USDA United States Department of Agriculture

Forest Service

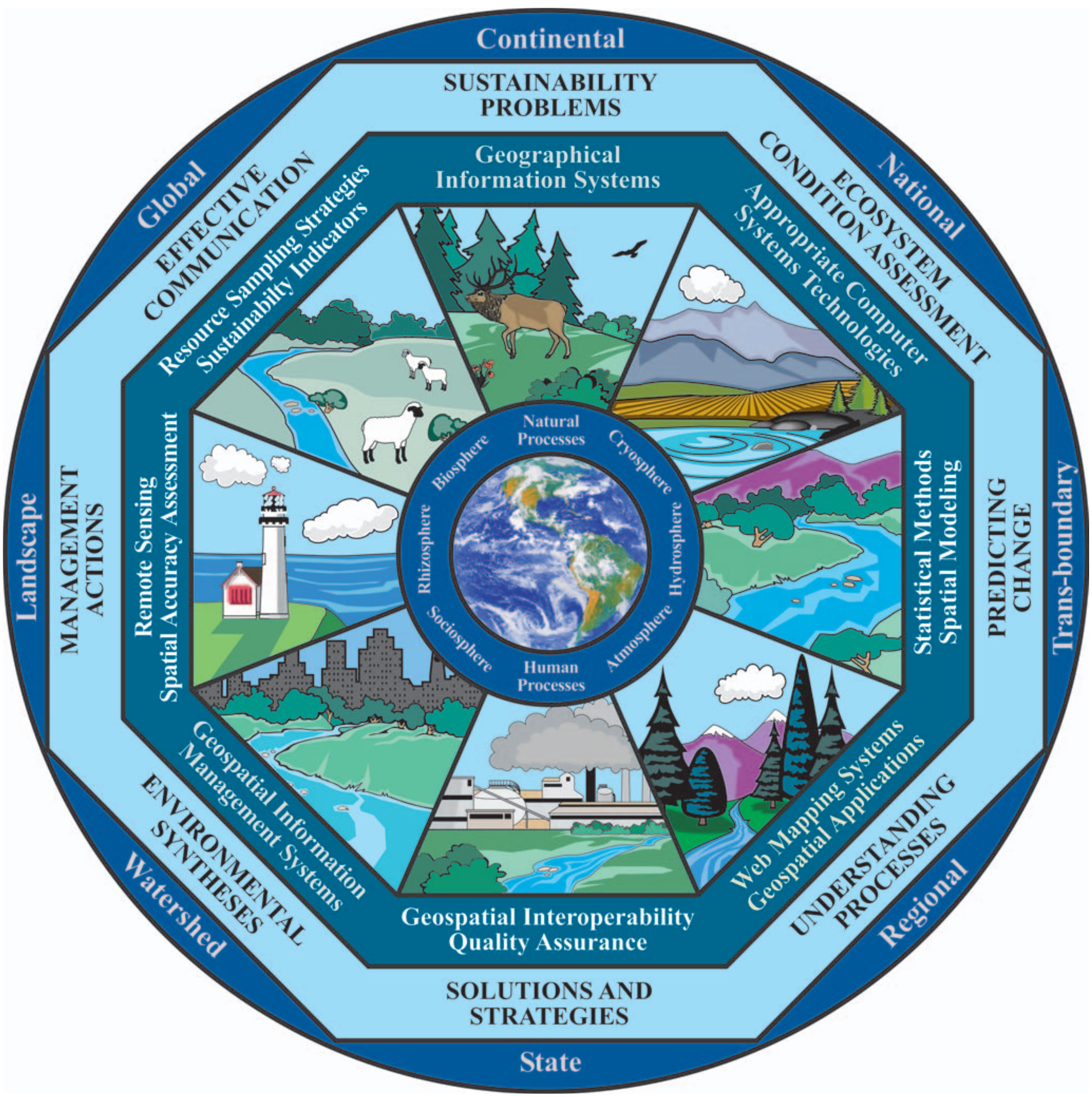
Rocky Mountain Research Station

General Technical Report RMRS-GTR-107

September 2003

Statistical Strategy for Inventorying and Monitoring the Ecosystem Resources of the Mexican States of Jalisco and Colima at Multiple Scales and Resolution Levels

H. T. Schreuder
M. S. Williams
C. Aguirre-Bravo
P. L. Patterson



Abstract

Schreuder, H. T.; Williams, M. S.; Aguirre-Bravo, C.; Patterson, P. L. 2003. **Statistical strategy for inventorying and monitoring the ecosystem resources of the Mexican States of Jalisco and Colima at multiple scales and resolution levels.** Gen. Tech. Rep. RMRS-GTR-107. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 15 p.

The sampling strategy is presented for the initial phase of the natural resources pilot project in the Mexican States of Jalisco and Colima. The sampling design used is ground-based cluster sampling with post-stratification based on Landsat Thematic Mapper imagery. The data collected will serve as a basis for additional data collection, mapping, and spatial modeling products to be described in detail in later documents. Estimation described in this document will be primarily useful for strategic planning at the state and national levels. Because it is a pilot study, the approach for the actual statewide and national inventories is likely to be modified as results are obtained and studied.

Keywords: spatial information, natural resources survey, strategic and management survey pilot study, pixel-based system

The Authors

H. T. Schreuder is a Mathematical Statistician, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO.

M. S. Williams is a Mathematical Statistician, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO.

C. Aguirre-Bravo is Research Coordinator for the Americas, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO.

P. L. Patterson is a Mathematical Statistician, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO.

You may order additional copies of this publication by sending your mailing information in label form through one of the following media. Please specify the publication title and number.

Telephone	(970) 498-1392
FAX	(970) 498-1396
E-mail	rschneider@fs.fed.us
Web site	http://www.fs.fed.us/rm
Mailing Address	Publications Distribution Rocky Mountain Research Station 240 West Prospect Road Fort Collins, CO 80526

Statistical Strategy for Inventorying and Monitoring the Ecosystem Resources of the Mexican States of Jalisco and Colima at Multiple Scales and Resolution Levels

H. T. Schreuder
M. S. Williams
C. Aguirre- Bravo
P. L. Patterson

Introduction

In Mexico, individual agencies address the extent, status, and trends of selected natural resources in response to agency missions. No agency or group of agencies examines the interactions and interdependence among multiple natural resource components from an integrated ecosystem perspective. Available information often is at scales that have limited value for state planning and policymaking. While useful for strategic national level planning, this information has limited use for supporting regional and local decision-making resource applications. Similarly, there is no integrated program to periodically assess multiple natural resources at regional scales, local scales, and multiple resolution levels. Accordingly, the ecosystem resource monitoring initiative, of which this document is a part, will enable comprehensive assessments of Jalisco's and Colima's natural resources for the management of their sustainability.

This ecosystem resource monitoring initiative addresses critical issues that stakeholders in the States of Jalisco, Colima, and elsewhere in Mexico are facing to ensure the environmental sustainability for present and future generations. Land management agencies charged with this responsibility have come to the realization that current data and information are insufficient to confront successfully the ecological and economic challenges of ecosystem resource sustainability. In light of the above, the purpose of this technical document is to describe the statistical approach for inventorying and monitoring the ecosystem resources of the Mexican States of Jalisco and Colima, at multiple scales and resolution levels. Only the design and estimation theory for large-scale national and statewide (strategic) surveys are presented here. Design-based inference is used to ensure a minimum number of assumptions.

Putting this strategy into action will enable us to develop quality assured data, knowledge, and assessment syntheses of ecosystem resources condition and trends. The information will be scientifically credible and technically defensible, and have the greatest social and economic utility for the purpose of planning and decisionmaking for sustainability.

Overview of the Sampling Design

Unlike many national level inventories, such as the U.S. Department of Agriculture's Forest Service Forest Inventory and Analysis program (FIA), the goal of the Jalisco-Colima pilot project (JCPP) is to design an integrated inventory system that provides both traditional assessments of population totals and means as well as spatially realistic maps that describe the location and distribution of various attributes of the population. In other words, an inventory is designed that meets strategic, management, and local needs. The products produced by the inventory include reliable estimates of means and totals for important ecosystem characteristics and maps describing the spatial and temporal properties of the ecosystem.

In designing an integrated multiresource inventory and monitoring system to evaluate the condition and change of variables and indicators for sustainable ecosystem resource management (forest, rangeland, agriculture, wildlife, water, soils, biodiversity, and so forth) one needs some baseline data for comparison. Because one is generally dealing with complex systems, it is not wise to focus on only one or two variables for ecological monitoring purposes. Also, analyzing these variables independently of one another may lead to incorrect conclusions because of their interdependencies. One approach is to model the spatial relationship between key indicator variables. In ecosystem resource management, for example, this information can then be used to identify forest habitats that are either conducive or a deterrent to the presence of ecologically important plant and/or animal species. Techniques commonly used in describing spatial relationships between two or more variables include regression analysis and a variety of geo-statistical procedures that take into consideration the spatial dependency (Cliff and Ord 1981). The proposed ecosystem resource monitoring system will rely on information collected at different spatial scales of resolution and sampling intensities (fig. 1) to provide detailed information at the local level for ecosystem resource planning and management purposes.

The JCPP is an integrated inventory and monitoring program that is designed to address numerous objectives at different spatial scales. The components of the inventory are broken down into five components, which are referred to as "levels," each addressing different spatial scale and information needs. This document provides a brief overview of all five levels of the inventory and a detailed description of levels 1 and 2. Other documents will address levels 3, 4, and 5 in more detail.

Spatially Continuous Monitoring (Level 1)

Landsat Thematic Mapper (TM) data will be used to provide a complete and uniform census of individual Environmental Accounting Units (EAU) across jurisdictional domains (private lands, federal and state lands, ejidos, communities, municipalities, regions, and so forth). An EAU is a watershed used to assess ecological and human activity assessment as well as for resource decisionmaking and planning applications. This approach will provide data or measurements collected as a series of contiguous and simultaneous measures across land tenure units. It will also provide the capability of monitoring EAU's for changes in spectral and spatial characteristics that can be applied over a range of spatial and temporal scales appropriate for addressing specific ecosystem resource issues.

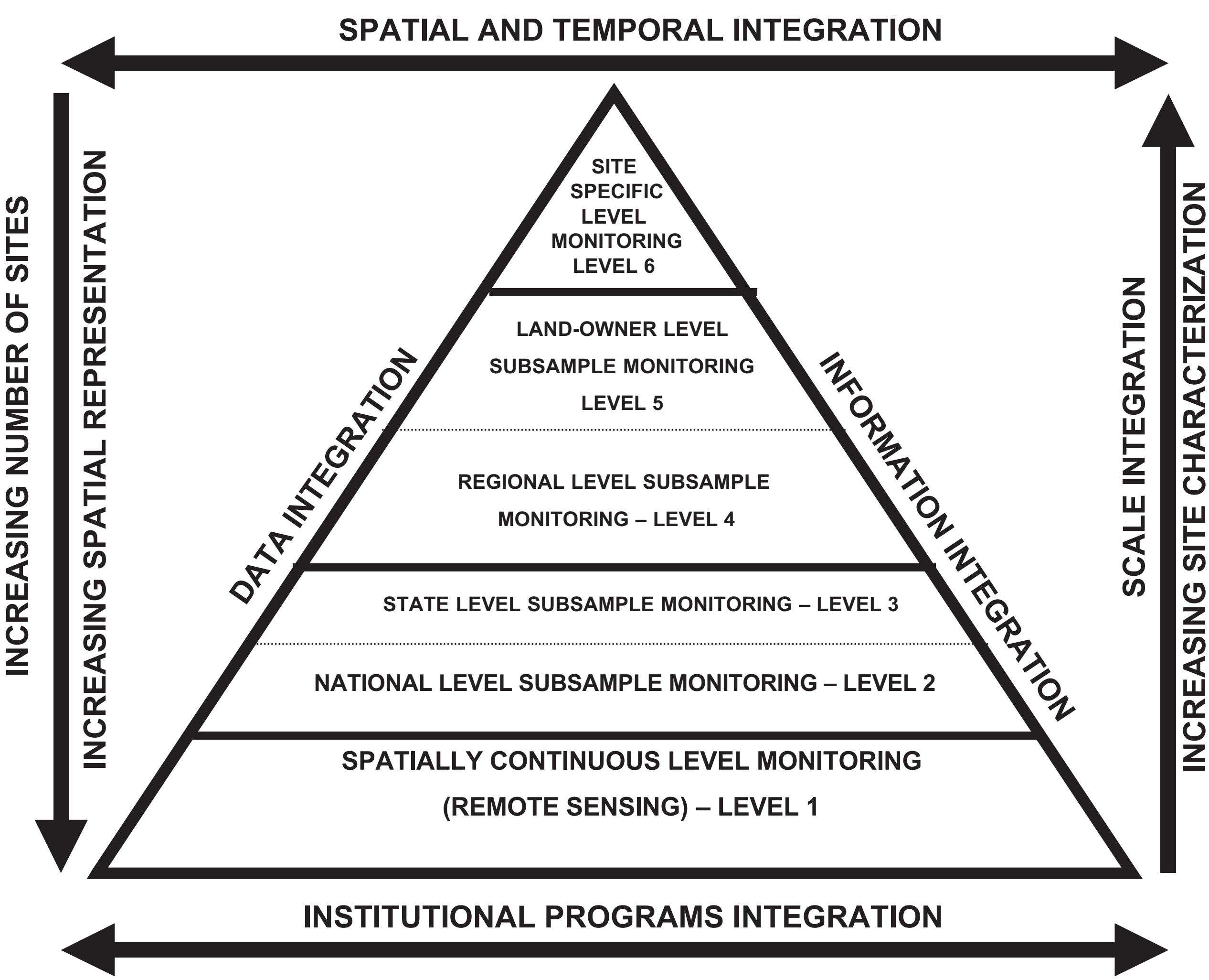


Figure 1—Conceptual model for integration of monitoring design and institutional processes.

Remote sensing data have been used primarily to develop static maps of vegetation cover or land use to guide management decisions in planning applications. However, a current component missing in ecosystem resource monitoring is its use for change detection and spatial modeling of ecosystem properties. When compared to traditional plot data, remotely sensed data can be quite helpful in detecting changes in the land base. Many critical indicator variables such as site productivity, presence or absence of threatened or endangered species, and biological diversity are simply not being measured with sufficient spatial coverage and frequency to allow evaluation of current and future trends. In our approach, Landsat TM data are treated as an auxiliary variable while the field data are treated as the primary variable of interest in describing specific properties of the region at any desired spatial scale (1 to 30 m). Appropriate use of the approach can greatly increase the value of outputs from monitoring programs (Metzger 1997).

The decision to use Landsat TM data over other remote sensors (for example, SPOT [10 m resolution], LEWIS [5 to 300 m resolution], Space Imaging [1 to 4 m resolution], Earthwatch [3 to 15 m resolution], Earthwatch Quickbird [1 to 4 m resolution], OrbView 2 [1,100 m resolution], OrbView-3 [1 to 8 m resolution], SPOT 4 Vegetation Sensor [1,000 m resolution], and MODIS [250 to 1,000 m resolution]) is based on the following reasons. The cost of acquiring data from these high resolution sensors is significantly higher than what it currently costs to acquire Landsat TM data. Users of these sensors have shown that in terms of modeling large-scale spatial variability (Metzger 1997), Landsat TM data are superior to currently available SPOT data, even though the latter has a finer resolution, but is also more expensive. Also the resolution associated with some of these new sensors are not the same across different spectral bands making it difficult to integrate with field data with the same level of precision. For this approach to work, it is important to have a wide range of spectral data at a consistent resolution. An additional benefit is that the spatial modeler involved in this work has considerable experience in working with TM. Therefore, Landsat TM data will be used as the remote sensing sensor of choice until some of these other sensors have been thoroughly evaluated.

In addition to the Landsat data, GIS grids of elevation, slope, and aspect will be developed from digital elevation models. Grid coverages for each topographic variable will be resampled (Resample function, nearest neighbor, Grid Module [ARC/INFO®, ESRI 1995]) to provide a 30 *m* spatial resolution so that an average elevation, slope, and aspect can be assigned to each pixel.

An independent ground sampling effort has been undertaken to collect data that will be used to construct a classified map of the cover types of Mexico. These ground data comprise approximately 2,600 purposively chosen points. At each point, the Global Positioning System (GPS) coordinates and the cover type are recorded. These data will be used to determine the spectral signature of the TM data and landscape characteristics associated with each cover type. This will allow us to generate a preliminary land cover map. Details on the techniques used to produce the map will appear in a later document.

Design-Based Inference for Inventory and Monitoring (Level 2)

The development of the sampling and plot designs is complicated by the variety of indicators to be assessed, the need to assess the ecosystem resources at a range of scales, the need to monitor the indicators over time, and the need to do so efficiently. To meet national level objectives for ecosystem monitoring and large-scale assessments, a traditional grid-based sampling design is used. The plot locations are treated as a simple random sample of the landscape, and post-stratification is employed as a variance reduction technique. This will be referred to as the Tier 1 inventory. This proportion of the inventory is funded by the national government.

The next level of the survey is designed to meet state level inventory objectives. This is referred to as the Tier 2 inventory. The only feature that distinguishes the Tier 1 and 2 inventories is the increased sampling intensity in Tier 2.

Model-Based Inference for Monitoring and Inventory (Levels 3–4)

For the remaining objectives involving estimation at local scales (levels 3–4), the plot data from the Tier 1 and 2 inventories will be enhanced with additional ground plots to provide information needed to develop geostatistical models. These models will be used to describe the location and distribution of various resource attributes and to estimate key attributes at all locations within the sampled population.

To ensure that the spatial variability in the study area is captured, additional sample plots will be located to enhance the spatial models for use in local management applications. The location of these plots can be based on a prestratification scheme that locates plots in areas with the highest errors in the spatial models or where the spectrally derived strata may not capture all of the unique spatial features on the landscape. The primary sampling unit (psu) is a 90 m x 90 m square.

Site Specific Monitoring (Level 5)

Because of the biological importance of certain areas in terms of threatened and endangered plant and animal species, there is a great need to initiate species and/or site specific research and monitoring. Very specific information will be collected at this stage through specific projects if funding is available. In addition to this information, data should also be collected that is compatible with key indicator variables collected at other monitoring levels. All data should be georeferenced to allow the integration with information collected at the different levels using spatially explicit models. Further development of this level of the inventory is not possible without knowing more specific details for each specific site and the characteristics of the population to be studied (for example, a study of animal species would require very different methods than would a study of endangered mosses and lichens).

Analytical Issues

The analysis of the level 1 and 2 inventory data is straightforward once the data have been edited and stored in an electronic file. Validation data sets will be made available using techniques described in Williams and Patterson (2002). An analysis program for level 1 and 2 data incorporating the estimators discussed later in this paper is now being developed in the computer language R (Thaka and Gentleman 1996). Analytical techniques for levels 3 and 4 of the inventory will be more complicated, but software and training will be made available. Analyses for level 5 of the inventory will be developed as needed.

Study Area

The Pilot Study Area consists of the Mexican Southwestern States of Jalisco and Colima with a continental area of approximately 9 million hectares (22.5 million acres). Though Jalisco is larger in area (90 percent), the state of Colima (10 percent) plays a very distinctive role in the economy of the whole region and diversifies the Pilot Study Area considerably. Four major ecological regions provide the natural resources and environmental conditions that make this region one of the most prosperous ones of Mexico (fig. 2). The eco-regions are: the Transversal Neo-Volcanic System, the Southern Sierra Madre, the Southern and Western Pacific Coastal Plain and Hills and Canyons, and the Mexican High Plateau. Linked to these ecological regions, there are several important hydrological regions (HR) that drain to the Pacific Ocean (HR12 Lerma-Santiago, HR13 Huicicila, HR14 Ameca, HR15 Costa de Jalisco, HR16 Armeria-Coahuayana, HR18 Balsas, and HR37 El Salado). One of the watersheds, the Lerma-Santiago Hydrological Region, is connected to Lake Chapala, the most important source of water for the City of Guadalajara.

Precipitation ranges from roughly 300 mm/year in some locations to more than 1,200 mm/year in the higher elevations, with the principal precipitation coming in summer monsoons. Spatially, the ecological systems of this region cut across the boundaries of other Mexican States. For example, several major watersheds drain through the tropical and subtropical forests of the State of Colima. Mostly in the State of Jalisco, water from surface and underground sources is heavily used for agriculture and industrial activities, though a significant portion goes to meet the domestic needs of approximately 10 million people. While on average Colima is wet, the lack of water in the State of Jalisco threatens the sustainability of urban and rural ecological and economic systems. Most of the land (85 percent) in the State of Jalisco is privately owned. Small private landowners are the main driving force of economic development in agriculture, forestry, and rangeland economic activities. In contrast to Colima, for example, a small portion of Jalisco's land is owned by ejidos (10 percent), communities (3 percent), and the government (2 percent). Recently, as a result of trade liberalization brought about by NAFTA policies, new industries have been established in these two States and natural resource utilization has increased due to high population growth rates.

The region's biophysical heterogeneity blends itself to bring about unique habitat conditions for a large diversity of plant and animal species. Within its boundaries, there are a significant number of species of mammals and birds, many of which are severely

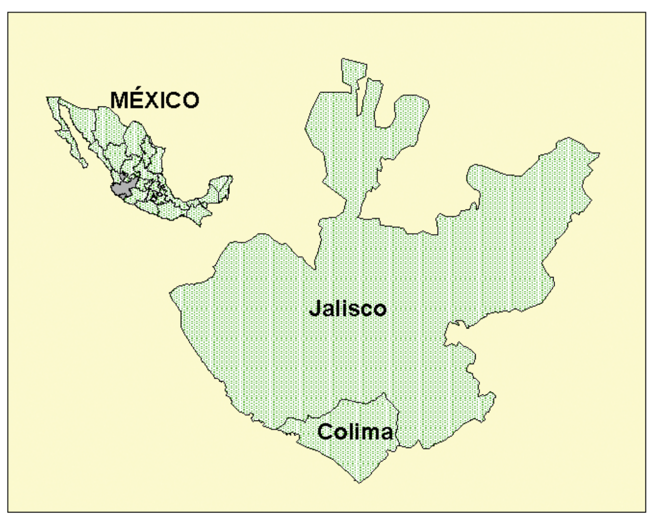


Figure 2—Geographic location of pilot study area in Mexico.

threatened by human activities. Some of the plant and animal species are endemic to specific locations within the ecological regions that comprise the Pilot Study Area. Extensive areas of pine-oak forest are home to "specialty" birds such as the thick-billed parrot, the Mexican-spotted owl, and woodpeckers. It is thought that habitat loss is the single most important element affecting bird populations in this ecosystem complex. Not much is known about how (in other words, what, when, where, why) plant and animal species are being impacted by human activities. Water and other biological resources are an integral part of these ecological regions, whose products transcend geopolitical domains and jurisdictions.

The Design and Estimation for the Tier 1 and 2 Inventories

There is no "perfect" solution for the sample design and estimators for a large-scale inventory. Limitations associated with remote sensing, GPS, and the amount of time required by the field crews for data collection influence the design of the survey (Williams and Eriksson 2002; Overton and Stehman 1996). The sampling strategy chosen for the JCPP is an efficient and flexible strategy.

An efficient sampling design for collecting information over large geographical areas is a systematic grid with equal spacing and a random start. It has the advantage of spreading the sample units uniformly throughout the population. However, the systematic grid design tends to be the least efficient for spatial modeling because the equal spacing between plots provides less information for modeling purposes than a design with uneven spacing. In order to avoid periodicity in the resource and to permit plots to occur at random distances for spatial modeling, the plots will be located randomly within hexagonal cells formed by a triangular grid. This grid for the Tier 1 inventory will uniformly cover all of Mexico with a spacing of 30 km x 30 km. To facilitate across border assessments, the grid for Mexico will be similar to the one developed for the FIA inventory in the U.S.A. In addition to the 30 km x 30 km grid, five more triangular grids of increasing intensities are available to support sampling decision-making processes (25 km x 25 km, 20 km x 20 km, 15 km x 15 km, 10 km x 10 km, and 5 km x 5 km). These grids can be used to intensify the Tier 1 sample for meeting state level inventory needs. This state level intensification is the Tier 2 component of the inventory. For the pilot study, the Tier 2 inventory uses a spacing of 5 km x 5 km with the Tier 1 ground plots being a subsample of the Tier 2 plots.

The field crews will locate all of the ground plots at the UTM coordinates at the center of the TM pixel given to them—accurate location of the points is important both for spatial modeling as well as future relocation of these permanent plots. The grid density will be set first to meet state needs. It will then be intensified to meet other multiple scale needs, based on available funding. The intensity can also be altered in each of the main hydrological regions or municipalities. Plot locations will be kept secret. The opportunity to intensify also exists for local areas within land tenure units, EAUs, or administrative units, as funding is made available.

The other major advantage of using Landsat TM data is that the pixels can be used to construct an area frame comprised of equal area sample units. The advantage of this area frame approach (Eriksson 1995; Husch and others 1982) is that it is a true equal probability sample regardless of the type of stratification used, with the inclusion probabilities of the sample units (for example, trees, coarse woody debris, and other vegetation) along the boundaries of different subpopulations not needing adjustment. This is especially advantageous when the sample data will be poststratified, because every division (stratification) of the population creates new subpopulations with new boundaries. These new boundaries would require adjustments to the inclusion probabilities of all population elements (trees) along the new boundaries, and it is unlikely the data needed for these could or would be collected (see Williams and Eriksson 2002 for a discussion of this topic).

For estimation, the plots will be poststratified using the classified level 1 Landsat TM data. At a minimum, the entire population and each EAU will be stratified into areas that are predominantly forest or nonforest. The plots will also be assigned to these strata, and then the appropriate results for poststratified random sampling will be applied (Cochran 1977; Schreuder and others 1993). The estimators are given below. For large areas, the sample size in each stratum will be approximately proportional to the size of each stratum.

Description of the Ground Sample Unit

The primary sampling unit (psu) measures 90 m x 90 m (fig. 3) and is constructed with a 3 x 3 block of Landsat TM pixels. The psu will be subsampled by five 10 m x 10 m secondary sampling units (ssus). Each primary sampling unit will be centered on the coordinates assigned to it and will be laid out in a north-south, east-west manner. PSU locations will be verified using a GPS with an estimated accuracy of within 3 m.

In order to maintain an equal probability sample design and because the locational and registration errors in remote sensing and GIS technologies prevent an accurate subdivision of a pixel, a psu is considered either totally within a population or totally out of a

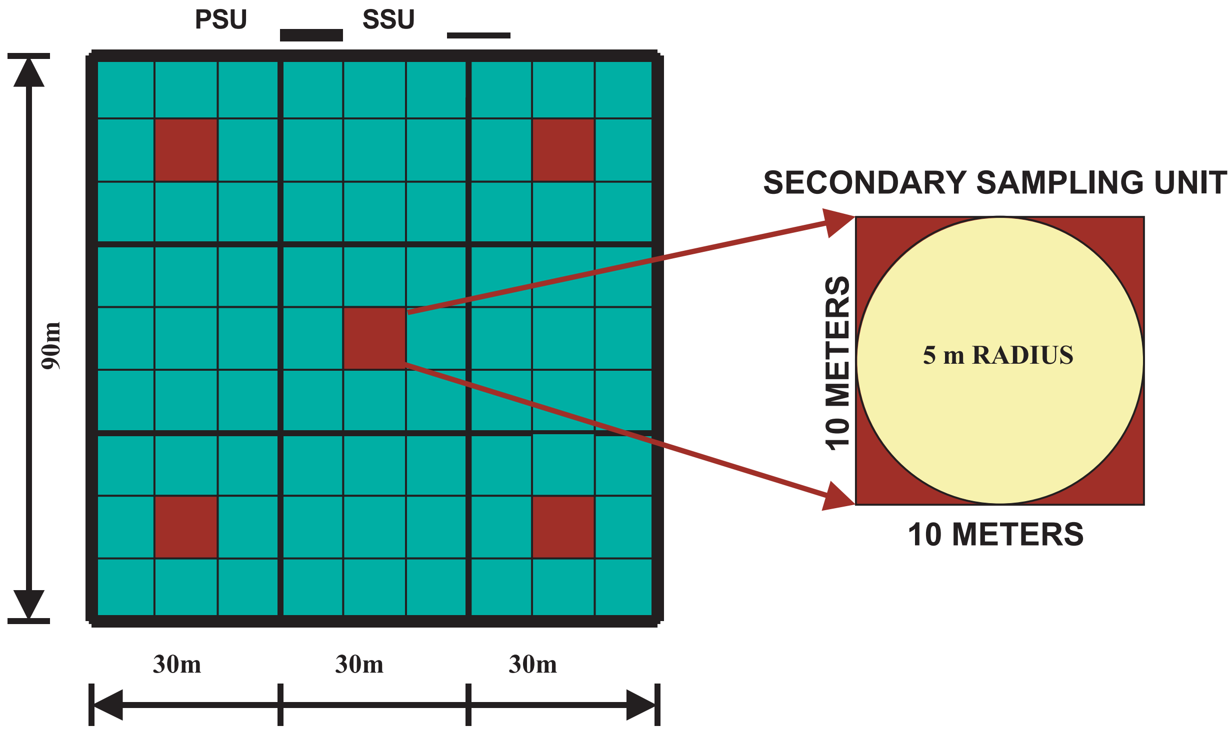


Figure 3—Plot layout for primary (psu) and secondary (ssu) sampling units.

population (or stratum). Thus the population boundaries are redefined to follow the area frame population boundaries created by the edges of the psus. A psu is "in" when the center of the psu falls inside the population or stratum boundary. This creates a jagged stair-stepped edge along the population boundaries, so a small portion of the actual landmass of an area segment will lie within the boundary of another area segment. The size of the difference should be minimal, generally less than 0.05 percent for the watersheds in Jalisco and Colima. For example, a minuscule quantity of the landmass of Colima may be included within the area frame for the Jalisco land mass.

Because these will be permanent plots, the psu center will be monumented on the ground. A sample of 5 of the 81 ssu's will be selected for measurement, using a circular plot of 5 m radius. This plot will be referred to as the 5 m plot. One 5 m plot will be located at the psu center. The other four 5 m plots will be located at fixed angle directions and constant distance with respect to the psu center (fig. 3).

Several kinds of subplots will be located within each of the 5 m plots (fig. 4), and different measurements will be made on each plot type. All large trees (≥ 12.5 cm d.b.h.) will be measured on each of the 5 m plots. Observed attributes will be specified in the field sampling and indicator measurement manuals. Saplings (2.5 cm ≤ d.b.h. < 12.5 cm) will be measured on a circular plot (3 m radius) colocated at the center of each tree subplot. The term "3 m plot" will be used to denote this plot. Within each of the 5 m plots will be three square plots, each measuring 1 m x 1 m. These will be referred to as 1 m quadrats. The first 1 m quadrat will be located at the center of the 5 m plot. The remaining two are located 6 m from the center plot on a diagonal of the 5 m plot (fig. 4). Seedlings (height ≥ 30 cm and d.b.h. < 2.5 cm) will be sampled on three

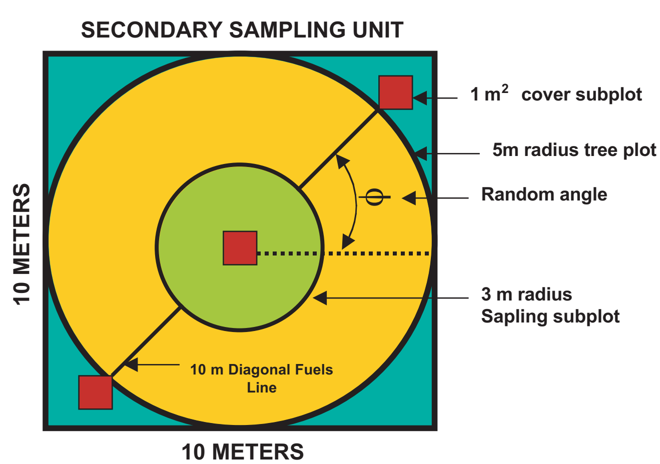


Figure 4—Layout of secondary sampling unit.

1 m quadrats. In addition to the number of seedlings, the percent cover of herbaceous plants, shrubs, and tree species < 30 cm tall will be recorded.

To estimate fuel loadings, a 10 m transect with a random orientation will be established. This will be referred to as the 10 m transect. Line intersect techniques will be used to estimate fuel loadings of large woody material (sound and rotten) ≥ 7.5 cm in diameter. All large woody material intersecting the 10 m transect will be counted and cross-sectional areas measured by genus. Medium size woody materials (2.5 to 7.4 cm in diameter) that are intersecting the first 5 m of the transect will be counted. In addition, the first, center, and last 1 m of the diagonal transect will be used to count fine woody materials (0.01–2.4 cm in diameter). In each case, the mean height of fuels in each sampled diameter class, as well as the slope of the diagonal transect, will be measured and reported.

Soil attributes will be observed on each 5 m radius plot. Any destructive soil samples will be collected on the west side (270 degree Azimuth) of the primary sampling unit at a distance of 5 m of the plot boundary line.

Most of the indicator variables are compatible with those used by the USDA Forest Service and Canadian ecosystem resource monitoring programs. Other indicator variables can be integrated into this pilot study as resources become available and the need dictates to ensure comparability and interoperability of indicators with participating government agencies from the United States of America and Canada.

Selection of the Ground Plot Locations and the Assignment of Measurement Periods

The system used to derive the Jalisco/Colima Pilot Project (JCPP) ground plot locations was developed to facilitate cross-border analyses between the United States and Mexico. The Environmental Monitoring and Assessment Program (EMAP) of the Environmental Protection Agency (EPA) originally developed it. EMAP developed a worldwide sampling grid to evaluate status, change, and trends in the condition of ecological resources. White and others (1992) used a truncated icosahedron as the basis for the grid of the Forest Health Monitoring Program of the U.S. Forest Service. The truncated icosahedron was made up of 20 hexagon faces and 12 pentagon faces that give it a "soccer ball" appearance (fig. 5).

The plot locations were arranged on a triangular systematic grid with a random start, with plot locations falling at the intersection of a triangular grid with approximately a 27 km point spacing. This grid allows for a 4-year time span between remeasurement of each ground plot on an interpenetrating rotating panel design. Thus, the entire grid and each individual panel is evenly distributed within the area, with a sampling intensity of approximately one ground plot for every 6,800 hectares (167,000 acres) for the FHM program.

The JCPP will adopt the same sampling intensity as FIA with one ground plot for every 2,400 hectares and maintain across-border consistency by meshing the hexagon grid covering Mexico with the existing grid covering the U.S.A. However, with the increased focus on spatial modeling, the placement of ground plots within each hexagon will not be on the systematic grid. Instead, the plot location within each hexagon will be completely random. This is done by generating a random (x, y) coordinate that falls within each hexagon. The justification for this practice is that the estimation of correlograms and variograms is difficult when ground plot locations are equally spaced, especially when the scale of spatial patterns is expected to be much finer than the spacing of the grid. For the purpose of estimating population means and totals, this sample is still treated as a simple random sample of the land base, as is common with most environmental surveys. The justification for this practice is summarized by Ripley (1981: 19–21).

The purpose of the hexagon grid is twofold. In the first place it is used to establish a pseudo-systematic sample (with a random start) of ground plots across the JCPP study area. Secondly, it is used to allocate ground plots to each yearly panel so that a good spatial representation of all conditions is represented in each panel. The hexagon grid has no other use in the estimation process. The resulting sample can be best described as a hybrid of centric- and nonaligned-systematic sampling (Ripley 1981: 19–21).

The assignment of each plot to a panel is loosely based on the original design used by Forest Health

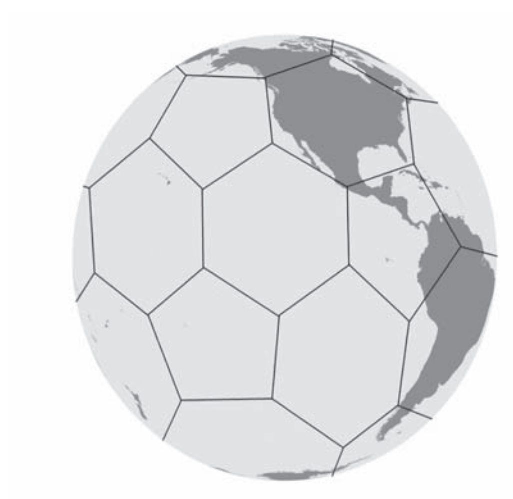


Figure 5—Hexagon tessellation of the Earth.

Monitoring (FHM) and the current FIA design. The EMAP grid was developed for interpenetrating rotating panels of 3, 4, 7, 9, and 11 years or any multiple. The FIA design is designed for sampling periods of 5, 7, or 10 years. The States of Jalisco and Colima are tentatively planning a 5-year period of remeasurement.

Estimators for the TIER 1 and 2 Inventories

The estimators for the Tier 1 and 2 inventories are simple. The purpose of this section is to present the estimators in detail. This allows the readers to progress only as far into the notation as necessary. Further justification for this approach and simulation results can be found in Williams and Patterson (2003).

One topic that will not be discussed in detail is the method used to assign stratum values to each 90 m x 90 m psu. Instead, it is assumed that an effective stratification algorithm will be developed for various regions that will combine the individual pixel attributes, cover map information, and other auxiliary information contained within the area frame. The algorithm may reflect regional forest and landscape characteristics and may depend on the type of analysis being performed. However, at this point in time there is too little known about available information to define a single technique for the study area.
Slightly different approaches are used to derive the estimators of various cover types versus the estimators associated with vegetation. Thus, these topics will be treated separately.

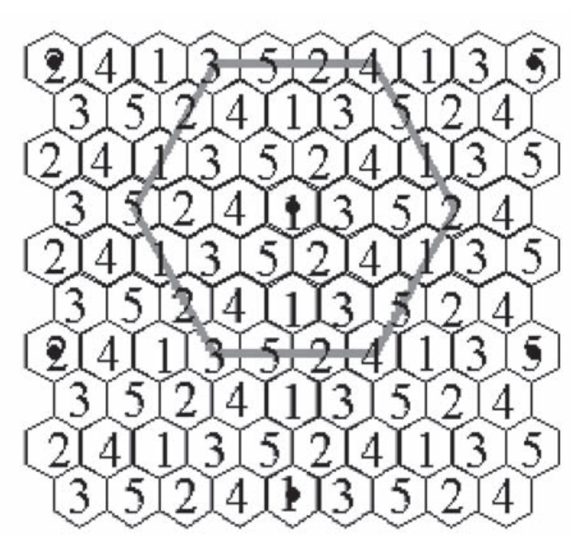


Figure 6—Numbering of each hexagon for the panel assignment.

Estimators for total area of land by cover type or condition classes

To estimate the total area of forest or other condition class, the five 5 m plots of the psu are viewed as the second stage of a two-stage cluster sample. Multistage cluster plot sampling results are used to divide each psu into $M = 81$ secondary sampling units, from which a sample of size $m = 5$ is chosen (fig. 3).

While the estimators derived from the smaller plot sizes (for example, the 3 m and 1 m plots) can be combined with the one derived from the 5 m plots, the nested structure of these plots will produce highly correlated estimates. This correlation between the estimators of forest area is such that essentially no reduction in variance exists when estimating forest areas on anything other than the largest plot used (Williams and Patterson 2003). Thus, the area of forest or condition class is estimated using only the information from the largest plot size implemented.

Regardless of the plot size or the method of assignment to a stratum for each psu, the goal is to estimate the proportion of specific condition classes in each stratum. To accomplish this, then n ground plots are treated as a poststratified sample with a random sample size of n_h in each stratum. For each ground plot, the proportion of the psu covered by forest is either estimated by a subsample of ssu's or by measuring the proportion of the psu covering the condition class of interest. Thus, given that the assumptions are reasonable, the estimator of the area of forest is a post-stratified, two-stage cluster sample (Cochran 1977), given by

$$\hat{A}_F = A\hat{p}_F = A\sum_{h=1}^{H}\frac{N_h}{N}\left[\sum_{i=1}^{n_h}\sum_{j=1}^{m}\frac{p_{hij}^F}{mn_h}\right], \qquad [1]$$

where p_{hij}^F is the proportion of forest in the j^{th} 5 m subplot of the i^{th} psu within the h^{th} stratum.

Since the variance of $\hat{A}_F$ is equal to $A^2 Var(\hat{p}_F)$, it suffices to determine $Var(\hat{p}_F)$. The derivation of the variance combines results from the variance of a simple random poststratified sample and the variance of a stratified two-stage sample (Cochran (1977), §5A.9 and §10.9). Conditioning on the n_h's, $Var(\hat{p}_F) = Var[E[\hat{p}_F | n_h]] + E[Var[\hat{p}_F | n_h]]$. Since for fixed stratum sample sizes, $\hat{p}_F$ is an unbiased estimator of p_F, the first term is zero. For fixed n_h,

$$Var(\hat{p}_F) = \sum_{h=1}^{H}\left(\frac{N_h}{N}\right)^2\left[\left(\frac{1}{n_h}-\frac{1}{N_h}\right)S_{1h}^2 + \frac{1-m/M}{mn_h}S_{2h}^2\right]$$

$$= \sum_{h=1}^{H}\frac{1}{n_h}\left\{\left(\frac{N_h}{N}\right)^2\left[S_{1h}^2 - \left(\frac{1}{m}-\frac{1}{M}\right)S_{2h}^2\right]\right\} - \frac{N_h}{N^2}S_{1h}^2 \qquad [2]$$

(Cochran 1977). The next task is to find the expectation of $1/n_h$. Assuming the sample size is sufficiently large to preclude $n_h = 0$ then to order n^{-2},

$$E\left[\frac{1}{n_h}\right] = \frac{1}{n(N_h/N)} + \frac{1 - N_h/N}{n^2(N_h/N)^2}.$$

Using results derived in Cochran (1977) it can be shown that an unbiased estimator for the above approximation of $Var\left[\hat{p}_F\right]$ is given by

$$var\left[\hat{p}_F\right] = \sum_{h=1}^{H} \frac{1}{n}\frac{N_h}{N}\left[\left(1-\frac{n}{N}\right)s_{1h}^2 + \frac{n}{N}\left(\frac{1}{m}-\frac{1}{M}\right)s_{2h}^2\right] + \left(1-\frac{N_h}{N}\right)\frac{1}{n^2}s_{1h}^2, \quad [3]$$

where s_{1h}^2 and s_{2h}^2 are the between cluster and within cluster sample variances for stratum h (Cochran 1977). Due to the large number of psu's (N) in comparison to sample size (n), the second term within the square brackets is small relative to the first term for both the Tier 1 and 2 inventories.

Vegetation or tree estimators

In this section, the estimator that utilizes the tree information from the entire ground plot is given. Unlike the estimator for forest area, where the assumption is made that the psu is divided into secondary sampling units, a different approach is taken. This is because the tree and other vegetation attributes are gathered from transect samples and across the 1 m, 3 m, and 5 m plots, which would require different M values and the calculation of the correlations between the estimators derived from each of the plot sizes. Thus, the proposed estimators of tree and other vegetation attributes within each psu are expressed as a triareal sampling design comprised of three different size subplots (Husch and others 1982). The estimator for the total within each psu is expressed using a Horvitz-Thompson estimator. Two assumptions are necessary. The first is that systematic nature of the three subplots within each psu is equivalent to having the points randomly located within the psu. In the literature, this assumption has been justified by assuming that the location of trees in the forest has a random pattern within the psu (de Vries 1986). When this assumption is not realistic, the variance estimator tends to overestimate the variance. The second assumption is that none of the points falls close enough to the boundary of the psu so no adjustments to the inclusion probabilities of trees are required. This allows a constant inclusion probability to be assigned to all trees tallied on each of the three types of subplots. This assumption is necessary because it is impractical to obtain the actual probabilities of selection for border trees.

The n ground plots are treated as a post-stratified sample of psu's, with a random sample size n_h in each stratum. Within each psu, a tree-centered circle, or 1 m quadrat in the case of seedlings (Husch and others 1982) is placed around each tree. The size of the circle is determined by the diameter at breast height limits, which in turn determine the inclusion probabilities. Then a tree is included in a sample drawn by a single point if the point lies within the tree-centered circle. Let a^{1m}, a^{3m}, a^{5m} denote the area of the 1 m quadrat-, 3 m-, and 5 m radius plots respectively and let a_{psu} = 8,100m^2 denote the area of the psu. The probability that a tree is included in the sample drawn by a single point within the psu depends on the size of the tree and is defined as:

$$\pi = \begin{matrix} a^{1m}/a_{PSU} \\ a^{3m}/a_{PSU}, \\ a^{5m}/a_{PSU} \end{matrix}$$

where a^{1m} is the area of the 1 m quadrat, a^{3m} is the area of the 3 m radius plot, and a^{5m} is the area of the 5 m radius plot.

The estimator given here is for the 3 m and 5 m radius plots, where m = 5 plots are sampled within each psu. The estimator for the 1 m quadrats is defined analogously with the exception being the change in sample size to m = 15. Estimation for the psu is based on the selection of 5 points within the psu and measurements of the subplots of 3 m and 5 m radius respectively centered at these points. Let T_{hij} be the number of trees tallied at the j^{th} point of psu i in stratum h. Then the estimator of the total of the tree attributes within the psu for the j^{th} point of psu i, in stratum h, is given by

$\hat{Y}_{hij} = \sum_{t=1}^{T_{hij}} \frac{y_{hijt}}{\pi_{hijt}}$, with y_{hijt} and π_{hijt} respectively the value of interest and the probability of selecting tree t at point j, psu i, in stratum h,

with the estimator for the total of the psu derived from the m = 5 points being

$$\tilde{Y}_{hi} = \frac{1}{m}\sum_{j=1}^{m} \hat{Y}_{hij}.$$

An unbiased estimator of $Var\left[\tilde{Y}_{hi}\right]$ is given by

$$var\left[\tilde{Y}_{hi}\right] = (m(m-1))^{-1}\sum_{j=1}^{m}\left(\hat{Y}_{hij} - \tilde{Y}_{hi}\right)^2.$$

This estimator is derived by summing across the psus and strata. Then the estimator is

$$\hat{Y} = \sum_{h=1}^{H} N_h \frac{1}{n_h}\sum_{i=1}^{n_h}\tilde{Y}_{hi}. \quad [4]$$

An approximation of the variance of the estimator is given by

$$Var[\hat{Y}] \approx \sum_{h=1}^{H} N^2 \left[\frac{N_h}{N} \frac{1}{n} + \left(1 - \frac{N_h}{N}\right) \frac{1}{n^2} \right] \quad [5]$$

$$\bar{V}[\tilde{Y}_{hi}] + N^2 \left[\frac{N_h}{N} \left(\frac{1}{n} - \frac{1}{N} \right) + \left(1 - \frac{N_h}{N}\right) \frac{1}{n^2} \right] S_h^2,$$

where $\bar{V}[\hat{Y}_{hi}] = \sum_{i=1}^{N_h} Var[\tilde{Y}_{hi}] / N_h$

and $S_h^2 = \sum_{i=1}^{N_h} (Y_{hi} - Y_{h\cdot})^2 / (N_h - 1).$

An unbiased estimate of the above approximation of $Var[\hat{Y}]$ is given by

$$var[\hat{Y}] = \sum_{h=1}^{H} N^2 \left[\frac{N_h}{N} \frac{1}{n} + \left(1 - \frac{N_h}{N}\right) \frac{1}{n^2} \right] \quad [6]$$

$$var[\tilde{Y}_{hi}] + N^2 \left[\frac{N_h}{N} \left(\frac{1}{n} - \frac{1}{N} \right) + \left(1 - \frac{N_h}{N}\right) \frac{1}{n^2} \right] \tilde{s}_h^2,$$

where $var[\hat{Y}_{hi}] = \sum_{i=1}^{n_h} var[\tilde{Y}_{hi}] / n_h$

and $\tilde{s}_h^2 = \sum_{i=1}^{n_h} \left(\tilde{Y}_{hi} - \tilde{\tilde{Y}}_{h\cdot} \right)^2 / \left(n_h - 1 \right).$

The above estimation assumes that there are no plots or subplots that cannot be measured. A nuisance that arises with almost any type of field sampling with plots is that parts of the psu or subplots will fall outside the population of interest or are inaccessible to sampling either because of difficulty of terrain or the necessity to trespass on private land. Not all landowners will allow the crews on or across their land even to measure other lands. How to treat inaccessible plots or subplots is described in the appendix. The following estimator deals with the situation of missing plots or subplots. Poststratification, as considered above, is not shown in the below formulation since it is a straightforward extension of what is given by simply applying the estimator shown to each stratum.

The most appropriate estimator to be used in estimating population totals with missing subplots is:

$$\hat{Y}_s = A \left[\sum_{i=1}^{n} \sum_{j=1}^{n_s} \sum_{k=1}^{n_{ss}} \hat{A}_{ijk} Y_{ijk} \right] / \left[\sum_{i=1}^{n} \sum_{j=1}^{n_s} \sum_{k=1}^{n_{ss}} \hat{A}_{ijk} \right] \quad [7]$$

and

$$\hat{Y} = A \left[\sum_{i=1}^{n} \sum_{j=1}^{n_s} \hat{A}_{ij} Y_{ij} \right] / \left[\sum_{i=1}^{n} \sum_{j=1}^{n_s} \hat{A}_{ij} \right] \quad [8]$$

where $\hat{A}_{ij}$ and $\hat{Y}_{ij}$ are the estimated sampled area and value of interest respectively in subplot j of plot i, n is the number of plots in the sample, n_s is the number of subplots in the sample for plot i where $\hat{Y}_s$ is the estimator for subplot size s and subplot ss, and m_i is the number of subplots in the sample for plot i (Max and others 1996). We recommend the use of bootstrap variance estimators here as discussed in Schreuder and others (1993).

Line intersect sampling estimators

The line intersect sample (LIS) is drawn using m = 5 randomly oriented lines with the centers covering each of the 5 m plots. The estimator employed is the LIS estimator as described in Kaiser (1983) and Gregoire (1998), with the key difference being that the sample locations are a systematic subsample of the psu rather than five random locations. There should be little if any detectable bias in the resulting estimators, but the variance estimator should overestimate the true sample variance. The LIS estimator for the amount of woody debris on the psu estimated from the m = 5 and L = 10 m lines with random orientation is given by

$$\tilde{Y}_{hi} = \frac{1}{m} \sum_{j=1}^{m} \hat{Y}_{hij} = \frac{1}{m} \sum_{j=1}^{m} \frac{a_{psu}\pi}{2L} \sum_{k=1}^{P_j} \frac{g_{hijk}}{\cos \delta_{hijk}}, \quad [9]$$

where g_{hijk} is the cross-sectional area of the log measured perpendicular to the long axis of the log (m^3) for the k^{th} piece of debris on the j^{th}transect, P_j is the number of pieces of debris intersected by the j^{th} line, and δ_{hijk} is the slope of the log. In relatively flat terrain, δ_{hijk}, may be set to zero. However, a bias in excess of 10 percent can occur whenever the slope exceeds 25 percent. The derivation and further details for this estimator can be found in Kaiser (1983 example 2c).

The variance estimator given in equation [6] applies.

Other estimators

While area and vegetation totals are the primary forest attributes that are estimated, other estimators are also of interest. The first class is mean per tree estimators (for example, average number of conks or disease agents per trees). The form of this estimator is $\hat{\bar{Y}} = \hat{Y}/\hat{N}$, where $\hat{N}$ is the estimator of the number of trees (eq. [4] with $y_{hijt} = 1$). Another class of estimators is used to assess change. These have the form $\Delta\hat{Y} = \hat{Y}_2 - \hat{Y}_1$, where $\hat{Y}_2$ and $\hat{Y}_1$ are the estimators at times 2 and 1, respectively.

Variance estimators for the mean per tree and difference estimators are

$$var(\hat{\bar{Y}}) \approx \frac{1}{\hat{N}}[var(\hat{Y}) + \frac{\hat{Y}}{\hat{N}} var(\hat{N}) - 2\frac{\hat{Y}^2}{\hat{N}^2} cov(\hat{Y},\hat{N})], \quad [10]$$

and

$$var(\Delta\hat{Y}) = var(\hat{Y}_2) + var(\hat{Y}_1) - 2cov(\hat{Y}_2,\hat{Y}_1). \quad [11]$$

Alternative estimators

At this point in time there are a number of alternative estimators that may offer advantages over the estimators given above.

In various circumstances we may have complete knowledge on a covariate associated with the variable of interest for which we know all the values in the population or we can get those with relative ease. Usually this information is combined with the information on the variable of interest measured on a subsample of the units in the population. Denoting by y = variable of interest and x = covariate, numerous estimators are possible. We focus only on the generalized regression and the ratio-of- means estimator, the others are generally not desirable.

A very general, efficient estimator is the generalized regression estimator developed by C. E. Sarndal:

$$\hat{Y}_{gr} = n\sum_{i=1}^{n} y_i / \pi_i + a_{gr}\left(N - \sum_{i=1}^{n} 1/\pi_i\right) + b_{gr}\left(X - \sum_{i=1}^{n} x_i/\pi_i\right)$$

$$= \sum_{i=1}^{N} \hat{y}_i + \sum_{i=1}^{n} e_i / \pi_i \quad [12]$$

where

$$\hat{y}_i = a_{gr} + b_{gr}x_i, e_i = y_i - \hat{y}_i,$$

$$a_{gr} = \left[\sum_{i=1}^{n} y_i/(\pi_i v_i) - b_{gr}\sum_{i=1}^{n} x_i/(\pi_i v_i)\right] / \sum_{i=1}^{n} 1/(\pi_i v_i)$$

$$b_{gr} = \left\{\sum_{i=1}^{n} 1/(\pi_i v_i) \sum_{i=1}^{n} x_i y_i/(\pi_i v_i) - \sum_{i=1}^{n} y_i/(v_i\pi_i) \sum_{i=1}^{n} x_i/(v_i\pi_i)\right\}$$

$$/\left\{\sum_{i=1}^{n} 1/(\pi_i v_i) \sum_{i=1}^{n} x_i^2/(v_i\pi_i) - \left[\sum_{i=1}^{n} x_i/(v_i\pi_i)\right]^2\right\}$$

with variance:

$$V(\hat{Y}_{gr}) = (1/2\sum_{i \neq j}^{N} (\pi_i\pi_j - \pi_{ij})(e_i/\pi_i - e_j/\pi_j)^2 \quad [13]$$

with 2 possible variance estimators given in Schreuder and others (1993) but we recommend using a bootstrap variance estimator here. This is the estimator to be used when possible.

Another estimator that is not a special case of the above is called the generalized ratio of means estimator:

$$\hat{Y}_{rm} = \left(\sum_{i=1}^{n} y_i/\pi_i / \sum_{i=1}^{n} x_i/\pi_i\right) X = (\hat{Y}_{HT} / \hat{X}_{HT}) X \quad [14]$$

with approximate variance

$$V(\hat{Y}_{rm}) = V(\hat{Y}_{HT}) - 2RCov(\hat{Y}_{HT}, \hat{X}_{HT}) + R^2 V(\hat{X}_{HT}). \quad [15]$$

There is a good discussion in Schreuder and others (1993) a variance estimation but, we recommend using a bootstrap variance estimator. Both the generalized regression and the ratio-of-means estimator are asymptotically unbiased.

There will be opportunities to use $\hat{Y}_{gr}$ in the pilot study. With the heavy emphasis on remote sensing platforms we will examine what we can do with that in relation to the ground sampling to obtain improved estimates for some parameters.

Variance estimators can also be derived that utilize resampling techniques, with the most common being bootstrap and jackknife estimators. Bootstrap samples are generated by repeating the sampling design with replacement at the same sample sizes as the actual sample design and then generating estimates for each bootstrap sample. The variance between those estimates is the variance estimate for the estimator of interest. Usually 2000 bootstrap samples are generated. This is applicable for all estimators above.

Strategy for the Remeasurement of Plots

During the first year, all plots will be measured. Following 2 years of reporting and analysis, a 5-year remeasurement cycle involving the remeasurement of a systematic and representative subsample will begin. The remeasurement cycle is not decided at this time but will be at most 5 years.

Landsat TM imagery of the study area will be obtained just prior to the periodic reports, such as every five years. The only criterion is that the image be cloud free to ensure no loss of information in modeling the spatial relationship between the field data and the Landsat imagery. This imagery must then be classified into forest and nonforest at a minimum. During the second measurement cycle, it may also be advantageous to add change strata (forest to nonforest and nonforest to forest).

All permanent plots will be remeasured at least once during a 5-year cycle. The plots will be divided into five or fewer systematically ordered panels, each representative of the whole study area. One panel will be measured each year.

For site-specific monitoring plots, the remeasurement cycle will depend on the specific objectives of the research study, or survey, as well as the funding available. Variables and measurement protocols will be compatible across plot types and monitoring levels.

When the same population is repeatedly sampled, the sampler is in a unique situation to make realistic cost and variance estimates and to use this information to apply techniques that lead to optimum efficiency of sampling (Cochran 1977). The current practice of remeasuring forest inventory plots on a yearly basis is to divide the sample plot into K groups (panels) and remeasure one group each year for T years. This allows us to estimate change on a T-year cycle, but the disadvantage of this approach is that the remeasured sample plots provide limited information on year-to-year changes. As an alternative, one could randomly sample, with replacement, P percent of the sample plots each of the T years. At the end of the T years some plots will have been remeasured zero times, one time, two times, three times, and so forth, at a 1 year interval, 2 year interval, and so forth. Constraints could be added to the sampling protocol to ensure that all sample plots are remeasured at least once during the T-year cycle. This would provide the information necessary to estimate current values as well as yearly changes.

To this point, the focus has been on the estimation of forest area and vegetation attributes for a single panel. In an annualized survey, the individual estimators for each panel are combined into a single overall estimator. Individual panel estimates will, however, have less precision due simply to the smaller sample size measured each year. With this in mind, it has been often suggested that the estimator of the current condition should take advantage of data collected in the previous years (panels).

There have been several different estimation strategies proposed that would take advantage of previous data. One estimator of the current population statistic, $\hat{Y}$, that appears often in the literature is the moving average (MA) given by

$$\hat{Y}_{MA,t} = \sum_{h=1}^{T} w_{t-T+h} \hat{Y}_{t-T+h}, \qquad [16]$$

where $\hat{Y}_{t-T+h}$, $h = 1,\ldots,T$, is an unbiased estimator of Y_{t-T+h} and w_{t-T+h} is a set of constants which sum to 1. The constant, T, is equal to the number of panels in the survey, or equivalently, the number of years until the entire inventory is completed. If 20 percent of the plots selected are sampled each year then $T = 5$. An advantage of the MA estimator is that it will have a substantially lower annual variation than any one of the yearly estimators. This fact may appeal to consumers of FIA data, who may distrust estimates that vary too much from year to year (Gillespie 1999).

The major problem with the MA estimator is that it is design biased as an estimator of the current year's population parameter. Particularly major changes in the population from one year to the next could cause serious bias. The rationale for the moving average estimator is that by using past data you are "borrowing strength" in terms of sample size, from previous years. While this creates a "lag" bias when estimating current conditions, it is more than compensated for by a reduction in variance. This unequal trade between variance and bias leads to a lower mean squared error for the moving average estimator.

Another problem is selection of the weights. Roesch and Reams (1999) suggest that equal weights could be used for each year but note that equal weighting might mask time trends. However, the equal weighting estimator can be thought of as an unbiased estimator for the population parameter at some time approximately in the middle of the rotation cycle, so if the time trends are long enough in duration, the equal weighting estimator will react to the trends. Eriksson (unpublished manuscript) has shown that if the population parameter is relatively constant over time, the equal weighting estimator is approximately optimal in terms of a squared error criterion. In order to track time trends more accurately, more weight could be placed on more recent estimates, but this will increase the variance of the estimator. Johnson and others (2002) also tested numerous estimators using a data set comprised of FIA plot data and found that the MA estimator has the smallest mean squared error in almost every case studied. Thus, in the absence of a catastrophic event or other large-scale change, the lag bias is expected to be small in comparison to the variance, and the MA estimator with equal weights can be considered a nearly unbiased estimator of current forest conditions. If a catastrophic event such as a large fire or a hurricane occurs, then the use of a different estimator will have to be considered. Johnson and others (2002) present several estimators that can be used in such a situation.

The panel system used by the FIA program produces independent yearly estimates from each panel, hence the variance of $\hat{Y}_{MA,t}$ is

$$V[\hat{Y}_{MAj}] = \sum_{h=1}^{T} w_{t-T+h}^2 V[\hat{Y}_{t-T+h}]. \qquad [17]$$

Summary and Recommendations

The statistical strategy presented in this document significantly benefits from the experience on national forest surveys in Canada and the United States. In light of how rapid technology has been evolving, this strategy incorporates new design elements so that remote sensing and land measurement protocol processes can be unified for purposes of making reliable statistical inferences. Consistency of data collection protocols at national and state levels is essential for this strategy to work properly and generate estimates of known confidence. Its implementation on the Pilot Study Area will produce valuable information for improving design elements and data analysis approaches. How this strategy can be integrated with spatial estimation strategies constitutes one of the questions to be addressed through the implementation of this project.

References

Cliff, A.D. and Ord, J.K. 1981. Spatial processes: models and applications. London: Pion. 266 p.

Cochran, W.G. 1977. Sampling techniques. 3rd ed. New York: John Wiley & Sons. 428 p.

De Vries, P. 1986. Sampling theory for forest inventory. New York: John Wiley & Sons. 399 p.

Efron, B and Tibshirani, R.J. 1993. An introduction to the bootstrap. New York: Chapman and Hall. 436 p.

Eriksson, M. 1995. Design-based approaches to horizontal point sampling. Forest Science 41: 890-907.

Gillespie, A.J.R. 1999. Rationale for a national annual inventory USDA Forest Service North Central Forest Experiment Station Research Bulletin NC-165.

Gregoire, T.G. and Scott, C.T. 1990. Sampling at the stand boundary: A comparison of the statistical performance among eight methods. In: Proc. XIX World Forestry Congress IUFRO; 1990 August 5-11; Montreal, Canada. Publ. No. FWS-3-90. Blacksburg, VA: Virginia Polytechnic Institute and University: 78-85.

Husch, B., C., Miller, C.I. and Beers, T.W. 1982. Forest mensuration. 3rd ed. New York: John Wiley & Sons. 402 p.

Johnson, D. S., Williams, M.S. and Czaplewski, R.L. 2003. A test of some estimators for rolling samples using Forest Inventory and Analysis data. Forest Science 49: 50-63.

Kaiser, L. 1983. Unbiased estimation in line-intersect sampling. Biometrics 39: 965-976.

LePage, R. and Billard, L., eds. 1992. Exploring the limits of bootstrap. (Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics Section). New York: John Wiley & Sons. 426 p.

Max, T.A., Schreuder, H.T., Hazard, J.W., Oswald, D.D., Teply, J. and Alegria, J. 1996. The Pacific Northwest Region vegetation and monitoring system. Research Paper RP-493. Portland, OR: USDA Forest Service, Pacific Northwest Research Station. 22 p.

Metzger, K. 1997. Modeling forest stand structure to a ten meter resolution using Landsat TM data. Fort Collins, CO: Colorado State University. 123 p. Thesis.

Overton, W.S. and Stehman, S.V. 1996. Desirable design characteristics for long-term monitoring of ecological variables. Environmental Ecology and Statistics 3: 349-361.

Schreuder, H.T., Gregoire, T.G. and Wood, G.B. 1993. Sampling methods for multiresource forest inventory. New York: John Wiley & Sons. 446 p.

Thaka, R. and Gentleman, R. 1996. R: a language for data analysis and graphics. Journal of Computational and Graphical Statistics 5: 299-314.

Williams, M.S. and Eriksson, M. 2002. Comparing the two paradigms for fixed-area sampling in large-scale inventories. Forest Ecology and Management 168: 135-148.

Williams, M.S. and Patterson. P.L. 2003. Comparison of area estimation techniques for established surveys using remotely sensed and ground data. Forest Science 49: 110-118.

Appendix

Definition of Symbols

A = Total area in population in ha
N = number of sample units in the population
n = number of sample units in the sample
ns = number of subsample units in each sample unit

$\overline{Y} = \sum_{i=1}^{N} Y_i / N$ = average per sample unit in the population for the variable of interest

Y = total for the variable of interest in the population, $Y = \sum_{i=1}^{N} Y_i = N\overline{Y}$

Population variance for y = $\sigma_y^2 = \sum_{i=1}^{N} (Y_i - \overline{Y})^2 / N$

π_i = probability of selecting sample unit I
π_{ij} = joint probability of selecting sample units I and j
N_h = number of possible sample units in stratum h
n_h = number of sample units in stratum h
p_{hij}^F = the proportion of forest in the j^{th} subplot of the i^{th} psu within the h^{th} stratum
$\hat{p}_{hij}^F$ = estimated proportion of forest in the j^{th} subplot of the i^{th} psu within the h^{th} stratum
$\hat{A}_F$ = estimated area in forest
$\hat{p}_F$ = estimated proportion of the total area in forest
$\hat{A}_{ij}$ = estimated sample area in subplot j of plot I
$\hat{Y}_{ij}$ = estimated sampled value of interest in subplot j of plot i
n = the number of plots in the sample
n_s = number of subplots in the sample for plot I
$\hat{Y}_s$ = estimator for subplot size s and subplot ss
m_i = number of subplots in the sample for plot i (Max and others 1996)

Treatment of Missing Subplots

The approach used in the pilot study for Jalisco and Colima closely resembles what is done in Region 6 of the U.S. Forest Service. The approach used there to treat missing subplots is presented here and will be used in the pilot study.

The basic rule for measuring plots that either intersect the boundary of the population or are partially inaccessible due to, say hazardous conditions, require adherence to the following basic principles:

(1) Decisions concerning whether or not to measure a subplot or what part of a subplot to measure are made independently on each subplot of a psu, regardless of subplot number, in other words, the psu center, subplot number 1, is no different than any other subplot center;

(2) If a subplot center is inaccessible, then no measurements are taken on this subplot, since the subplot center cannot be established and referenced by standard field methods. In cases where part of the subplot is accessible or within the population boundary, even though the subplot center is inaccessible or outside the population, then this results in a part of the population that is selected to be included in the sample but is not measured. Hence this is a source of potential bias in the estimation process.

(3) If the subplot center is accessible and within the population boundary, then all regular measurements are taken on that part of the plot that is accessible. The subplot must be mapped, with relevant measurements, in sufficient detail so that the amount of area of the subplot that was actually measured can be calculated accurately. This area provides the information essential for computing the proper area weight for the partially measured subplot.

(4) For purposes of making decisions about measurement, a psu can be thought of as consisting of distinct and separate, although contiguous, subplots. Each subplot has nested area-based measurement units as described in the text.

Inaccessible PSU's

In this section we address the situation where some, or possibly all, of a psu is inaccessible. Inaccessibility is usually caused by hazardous conditions, for example the existence of cliffs, that prevents safe access to the area covered by the plot.

1. Subplots where subplot center is inaccessible.

   a.- Entire area of subplot is inaccessible. Follow the basic rule, part (2). The entire subplot is not measured. This lack of measuring a part of the selected sample area is a source of potential bias in estimation. If the subplot center that is inaccessible is subplot 1, the psu center, then the location of the psu must be monumented with respect to a subplot whose center is accessible. This is a deviation from standard field procedure in which the location of the psu is referenced with respect to the psu center coincident with the center of subplot 1.

   b.- Some area of subplot is inaccessible. Follow the basic rule, part B (2). The entire subplot is not measured because the center of the subplot is inaccessible and cannot be establish using usual field procedures. This lack of measuring a part of the selected sample area is, again, a source of potential bias in estimation. If the subplot center that is inaccessible is subplot 1, the psu center, then the location of the psu must be monumented with respect to a subplot whose center is accessible. This is a deviation from standard field procedure in which the location of the psu is referenced.

2. Subplots where subplot center is accessible.

   The basic rule is followed, in this case, without exception. Follow the basic rule, part B (2). The part of this subplot within the population boundary is really part of the population. To access and measure this part of the subplot requires establishing the subplot center that actually is located on the adjacent property. To avoid gaining access to this adjacent property, we forgo measuring the part of the subplot on NFS land. There is a potential bias associated with not measuring the part of this subplot that is within the population boundary. If the subplot center that is outside the population boundary is subplot 1, the psu center, then the location of the psu must be monumented with respect to a subplot whose center is within the population boundary. This is a deviation from standard field procedure in which the location of the psu is referenced with respect to the psu center, coincident with the center of subplot 1.

3. Subplots where center is within the population.

   The basic rule is followed, in this case, without exception.

RMRS
ROCKY MOUNTAIN RESEARCH STATION

The Rocky Mountain Research Station develops scientific information and technology to improve management, protection, and use of the forests and rangelands. Research is designed to meet the needs of National Forest managers, Federal and State agencies, public and private organizations, academic institutions, industry, and individuals.

Studies accelerate solutions to problems involving ecosystems, range, forests, water, recreation, fire, resource inventory, land reclamation, community sustainability, forest engineering technology, multiple use economics, wildlife and fish habitat, and forest insects and diseases. Studies are conducted cooperatively, and applications may be found worldwide.

Research Locations

Flagstaff, Arizona
Fort Collins, Colorado*
Boise, Idaho
Moscow, Idaho
Bozeman, Montana
Missoula, Montana
Lincoln, Nebraska
Reno, Nevada
Albuquerque, New Mexico
Rapid City, South Dakota
Logan, Utah
Ogden, Utah
Provo, Utah
Laramie, Wyoming

*Station Headquarters, Natural Resources Research Center,
2150 Centre Avenue, Building A, Fort Collins, CO 80526