and parameter set relationships.

4.2.5 The Future

As indicated in the introduction, the trend in management seems to be toward much more detailed evaluation of management alternatives and the consequencesof management actions. Associated with this analysis are increasingly complex decision models. More detailed information is a key element in the success of these activities, and in the case of wildlands most information is obtained from sample surveys. With more and more information on hand in the information system data base, sample surveys need to be and can be more efficient, in the sense that maximum advantage can be taken of available information. Further, improvements are possible in both the planning phase and the implementation phase of sample surveys. The multivariate survey planning model described here represents a significant first step towards taking advantage of available information for the design and implementation of wildland sample surveys.

REFERENCES

Arvanitis, L.G. & B. Alfonza, 1971. Use of the generalized variance and the gradient projection method in multivariate stratified sampling. Biometrics 1971 27:119-127.

Arvanitis, L.G. & W.G. O'Regan, 1967. Computer simulation and economic efficiency in forest sampling. Hilgardia 38(2):133-164.

Baker, R.D., 1975. Personal communication. Department of Forest Science, School of Natural Biosciences, College of Agriculture, Texas A&M University, College Station, Texas 77843.

Bonner, G.M., 1972. Forest sampling and inventories: a bibliography. Canada FMI internal report FMR-24, April 1972. Multistage refs: 98, 140, 141.

Chakravarti, I.M., 1955. On the problem of planning a multistage survey for multiple correlated characters. Sankhya 14:211-216.

Chatterjee, S., 1968. Multivariate stratified surveys. Journal American Statistical Association 63:530-534.

Cochran, W.G., 1963. Sampling techniques, 2nd edition. John Wiley & Sons, Inc. New York, 413 pp.

Colwell, R.N. et al., 1974. ERTS-1 data as an aid to wildland resource management in Northern California. A report of work done by scientists at the University of California, Berkeley, under NASA contract no. NAS 5-21827. Space Sciences Laboratory, University of California, Berkeley. 472 pp.

Dalenius, T., 1953. The multivariate sampling problem. Skard Aktmar-Tidskr 36:92-102.

Dalenius, T. Sampling in Sweden. Stockholm: Almquist and Wikrell. (1) symposia of sample survey methods and theories; (8) determining the optimum number of strata; (9) multiparametric stratified sampling; (10) the problem of regional statistics (design for total and domains of interest)

Dell, T.R. & J.L. Clutter, 1972. Ranked set sampling theory with order statistics background. Biometrics 28:545,555.

Duerr, W.A., D.E. Teeguarden, S. Guttenberg, N.B. Christiansen, 1975. Forest resource management, decision-making principles and cases. 2 volumes. Oregon State University Book Stores, Inc., Corvallis, Oregon.

Durbin, 1956. Sampling theory for estimates based on fewer individuals than the number selected. Bulletin of the International Statistical Institute 36:113-119.

Ek, A.R., 1971. A comparison of some estimates in forest sampling. Forest Science 17(1):2-13.

Fiacco, A.V. & G.P. McCormick, 1964. The sequential unconstrained minimization technique for non-linear programming, a primal, dual method. Management Science 10:360-366.

Fiacco, A.V. & G.P. McCormick, 1968. Non-linear programming: sequential unconstrained minimization techniques. Wiley & Sons, New York. 210 pp.

Ghosh, S.P., 1958. A note on stratified random sampling with multiple characters. Bulletin Calcutta Statistical Association 8:81-90.

Hajek, J., 1957. Some contributions to the theory of probability sampling. Bulletin of International Statistical Institute 36(3):127-133.

Hajek, J., 1964. Asymptotic theory of rejective sampling with varying probabilities from a finite population. Annual Mathematical Statistics 35:1491-1523.

Hansen, M.H., W.N. Hurwitz, & W.G. Madow, 1953. Sample survey methods and theory. Volume I, Methods and Applications, John Wiley & Sons, Inc., New York. 638 pp.

Hartley, H.O. & R.R. Hocking, 1963. Convex programming by tangential approximation. Management Science 9:600-612.

Hartley, H.O., 1965. Multiple purpose optimum allocation in stratified random sampling. Proceedings American Statistical Association, 1965. Social Sciences Section 258-261.

Hazard, J.W., 1974. Design of successive forest inventories: optimization by convex mathematical programming. Forest Science 20(2):117-127.

Hillier, F.S. & G.J. Lieberman, 1967. Introduction to operations research. Holden, Day Inc., San Francisco. 639 pp.

Huddleston, H.F., P.L. Claypool, & R.R. Hocking, 1970. Optimal allocation to strata using convex programming. Applied Statistics 19:273-278.

Hyde, W.F., 1976. Resources planning act: critique and alternative approach. Journal of Forestry 74(5):282-284.

Kendall, M.G. & A. Stuart, 1968. The advanced theory of statistics. Volume 3, Design and Analysis and Time Series. Hafner Publishing Co., New York. 557 pp.

Kokan, A.R., 1963. Optimum allocation in multivariate surveys. J.R. Statistical Society, Series A 126:557-565.

Kokan, A.R. & S. Khan, 1967. Optimum allocation in multivariate surveys: an analytical solution. J.R. Statistical Society, Series B 29:115-125.

Lahiri, D.B., 1965. Multisubject sample survey system - some thoughts based on Indian experience in contributions to statistics. Presented to Professor P.C. Mahalanobis on the occasion of his 70th birthday. Perganion Press. pp. 175-220.

Langley, P.G., 1969. Multistage sampling of forest resources by using space photography - an Apollo 9 case study. Second Annual Earth Resources Aircraft Program Status Review, Volume 2. NASA Manned Spacecraft Center, Houston, Texas. 28 pp.

Langley, P.G., 1969. New multistage sampling techniques using space and aircraft imagery for forest inventory. Proceedings Sixth International Symposium on Remote Sensing of Environment, University of Michigan. pp. 1179-1192.

Langley, P.G., 1971. Multistage sampling of Earth resources with aerial and space photography. Monitoring Earth Resources from Aircraft and Spacecraft. NASA SP-275. (Scientific and Technical Information Office)

Leopold, L.B., F.E. Clarke, B.B. Hanshaw, & J.R. Balskey, 1971. A procedure for evaluating environmental impact. USGS Circular 645.

Mahalanobis, P.C., 1952. Some aspects of the design of sample surveys. Sankhya 12:1-7.

Miller, R.G., 1966. Simultaneous statistical inference. McGraw-Hill Book Co., New York. 272 pp.

Morrison, D.F., 1967. Simultaneous statistical inference. McGraw-Hill Book Co., New York. 338 pp.

Murthy, M.N., 1967. Sampling theory and methods. Statistical Publishing Society, 204/1 Barrockpore Truck Road, Calcutta 35, India.

Mylander, W.C., R.L. Holmes, & G.P. McCormick, 1971. A guide to SUMT-Version 4: the computer program implementing the sequential unconstrainted minimization technique for nonlinear programming. Research Analysis Corporation, McLean, Virginia. 215 pp.

NASA Goddard Space Flight Center, 1971. Earth resources technology satellite data users handbook. NASA Document Number 71 SD 4249.

Naylor, T.H., J.L. Balintfy, D.S. Burdick, K. Chu, 1966. Computer simulation techniques. John Wiley & Sons, Inc., New York. 352 pp.

Raj, D., 1968. Sampling theory. McGraw-Hill Book Co., New York. 302 pp.

Rao, J.N.K., 1966. On the relative efficiency of some estimators in PPS sampling for multiple characteristics. Sankhya A28:61-70.

Remote Sensing Research Program, 1976. Mapit users handbook. Internal software documentation. University of California, Berkeley.

Sampford, M.R., 1962. An introduction to sampling theory. Oliver & Boyd, London. 292 pp.

Schneeberger, H., 1971. Optimum stratification and allocation in the theory of sampling - in German. Unternehmens Jorschung 1971, 15:240-254. (Abstract found in Statistical Theory and Method Abstract 1972(13):1468.

Schreuder, H.T., J. Sedransk, K.D. Ware, & D. Hamilton, 1968. 3-P sampling and some alternatives, I. Forest Science 14:429-454.

Schreuder, H.T., J. Sedransk, K.D. Ware, & D. Hamilton, 1971. 3-P sampling and some alternatives, II. Forest Science 17:103-118.

Sedransk, J., 1967. Designing some multifactor analytical studies. Journal of American Statistical Association 62:1121-39.

Singh, M.P. & R. Tessier, 1976. Some estimators for domain totals. Journal of American Statistical Association 71 (354):322-325.

Stuart, A., 1963. Some remarks on sampling with unequal probabilities. Statistical Methodology 4:773-780.

Tikkiwal, D.B., 1960. On the theory of classical regression and double sampling estimation. Journal of the Royal Statistical Society Series B, 22:131-138.

Titus, S.J., M.J. Gialdini & J.D. Nichols, 1975. A total timber resource inventory based upon manual and automated analysis of Landsat-1 and supporting aircraft data using stratified multi-stage sampling techniques. In Tenth International Symposium on Remote Sensing of Environment, 6-10 October 1975, Ann Arbor, Michigan.

Tomlinson, R.F., 1972. Geographic data handling. Symposium Edition, 2 vols. IGU Commission of Geographical Data Sensing, UNESCO/IGU Second Symposium on Geographical Information Systems, Ottawa.

APPENDICES

A. Sam Houston National Forest Inventory description

B. Sam Houston National Forest Inventory Estimates

C. Survey Planning Model Version 1, Summary

D. Programming Formulations for Sampling Strategies

E. Plumas National Forest Cost Data

F. Mean-Covariance Summaries for Classification of Plumas NF

G. Sampling Simulation Results for Two Populations

H. Quincy Ranger District Planning Model Application Results

I. Planning Model, Input and Solution Output

J. Implementation plan for the optimal Quincy Ranger District sampling system.

APPENDIX A

Sam Houston National Forest Inventory Description

1. Population

    1.1 Entire Sam Houston National Forest included

    1.2 Units are circular plots of .4 acre size

2. Parameters - totals of variables of interest within plot

3. Sampling technique and estimators

    3.1 Sampling Technique

    -- Sets of (X,y) coordinates chosen at random with replacement

    -- Flight lines fixed by sets of coordinates

    -- Plots located along flight lines

    -- Stratum (USFS type condition class) of each plot identified

    -- Simple random sample without replacement of plots within each stratum for photo interpretation

    -- Auxiliary variable measured on photo for this set of plots.

    -- Subset of these plots chosen within each stratum by simple random sampling without replacement; variable of interest measured for these units on ground

    3.2 Parameter estimator

    X_{hig} = photo measurement for plot j of flight line i in stratum h

    X_{hij} = ground measurement for plot j of flight line i in stratum h

    $\hat{R} = \frac{\sum y_{hij}}{\sum X_{hij}}$, where summations are over all plots for which both photo and ground measurements exist

    $\overline{X}_h$ = average photo measurement for stratum h

    $\overline{y}_h$ = stratum h estimate $= \overline{X}_h \hat{R}$

    L = number of strata

    N_h = area expansion factor for stratum h (stratum size/average plot size)

    $\overline{y}$ = parameter estimator $= \sum_h^L N_h \overline{y}_h$

3.3 Variance estimator

$\hat{V}(X)$ = variance of X for photo plots matching ground plots

$\tilde{V}(X)$ = variance of X for all photo plots

$\hat{V}(y)$ = variance of y for all ground plots

$\hat{c}(X,y)$ = covariance of X and y

n_1 = number of photo plots in stratum

n_2 = number of ground plots in stratum

$\hat{V}(\bar{y}_h)$ = stratum h variance estimator

$$\hat{V}(\bar{y}_h) = \left\{ \frac{\hat{V}(y) + \hat{R}^2\,\hat{V}(X) - 2\hat{R}\,\hat{C}(X,y)}{n_2} + \frac{2\hat{R}\,\hat{C}(X,y) - \hat{R}^2\,\hat{V}(X)}{n_1} \right\} S$$

where $S = 1 + \frac{\tilde{V}(X)}{\left(\sum X_{hij}/n_1\right)^2}$

and where the summation is over all photo plots

$\hat{V}(\bar{y})$ = over-all variance estimator

$$\hat{V}(\bar{y}) = \sum_{H}^{L} N_h^2\,\hat{V}(\bar{y}_h)$$

4. Sample size and allocation

4.1 30 flight lines were flown over the forest

4.2 264 photo plots were acquired by post stratification with approximately proportional allocation among strata

4.3 29 field plots were selected proportionally to stratum area with a minimum of 1 plot per strata.

5. MEASUREMENT PROCEDURES

5.1 Preliminary Testing

5.1.1 Photo Acquisition and Field Work

On March 14, 1975, twelve flight lines of photography were acquired over the Sam Houston National Forest. These lines contained photography at two different scales, 1:4000 (wide angle) and 1:1000 (on a 3 x 5 print). The photography was acquired using two 35mm cameras. The wide angle photography was acquired using a manual advance Nikkormat with a 50mm lens. Forward lap of the wide angle photos was approximately 60% and these photos were used primarily for location of the flight lines on maps and navigation in the field. The larger scale photos, 1:1000, were acquired as stereo triplets located at the centers of the wide angle photos. These stereo triplets were acquired using a motor drive Nikon F and a 200mm lens. The forward lap of the triplet photos was 70 - 80%. The flight lines were located so as to adequately image vegetation types found on the forest and to be easily accessed for ground work.

Ground data collection consisted of locating the plot centers of eleven stereo triplets and identifying tree and brush individuals which were apparent on both the photo and the ground. The individuals were located on the center photo of the triplet, pinpricked and numbered. The individual was identified to species when possible, and when appropriate the diameter (DBH) and height were measured and recorded. Diameters were measured with a standard tape and heights were measured with a Sunto clinometer. The diameter and height information was to be used in defining photo volume estimation procedures and the species identification was used to train interpreters for the forest inventory photo interpretation.

Field work was completed in four days by one man. Forest Service personnel from the Raven District office provided excellent field support by making a man available for field orientation and aiding in species identification. An important consideration became evident during the last four days of field work and that was the problems of determining ground location fro. the photos for navigation to specific plots and establishing, with a high degree of accuracy, the position of those plots. The problem is caused by the uniformity of the terrain and the dense forest canopy. As a result, a majority of the training plots were chosen because of proximity to well defined landmarks.

5.1.2 Photo - Ground Relationships

Extreme uniformity and density of vegetation made individual tree measurements difficult or impossible. Instead average photo variables for a plot were related to ground characteristics. Average difference of parallax, a proxy for height, was related to average plot DBH for the eleven test plots and interpretations were made by three different interpreters. Correlations obtained were 0.838, 0.876, and 0.769. With these results and the well established high correlation between diameter and volume it was concluded that difference in parallax should be reasonably well related to stand volume.

5.2 Photo Interpretation

5.2.1 Photo Acquisition

The forest wide volume estimation procedure utilized 30 flight lines of photography which were located at random over the Sam Houston National Forest. The photos were acquired with the same specification discussed in the previous section. The photos were acquired in December, 1975. The photos were labeled by flight line, circular plot boundaries for 4/10 acre plots were inked onto the middle photo of each stereo triplet and the photo plot centers located on Forest Service maps and photo mosaics. The photos were printed in 3 x 5 format for both scales of photography.

5.2.2 Photo Interpretation Procedures

Two interpreters analyzed 264 large scale photos for a multiple of variables. All of the interpretation was made using an Abrams Stereoscope. One interpreter viewed the entire set and recorded for each plot locational data, Forest Service condition class, numbers of trees in the plot and hardwoods, canopy percentage for both hardwood and conifers and the number of snags present.

The second interpreter measured parallax difference. Where possible five trees representative of conditions in the plot were chosen. Only conifers were measured. Three base to top parallax measurements were made for each tree using a Wild mirror stereoscope and parallax bar. Thirty to forty percent of the plots had three or only one tree measured due to extremely dense canopies, deep shadowing in openings which were present or very sparse tree canopy. The parallax difference measurements were accumulated and the average determined for each plot.

5.2 Field Work

After the photo plots were located on the Forest Service maps it was determined for each plot whether it was on Forest Service land and what type condition class in which it lay. The number of ground sample plots chosen in each type condition class was determined by the relative area of the class in relation to the total forest area. Table 1 shows the number of plots chosen for ground sampling in each type condition class. The specific plots were chosen at random without replacement from the set of photo plots falling in the type condition class. The large scale photos of the specific plots were duplicated for use by the field crew. In the field the crew of two men used NASA supplied high altitude photography (1:60,000) for general navigation. In the area of the plot the

the wide angle coverage was used to navigate to the specific plot area and finally the center point of the circular plot was located by interpreting the large scale stereo triplet. The crew measured the actual ground distance of the radius of the inked circular plot boundary to determine exact scale of the photos. Within the plot boundaries all conifer and hardwood trees 5" and greater in diameter were measured. Data recorded for the conifer trees included DBH, length of last five years growth, bark thickness, and Forest Service tree class. For the hardwoods, only DBH was recorded. In addition five dominant or codominant trees were selected and marked on the photos. Data for these trees included DBH, total age and height. Recent cutting was evaluated by recording stump diameters and the type of removal, thinning or harvest. Type-condition of the plot was recorded and compared with Forest Service designation. Travel times and time on plot were recorded. The ground data collection effort for a total of 29 plots required 54 man days for completion.

Table A-1. Number of ground plots sampled in each Forest Service designated type condition classes. Number of plots sampled was determined by relative areas of the various classes. Type condition class codes are consistent with Compartment Prescription Handbook, USFS-FSH 2409.21d.

Type Condition Class	Number of Field Plots
3101	1
3106 and 3110	5
3111	1
3112	14
3113 and 3114	1
3200	5
46	1
Hardwoods	1

APPENDIX B

SHNF Estimates Based on 1976 Survey
Notes for the Tables

1) All estimates are in thousands of units

2) Diameter class is inches at breast height

3) Relative error is the standard error of an estimate divided by the estimate

Table 10 -- Number of Growing Stock Trees on Commercial Forest Land by Species and Diameter Class, SHNF, 1976.

FINAL ESTIMATE

Diameter Class

	TOTAL	5-7	7-9	9-11	11-13	13-15	15-17	17-19	19-21	21-29	OVER 29
SOFTWOOD	0.158E 5	0.400E 4	0.313E 4	0.270E 4	0.235E 4	0.160E 4	764.	524	300.	404.	12.0
HARDWOOD	0.854E 4	0.508E 4	0.181E 4	809.	404.	135.	135.	59.6	30.8	74.9	0.000
TOTAL	0.243E 5	0.906E 4	0.494E 4	0.350E 4	0.276E 4	0.174E 4	899.	584	330.	479.	12.0

RELATIVE ERROR

	TOTAL	5-7	7-9	9-11	11-13	13-15	15-17	17-19	19-21	21-29	OVER 29
SOFTWOOD	0.793E -1	0.101	0.122	0.104	0.142	0.133	0.173	0.264	0.280	0.442	1.12
HARDWOOD	0.126	0.126	0.152	0.193	0.261	0.417	0.415	0.981	0.845	0.984	0.000
TOTAL	0.493E -1	0.894E -1	0.767E -1	0.746E -1	0.116	0.125	0.154	0.255	0.282	0.421	1.12

Table 11. Volume of Timber on Commercial Forest Land, by Class of Timber and Softwoods and Hardwoods, SHNF, 1976.

FINAL ESTIMATE

Timber Class

	TOTAL	CODE 05	CODE 10	CODE 20	CODE 30	CODE 40
SOFTWOOD	0.322E 6	0.532E 4	0.217E 6	0.973E 5	0.180E 4	713.
HARDWOOD	0.611E 5	689.	0.183E 5	0.311E 5	0.513E 4	0.578E 4
TOTAL	0.383E 6	0.601E 4	0.235E 6	0.128E 6	0.693E 4	0.649E 4

RELATIVE ERROR

	TOTAL	CODE 05	CODE 10	CODE 20	CODE 30	CODE 40
SOFTWOOD	0.772E -1	0.340	0.943E -1	0.115	0.468	0.485
HARDWOOD	0.303	0.485	0.278	0.240	0.820	0.690
TOTAL	0.781E -1	0.742	0.795E -1	0.108	0.679	0.605

Code	Description (USFS definitions)
5	Mortality trees
10	Desirable trees
20	Acceptable trees
30	Rough trees
40	Rotten cull trees

ORIGINAL PAGE IS
OF POOR QUALITY

Table 12. Volume of Growing Stock and Sawtimber on Commercial forest land, by Ownership (all USFS) and Softwoods and Hardwoods, SHNF, 1976

FINAL ESTIMATE

	Diameter Class		
	TOTAL	5 TO 11	11 AND UP
SOFTWOOD	0.323E 6	0.672E 5	0.256E 6
HARDWOOD	0.611E 5	0.296E 5	0.314E 5
TOTAL	0.384E 6	0.968E 5	0.287E 6

RELATIVE ERROR

	TOTAL	5 TO 11	11 AND UP
SOFTWOOD	0.766E -1	0.123	0.964E -1
HARDWOOD	0.303	0.159	0.534
TOTAL	0.776E -1	0.942E -1	0.108

Table 13. Volume of Growing Stock on Commercial Forest Land, by Species and Diameter Classes, SHNF, 1976.

FINAL ESTIMATE

		Diameter Class									
	TOTAL	5-7	7-9	9-11	11-13	13-15	15-17	17-19	19-21	21-29	OVER 29
SOFTWOOD	0.323E 6	0.121E 5	0.210E 5	0.340E 5	0.490E 5	0.528E 5	0.354E 5	0.336E 5	0.270E 5	0.537E 5	0.440E 4
HARDWOOD	0.611E 5	0.117E 5	0.980E 4	0.814E 4	0.690E 4	0.362E 4	0.502E 4	0.325E 4	0.237E 4	0.103E 5	0.000
TOTAL	0.364E 6	0.238E 5	0.308E 5	0.421E 5	0.559E 5	0.565E 5	0.404E 5	0.360E 5	0.294E 5	0.640E 5	0.440E 4

RELATIVE ERROR

	TOTAL	5-7	7-9	9-11	11-13	13-15	15-17	17-19	19-21	21-29	OVER 29
SOFTWOOD	0.766E -1	0.179	0.152	0.137	0.186	0.131	0.177	0.265	0.258	0.423	1.21
HARDWOOD	0.303	0.121	0.175	0.217	0.265	0.422	0.439	0.714	0.879	1.03	0.000
TOTAL	0.776E -1	0.109	0.111	0.114	.158	.119	.157	.246	.257	.4 4	1.21

Table 14. Volume of Sawtimber on Commercial Forest Land, by Species and Diameter Classes, SHNF, 1976.

FINAL ESTIMATE

	TOTAL	11-13	13-15	15-17	17-19	19-21	21-23	23-25	25-27	27-29	OVER 29
						Diameter Class					
SOFTWOOD	0.256E 6	0.490E 5	0.528E 5	0.354E 5	0.336E 5	0.270E 5	0.224E 5	0.236E 5	0.788E 4	0.000	0.440E 4
HARDWOOD	0.314E 5	0.670E 4	0.362E 4	0.502E 4	0.325E 4	0.237E 4	0.464E 4	0.000	0.000	0.565E 4	0.000
TOTAL	0.287E 6	0.559E 5	0.565E 5	0.404E 5	0.369E 5	0.294E 5	0.270E 5	0.236E 5	0.788E 4	0.565E 4	0.440E 4

RELATIVE ERROR

	TOTAL	11-13	13-15	15-17	17-19	19-21	21-23	23-25	25-27	27-29	OVER 29
SOFTWOOD	0.964E -1	0.196	0.131	0.177	0.265	0.258	0.578	0.435	0.596	0.000	1.21
HARDWOOD	0.534	0.265	0.422	0.439	0.714	0.879	0.873	0.000	0.000	1.26	0.000
TOTAL	0.108	0.128	0.119	0.157	0.249	0.257	0.521	0.435	0.596	1.26	1.21

Table 16A and 18A. 10 Year Volume Growth of Growing Stock and Sawtimber on Commercial Forest Land, by Species, SHNF, 1976.

FINAL ESTIMATE

	TOTAL	5 TO 11	11 AND UP
SOFTWOOD	0.111E 5	0.315E 4	0.796E 4
HARDWOOD	0.000	0.000	0.000
TOTAL	0.111E 5	0.315E 4	0.796E 4

RELATIVE ERROR

	TOTAL	5 TO 11	11 AND UP
SOFTWOOD	0.684E -1	0.132	0.779E -1
HARDWOOD	0.000	0.000	0.000
TOTAL	0.684E -1	0.132	0.779E -1

Table 16B and 18B. 10 Year Volume Removals of Growing Stock and Sawtimber on Commercial Forest Land, by Species, SHNF, 1976.

FINAL ESTIMATE

	TOTAL	5 TO 11	11 AND UP
SOFTWOOD	0.785E 4	222.	0.100E 4
HARDWOOD	222.	222.	0.000
TOTAL	0.807E 4	443.	0.100E 4

RELATIVE ERROR

	TOTAL	5 TO 11	11 AND UP
SOFTWOOD	0.569	1.13	0.775
HARDWOOD	1.15	1.15	0.000
TOTAL	0.551	0.766	0.775

Table 20. Mortality of Growing Stock and Sawtimber on Commercial Forest Land, by Species, SHNF, 1976.

FINAL ESTIMATE

Diameter Class

	TOTAL	5 TO 11	11 AND UP
SOFTWOOD	0.567E 4	421.	724.
HARDWOOD	348.	348.	0.000
TOTAL	0.602E 4	772.	724.

RELATIVE ERROR

	TOTAL	5 TO 11	11 AND UP
SOFTWOOD	0.789	1.28	1.28
HARDWOOD	1.40	[illegible].40	0.000
TOTAL	0.744	0.901	1.28

APPENDIX C

Survey Planning Model Version 1, Summary

1. SPM is a coordinated set of computer programs written in FORTRAN IV and implemented on a CDC 7600 computer.

2. It functions to carry out major tasks associated with components of the model control and allows execution of each program singly or in sequence.

3. Four major programs are involved

   .1 SIMULATE reads data describing the population and simulates individual population elements. Results are saved for later use.

   .2 CLUSTER reads data produced by SIMULATE and prepares the population representation for sampling. This includes stratification, clustering into sampling units, definition of selection probabilities and computation of population variance components.

   .3 SUMT utilizes data output from CLUSTER as well as user specification of sampling strategy, cost coefficient, precision desires and other parameters. A nonlinear programming problem formulation is used to determine the sampling allocation which will meet the precision constraints with the least cost.

   .4 COMPARE evaluates output from several alternative formulations of SUMT and summarizes each beginning with the least cost alternative.

4. A flow chart is chown in Figure C-1 to indicate interactions, inputs, and outputs of the model.

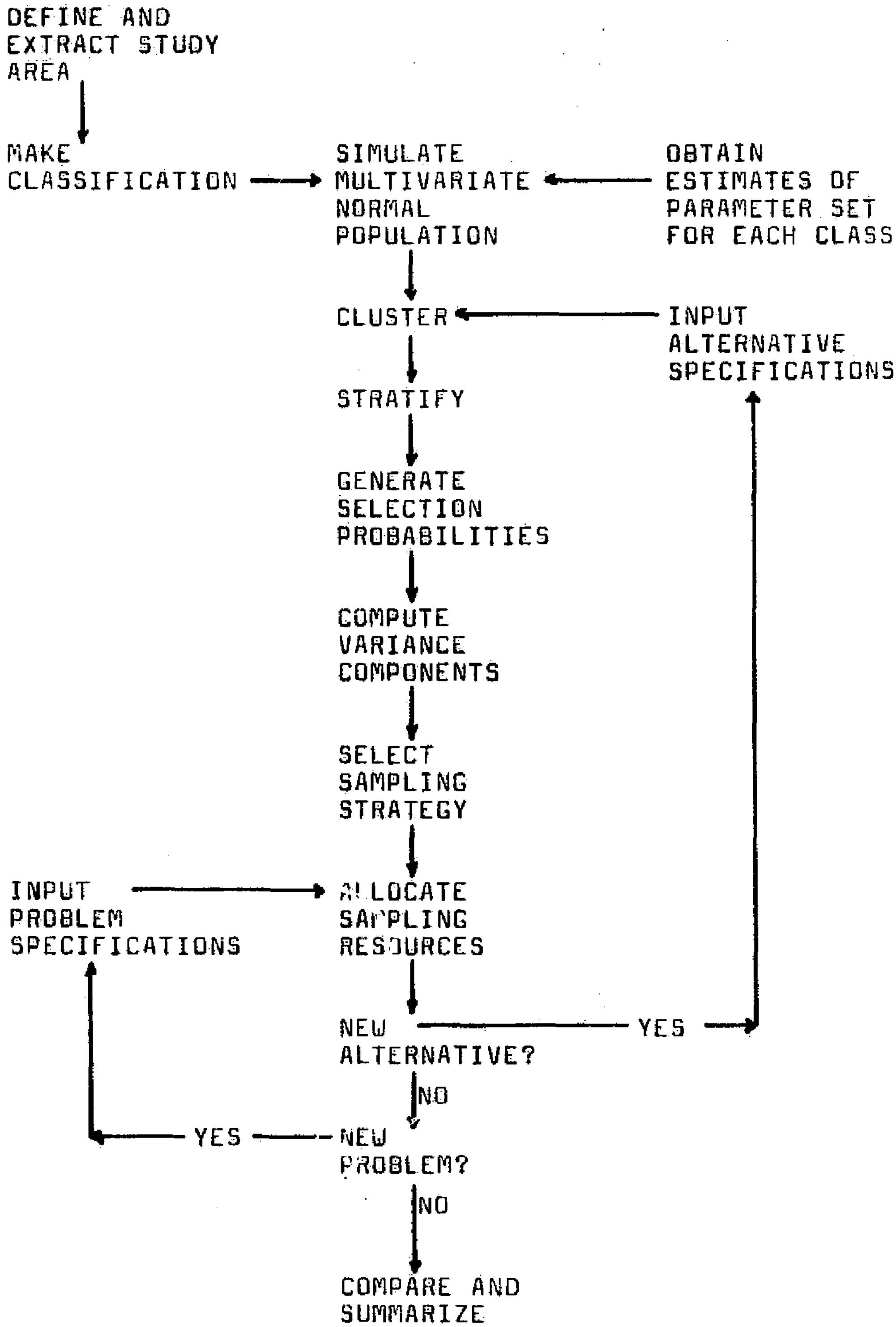


Figure C-1. Planning model flow chart

APPENDIX D.

Programming Formulations for Sampling Strategies

Cost and variance functions as well as other constraints on sample size are included here for 1) stratified sampling, 2) stratified two stage sampling, and 3) stratified two stage with double sampling. First and second partial derivations and evaluated for both cost and variance functions.

Notation, except where otherwise defined, follows that of Cochran, 1963.

1. Stratified Sampling Allocation

$$\text{Min } \pi = \sum_{h=1}^{L} \left\{ C_{h1} n_h + C_{h2} n_h^{\frac{1}{2}} \right\}$$

$$\text{Subject to } V(\hat{Y}_j) = \sum_{h=1}^{N} N_h^2 S_{h_j}^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) \leq V_j, \; 0 \leq n_h \leq N_h \;, j = 1 \ldots M$$

$$\frac{\partial \pi}{\partial n_h} = C_{h1} + \frac{C_{h2}}{2} n_h^{-\frac{1}{2}} = 1 \ldots L$$

$$\frac{\partial^2 \pi}{\partial n_h^2} = - \frac{C_{h2}}{4} n_h^{3/2} \quad h = 1 \ldots L$$

$$\frac{\partial^2 \pi}{\partial n_h \; \partial n_h'} = 0$$

$$\frac{\partial V(\hat{Y}j)}{\partial n_h} = \frac{N_h^2 S_{hj}^2}{n_h^2} \quad h = 1 \ldots L, \; j = 1 \ldots m$$

$$\frac{\partial^2 V(\hat{Y}_j)}{\partial n_h^2} = \frac{-2N_h^2 S_{hj}^2}{n_h^3} \quad h = 1 \ldots L, \; j = 1 \ldots M$$

$$\frac{\partial^2 V(\hat{Y}_j)}{\partial n_h \partial n_{h'}} = 0 \qquad h = 1 \ldots L, \; j = 1 \ldots M$$

$$h' \geq h$$

2. Stratified Two Stage Sample Allocation NLP Formulation

$$\text{Min} \quad = \sum_{h=1}^{L} \left\{ C_{1h}\, n_h^* + C_{2h}\, n_h^* n'_h + C_{3h}\sqrt{n_h^*} \right\} + C_o$$

Subject to

(1) $V(\hat{Y}_j) \leq Vj = 1 \ldots \text{NPAR}$

$$V(\hat{Y}_j) = \sum_{h=1}^{L} \left\{ \frac{1}{n^*_h} \left\{ \sum P_i \left(\frac{y_{ij}}{P_i} - Y_j \right)^2 + \sum \frac{\sigma_{ij}{}^2}{P_i} \right\} \right\}$$

$$\sigma_i{}^2 = M_{hi}^2 \quad S_{hij}^2 \left(\frac{1}{n'_h} - \frac{1}{M_h} \right)$$

(2) $n'_h < M_h$

(3) $n_h^* < N_h$

maximum sample size

(4) $n_h^* \geq 0.01$

$n'_h \geq 0.01$

minimum sample size

$$\text{Min}\pi = \sum_{h=1}^{L} \left\{ C_{h1}\, n_h + C_{h2}\, m_h n + C_{h3}\, n_h^{\frac{1}{2}} \right\} + C_o$$

$$\text{Subject to } V(\hat{Y}_j) = \sum_{h=1}^{L} \left\{ N_h^2\, S_{hbj}^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) + \frac{N_h^2 M_h^2 S_{hwj}^2}{n_h} \quad \frac{1}{m_h} - \frac{1}{n_h} \right\} \leq \quad J = 1 \ldots \text{NPAR}$$

$$\frac{\partial \pi}{\partial n_h} = C_{h1} + C_{h2}\, m_h + \frac{Ch3}{2}\, n_h^{-\frac{1}{2}}$$

$$\frac{\partial \pi}{\partial m_h} = C_{h2}\, n_h$$

$$\frac{\partial^2 \pi}{\partial n_h^2} = \frac{-C_{h3}}{4}\, n_h^{-3/2}$$

$$\frac{\partial^2 \pi}{\partial m_h^2} = 0$$

2. <u>Stratified Two Stage Sample Allocation</u> (cont.)

$$\frac{\partial^2 \pi}{\partial n_h \partial n_h'} = 0$$

$$\frac{\partial^2 \pi}{\partial n_h \partial m_h} = C_{h2}$$

$$\frac{\partial \pi}{\partial n_h \partial m_h'} = 0$$

$$\frac{\partial V_h(\hat{Y}_j)}{\partial n_h} = \frac{N_h^2 S_{hbj}^2}{n_h^2} + \frac{N_h^2 M_h^2 S_{hwj}^2}{n_h^2} + (\frac{1}{m_h} - \frac{1}{M_h})$$

$$\frac{\partial V_h(\hat{Y}j)}{\partial n_h} = \frac{N_h^2 M_h^2 S_{hwj}^2}{n_h \; m_h^2}$$

$$\frac{\partial^2 V_h(\hat{Y}j)}{\partial n_h{}^2} = \frac{-2 \; N_h^2 S_{hbj}^2}{n_h^3} - \frac{2 \; N_h^2 \; M_h^2 \; S_{hwj}^2}{n_h^3} \; (\frac{1}{m_h} - \frac{1}{M_n})$$

$$\frac{\partial^2 V_h(\hat{Y}j)}{\partial m_h^2} = \frac{-2 \; N_h^2 \; M_h^2 \; S_{hwj}^2}{n_h \; m_h^3}$$

$$\frac{\partial^2 V_h(Yj)}{\partial n_h \; \partial n_h'} = 0$$

$$\frac{\partial^2 V_h(\hat{Y}_j)}{\partial n_h \partial m_h} = \frac{-N_h^2 \; M_h^2 \; S_{hwj}^2}{n_h{}^2 m_h{}^2}$$

$$\frac{\partial^2 V_h(\hat{Y}_j)}{\partial m_h \; \partial m_h'} = 0$$

3. Two Stage with Double Sampling NLP Formulation (ignoring Stratification)

$$\underline{\text{Min}}\ \pi = C_1 n^* + C_2 n^* n' + C_3 \sqrt{n^*} + C_4\, n^* n + C_5 \sqrt{n^* n}$$

Subject to

(1) $V(\hat{Y}_j) \leq V_j \quad j = 1 \ldots$ NPAR (The number of parameters to be estimated)

$$V(\hat{Y}_j) = \frac{1}{n^*}\left\{\sum P_i \left(\frac{y_{ij}}{P_i} - Y_j\right)^2 + \sum \frac{\sigma_{ij}^2}{P_i}\right\}$$

$$\sigma_{ij}^2 = M_i^2 \left[\left(\frac{n^* n' - n^* n}{n^* n'}\right)\left(\frac{1}{n^* n}\right) \sigma_{yij}^2 \ (1-\rho^2) \ \left(1 + \frac{1}{n^*(n-3)}\right) + \frac{M-n'}{Mn'}\ \sigma_{yij}^2\right]$$

(assumes combined sampling for regression)
(Reference: Tikkiwal 1960, p. 137)

(2) $\left.\begin{array}{l} n^* \geq 0.01 \\ n' \geq 0.01 \\ n \geq 0.01 \end{array}\right\}$ Min sample size

(3) $n^* * n \geq 3$ to assure sufficient degrees of freedom for variance

(4) $\left.\begin{array}{l} n' > n \\ M > n' \\ N > n^* \end{array}\right\}$ max sample size

To obtain partials for variance constraints:

$$V(\hat{Y}) = \frac{1}{n^*}\underbrace{\sum P_i\left(\frac{y_i}{P_i} - Y\right)^2}_{\text{Define} \longrightarrow B} + \left(\frac{M^2}{m^*}\right)\left(\frac{n'-n}{n'n}\right)\left(\frac{1-\rho^2}{n^*}\right)\left(1 + \frac{1}{n^*n-3}\right)\sum\frac{\sigma^2 yi}{P_i} + \left(\frac{M^2}{n^*}\right)\left(\frac{M-n'}{Mn'}\right)\sum\frac{\sigma^2 yi}{P_i}$$

$$V(\hat{Y}) = \frac{B}{n^*} + \underbrace{M^2(1-\rho^2)}_{\text{Define} \longrightarrow D}\left(\frac{1}{n^*n} - \frac{1}{n^*n'}\right)\left(1 + \frac{1}{n^*n-3}\right)\underbrace{\sum\frac{\sigma^2 yi}{P_i}}_{F} + \underbrace{M^2}_{E}\left(\frac{1}{n^*n'} - \frac{1}{n^*M} - \frac{1}{n^*M}\right)\underbrace{\sum\frac{\sigma^2 yi}{P_i}}_{F}$$

$$V(\hat{Y}) = \frac{B}{n^*} + \frac{DF}{n^{*2}n} - \frac{DF}{n^{*2}n'} + \frac{DF}{n^{*2}n(n^*n-3)} - \frac{DF}{n^{*2}n'(n^*n-3)} + EF\left(\frac{1}{n^*n'} - \frac{1}{n^*M}\right)$$

Define $X_1 = \frac{1}{n^*}$

$x_2 = \frac{1}{n'}$

$x_3 = \frac{1}{n}$

$$V(\hat{Y}) = B\,X_1 + DF\,X_1^2\,X_3 - DF\,X_1^2X_2 + \frac{DF\,X_1^2X_3}{\frac{1}{X_1X_3} - 3} - \frac{DF\,X_1^2X_2}{\frac{1}{X_1X_3} - 3} + EF\left(X_1X_2 - \frac{X_1}{M}\right)$$

$$\Downarrow$$

$$B\,X_1 + DF\,X_1^2(X_3 - X_2) + \frac{DF\,X_1^3X_3^2}{1-3X_1X_3} - \frac{DF\,X_1^3X_2X_3}{1-3X_1X_3}$$

$$\Downarrow$$

$$\frac{DF\,X_1^3(X_3^2 - X_2X_3)}{1 - 3X_1X_3}$$

$$\frac{\partial v}{\partial X_1} = B + 2DF\ X_1X_3 - 2\ DFX_1X_2 + \underbrace{3DF\ (X_3^2-X_2X_3)\left[\frac{X_1^2}{(1-3X_1X_3)}\frac{X_1^3}{(1-3X_1X_3)}\right]}_{\text{For the derivation of this term see 1 below}} + EF\ X_2 - \frac{EF}{M}$$

$$\frac{\partial v}{\partial X_2} = -\ DFX_1{}^2 - \frac{DFX_1^3\ X_3}{1-3X_1X_3} + EFX_1$$

$$\frac{\partial v}{\partial X_3} = DF\ X_1^2 + \underbrace{\frac{DF\ X_1^3\ (2X_3-X_2)}{1-3X_1X_3} + \frac{3\ DF\ X_1^4\ (X_3^2 - X_2X_3)}{(1-3X_1X_3)^2}}_{\text{For the derivation of this term see 2 below}}$$

Second partials were evaluated using numerical differencing techniques.

(1) $$f = \frac{D\ F\ X_1^3}{(1-3\ X_1X_3)}\ (X_3^2 - X_2X_3) = \frac{u}{v}$$

$$\frac{\partial f}{\partial X_1} = \frac{1}{v}\ \frac{d\ u}{d\ X_1} - \frac{u}{v^2}\ \frac{d\ v}{d\ X_1}$$

$$= \frac{3\ DFX_1^2\ (X_3^2 - X_2X_3)}{1-3\ X_1X_3} - \frac{DFX_1^3\ (X_3^2 - X_2X_3)}{(1-3\ X_1X_3)^2}\quad(-3\ X_3)$$

$$= 3\ DF\ (X_3^2 - X_2X_3)\quad \frac{X_1^2}{1-3X_1X_3} + \frac{X_1^3X_3}{(1-3X_1X_2)}{}^2$$

---

(2) $$f = \frac{DF\ X_1^3\ (X_3^2 - X_2X_3)}{1 - 3X_1X_3}$$

$$\frac{\partial f}{\partial X_3} = \frac{D\ F\ X_1^3\ (2X_3-X_2)}{1-3\ X_1X_3} + \frac{3D\ F\ X_1{}^4(X_3^2 - X_2X_3)}{(1 - 3X_1X_3)^2}$$

APPENDIX E

Cost Figures - Plumas Inventory 1974

I. Pre-photo/ground		Single Date
Tape acquisition (1021-18163)		360.00
Tape Reformatting		
Tape 3 tapes/date @7.75 each	23.25	
Computer Time 1 hour per date @$40/hour	40.00	
Operator 1½ hours @$5/hour	7.50	70.75
Test Area Extraction		
Tape 3 tapes/date @7.75 each	23.25	
Computer 1 hour per date @40/hour	40.00	
Operator 1½ hours per date @5/hour	7.50	70.75
Delineation and extraction of stratification (initial)		
Photo reduction of map	15.75	
Digitizer with operator 2 hours @$43/hour	86.00	
LSR. fit of co-ordinates: Computer 6 runs @$1.80	10.80	
Operator 8 hours @$3.00/hour	24.00	136.55
Computer Mask Generation: Computer	12.50	
Operator 3 hours @$5/hour	15.00	27.50
Delineation and Extraction of Administrative Boundaries		
Photo Reduction of Map	15.75	
Digitizer with Operator 8 hours @$43/hour	344.00	
LSR Fit of coordinates: Computer 9 runs $1.80/run	16.20	
Operator 10 hours $3/hour	30.00	406.00
Computer Mask Generation: Computer	12.50	
Operator 3 hours @$5/hour	15.00	27.50
Training of Classifier = 60 classes/strip, 3 strips/date*		
Computer Display terminal 24 hours @$40/hour	960.00	
Image analysts 40 hours @$6/hour	240.00	
Statistical analysis: Computer (LBL $25.00)	25.00	
Operator 10 hours @$3/hour	30.00	1255.00
Selection of Channels & classes: 20 hours @$5.00		100.00

Discriminant analysis run

Single date:	Computer (3 runs totaling)	918.00	
	Operator/analyst 24 hours @$7.60/hour	182.00	
			1100.00

Generation and Selection of PSU's

Computer 3 hours @$40/hour	120.00	
Analyst 4 hours @$3.40/hour	14.00	
		134.00

Location of PSU's (for aerial photography)

Computer 10 hours @$40/hour	400.00	
Analyst 80 hours @$5/hour	400.00	
		800.00
Sub-total		4488.00

II. Aerial Photography/Interpretation

Photo acquisition

Aircraft 25 hours @$32/hour	800.00	
Pilot 32 hours @$6/hour	192.00	
Photographer 32 hours @$3.40/hour	109.00	
Film 60 rolls @$3.60/roll	216.00	
Processing 60 rolls @$1.90/roll	114.00	
Printing Ave. 30 images per roll @$63/5" x 7" print (1800 prints)	1134.00	
		2565.00

Photo Interpretation

		1244.00
Image analyst 360 hours @$3.40/hour	Sub-total II	3789.00

III. Ground Data Collection

Travel

Mileage 4590 @ 15¢/mile		689.00

Crew

Wages 46 boys x 4 men = 184 man days (1472 hrs. @ 3.86/hour	5682.00	
Per diem @ $21/day	3800.00	9482.00
Sub-total III		10,171.00

IV. Data Summary and Map Generation

Computer Analysis ground and photo data	$ 280.00	
Combining ERTS and ground data 8 hours @$5.00/hour	40.00	
Generation of summary statistics 16 hours @$5/hour	80.00	
Report preparation and reproduction	650.00	
Computer time	250.00	
Sub-total IV		1300.00
Sub-total	$19,748.00	
V. Administrative (27%)	5332.00	
VI. Overhead (30.2% of I-V)	7574.00	
Total Cost	$32,654.00	

APPENDIX F

Mean-covariance summaries for the classification of the Plumas NF.

This appendix summarizes the basic population estimates for three parameters by vegetation type-condition class (TCC) category as follows:

TCC	Description
1	True fir - regeneration
2	" - immature
3	" - mature
4	" - overmature
5	" - poorly stocked
6	Mixed conifer - regeneration
7	" - immature
8	" - mature
9	" - overmature
10	" - poorly stocked
11	East side pine - regneration
12	" - immature
13	" - mature
14	" - overmature
15	" - poorly stocked
16	Hardwoods
17	nonforest

SUBROUTINE CCVMAT N = 4 M = 3

	BASAL AREA	NO. TREES	BAGR
SUM	4.53424E+02	3.66636E+02	3.88350E+01
MEAN	1.13356E+02	9.16589E+01	9.70874E+00
VAR	8.91645E+03	2.38600E+03	2.32859E+01
CV	8.33025E+01	5.32918E+01	5.54104E+01

COVARIANCE MATRIX

	BASAL AREA 1	NO. TREES 2	BAGR 3
BASAL AREA 1	8.91645E+03		
NO. TREES 2	2.34841E+03	2.38600E+03	
BAGR 3	4.48906E+02	1.52901E+02	2.32859E+01

CORRELATION MATRIX

	BASAL AREA 1	NO. TREES 2	BAGR 3
BASAL AREA 1	1.0000		
NO. TREES 2	.5091	1.0000	
BAGR 3	.9852	.6487	1.0000

NEW CY...E...TCC 2

SUBROUTINE CCVMAT N = 17 M = 3

	BASAL AREA	NO. TREES	BAGR
SUM	1.38623E+03	9.40346E+02	8.42823E+01
MEAN	8.15430E+01	5.53145E+01	4.95778E+00
VAR	6.17838E+03	9.82957E+02	1.04427E+01
CV	9.63942E+01	5.66798E+01	6.51807E+01

COVARIANCE MATRIX

	BASAL AREA 1	NO. TREES 2	BAGR 3
BASAL AREA 1	6.17838E+03		
NO. TREES 2	8.92356E+02	9.82957E+02	
BAGR 3	1.82102E+02	8.15856E+01	1.04427E+01

CORRELATION MATRIX

	BASAL AREA 1	NO. TREES 2	BAGR 3
BASAL AREA 1	1.0000		
NO. TREES 2	.3621	1.0000	
BAGR 3	.7169	.8053	1.0000

```
SUBROUTINE COVMAT          N =    2   , M =    3

              BASAL AREA  NO. TREES   BAGR
SUM           3.16459E+02 2.33912E+02 2.53900E+01
MEAN          1.58249E+02 1.16956E+02 1.26950E+01
VAR           2.05662E+03 9.57740E+02 4.17577E+00
CV            2.86575E+01 2.64607E+01 1.60966E+01

COVARIANCE MATRIX

                  BASAL AREA  NO. TREES   BAGR
                      1           2           3
BASAL AREA1 2.05662E+03
NO. TREES 2 1.40346E+03 9.57740E+02
BAGR      3 9.26714E+01 6.32400E+01 4.17577E+00

CORRELATION MATRIX

                 BASAL AREA  NO. TREES   BAGR
                      1           2           3
BASAL AREA1         1.0000
NO. TREES 2         1.0000      1.0000
BAGR      3         1.0000      1.0000      1.0000
```

```
NEW CV" E...TCC  5
```

```
SUBROUTINE COVMAT          N =   14     M =    3

              BASAL AREA  NO. TREES   BAGR
SUM           5.12901E+02 3.90828E+02 3.55474E+01
MEAN          3.66358E+01 2.79163E+01 2.53910E+00
VAR           1.08304E+03 5.12841E+02 5.70561E+00
CV            8.98292E+01 8.11211E+01 9.40744E+01

COVARIANCE MATRIX

                  BASAL AREA  NO. TREES   BAGR
                      1           2           3
BASAL AREA1 1.08304E+03
NO. TREES 2 6.92130E+02 5.12841E+02
BAGR      3 7.24924E+01 5.25794E+01 5.70561E+00

CORRELATION MATRIX

                 BASAL AREA  NO. TREES   BAGR
                      1           2           3
BASAL AREA1         1.0000
NO. TREES 2          .9287      1.0000
BAGR      3          .9222       .9720      1.0000
```

SUBROUTINE COVMAT N = 21 M = 3

	BASAL AREA	NO. TREES	BAGR
SUM	1.75810E+03	1.67817E+03	1.21426E+02
MEAN	8.37192E+01	7.99128E+01	5.78221E+00
VAR	3.48194E+03	2.33400E+03	1.29482E+01
CV	7.04832E+01	6.04553E+01	6.22317E+01

COVARIANCE MATRIX

	BASAL AREA 1	NO. TREES 2	BAGR 3
BASAL AREA 1	3.48194E+03		
NO. TREES 2	2.00754E+03	2.33400E+03	
BAGR 3	1.44168E+02	1.51571E+02	1.29482E+01

CORRELATION MATRIX

	BASAL AREA 1	NO. TREES 2	BAGR 3
BASAL AREA 1	1.0000		
NO. TREES 2	.7042	1.0000	
BAGR 3	.6790	.8719	1.0000

NEW CYC[illegible]...TCC 7

SUBROUTINE COVMAT N = 70 M = 3

	BASAL AREA	NO. TREES	BAGR
SUM	7.52329E+03	5.28012E+03	4.90954E+02
MEAN	1.07476E+02	7.54303E+01	7.01363E+00
VAR	4.19527E+03	2.43081E+03	1.53319E+01
CV	6.02657E+01	6.53627E+01	5.58284E+01

COVARIANCE MATRIX

	BASAL AREA 1	NO. TREES 2	BAGR 3
BASAL AREA 1	4.19527E+03		
NO. TREES 2	1.99637E+03	2.43081E+03	
BAGR 3	1.67625E+02	1.53224E+02	1.53319E+01

CORRELATION MATRIX

	BASAL AREA 1	NO. TREES 2	BAGR 3
BASAL AREA 1	1.0000		
NO. TREES 2	.6252	1.0000	
BAGR 3	.6609	.7957	1.0000

```
SUBROUTINE CCVMAT             N =    58      M =      3

              BASAL AREA  NO. TREES    BAGR
SUM           7.61380E+03 3.75318E+03 4.57008E+02
MEAN          1.31272E+02 6.47100E+01 7.87945E+00
VAR           7.72555E+03 2.45028E+03 4.56474E+01
CV            6.69563E+01 7.64956E+01 8.57457E+01


COVARIANCE MATRIX

                    BASAL AREA  NO. TREES    BAGR
                        1           2          3
BASAL AREA1 7.72555E+03
NO. TREES 2 2.62456E+03 2.45028E+03
BAGR      3 4.61346E+02 2.57108E+02 4.56474E+01


CORRELATION MATRIX

                   BASAL AREA  NO. TREES    BAGR
                       1           2          3
BASAL AREA1         1.0000
NO. TREES 2          .6032       1.0000
BAGR      3          .7769        .8684      1.0000
```

```
NEW CY' '...TCC  9


SUBROUTINE CCVMAT             N =     7      M =      3

              BASAL AREA  NO. TREES    BAGR
SUM           8.95269E+02 4.76321E+02 6.19120E+01
MEAN          1.27896E+02 6.80459E+01 8.84458E+00
VAR           3.33567E+03 1.77853E+03 2.15615E+01
CV            4.51581E+01 6.19767E+01 5.25004E+01


COVARIANCE MATRIX

                    BASAL AREA  NO. TREES    BAGR
                        1           2          3
BASAL AREA1 3.33567E+03
NO. TREES 2 1.57507E+03 1.77853E+03
BAGR      3 2.43058E+02 1.50116E+02 2.15615E+01


CORRELATION MATRIX

                   BASAL AREA  NO. TREES    BAGR
                       1           2          3
BASAL AREA1         1.0000
NO. TREES 2          .6467       1.0000
BAGR      3          .9063        .7666      1.0000
```

```
SUBROUTINE COVMAT            N =   61      M =    3

               BASAL AREA  NO. TREES   BAGR
SUM            2.47052E+03 1.35790E+03 1.47606E+02
MEAN           4.05004E+01 2.22607E+01 2.41976E+00
VAR            1.46910E+03 8.90614E+02 7.99093E+00
CV             9.46381E+01 1.34062E+02 1.16822E+02

COVARIANCE MATRIX

                   BASAL AREA  NO. TREES   BAGR
                        1          2           3
BASAL AREA1 1.46910E+03
NO. TREES 2 5.22672E+02 8.90614E+02
BAGR      3 6.80274E+01 7.54128E+01 7.99093E+00

CORRELATION MATRIX

                  BASAL AREA  NO. TREES   BAGR
                       1          2          3
BASAL AREA1         1.0000
NO. TREES 2          .4569       1.0000
BAGR      3          .6279        .8939      1.0000
```

```
NEW CYCLE...TCC 11

SUBROUTINE COVMAT            N =   14      M =    3

               BASAL AREA  NO. TREES   BAGR
SUM            7.56274E+02 1.06543E+03 9.45534E+01
MEAN           5.40196E+01 7.61022E+01 6.75381E+00
VAR            8.04485E+02 2.28589E+03 2.02639E+01
CV             5.25059E+01 6.28247E+01 6.66519E+01

COVARIANCE MATRIX

                   BASAL AREA  NO. TREES   BAGR
                        1          2           3
BASAL AREA1 8.04485E+02
NO. TREES 2 8.42558E+02 2.28589E+03
BAGR      3 9.14871E+01 2.00629E+02 2.02639E+01

CORRELATION MATRIX

                  BASAL AREA  NO. TREES   BAGR
                       1          2          3
BASAL AREA1         1.0000
NO. TREES 2          .6213       1.0000
BAGR      3          .7165        .9322      1.0000
```

SUBROUTINE COVMAT N = 43 M = 3

	BASAL AREA	NO. TREES	BAGR
SUM	3.85850E+03	3.63435E+03	3.05035E+02
MEAN	8.97325E+01	8.45199E+01	7.09383E+00
VAR	2.20578E+03	1.16506E+03	8.67339E+00
CV	5.23397E+01	4.03846E+01	4.15158E+01

COVARIANCE MATRIX

		BASAL AREA 1	NO. TREES 2	BAGR 3
BASAL AREA	1	2.20578E+03		
NO. TREES	2	7.99784E+02	1.16506E+03	
BAGR	3	8.15627E+01	2.64883E+01	8.67339E+00

CORRELATION MATRIX

		BASAL AREA 1	NO. TREES 2	BAGR 3
BASAL AREA	1	1.0000		
NO. TREES	2	.4939	1.0000	
BAGR	3	.5957	.8644	1.0000

NEW CYC[illegible]..TCC 13

SUBROUTINE COVMAT N = 19 M = 3

	BASAL AREA	NO. TREES	BAGR
SUM	1.83331E+03	1.38663E+03	1.08366E+02
MEAN	9.64859E+01	7.29008E+01	5.70345E+00
VAR	2.31504E+03	1.81400E+03	1.53938E+01
CV	4.98652E+01	5.83594E+01	6.87916E+01

COVARIANCE MATRIX

		BASAL AREA 1	NO. TREES 2	BAGR 3
BASAL AREA	1	2.31504E+03		
NO. TREES	2	1.17663E+03	1.81400E+03	
BAGR	3	1.36868E+02	1.33869E+02	1.53938E+01

CORRELATION MATRIX

		BASAL AREA 1	NO. TREES 2	BAGR 3
BASAL AREA	1	1.0000		
NO. TREES	2	.5742	1.0000	
BAGR	3	.7250	.8011	1.0000

SUBROUTINE COVMAT N = 49 M = 3

	BASAL AREA	NO. TREES	BAGR
SUM	1.13662E+03	9.46532E+02	8.88958E+01
MEAN	2.31963E+01	1.93170E+01	1.81420E+00
VAR	5.15044E+02	6.87027E+02	4.75871E+00
CV	9.78373E+01	1.35690E+02	1.20243E+02

COVARIANCE MATRIX

		BASAL AREA 1	NO. TREES 2	BAGR 3
BASAL AREA	1	5.15044E+02		
NO. TREES	2	4.51626E+02	6.87027E+02	
BAGR	3	3.58833E+01	5.29084E+01	4.75871E+00

CORRELATION MATRIX

		BASAL AREA 1	NO. TREES 2	BAGR 3
BASAL AREA	1	1.0000		
NO. TREES	2	.7552	1.0000	
BAGR	3	.7246	.9253	1.0000

NEW CYCLE...TCC 16

SUBROUTINE COVMAT N = 20 M = 3

	BASAL AREA	NO. TREES	BAGR
SUM	5.70427E+02	3.52741E+02	3.56983E+01
MEAN	2.85214E+01	1.76370E+01	1.78492E+00
VAR	4.79377E+02	4.33290E+02	4.37599E+00
CV	7.67659E+01	1.18022E+02	1.17198E+02

COVARIANCE MATRIX

		BASAL AREA 1	NO. TREES 2	BAGR 3
BASAL AREA	1	4.79377E+02		
NO. TREES	2	3.32868E+02	4.33290E+02	
BAGR	3	3.31351E+01	4.22859E+01	4.37599E+00

CORRELATION MATRIX

		BASAL AREA 1	NO. TREES 2	BAGR 3
BASAL AREA	1	1.0000		
NO. TREES	2	.7304	1.0000	
BAGR	3	.7235	.9711	1.0000

Sampling Simulation Results for Two Populations

Results of a sampling simulation using three variations of probability proportional to size sampling are summarized here for two small finite populations. Population one has 115 elements and population two has 12 elements.

Notation

N = population size

R = Number of replications of the sampling procedure

$$Y = \sum_{i=1}^{N} Y_i$$

$\hat{Y}$ = estimate of Y, depends on sampling procedure and may be one of the following:

$$\hat{Y}_{PPSOR} = \frac{1}{n} \sum_{i=1}^{n} t_i \quad \text{where } t_i = y_1/p_1$$

$$\text{and } t_i = \sum_{j=1}^{k-1} y_i + y_i \frac{1 - \sum_{j=1}^{k-1} p_i}{p_i}, \; k = 2,\ldots,n$$

$$\hat{Y}_{PPS/WR} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p_i}$$

$$\hat{Y}_{PPS/WOR} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p_i}$$

where PPSOR = Sampling with probability proportional to size for the remaining elements in the population (Raj. 1973, p. 57)
PPS/WR = Sampling with probability proportional to size, with replacement
PPS/WOR = Sampling with probability proportional to size, without replacement

$$E(\hat{Y}) = \frac{1}{R} \sum_{j=1}^{R} Y_j, \text{ where } \hat{Y} \text{ is one of the estimators given above}$$

$$MIN(\hat{Y}) = \min(\hat{Y}_j, j = 1,\ldots,R)$$

$$MAX(\hat{Y}) = \max(\hat{Y}_j, j = 1,\ldots,R)$$

PROP±1 SE = proportion of the R estimates $\hat{Y}$ which fall within ± 1 standard error of the estimate

PROP±2 SE = proportion of the R estimates $\hat{Y}$ which fall within ± 2 standard error

$V(\hat{Y})$ = population variance of $\hat{Y}$, depends on the sampling procedure and may be one of the following:

$$V(\hat{Y}_{PPSOR}) = \frac{1}{m\,(m-1)} \sum_{i=1}^{m} (t_i - \hat{Y}_{PPSOR})^2$$

$$V(\hat{Y}_{PPS/WR}) = \frac{1}{m\,(m-1)} \sum_{i=1}^{m} \left(\frac{y_i}{p_i} - \hat{Y}_{PPS/WR}\right)^2$$

$$V(\hat{Y}_{PPS/WOR}) = V\,(\hat{Y}_{PPS/WR})$$

$$E(\hat{V}(\hat{Y})) = \frac{1}{R} \sum_{j=1}^{R} \hat{V}(\hat{Y}_j)$$

$$MIN\ (\hat{V}(\hat{Y})) = \min\ (\hat{V}(\hat{Y}_j),\ j = 1,\ldots,R)$$

$$MAX\ (\hat{V}(\hat{Y})) = \max\ (\hat{V}(\hat{Y}_j),\ j = 1,\ldots,R)$$

$$MINRSE = \min\ (RSE = \frac{\hat{Y}_j}{\sqrt{\hat{V}(\hat{Y}_j)}},\ j = 1,\ldots,R)$$

$$MAXRSE = \max\ (RSE_j\ ,\ j = 1,\ldots,R)$$

$$CV\ (V(\hat{V}(\hat{Y})) = \frac{\sqrt{V(\hat{V})\,(\hat{Y})}}{E(Y)} \quad \text{where } V(\hat{V}(\hat{Y})) = \frac{1}{R-1}\sum \left[V(Y) - \frac{1}{R}\sum^{R} V(Y)\right]^2$$

RUN = computer run identifier

SAMPLING WITHOUT REPLACEMENT

Population #1

NUMBER OF REPLICATES

200

NUMBER OF UNITS IN POPULATION

115

Table G-1. Population #1 data, population size is 115 units

X = AUXILIARY VARIABLE
Y = VARIABLE OF INTEREST
P = SELECTION PROBABILITY OBTAINED FROM X

	UNIT ID	X	Y	P
1	8	24.2000	17.2800	.0009
2	9	119.1000	53.7600	.0046
3	10	167.9000	86.4000	.0065
4	11	7.7000	5.7600	.0003
5	14	59.8000	36.4800	.0023
6	15	231.0000	101.7600	.0090
7	16	349.7000	149.7600	.0136
8	20	85.6000	63.3600	.0033
9	21	278.7000	132.4800	.0108
10	22	358.0000	153.6000	.0139
11	26	97.2000	67.2000	.0038
12	27	508.0000	192.0000	.0197
13	28	287.2000	138.2400	.0111
14	32	132.9000	97.9200	.0052
15	33	333.1000	157.4400	.0129
16	34	188.0000	111.3600	.0073
17	38	108.8000	115.2000	.0042
18	39	49.6000	46.0800	.0019
19	40	33.6000	23.0400	.0013
20	44	69.1000	90.2400	.0027
21	50	42.2000	61.4400	.0016
22	56	42.2000	61.4400	.0016
23	62	42.2000	61.4400	.0016
24	68	24.0000	38.4000	.0009
25	74	18.6000	30.7200	.0007
26	80	65.2000	44.1600	.0025
27	81	33.0000	19.2000	.0013
28	86	198.8000	76.8000	.0077
29	87	169.9000	71.0400	.0066
30	92	233.8000	92.1600	.0091
31	93	154.3000	65.2800	.0060
32	98	354.0000	136.3200	.0137
33	99	362.0000	99.8400	.0140
34	100	184.0000	61.4400	.0071
35	104	460.0000	153.6000	.0178
36	105	442.0000	117.1200	.0171
37	106	162.4000	59.5200	.0063
38	102	19.2000	7.6800	.0007
39	108	74.5000	26.8800	.0029
40	110	428.0000	[illegible]	[illegible]
41	111	502.0000	130.5600	.0195
42	112	107.5000	53.7600	.0042
43	114	136.0000	32.6400	.0053
44	116	432.0000	155.5200	.0167
45	117	610.0000	182.4000	.0236
46	118	121.0000	74.8800	.0047
47	120	154.3000	63.3600	.0060
48	122	422.0000	157.4400	.0164
49	123	710.0000	186.2400	.0275
50	124	138.6000	63.3600	.0054
51	126	150.0000	48.0000	.0058
52	128	145.2000	71.0400	.0056
53	129	732.0000	192.0000	.0284
54	130	130.2000	59.5200	.0050
55	132	17.9000	9.6000	.0007
56	134	109.2000	51.8400	.0042
57	135	734.0000	192.0000	.0285
58	136	133.4000	44.1600	.0052
59	140	137.8000	57.6000	.0053
60	141	94.0000	192.0000	.0036
61	142	139.6000	63.3600	.0054
62	146	169.0000	61.4400	.0066
63	147	728.0000	192.0000	.0282
64	148	526.0000	165.1200	.0204
65	149	29.4000	17.2800	.0011
66	152	145.8000	55.6800	.0057
67	153	758.0000	192.0000	.0294
68	154	504.0000	161.2800	.0195
69	155	368.7000	117.1200	.0143
70	158	382.7000	161.2800	.0148
71	159	534.0000	145.9200	.0207
72	160	342.3000	124.8000	.0133
73	161	313.6000	101.7600	.0122
74	164	562.0000	192.0000	.0218
75	165	416.5000	134.4000	.0161
76	166	278.0000	97.9200	.0108
77	167	229.1000	80.6400	.0089
78	170	378.0000	192.0000	.0147
79	171	357.0000	134.4000	.0138
80	172	154.1000	74.8800	.0060
81	173	240.5000	84.4800	.0093
82	176	352.0000	126.7200	.0136
83	177	409.0000	132.4800	.0159
84	178	318.3000	124.8000	.0123
85	179	3.0000	1.9200	.0001
86	182	122.8000	72.9600	.0048
87	183	379.7000	130.5600	.0147
88	184	334.6000	119.0400	.0130
89	188	184.1000	107.5200	.0071
90	189	110.1000	53.7600	.0043
91	190	139.4000	63.3600	.0054
92	194	54.6000	57.6000	.0021
93	195	52.4000	26.8800	.0020
94	196	78.1000	40.3200	.0030
95	200	64.8000	51.8400	.0025
96	201	252.0000	71.0400	.0098
97	202	260.5000	97.9200	.0101
98	206	112.8000	76.8000	.0044
99	207	301.8000	97.9200	.0117
100	208	178.2000	67.2000	.0069
101	212	169.4000	99.8400	.006
102	213	283.4000	71.0400	.011
103	214	39.6000	19.2000	.001
104	218	212.2000	107.5200	.008
105	219	130.7000	55.6800	.005
106	220	6.1000	3.8400	.000
107	224	81.3000	34.5600	.003
108	225	59.3000	99.8400	.002
109	230	49.5000	24.9600	.001
110	231	205.6000	94.0800	.008
111	236	51.8000	28.8000	.002
112	237	193.6000	105.6000	.007
113	242	12.7000	11.5200	.000
114	243	250.5000	90.2400	.009
115	249	66.4000	38.4000	.002
	TOTAL	25793.2000	10250.8800	1.000
	AVERAGE	224.2887	89.1381	.008

ORIGINAL PAGE IS OF POOR QUALITY

SAMPLING WITH REPLACEMENT

NUMBER OF REPLICATES

Population #2

2000

NUMBER OF UNITS IN POPULATION

12

X = AUXILIARY VARIABLE
Y = VARIABLE OF INTEREST
P = SELECTION PROBABILITY OBTAINED FROM X

	UNIT ID	X	Y	P
1	PGE 1-16	18.4000	42.7098	.1258
2	PGE 1-19	15.7000	17.7148	.1073
3	PGE 1-22	9.2000	27.9630	.0629
4	PGE 2-7	6.5000	35.8611	.0444
5	PGE 2-19	1.6000	9.5904	.0109
6	PGE 2-31	13.8000	36.5842	.0943
7	PGE 3-4	15.7000	65.3495	.1073
8	PGE 3-7	16.4000	58.1705	.1121
9	PGE 3-13	18.2000	70.6177	.1244
10	PGE 4-7	9.2000	32.9078	.0629
11	PGE 4-17	11.2000	11.5472	.0766
12	PGE 4-20	10.4000	18.8801	.0711
	TOTAL	146.3000	427.8961	1.0000
	AVERAGE	12.1917	35.6580	.0833

Table G-2. Population #2 data, population size is 12 units

POPULATION #1 N = 115 n = 15 200 REPLICATIONS

	PPSOR ESTIMATE SAMPLING WITHOUT REPLACEMENT			PPS ESTIMATE SAMPLING WITH REPLACEMENT		
	RN SEED			RN SEED		
	1	2	3	1	2	3
Y	10251	10251	10251	10251	10251	10251
E($\hat{Y}$)	10198	10290	10308	10195	10325	10357
MIN $\hat{Y}$	8404	8262	8330	7806	8187	8162
MAX $\hat{Y}$	15169	14559	14729	16065	15105	15172
PROP ±1SE	.7400	.8150	.7300	.7350	.7150	.7100
PROP ±2SE	.9700	.9600	.9400	.9500	.9550	.9500
V($\hat{Y}$)	1,460,354	1,350,601	1,765,733	1,857,933	1,700,433.4	1,812,955
E ($\hat{V}$($\hat{Y}$))	1,407,140	1,637,944	1,593,888	1,597,784	1,962,819	1,899,458
MIN $\hat{V}$($\hat{Y}$)	91388	103,398	81123	101,001	126,192	136,022
MAX $\hat{V}$($\hat{Y}$)	10,669,310	9,867,814	10,837,540	12,447,447	16,021,953	10,872,164
MIN RSE	.0353	.0336	.0327	.0367	.0385	.0411
MAX RSE	.2344	.2418	.2343	.2606	.2650	.2481
CV V($\hat{V}$($\hat{Y}$))	1.5070	1.2945	1.3850	1.6175	1.3340	1.3409
RUN	J1979CZ	03/19/76		FAPCAØ1	03/19/76	

Table G-3. Results of repeated sampling from Population #1 with a sample size of 15 using three different sampling procedures and three different random number sequences.

POPULATION #1

PPS Sampling Without Replacement

Random Number Seed

	Random Number Seed		
	1	2	3
Y	10,251	10,251	10,251
$E(\hat{Y})$	10,360	10,469	10,483
MIN $\hat{Y}$	8,469	8,393	8,463
MAX $\hat{Y}$	16,190	15,192	15,405
PROP $\pm$ 1 SE	.7450	.7550	.7250
PROP $\pm$ 2 SE	.9550	.9550	.9250
$V(\hat{Y})$	1,592,631	1,572,780	2,023,960
$E(\hat{V}(\hat{Y}))$	1,706,046	2,056,240	1,969,355
MINV$(\hat{Y})$	116,831	121,059	103,009
MAX $\hat{V}(\hat{Y})$	14,444,207	12,409,517	13,625,708
MIN RSE	.0401	.0361	.0369
MAX RSE	.2451	.2567	.2444
$CV(\hat{V}(\hat{Y}))$			

RUN FAPCA00 06/23/76

Table G-3, continued.

Table G-4. Results of repeated sampling from Population #2 with a sample size of 4 using three different sampling procedures and four different random number sequences.

POPULATION #2 N = 12 n = 4 2000 REPLICATIONS

	PPSOR ESTIMATE SAMPLING WITHOUT REPLACEMENT				PPS ESTIMATE SAMPLING WITH REPLACEMENT			
	RN SEED				RN SEED			
	1	2	3	4	1	2	3	4
$\hat{Y}$	427.90	427.90	427.90	427.90	427.90	427.90	427.90	427.90
E $(\hat{Y})$	427.32	426.87	428.21	425.35	427.40	426.46	427.80	425.30
MIN $\hat{Y}$	220.39	216.15	216.15	215.50	183.08	154.40	154.40	154.40
MAX $\hat{Y}$	702.89	701.17	679.43	666.28	747.28	716.53	715.17	742.94
PROP$\pm$1SE	.7520	.7350	.7465	.7445	.6905	.6715	.6820	.6860
PROP$\pm$2SE	.9830	.9785	.9825	.9810	.9625	.9540	.9585	.9515
V$(\hat{Y})$	6017	6272	6045	6015	7765	8234	7926	8026
E$(\hat{V}(\hat{Y}))$	6230	6017	6075	6247	8207	7987	8060	8348
MIN $\hat{V}(\hat{Y})$	234	239	235	251	0	0	0	2
MAX $\hat{V}(\hat{Y})$	27433	29019	30874	28154	35130	34355	32532	32532
MIN RSE	.0277	.0278	.0277	.0288	0	0	0	.0024
MAX RSE	.4132	.4605	.4113	.4183	.5188	.5116	.5369	.5369
CV$(\hat{V}(\hat{Y}))$	.6929	.6815	.6934	.6691	.7294	.7137	.7163	.7051
RUN	FAPCAØ4	03/19/76			FAPCAØ9	04/14/76		

POPULATION #2 N = 12 n = 4

2000 REPLICATIONS

	PPS ESTIMATE SAMPLING WITHOUT REPLACEMENT			
	RN SEED 1	2	3	4
Y	427.90	427.90	427.90	427.90
$E(\hat{Y})$	428.78	427.76	428.91	426.62
MIN $\hat{Y}$	230.27	230.27	230.27	230.27
MAX $\hat{Y}$	715.17	715.17	704.08	663.01
PROP $\pm$1SE	.7375	.7315	.7495	.7470
PROP $\pm$2SE	.9825	.9755	.9810	.9835
$V(\hat{Y})$	6112	6345	6100	5980
$E(\hat{V}(\hat{Y}))$	9195	9029	9054	9323
MIN $\hat{V}(\hat{Y})$	448	448	448	448
MAX $\hat{V}(\hat{Y})$	34229	33354	39209	31518
MIN RSE	.0381	.0382	.0382	.0382
MAX RSE	.4477	.4736	.4736	.4477
CV$(\hat{V}(\hat{Y}))$	.6166	.6178	.6274	.6117
RUN				

Table G-4, continued.

APPENDIX H

Quincy Ranger District Planning Model Application Results

The planning model applications results consist of six tables H-1 through H-6 which show sampling system cost for various factors.

Table H-1. Allocation Results, Cell Entries Represent System Cost

Strategy = STR2SD   PSU Size = 60 X 6

PROB	Cost	CORREL	Seed	PRECISION			
				t = .100		t = 1.96	
				AE = .2	AE = .1	AE = .2	AE = .1
PPS	$C_5 = 175$	.8	1	3714	5747	5652	12044
			2	3718	5760	5665	12075
	$C_4 = 11$	.9	1	3589	5396	5309	11234
			2	3593	5404	5317	11261
	$C_5 = 100$	.8	1	2558	4467	4373	10428
			2	2560	4476	4380	10459
	$C_4 = 11$	.9	1	2455	4181	4095	9762
			2	2462	4190	4103	9790
	$C_5 = 0$	.8	1	3852			13646
			2				
	$C_4 = 111$	.9	1				
			2				
	$C_5 = 778$	.8	1	12429			23190
			2				
	$C_4 = 11$	.9	1				
			2				
EQUAL	$C_5 = 175$	.8	1	4570	8956	8728	23845
			2	4577	8978	8759	23895
	$C_4 = 11$	.9	1	4415	8560	8351	23031
			2	4421	8578	8369	23075
	$C_5 = 100$	.8	1	3430	7660	7446	22259
			2	3435	7679	7465	22305
	$C_4 = 11$	.9	1	3304	7340	7134	21594
			2	3309	7356	7150	21635
	$C_5 = 0$	.8	1	4722			25341
			2				
	$C_4 = 111$	.9	1				
			2				
	$C_5 = 778$	.8	1				
			2	13387			34944
	$C_4 = 11$	.9	1				
			2				

Table H-2. Allocation Results, Cell Entries Represent System Cost

STRATEGY = STR2SD PSU SIZE = 40 X 4

PROB	COST	CORREL	t = 1.00		t = 1.96	
			AE = .2	AE = .1	AE = .2	AE = .1
PPS	C_5 = 175	.8				
	C_4 = 11	.9				
	C_5 = 100	.8	2558			10749
	C_4 = 11	.9				
	C_5 = 0	.8	3793			13875
	C_4 = 111	.9				
	C_5 = 778	.8	12498			23486
	C_4 = 11	.9				
EQUAL	C_5 = 175	.8				
	C_4 = 11	.9				
	C_5 = 100	.8	3205			19697
	C_4 = 11	.9				
	C_5 = 0	.8	4500			22768
	C_4 = 111	.9				
	C_5 = 778	.8	13193			32385
	C_4 = 11	.9				

Table H-3. Allocation Results, Cell Entries Represent System Cost

STRATEGY = STR2SD PSU SIZE = 10 X 10

PROB	COST	CORREL	t = 1.00		t = 1.96	
			AE = .2	AE = .1	AE = .2	AE = .1
PPS	C_5 = 175	.8				
	C_4 = 11	.9				
	C_5 = 100	.8	2534			11166
	C_4 = 11	.9				
	C_5 = 0	.8	3758			14370
	C_4 = 111	.9				
	C_5 = 778	.8	12477			24075
	C_4 = 11	.9				
EQUAL	C_5 = 175	.8				
	C_4 = 11	.9				
	C_5 = 100	.8	3119			18869
	C_4 = 11	.9				
	C_5 = 0	.8	4402			22003
	C_4 = 111	.9				
	C_5 = 778	.8	13085			31711
	C_4 = 11	.9				

Table H-4. Allocation Results, Cell Entries Represent System Cost

STRATEGY = STRAT2S PSU SIZE = 60 X 6

PROB	COST	SEED	t = 1.00 AE = .2	t = 1.00 AE - .1	t = 1.96 AE = .2	t = 1.96 AE - .1
PPS	C_3 = 200 C_2 = 23	1 2	X 1542 X 1548	4782 4813	4620 4650	15537 15618
	C_3 = 125 C_2 = 23	1 2	1295	4256	4106	14426
	C_3 = 0 C_2 = 147	1 2	4479			67892
	C_3 = 575 C_2 = 23	1 2	2630			20505
EQUAL	C_3 = 200 C_2 = 23	1 2	3344 3360	9769 8820	9461 9510	28196 28342
	C_3 = 125 C_2 = 23	1 2	2718 2732	8446 8490	8166 8209	25593 25726
	C_3 = 0 C_2 = 147	1 2	6729			90890
	C_3 = 575 C_2 = 23	1 2	6191			40249

Table H-5. Allocation Results, Cell Entries Represent System Cost

STRATEGY = STRAT2S PSU SIZE = 40 X 4

PROB	COST	t = 1.00		t = 1.96	
		AE = .2	AE - .1	AE = .2	AE - .1
PPS	C_3 = 200 C_2 = 23				
PPS	C_3 = 125 C_2 = 23	1362			14929
PPS	C_3 = 0 C_2 = 147	4461			67953
PPS	C_3 = 575 C_2 = 23	2826			21983
EQUAL	C_3 = 200 C_2 = 23				
EQUAL	C_3 = 125 C_2 = 23	2489			24887
EQUAL	C_3 = 0 C_2 = 147	6200			90404
EQUAL	C_3 = 575 C_2 = 23	5713			39250

Table H-6. Allocation Results, Cell Entries Represent System Cost

STRATEGY = STR2SD PSU SIZE = 10 x 10

PROB	COST	t = 1.00		t = 1.96	
		AE = .2	AE - .1	AE = .2	AE - .1
PPS	C_3 = 200				
	C_2 = 23				
	C_3 = 125	1415			15274
	C_2 = 23				
	C_3 = 0	4404			67138
	C_2 = 147				
	C_3 = 575	3027			22918
	C_2 = 23				
EQUAL	C_3 = 200 C_2 = 23				
	C_3 = 125 C_2 = 23	2440			24288
	C_3 = 0 C_2 = 147	5879			87251
	C_3 = 575 C_2 = 23	5729			39129

APPENDIX I

Planning Model Example Output

The output reproduced here represents a typical detailed output summary for a single model run.

The 59 constraints may be evaluated as follows:

1-8	max n^*
9-16	max n'
17-24	max n
25-27	variance constraint
28-35	min n^*
36-43	min n'
44-51	min n
52-59	minimum degrees of freedom constraint.

The current value of x entries represent $1/n^*$, $1/n'$, $1/n$ except for entries under SUBROUTINE SUMRIZE which show n^*, n', n. The sequence of entries is

$$\left.\begin{array}{l}(n^*_h,\ h=1,\ \ldots,8)\\(n'_h,\ h=1,\ \ldots,8)\\(n_h,\ h=1,\ldots,8)\end{array}\right\}\ \text{24 total values of "X"}$$

```
SUBR  TINE READIN  --  STRATIFIED TWO STAGE WITH DOUBLE SA   E ALLOCATION

NCLS =   17            NPAR =    3             X =   40   Y =    4
READIN PRINTON        T =     1.9600    RHO =       .8000
 20.   7.  15.  11. 778.
CLASS NAME,COUNT,(AMEAN(I),I=1,NPAR),COVMATRIX LOWER TRIANGLE
ONE          0/      0.00       0.00       0.00/ 0.          0.          0.          0.          0.          0.
                                                 0.          0.          0.          0.          0.          0.
TWO          0/      0.00       0.00       0.00/ 0.          0.          0.          0.          0.          0.
                                                 0.          0.          0.          0.          0.          0.
THREE       52/  12223.55    8578.22     904.49/ .34265E+08  .24210E+08  .25725E+07  .17175E+08  .18236E+07  .19430E+06
                                                 .96837E+08  .63205E+08  .65245E+07  .52311E+08  .50634E+07  .54579E+06
             0/      0.00       0.00       0.00/ 0.          0.          0.          0.          0.          0.
                                                 0.          0.          0.          0.          0.          0.
FIVE         9/   8379.87    5835.93     561.00/ .93557E+07  .64332E+07  .66109E+06  .45181E+07  .45883E+06  .47747E+05
                                                 .97248E+08  .60983E+08  .59453E+07  .49515E+08  .44292E+07  .45776E+06
SIX         56/   8371.96    6300.73     538.28/ .14470E+08  .11007E+08  .92974E+06  .80443E+07  .71848E+06  .60709E+05
                                                 .74584E+08  .39126E+08  .38752E+07  .40068E+08  .30271E+07  .31320E+06
SEVEN      707/  10337.71    6775.31     677.81/ .18667E+08  .11883E+08  .12326E+07  .77810E+07  .79540E+06  .82820E+05
                                                 .94595E+08  .50042E+08  .53864E+07  .45535E+08  .40059E+07  .44753E+06
EIGHT       22/   8072.64    5126.53     519.47/ .23026E+08  .14273E+08  .14538E+07  .09774E+07  .90965E+06  .93082E+05
                                                 .65330E+09  .28404E+08  .35594E+07  .27538E+08  .24928E+07  .31031E+06
             0/      0.00       0.00       0.00/ 0.          0.          0.          0.          0.          0.
                                                 0.          0.          0.          0.          0.          0.
TEN         16/   2838.34    1841.53     182.71/ .11758E+09  .74570E+07  .75285E+06  .49176E+07  .49540E+06  .50025E+05

                                                 .24010E+08  .13771E+08  .13992E+07  .12537E+08  .10931E+07  .11233E+06
             0/      0.00       0.00       0.00/ 0.          0.          0.          0.          0.          0.
                                                 0.          0.          0.          0.          0.          0.
             0/      0.00       0.00       0.00/ 0.          0.          0.          0.          0.          0.
                                                 0.          0.          0.          0.          0.          0.
             0/      0.00       0.00       0.00/ 0.          0.          0.          0.          0.          0.
                                                 0.          0.          0.          0.          0.          0.
             0/      0.00       0.00       0.00/ 0.          0.          0.          0.          0.          0.
                                                 0.          0.          0.          0.          0.          0.
             0/      0.00       0.00       0.00/ 0.          0.          0.          0.          0.          0.
                                                 0.          0.          0.          0.          0.          0.
SIXTEEN    212/   7136.99    4566.44     463.04/ .14193E+08  .89176E+07  .91813E+06  .57158E+07  .56457E+06  .60378E+05
                                                 .67276E+08  .36941E+08  .39952E+07  .32646E+08  .29945E+07  .33282E+06
SEVENTEEN  428/   4988.34    3516.98     324.02/ .86971E+07  .60191E+07  .56260E+06  .43020E+07  .39426E+06  .37034E+05
                                                 .58970E+08  .34605E+08  .33420E+07  .31358E+08  .25854E+07  .25930E+06
ALLOWABLE ERROR(I),I=1,NPAR
        1         .100
        2         .100
        3         .100
COMMON BLOCK STRAT -- SUMT INPUTS
NPINV(1) ,CH(3),D(3),B(3),MHINV,NSTR
 .19231E-01  .94265E+11  .50922E+11  .53129E+09  .92654E+11  .46442E+11  .52539E+09  .26185E+12  .14145E+12  .14793E+10
 .11111E+00  .28358E+10  .14439E+10  .13348E+08  .75862E+09  .36597E+09  .38675E+07  .78771E+10  .40107E+10  .37079E+08
 .17857E-01  .84203E+11  .45235E+11  .35359E+09  .45407E+11  .27109E+11  .19038E+09  .23390E+12  .12565E+12  .98220E+09
 .14144E-02  .17022E+14  .81938E+13  .80532E+11  .93300E+13  .38893E+13  .41417E+11  .47283E+14  .22760E+14  .22370E+12
 .45455E-01  .11383E+11  .47982E+10  .54069E+08  .11145E+11  .43451E+10  .45052E+08  .31620E+11  .13328E+11  .15017E+09
 .62500E-01  .22127E+10  .11555E+10  .10052E+08  .29077E+10  .12589E+10  .12806E+08  .61465E+10  .32096E+10  .28759E+08
 .47170E-02  .10085E+13  .59821E+12  .53850E+10  .63791E+12  .25689E+12  .27136E+10  .30237E+13  .14672E+13  .14958E+11
 .23364E-02  .38889E+13  .20679E+13  .17004E+11  .15932E+13  .78806E+12  .67786E+10  .10802E+14  .57442E+13  .47511E+11
 .62500E-02
       2
V(I),I=1,NPAR
 .93000E+06  .42120E+06  .41461E+05
```

```
NONLINEAR PROGRAMMING ROUTINE-SUMT VERSION 4  8/10/71

    N= 24       M= 59       MZ=  0

      MAX. TIME=  .2000000E+02     R=       1.0000000     RATIO=  .1600000E+02      EPSILON=  .1000000E-01     THETA=  .1000000E-0

OPTIONS SELECTED

    2      1      1      1      1      1      1     -0      1      1

TOLERANCES
     .1000000E-03          .1000000E-03

SECOND SET OF OPTIONS
    1       1
    TIME=       .018 SECONDS
     F=    .2146156E+07       P=  0.                 G=  0.                  RSIGMA=  0.                      H=  0.
    THE CURRENT VALUE OF X IS
     .2115385E-01        .1222222E+00        .1964286E-01        .1555870E-02        .5000000E-01        .6875000E-01
     .2188679E-02        .2570093E-02        .6875000E-02        .6875000E-02        .6875000E-02        .6875000E-02
     .6875000E-02        .6875000E-02        .6875000E-02        .6875000E-02        .4125000E-01        .4125000E-01

     .4125000E-01        .4125000E-01        .4125000E-01        .4125000E-01        .4125000E-01        .4125000E-01
    THE CONSTRAINT VALUES
                         NOT INCLUDING THE NON-NEGATIVITIES
     .1923077E-02        .1111111E-01        .1785714E-02        .1414427E-03        .4545455E-02        .6250000E-02
     .4716981E-03        .2336449E-03        .6250000E-03        .6250000E-03        .6250000E-03        .6250000E-03
     .6250000E-03        .6250000E-03        .6250000E-03        .6250000E-03        .3437500E-01        .3437500E-01
     .3437500E-01        .3437500E-01        .3437500E-01        .3437500E-01        .3437500E-01        .3437500E-01
     .3719835E+12        .1661628E+12        .1604191E+10        .9997885E+02        .9987778E+02        .9998030E+02
     .9999844E+02        .9995000E+02        .9993125E+02        .9999481E+02        .9999743E+02        .9999312E+02
     .9999312E+02        .9999312E+02        .9999312E+02        .9999312E+02        .9999312E+02        .9999312E+02

     .9999312E+02        .9995875E+02        .9995875E+02        .9995875E+02        .9995875E+02        .9995875E+02
     .9995875E+02        .9995875E+02        .9995875E+02        .9973822E+00        .9848750E+00        .9975692E+00
     .9998475E+00        .9938125E+00        .9914922E+00        .9993579E+00        .9996820E+00
    TIME=       .023 SECONDS

*****THE FEASIBLE STARTING POINT TO BE USED IS ...
     F=    .2146156E+07       P=  0.                 G=  0.                  RSIGMA=  0.                      H=  0.
    THE CURRENT VALUE OF X IS
     .2115385E-01        .1222222E+00        .1964286E-01        .1555870E-02        .5000000E-01        .6875000E-01
     .2188679E-02        .2570093E-02        .6875000E-02        .6875000E-02        .6875000E-02        .6875000E-02
     .6875000E-02        .6875000E-02        .6875000E-02        .6875000E-02        .4125000E-01        .4125000E-01
     .4125000E-01        .4125000E-01        .4125000E-01        .4125000E-01        .4125000E-01        .4125000E-01
    THE CONSTRAINT VALUES
                         NOT INCLUDING THE NON-NEGATIVITIES
     .1923077E-02        .1111111E-01        .1785714E-02        .1414427E-03        .4545455E-02        .6250000E-02
     .4716981E-03        .2336449E-03        .6250000E-03        .6250000E-03        .6250000E-03        .6250000E-03
     .6250000E-03        .6250000E-03        .6250000E-03        .6250000E-03        .3437500E-01        .3437500E-01
     .3437500E-01        .3437500E-01        .3437500E-01        .3437500E-01        .3437500E-01        .3437500E-01
     .3719363E+12        .1661628E+12        .1604191E+10        .9997885E+02        .9987778E+02        .9998030E+02
     .9999844E+02        .9995000E+02        .9993125E+02        .9999481E+02        .9999743E+02        .9999312E+02
     .9999312E+02        .9999312E+02        .9999312E+02        .9999312E+02        .9999312E+02        .9999312E+02
     .9999312E+02        .9995875E+02        .9995875E+02        .9995875E+02        .9995875E+02        .9995875E+02
     .9995875E+02        .9995875E+02        .9995875E+02        .9973822E+00        .9848750E+00        .9975692E+00

     .9998475E+00        .9938125E+00        .9914922E+00        .9993579E+00        .9996820E+00
```

ORIGINAL PAGE IS OF POOR QUALITY

TIME= 11.970 SECONDS

****

POINT= 194 DOTT= .8469036E-02 RHO= .8260296E+00 MAGNITUDE= .1768893E+04 PHASE= 2

F= .3238505E+05 P= .3231896E+05 G= .3231649E+05 RSIGMA= -.8512222E+02 H= 0.

THE CURRENT VALUE OF X IS

.8217717E-01	.9180437E+00	.1030874E+00	.1001082E-01	.2485985E+00	.6322822E+00
.3515557E-01	.2105114E-01	.2719671E+00	.3620364E+00	.2372867E+00	.1711030E+00
.3438259E+00	.5256335E+00	.1927707E+00	.1640564E+00	.3155813E+01	.3620890E+00
.2496088E+01	.1073897E+02	.1188245E+01	.5257749E+00	.5474022E+01	.6883091E+01

THE CONSTRAINT VALUES NOT INCLUDING THE NON-NEGATIVITIES

.6294640E-01	.8069726E+00	.8523027E-01	.8590391E-02	.2031440E+00	.5697822E+00
.3043858E-01	.1871469E-01	.2657171E+00	.3557864E+00	.2310367E+00	.1648230E+00
.3375759E+00	.5193835E+00	.1865407E+00	.1584064E+00	.2883846E+01	.5319307E-04
.2259401E+01	.1056787E+02	.8444194E+00	.1413987E-03	.5281231E+01	.6718432E+01
.2203479E+11	.1193387E+08	.4230790E+06	.9991782E+02	.9908192E+02	.9989691E+02
.9998999E+02	.9975140E+02	.9936771E+02	.9996484E+02	.9997895E+02	.9972803E+02
.9963796E+02	.9976271E+02	.9982890E+02	.9965617E+02	.9947437E+02	.9980721E+02
.9933534E+02	.9684419E+02	.9963791E+02	.9750331E+02	.8926103E+02	.9861172E+02
.9947423E+02	.9452598E+02	.9311691E+02	.2219928E+00	.2714483E-02	.2278689E+00
.6774624E+00	.1138120E+00	.2680579E-02	.4220730E+00	.5653092E+00	

2ND ORDER ESTIMATES

F= .3238100E+05 P= .3239718E+05 G= .3237788E+05 RSIGMA= -.8515222E+02 H= 0.

THE CURRENT VALUE OF X IS

.8232298E-01	.9268051E+00	.1031171E+00	.1001878E-01	.2492121E+00	.6413809E+00
.3519559E-01	.2105047E-01	.2709322E+00	.3603742E+00	.2363255E+00	.1709624E+00
.3402855E+00	.5231802E+00	.1924895E+00	.1644756E+00	.3150335E+01	.3594696E+00
.2495630E+01	.1072326E+02	.1185599E+01	.5183265E+00	.5466296E+01	.6876909E+01

THE CONSTRAINT VALUES NOT INCLUDING THE NON-NEGATIVITIES

.6309221E-01	.8156940E+00	.8525991E-01	.8604352E-02	.2037576E+00	.5788809E+00
.3047861E-01	.1871402E-01	.2646822E+00	.3540242E+00	.2300755E+00	.1647124E+00
.3345355E+00	.5169302E+00	.1862395E+00	.1582256E+00	.2879403E+01	-.1404614E-02
.2259304E+01	.1055230E+02	.8447138E+00	-.4853710E-02	.5273806E+01	.6712433E+01
.2470089E+11	.1390844E+10	.1296068E+08	.9991768E+02	.9907319E+02	.9989688E+02
.9998998E+02	.9972079E+02	.9935862E+02	.9996480E+02	.9997895E+02	.9972907E+02
.9963913E+02	.9976367E+02	.9982904E+02	.9965011E+02	.9947682E+02	.9980751E+02
.9983552E+02	.9684967E+02	.9964053E+02	.9750437E+02	.8927674E+02	.9881440E+02
.9948167E+02	.9453370E+02	.9312309E+02	.2219651E+00	.5253716E-03	.2279740E+00
.6776981E+00	.1136028E+00	.2665928E-02	.4228315E+00	.5657135E+00	

1ST ORDER ESTIMATES

F= .3238201E+05 P= .3239822E+05 G= .3237788E+05 RSIGMA= -.8512222E+02 H= 0.

THE CURRENT VALUE OF X IS

.9231873E-01	.9264296E+00	.1031227E+00	.1001852E-01	.2492826E+00	.6409269E+00
.3519458E-01	.2105110E-01	.2709181E+00	.3600291E+00	.2362988E+00	.1709663E+00
.3408657E+00	.5231776E+00	.1924895E+00	.1644775E+00	.3150447E-01	.3594058E+00
.2495460E+01	.1072362E+02	.1185524E+01	.5184872E+00	.5466410E+01	.6876864E+01

THE CONSTRAINT VALUES
NOT INCLUDING THE NON-NEGATIVITIES

.6308796E-01	.8153185E+00	.5265510E-01	.8604096E-02	.2037480E+00	.5784269E+00
.3047760E-01	.1871465E-01	.2646681E+00	.3545791E+00	.2300488E+00	.1647163E+00
.3346157E+00	.5169276E+00	.1862395E+00	.1582275E+00	.2879529E-01	-.1423256E-02
.2259162E+01	.1055266E+02	.8446584E+00	-.4690422E-02	.5273920E+01	.6712339E+01
.2349345E+11	.7742884E+09	.7281074E+07	.9991768E+02	.9907357E+02	.9939688E+02
.9998998E+02	.9975080E+02	.9935907E+02	.9904481E+02	.9997895E+02	.9972908E+02
.9963917E+02	.9976370E+02	.9982903E+02	.9965913E+02	.9947682E+02	.9960751E+02
.9983552E+02	.9684955E+02	.9964050E+02	.9750454E+02	.8927638E-02	.9881446E+02
.9948151E+02	.9453359E+02	.9312316E+02	.2219777E+00	.1107355E-02	.2279875E+00
.6776954E+00	.1136931E+00	.3062922E-02	.4228360E+00	.5657033E+00	

LAGRANGE MULTIPLIERS

F= .3238201E+05 P= .3239822E+05 G= .3237788E+05 RSIGMA= -.8512222E+02 H= 0.

THE CURRENT VALUE OF X IS

1/m	.1005181E+02	.8997324E+00	.8012905E+01	.8251370E+02	.3322746E+01	.1306419E+01
	.2349641E+02	.3923918E+02	.3037241E+01	.2280621E+01	.3481145E+01	.3827675E+01
1/m'	.2402465E+01	.1571494E+01	.4284593E+01	.5011668E+01	.2617466E+00	.2281785E+01
1/m	.3308502E+00	.7691889E-01	.6951676E+00	.1571071E+01	.1509000E+00	.1200085E+00

THE CONSTRAINT VALUES
NOT INCLUDING THE NON-NEGATIVITIES

1	.1312275E+02	.1023616E+01	.9691740E+01	.9600828E+02	.4066228E+01	.1449720E+01
7	.2713758E+02	.4413802E+02	.3108681E+01	.2321701E+01	.3575317E+01	.5010705E+01
13	.2446945E+01	.1590404E+01	.4428140E+01	.5214624E+01	.2864334E+00	.5528905E+02
19	.3655967E+00	.7816427E-01	.9782221E+00	.5841847E+04	.1564085E+00	.1229497E+00
25	.3748752E-10	.6921722E-07	.1952424E-05	.8267098E-02	.8336835E-02	.8268820E-02
31	.8261123E-02	.8280887E-02	.8312857E-02	.8264201E-02	.8262036E-02	.8282823E-02
37	.8290310E-02	.8279944E-02	.8274654E-02	.8288795E-02	.8303945E-02	.8276225E-02
43	.8273920E-02	.8529470E-02	.8290315E-02	.8471811E-02	.9254090E-02	.8359629E-02
49	.8303956E-02	.8738652E-02	.8870886E-02	.3721975E+01	.3043046E+03	.3625025E+01
55	.1219264E+01	.7257843E+01	.3081534E+03	.1954300E+01	.1461200E+01	

SUBROUTINE SUMRIZE -- STRATIFIED TWO STAGE WITH DOUBLE SAMPLING

```
IX,IY,T,RHO,AERR(3)
      40           4        1.9600        .8000        .1000        .1000        .1000
      50.000          7.000         15.000         11.000         778.000

I,FTOT,VAR,AVAR,RSE,CIHWPC
      1    .123597E+08    .375617E+12    .612876E+06    .495867E+01    .971809E+01
      2    .825716E+07    .177469E+12    .421270E+06    .510187E+01    .999966E+01
      3    .812636E+06    .171859E+10    .414559E+05    .510141E+01    .999877E+01
```

STRATA	m^*	m'	m
1	12.169	3.677	.317
2	1.559	2.762	2.762
3	9.701	4.214	.401
4	99.892	2.844	.093
5	4.023	2.908	.842
6	1.902	1.902	1.902
7	28.445	2.187	.183
8	47.503	5.073	.145
TOTALS	204.453	1123.197	38.543

```
**********************************************
   POINT= 194      DOTT=  .8469036E-02      RHO=  .8260296E+00      MAGNITUDE=  .1768893E+04      PHASE= 2
 F=  .3235505E-05      P=  .3231896E+05      G=  .3231649E+05      RSIGMA= -.8515222E+02      H= 0.
THE CURRENT VALUE OF X IS
  .1216683E+02    .1089225E+01    .9700505E+01    .9989194E+02    .4022550E+01    .1581564E+01
  .2844500E+02    .4750335E+02    .3676916E+01    .2762153E+01    .4214310E+01    .5844434E+01
  .2938449E+01    .1902466E+01    .5186973E+01    .6073254E+01    .3168756E+00    .2761748E+01
  .4005306E+00    .9311880E-01    .8415771E+00    .1901955E+01    .1826810E+00    .1452839E+00
THE CONSTRAINT VALUES
                  NOT INCLUDING THE NON-NEGATIVITIES
  .6294640E-01    .8069726E+00    .8523027E-01    .8595391E-02    .2031440E+00    .5697855E+00
  .3043956E-01    .1871469E-01    .2657171E+00    .3557864E+00    .2310367E+00    .1648530E+00
  .3375759E+00    .5193835E+00    .1865407E+00    .1584064E+00    .2883846E+01    .5319307E-01
  .2259401E+01    .1056737E+02    .8444194E+00    .1413987E-03    .5281231E+01    .6718432E+01
  .2203479E+11    .1193387E+08    .4230790E+06    .9991782E+02    .9908192E+02    .9989691E+02
  .9994599E+02    .9975140E+02    .9936771E+02    .9996484E+02    .9997895E+02    .9972803E+02
  .9963795E+02    .9976271E+02    .9982890E+02    .9965617E+02    .9947437E+02    .9980721E+02
  .9983534E+02    .9684419E+02    .9963791E+02    .9751331E+02    .8926103E+02    .9881175E+02
  .9947423E+02    .9452598E+02    .9311691E+02    .2211928E+00    .2714483E-02    .2278680E+00
  .6774824E+00    .1138120E+00    .2680579E-02    .4226730E+00    .5653092E+00
```

Appendix J

Implementation Plan for the optimal Quincy Ranger District sampling system

1.0 Introduction

The plan presented here for the Quincy Ranger District is intended only as a brief summary of a plan which would be in considerably more detail were the system actually to be implemented. The intent of this summary version is to identify the major topics which must be addressed prior to conducting the sample survey itself.

2.0 Estimation objectives

2.1 Population specifications -- The sampling population is defined as all the land owned and administered by the U.S. Forest Service within the Quincy Ranger District of the Plumas National Forest.

2.2 parameter specifications -- Three major parameters are to be estimated by the sampling system: 1) number of trees, 5.0" DBH and larger coniferous species; 2) total basal area in square feet inside bark, trees 5.0" DBH and larger, coniferous species; and 3) total basal area growth, annual, in square feet inside bark, trees 5.0" DBH and larger, coniferous species. These are the parameters upon which design is based. Other parameters could be incorporated, if desired, but without any opportunity to control precision obtained.

2.3 Desired format for summarizing estimates -- A tabular output is required showing the estimate, the sample estimates of variance, the standard error and the relative standard error in per cent.

3.0 Summary of the planning phase

3.1 Data utilized -- Spatially referenced land use/vegetation class data for grid cells of roughly 1.1 acre were obtained from analysis of LANDSAT multispectral data. Supplementary data from large scale photographs and field samples were used to estimate the parameter set and its associated covariance matrix for each vegetation class. Ownership data was used to eliminate all non Forest Service lands. LANDSAT data was acquired in 1972, sample data in 1974, and ownership data is from 1971.

3.2 alternatives, considered and selection of optimum System -- The alternatives considered have been described previously in section 3.3.3, Application Results, of the main report. That material and the development of the best system would normally be included here.

Particular sources of auxiliary data are summarized here since they effect the implementation, especially measurement techniques. The stratification and selection probabilities use the cell-by-cell vegetation class assignments (see section 3.2.2.2). Correlation coefficient for the double sampling strategy is a reflection of the strength of the assumed linear regression relationship between a photo variable, x, and the variable of interest, y. In this application y is actually a vector consisting of three variables (number of trees, basal area, and basal area growth) the x vector would be the photo variable which best approximates these variables, or a proxy variable which is strongly related to the y variable. The correlation would also reflect the impact of other factors in the measurement processes including photo variables (scale, date, image quality, etc.) and interpretation variables such as the experience level of interpreters. In this application the same correlation is assumed for all variables.

4.0 Sampling System Description

4.1 Sampling model -- The population to be sampled is specified in 2.1. Sampling units are to be formed by partitioning the area into clusters of 60 x 6 pixels. The sampling frame is the set of sampling units defined by the partitioning of the population.

4.2 Parameters -- The three parameters are defined in 2.2.

4.3 Auxiliary data -- The land use/vegetation classification results are to be used for stratification purposes. Interpretation of large scale photographs provides supplementary estimates of basal area and number of trees. These variables have shown strong correlation with number of trees, basal area, and basal area growth as measured directly at field plots.

4.4 Sampling technique and estimators -- Stratification will be accomplished by assigning each sampling unit to a land use/vegetation class based on the class with a plurabity of "good" pixels within each unit. Specifically there are to be 8 strata as indicated in Table 12 of the text. The sampling system is to be two stage with double sampling within each stratum. At stage I n_h^* sample units will be selected in each stratum. Selection probabilities for stage I units are proportional to size (see text). Within each unit n_h' second stage sample units will be selected. They will consist of 0.4 acre circular plots located by systematic coverage of large scale photographs at a scale of 1:1,000.

Measurements of auxiliary variables (4.3) will be made for each selected plot. From the set of all photo plots n_h will be selected from each stratum for direct measurement of the parameter of interest.

4.5 Sample size and allocation -- Sample sizes based on prior analysis (See text table 12) is as follows:

Stratum (h)	m_h^*	n_h'	n_h	$n_h^* n_h'$	$n_h^* n_h'$
1	5	6	1	30	5
2	1	9	3	9	3
3	3	5	2	15	6
4	36	11	1	396	36
5	1	9	9	9	9
6	1	10	10	10	10
7	11	8	1	88	11
8	23	10	1	230	23
Total	81			787	103

5.0 Implementation procedures and specifications

This section would deal with the "how-to-do-it" aspects of the survey and three major topics would be addressed: 1) sample selection procedure, 2) measurement proceudre, and 3) data recording procedure.

6.0 Analysis procedures

This section would deal with the specification of method and procedure necessary to analyse the data, construct estimates, and summarize the results.